```
901  TCCCTCCCCT TTGTCAGTGT ATAGAGGGTT CTTTGGATGT AGACACTTTT CAAAGACCAA
     H  P  S  P  L  S  V  Y  R  G  F  F  G  C  R  H  F  S  K  T

961  TGAGCTGCTG CAGAAGTCTG GCAAGAAGCC CATTGACTGG AAGGAGCTGT GATCATCAGC
     N  E  L  L  Q  K  S  G  K  K  P  I  D  W  K  E  L

1021 TGAGGGGTGG CCTTTGAGAA GCTGCTGTTA ACGTATTTGC CAGTTACGAA GTTCCACTGA

1081 AAATTTTCCT ATTAATTCTT AAGTACTCTG CATAAGGGGG AAAAGCTTCC AGAAAGCAGC

1141 CATGAACCAG GCTGTCCAGG AATGGCAGCT GTATCCAACC ACAAACAACA AAGGCTACCC

1201 TTTGACCAAA TGTCTTTCTC TGCAACATGG CTTCGGCCTA AAATATGCAG AAGACAGATG

1261 AGGTCAAATA CTCAGTTGGC TCTCTTTATC TCCCTTGCCT TTATGGTGAA ACAGGGGAGA

1321 TGTGCACCTT TCAGGCACAG CCCTAGTTTG GCGCCTGCTG CTCCTTGGTT TTGCCTGGTT

1381 AGACTTTCAG TGACAGATGT TGGGGTGTTT TTGCTTAGAA AGGTCCCCTT GTCTCAGCCT

1441 TGCAGGGCAG GCATGCCAGT CTCTGCCAGT TCCACTGCCC CCTTGATCTT TGAAGGAGTC

1501 CTCAGGCCCC TCGCAGCATA AGGATGTTTT GCAACTTTCC AGAATCTGGC CCAGAAATTA

1561 GGGCTCAATT TCCTGATTGT AGTAGAGGTT AAGATTGCTG TGAGCTTTAT CAGATAAGAG

1621 ACCGAGAGAA GTAAGCTGGG TCTTGTTATT CCTTGGGTGT TGGTGGAATA AGCAGTGGAA

1681 TTTGAACAAG GAAGAGGAGA AAAGGGAATT TTGTCTTTAT GGGGTGGGGT GATTTTCTCC

1741 TAGGGTTATG TCCAGTTGGG GTTTTTAAGG CAGCACAGAC TGCCAAGTAC TGTTTTTTTT

1801 AACCGACTGA AATCACTTTG GGATATTTTT TCCTGCAACA CTGGAAAGTT TTAGTTTTTT

1861 AAGAAGTACT CATGCAGATA TATATATATA TATTTTTCCC AGTCCTTTTT TTAAGAGACG

1921 GTCTTTATTG GGTCTGCACC TCCATCCTTG ATCTTGTTAG CAATGCTGTT TTTGCTGTTA

1981 GTCGGGTTAG AGTTGGCTCT ACGCGAGGTT TGTTAATAAA AGTTTGTTAA AAGTTCAAAA

2041 AAAAAAAAAA AAA
```

or a fragment thereof encoding a catalytically active product comprising at least nucleotides 121 to 130, preferably 71 to 202 in addition to the catalytic domain, or a sequence which is degenerate, substantially homologous with or which hybridizes with at least nucleotides 121 to 130, preferably 71 to 202 of any such aforesaid sequence. Preferably such degeneracy, homology or hybridization applies to the entire

sequence.

"Catalytically active product" as used herein refers to any product encoded by said sequence which exhibits uracil DNA glycosylase activity.

"Substantially homologous" as used herein includes those sequences having a sequence homology of approximately 60% or more, eg. 70% or 80% or more, and also functionally-equivalent allelic variants and related sequences modified by single or multiple base substitution, addition and/or deletion. By "functionally equivalent" in this sense is meant nucleotide sequences which encode catalytically active polypeptides, ie. having uracil DNA glycosylase activity.

Sequences which "hybridize" are those sequences binding under non-stringent conditions (eg. 6 x SSC 50% formamide at room temperature) and washed under conditions of low stringency (eg. 2 x SSC, room temperature, more preferably 2 x SSC, 42°C) or conditions of higher stringency (eg. 2 x SSC, 65°C) (where SSC = 0.15M NaCl, 0.015M sodium citrate, pH 7.2). Generally speaking, sequences which hybridize under conditions of high stringency are included within the scope of the invention, as are sequences which, but for the degeneracy of the code, would hybridize under high stringency conditions.

The significance of the UNG1, UNG2 presequences has also been investigated in the present invention, by the use of constructs that express fusion products of UNG1 or UNG2 and green fluorescent protein (EGFP). Surprisingly, significant effects on subcellular targeting were observed and after transient transfection of HeLa cells, the pUNG1-EGFP-N1 product co-localized with mitochondria whereas the pUNG2-EGFP-N1 product targeted exclusively to nuclei. Whilst not wishing to be bound by theory, it appears that these sequences may be instrumental in the localization of the enzymes. The

putative nuclear signal was identified as RKRH which also appears in the catalytic domain of both UNG1 and UNG2. Whilst it was recognized previously by Slupphaug et al., 1993, Nucl. Acids Res., 21(11), p2579-2584, that the signal for mitochondrial translocation resides in the UNG1 presequence, it was believed that the signal for nuclear import lay within the mature protein as in the absence of the presequence, UNG1 was transported to the nucleus. However, UNG2 has now been identified which has a presequence and which localizes to the nucleus. These presequences thus have utility for directing the subcellular localization of molecules attached to them.

Thus, viewed from a further aspect, the invention provides nuclear localization peptides encoded by a nucleic acid molecule comprising the sequence (SEQUENCE I.D. Nos 3 and 4):

```
ATGATCGGCC AGAAGACGCT CTACTCCTTT TTCTCCCCCA GCCCCGCCAG
  M  I  G   Q  K  T   L  Y  S  F   F  S  P   S  P  A

GAAGCGACAC GCCCCCAGCC CCGAGCCGGC CGTCCAGGGG ACCGGCGTGG CTGGGGTGCC
R  K  R  H   A  P  S   P  E  P   A  V  Q  G   T  G  V   A  G  V

TGAGGAAAGC GGAGATGCGG CG
P  E  E  S   G  D  A   A
```

or a fragment thereof encoding a functional equivalent or a sequence which is degenerate, substantially homologous with or which hybridizes with any such aforesaid sequence.

Functionally equivalent fragments refer to products which may serve as appropriate localization peptides. Especially preferred nuclear localizing peptides are those which include the amino acid sequence RKRH.

A further preferred feature of the invention comprises DNA glycosylases of the invention which additionally comprise at least one of the aforesaid nuclear localization peptide sequences or at least one

mitochondrial localization peptide sequence encoded by a nucleic acid molecule comprising the sequence (SEQUENCE I.D. Nos 5 and 6):

```
ATGGGCGTCT TCTGCCTTGG GCCGTGGGGG TTGGGCCGGA AGCTGCGGAC  GCCTGGGAAG
  M  G  V   F  C  L   G  P  W  G   L  G  R   K  L  R   T  P  G  K

GGGCCGCTGC AGCTCTTGAG CCGCCTCTGC GGGGACCACT TGCAG
  G  P  L   Q  L  L   S  R  L  C   G  D  H   L  Q
```

or a fragment thereof encoding a functional equivalent or a sequence which is degenerate, substantially homologous with or which hybridizes with any such aforesaid sequence, e.g. CDG or TDG with a localization peptide. Such a composite may be prepared for example by appropriate modification of UNG1 or UNG2.

The novel DNA glycosylases of the invention conveniently may be obtained by modification of existing DNA glycosylase enzymes, such as the human UDG mentioned above. Such modification, for example by replacement, addition or deletion of one or more amino acid residues, or indeed chemical modification of amino acid residues, may readily be achieved using methods well known in the art and include modifications both at the protein level and also at the level of the encoding nucleic acid. For example, site-directed mutagenesis techniques are widely described in the literature. Other conventional mutagenesis treatments which may be used to obtain enzymes according to the invention include random or regional random mutagenesis by chemical agents, such as N-nitroso compounds, or physical agents, such as ultraviolet light, as well as random or regional random mutagenesis by polymerase chain reaction (PCR) methods. Regional random mutagenesis may be carried out by subcloning one or more relevant DNA sequences encoding segments of the starting protein e.g. UDG, followed by random mutagenesis on this fragment or fragments. After the fragments have been mutagenized they may be reinserted into a DNA sequence encoding the starting

protein e.g. UDG. Screening of individual colonies for novel DNA glycosylases of the invention may then be performed using assay methods described herein.

Alternatively, the novel DNA glycosylases of the invention may be obtained by other techniques, for example polypeptide synthesis, construction of fusion proteins etc.

DNA glycosylase activity may readily be assayed according to techniques well known in the art, see for example Slupphaug et al. (1995) Biochemistry, 34: 128-138, and Nedderman & Jiricny, supra. Assays for DNA glycosylase may be used for identifying enzymes according to the invention. The enzymes may be naturally occurring or formed as the result of manipulations of naturally occurring gene sequences or products. Thus, for example, a cell-free extract may be assayed using a thymine or cytosine-containing substrate to identify enzymes which perform excision of one or more of the bases. For the purposes of assessment, the cytosine and thymine bases in the substrates are conveniently labelled, for example fluorescent or radiolabelled e.g. with ^{3}H. Suitable substrates may be prepared by methods known in the art e.g. by nick translation, random priming, PCR or chemical synthesis. To ascertain if the enzymes are also capable of excising uracil, substrates including uracil may also be used. Conveniently, the uracil bases should be labelled to allow detection. Assays for the excision of different bases are preferably performed independently.

Thus, viewed from a yet further aspect, the invention provides an assay for the identification of DNA glycosylases of the invention in a sample, in which said assay comprises at least the step of assaying for activity in the sample which is capable of excising thymine or cytosine and optionally also uracil from an introduced ssDNA and/or dsDNA substrate. Optionally, the moiety responsible for such activity may be

isolated. Suitable assays are described herein and are also known in the art.

DNA glycosylases of the invention include modifications of human UDG by amino acid replacement, as mentioned above, especially at positions 204 and 147. Such amino acid-substituted mutants of human UDG may also comprise additional modifications, for example truncation from the N- and/or C-terminal, or chemical derivation of amino acid residues and/or addition, deletion or mutation of constituent residues which do not affect the overall specificity of the enzyme.

Derivatives of UDG or other DNA glycosylase enzymes from other genera or species, having the CDG or TDG functional activity mentioned above, are also included within the scope of the invention. It will be appreciated that appropriate modification of such enzymes would be performed on comparable residues to those in the human enzyme which form part of the active site and which could be identified by methods known in the art, e.g. by sequence comparison to human UDG and/or by mutation of residues which are identified as potentially conferring specificity to the enzyme and subsequent substrate specificity analyses of the mutant enzymes thus obtained.

The novel DNA glycosylases of the invention may have a number of uses, for example as tools in molecular biology procedures, most notably in mutagenesis, both <u>in vitro</u> and <u>in vivo</u>, but also in other areas such as cell killing, removal of contaminating DNA, random degradation of DNA, enzymatic DNA sequencing etc.

In light of the identification of mitochondrial and nuclear localizing peptides, it is now possible to direct human uracil-DNA glycosylase either to nuclei or to mitochondria by making constructs containing either a nuclear localization signal, such as in UNG2, or a mitochondrial localization signal, such as in UNG1, as mentioned above. Whilst this alone may be used to

mutate RNA in the cells, this is particularly useful in combination with site directed mutations that give rise to mutants that have either TDG activity or CDG activity because it allows for selective mutagenesis of nuclear DNA or mitochondrial DNA. Furthermore, it is useful in a system where either nuclear or mitochondrial DNA is the target for degradation for the purpose of killing cells, eg. cancer cells.

As mentioned above, DNA glycosylases according to the invention may be used in a mutagenesis system both in vitro and in vivo. These proteins have numerous advantages over typical chemical mutagens, particularly regarding their ease of use. Small molecular mutagens, such as methylnitrosurea (MNU), methylmethanesulfonate (MMS) or methylnitrosoguanidine (MNNG) are very toxic on contact with eyes, skin or mucosal membranes and may decompose to explosive and volatile toxic compounds. Other mutagens, such as dimethylnitrosamine and benzo(a)pyrene require metabolic activation by special enzymes that are only present in some cells. They can therefore only be used under certain experimental conditions and will often require the addition of a fraction containing activating enzymes. All these chemical mutagens therefore require specialised precautions in order to protect the user. One major advantage of DNA glycosylases according to the invention is that they are not volatile and are not harmful to the user, for example, by mere skin contact.

Mutagenesis in vitro may be performed on a complex sample, e.g. a cell-free extract, a partially refined sample, e.g. nucleic-acid enriched or purified sample or on a single population of nucleic acid material, e.g. amplified nucleic acid material. Random mutation may be performed using selected DNA glycosylases of the invention (possibly in combination with one another and/or with known DNA glycosylases), to release particular bases or combinations of bases from the

nucleic acid substrate. Removal of the resulting abasic site and replacement of the removed base with another base may be performed by provision of appropriate enzymes and bases.

Specific mutagenesis may be performed in a number of ways. Depending on the specificity of the DNA glycosylase for ssDNA or dsDNA, either one or the other type of DNA may be targeted. One application of such a method may be to introduce labelled bases into the target DNA to identify its presence or amount in the total nucleic acid material. Alternatively, the substrate which is uniquely recognizable (e.g. dsDNA) may be made sensitive to digestion or degradation after release of the appropriate base by DNA glycosylase activity when replacement of the base has not been performed. This may then be used to remove certain ss- or ds-DNA from a sample. Such an application is discussed in more detail hereinafter.

Another application involves the introduction of selected bases after release of the specific bases recognized by the DNA glycosylase. In this way, replacement of specific bases by specific other bases may be performed. It is known from the art that the human UDG has sequence specificity for uracil excision in the sequence surrounding the uracil base (Slupphaug et al., 1995, supra). Appropriate selection of enzyme concentrations and other determinants may be employed to excise specific bases from known sequences or alternatively, by replacement with appropriately labelled bases, to determine the presence of such sequences in nucleic acid samples.

For mutagenesis <u>in vivo</u>, e.g. in a cell, a nucleotide sequence encoding a DNA glycosylase according to the invention under the control of an suitable expression vector may be introduced into the cell by any suitable means, for example, by transformation or through the use of liposomes.

A further aspect of the present invention thus provides a nucleic acid molecule comprising a nucleotide sequence which encodes a DNA glycosylase and/or nuclear localizing peptide of the invention as defined above. Such nucleic acid molecules may readily be prepared using conventional techniques well known in the art. Thus, for example, as already mentioned above, known gene sequences coding for DNA glycosylases, e.g. the UNG gene mentioned above, may be modified e.g. by nucleic acid substitution using standard techniques such as site-directed mutagenesis.

In further aspects the invention also provides an expression vector containing a nucleic acid molecule of the invention, and transformed or transfected host cells carrying a nucleic acid molecule of the invention.

The expression vector may be any conventional expression vector known in the art or described in the literature, including both phage and plasmid vectors. In general, these will comprise suitable regulatory sequences e.g. a promoter and/or enhancer operably connected to a gene expressing the enzyme. Suitable promoters include SV40 early or late promoter, e.g. PSVL vector, cytomegalovirus (CMV) promoter and mouse mammary tumour virus long terminal repeat, although preferably inducible promoters are used, e.g. mouse metallothionein I promoter. The vector preferably includes a suitable marker such as a gene for dihydrofolate reductase or glutamine synthetase. The expression vector may for example be an inducible vector, such as the E. Coli vector pTrc99A (See Slupphaug (supra)) inducible with isopropyl β-D-thiogalactopyranoside (IPTG). Other suitable expression vectors include any vector carrying an inducible promoter, such as *lac*, or bacteriophage lambda λ P_L, in which the promoter is under the control of a temperature sensitive repressor (*cI*). Examples of such vectors are pKK223-2 and pP_L-Lambda Inducible (from Pharmacia). The DNA glycosylases of the invention may

also be expressed as fusion proteins. The expression of such fusion proteins may facilitate purification e.g. by using a system such as the GST-gene fusion systems, exemplified by the pGEX vector systems (Pharmacia) or the fusion proteins with peptide sequences that are recognized by specific antibodies, exemplified by the FLAG Expression vectors (Kodak).

The host cell may likewise be any suitable host cell known in the art, including both eukaryotic e.g. yeast, mammalian and plant cells, and prokaryotic cells, e.g. bacteria.

Transfection and transformation techniques are also well known in the art as described for example in Sambrook *et al.* (1989), Molecular Cloning : A laboratory manual, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring harbor, N.Y.) as indeed are other techniques for introducing nucleic acids into cells, for example using calcium phosphate, DEAE dextran, polybrene, protoplast fusion, liposomes, electroporation, direct microinjection, gene cannon etc.

Expression of the DNA glycosylase according to the invention results in the release of C or T from the cellular DNA, which may lead to transition mutations upon replication.

Mutagenesis of cells, e.g. mammalian cells, may also be performed by introduction of the DNA glycosylase protein of the invention into the cell. This may be performed using for example liposomes or other appropriate techniques known in the art.

TDG or CDG may also be used to specifically induce mutations either in the cell nucleus or mitochondria of eukaryotic cells. This may be carried out by expressing cDNA with the complete open reading frame of UNG2, but with a site directed mutation in codon 204 (preferably Asn204Asp) or in codon 147 (preferably Tyr147Ala), in which the N-terminal amino acid sequence contains a nuclear localization signal, as described previously, to

obtain mutations in the nuclear DNA, or by expressing a cDNA expressing the complete reading frame of UNG1, in which the N-terminal amino acid sequence contains a mitochondrial localization signal, as described previously, with similar site directed mutations to those mentioned above, to specifically obtain mitochondrial mutations. For this purpose any expression vector applicable to eukaryotic cells may be used, but preferably the vector system should be inducible. To introduce the expression vectors into the cells, any method for transfection my be used. Alternatively, the same proteins may be expressed and purified and then introduced into the cells by liposome technology or other appropriate techniques in the art as mentioned previously.

Combined in vitro/in vivo mutagenesis may also be performed. For example, an isolated restriction fragment of interest (or possibly the whole plasmid) may be treated with limited amounts of cytosine-DNA glycosylase or thymine-DNA glycosylase. Subsequently, the treated fragment may be reinserted into a vector and transformed into E. coli cells (the cells may also be pre-treated with a DNA damaging agent to ensure an error-prone SOS-repair). As a result of the mutagenicity of AP-sites, this should yield random mutations.

The Examples below describe the induction of mutations in bacterial cells by the expression within such cells of a DNA glycosylase ie. a CDG or TDG according to the present invention. Expression of the DNA glycosylases of the invention in the transformed cells causes an increase in mutation frequencies. Similar results may be obtained with other cells. To enhance mutagenesis, strains may be used, including both prokaryotic and eukaryotic strains, which are defective in the repair of AP-sites or are otherwise hypermutatable e.g. bacterial mutants that are defective

in endonuclease IV or exonuclease III, or both, or other mutants that similarly enhance the yield of mutations.

Thus, the use of one or more DNA glycosylases according to the invention in in vitro and/or in vivo mutagenesis systems provide yet further aspects of this invention.

Another use of DNA glycosylases of the invention involves DNA modification. By treating any type of DNA (single or double-stranded) in vitro with a DNA glycosylase according to the invention, naturally-occurring C or T will be released, thus leaving an apyrimidic site (AP-site). Subsequent treatment of this DNA with alkaline solutions or enzymes such as apurinic/apyrimidinic-site endonucleases (AP-endonucleases) recognising AP-sites will cause breaks in the DNA at the AP-sites. This method may therefore be used for the random cleavage of DNA. The number of cleaved sites will depend on the amount of the DNA glycosylases according to the invention used, thus allowing the number of AP-sites and hence breaks to be controlled. Uses of such methods include the removal of possible contaminating DNA prior to PCR amplification and for the enzymatic sequencing of DNA. The random cleavage of DNA can also be used for producing randomly fragmented DNA of defined size ranges for different purposes, for example for efficient hybridization of DNA, for preparing genomic libraries or for removal of high-molecular weight viscous DNA.

One advantage of using a DNA glycosylase according to the invention in such methods is that in contrast to nucleases, DNA glycosylases do not require divalent cations and this is advantageous when buffers containing divalent cations are not desirable. A further advantage is that the DNA glycosylase may be inactivated by heating the reaction mixture to 80°C for 15 minutes, thus eliminating or substantially reducing its activity.

Uracil-DNA glycosylase has previously been shown to

be efficient in removing contaminating DNA prior to PCR amplification. This method has the disadvantage that only DNA containing uracil could be removed and meant that uracil-containing DNA had to be prepared using appropriate uracil-containing primers to obtain DNA which could be removed prior to amplification. One advantage of the DNA glycosylases according to the present invention is that they do not have this requirement as any contaminating DNA would be likely to contain cytosine or thymine bases. Thus, CDG and/or TDG according to the invention may be added to a reaction mix and allowed to digest contaminating DNA. After treatment the enzymes/s are inactivated prior to the addition of the DNA sample and amplification to avoid degradation of the template or product.

Thus a further aspect of the invention provides the use of one of more DNA glycosylases according to the present invention for removing contaminating DNA prior to PCR amplification. The use of one or more DNA glycosylases according to invention in DNA modification provides a further aspect of the invention. The term "modification" as used herein refers to all forms of modifying or manipulating DNA, including cleavage, base substitution or insertion etc.

A DNA glycosylase according to the present invention may also be used in a method for the killing of cells. A DNA glycosylase according to the present invention may be introduced into specific target cells by means of known transformation techniques, liposomes, specific targeting systems such as ligands that bind to specific receptors, or any other suitable techniques. The DNA glycosylase may be expressed in a tissue-specific manner by placing a tissue-specific promoter upstream of the DNA sequence encoding a DNA glycosylase according to the present invention. Examples of such tissue-specific promoters are well known and are for example found in genes for a number of liver specific proteins such as

albumin, blood clotting factors and apolipoproteins; several hormones, such as human growth hormone from the pituitary gland and insulin from Langerhans islands in pancreas, as well as aromatase involved in the estrogen biosynthetic pathway; porphobilinogen deaminase which is the third enzyme in the heme biosynthetic pathway; glycoprotein IIb/IIIa which is expressed in maturing megakaryocytes; the Zeta subunit of T-cell antigen receptor (TCR) which is expressed in T-cells; CD14 expressed in monocytes and macrophages; villin expressed in certain epithelial tissues and tyrosinase expressed in melanocytes and melanomas. In some cases abnormal expression from tissue specific promoters has been observed in tumour cells, and this may be exploited by using constructs of novel DNA glycosylases and the relevant tissue specific promoter.

When the DNA glycosylase is expressed it may fragment the DNA in the cell and therefore kill the cell. Specific cells may also be targeted through the use of promoters containing other control elements, for example, promoters which are controlled in a cell-cycle or temporal manner or those possessing regulatory elements responsive to internal or external factors, e.g. promoters activatable by specific inducers, e.g. the inducer IPTG, which induces the *lac* promoter or *lac* derivatives such as *trc*, by certain metals (e.g. metallothionein promoter), by certain hormones such as dexamethasone, androgens (on for example the promoter of the gene for prostate specific antigen which is tissue specific), retinic acid and certain cytokines.

Conceivably, where enzymes of the invention exhibit specific substrate requirements in the sequence surrounding the base for excision, this specificity may be employed by appropriate low level expression of the DNA glycosylase such that only DNA with the specific sequence is made susceptible to degradation.

Thus a further aspect of the invention provides a

method of killing cells, comprising the steps of introducing a DNA glycosylase according to the present invention into a cell and expressing said DNA glycosylase in the cell to an extent which results in the killing of that cell. Preferably, the DNA glycosylase according to the present invention is contained within an expression vector, most preferably, a tissue-specific expression vector.

A further use of DNA glycosylases of the invention is for performing enzymatic DNA sequencing. This may be performed in a manner analogous to the chemical sequencing method of Maxam and Gilbert (Maxam and Gilbert (1980) Methods in Enzymology, 65: 499). However, the Maxam-Gilbert procedure involves the use of several very toxic chemicals, such as dimethylsulfate (DMA) and hydrazine (the latter is also explosive) and use of the glycosylases of the invention present a considerable advantage. Enzymatic sequencing may be performed for example by end-labelling the sample DNA fragment appropriately, for example with ^{32}P, ^{33}P or ^{35}S. For identifying the positions of cytosines and thymines in the DNA, the DNA is treated with limiting amounts of cytosine-DNA glycosylase and thymine-DNA glycosylase according to the invention, respectively. The resulting AP-sites are then cleaved, e.g. by alkaline solution (pyridine) or by an AP-endonuclease. The resulting end-labelled fragments are subsequently separated e.g. by electrophoresis and the position of fragments of varying length identified appropriately, e.g. by autoradiography. Ideally, the positions of adenines and guanines should be determined in the same way using adenine- or guanine-DNA glycosylases. At the present time such enzymes are not available. However, the E. coli DNA repair enzymes Tag and AlkA recognize adenine alkylated in the 3-position (Tag, AlkA) and guanine alkylated in the 3-position (AlkA). Thus, one way of determining the positions of adenines and guanines may

be after alkylation of DNA with DMS, followed by treatment with AlkA and Tag. Subsequent experimental procedure may be performed as for determining the C and T positions.

Thus, a further aspect of the invention provides a method of performing enzymatic DNA sequencing to determine the position of cytosine and/or thymine bases by treating said DNA with at least one CDG and/or TDG of the invention.

The invention will now be described more specifically in the following non-limiting Examples with reference to the following drawings in which:

FIGURE 1 comprises graphs showing <u>in vitro</u> excision of radiolabelled material from double stranded (ds) or single stranded (ss) [^{3}H]cytosine-labelled DNA substrate (C-substrate) and [^{3}H]thymine-labelled DNA substrate (T-substrate) by human UDG-mutants (CDG : Panel A, TDG : Panel B). The data represent mean values from two independent experiments each in duplicate for each time point. Symbols in panel B of Figure 1 are as indicated in panel A;

FIGURE 2 comprises graphs showing analysis of the radioactive excision products of substrate DNA by UDG (panel A) and UDG mutants TDG (Panel B) and CDG (Panel C), performed by thin layer chromatography. U-substrate is indicated by stars (★), other symbols are as in Figure 1. The migration of unlabelled standards (the free bases uracil, cytosine or thymine) is indicated as rectangles, marked respectively U-marker, C-marker and T-marker over the relevant fraction numbers;

FIGURE 3 shows a revised organisation of the human *UNG* gene. The restriction maps with *EcoRI*, *HindIII*, *SacI* and *XbaI* are indicated. Exons are shown as black boxes and are numbered by Roman numbers. Exon 1A is a

previously unrecognised exon. Interspersed repeats are indicated (-: Alu, •: MER, ♦: MIR, *: position of a 300 bp TA dinucleotide repeat);

FIGURE 4 shows the generation of human UNG1 and UNG2 by transcription from two promoters and alternative splicing. P2 is the previously recognised promoter for transcription of UNG1 (Haug et al., 1994, FEBS Letters, 353, p180-184) and P1 the promoter from which UNG2 is transcribed. Exon 1A encodes 44 amino acids present in UNG2, but absent in UNG1. The 35 N-terminal codons of exon 1B are only present in UNG1. The presequence of UNG2 is shown on top with the putative nuclear localization signal underlined. The presequence of UNG1 directing mitochondrial import is shown in the bottom line;

FIGURE 5 shows the structure of the 5'-terminal part of the human *UNG* gene (SEQUENCE I.D. No. 7). Bold letters indicate exons (1A and 1B);

FIGURE 6 shows the alignment of UNG proteins from man and mouse (SEQUENCE ID Nos 8 (hUNG1), 9 (mUNG1), 2 (hUNG2) and 10 (mUNG2)). Note that UNG1 and UNG2 proteins have been aligned separately down to the common splice corresponding to codon 44 in human UNG2. The presequence not present in the catalytically active form of human placental uracil-DNA glycosylase originally isolated, residues 1-77 in human UNG1 (Wittwer et al., 1989, Biochemistry, 28, p780-784) is shown in bold letters. Downstream of the alternative splice site (⬇) used for generating UNG2 forms (from 45 in human UNG2), the sequences of the two forms are identical in each species. Residues that make up walls of the uracil-binding pocket or which are directly involved in catalysis are marked with a star (★). Residues that are involved in DNA-binding (except those involved in

uracil-binding) are marked with a triangle (▾); and

FIGURE 7 shows the subcellular localization in HeLa cells of UNG2-EGFP-N1 and UNG1-EGFP-N1 fusion products. HeLa cells were transfected with constructs expressing pUNG2-EGFB-N1 (C), pUNG1-EGFP-N1 (D) or the control pEGFP-N1 (A), all expressed from the CMV promoter, and processed for confocal microscopy. Panel B shows staining of mitochondria with Texas red.

Example 1

Site directed mutagenesis of human UDG codons

Site directed mutagenesis was performed on the relevant codons in human UDG and the proteins expressed in Escherichia coli.

Methods

Site-directed mutagenesis was carried out as in Mol *et al.*, 1995, supra. To obtain the Tyr147Ala mutant, codon 147, TAT→Tyr, was changed to GCT→Ala, and to obtain the Asn204Asp mutant, codon 204, AAC→Asn, was changed to GAC→Asp. Mutated DNA fragments were subcloned into human UDG expression construct pTUNGΔ84 by replacing restriction fragments in the expression construct by fragments containing the respective mutations. In Escherichia coli pTUNGΔ84 expresses high levels of a fully active human UDG (UNGΔ84) lacking 7 non-essential and non-conserved NH_2-terminal residues of the mature form of UDG (Mol *et al.*, 1995, supra; Slupphaug *et al.*, 1995, supra). Expression of mutant proteins in Escherichia coli and purification of the mutant proteins to apparent homogeneity were carried out as described previously (Mol *et al.*, 1995, supra; Slupphaug *et al.*, 1995, supra). Relevant fractions were assayed for DNA glycosylase activity during each step in the purification. As a result of high expression, purification may also take advantage of the UV

absorption of the enzymes. Peaks of UV absorption corresponding to the enzyme of interest could already be observed after only the first two column steps.

To test enzymatic substrate specificities, 250 ng purified human "wild type" UDG (UNGΔ84), UNGΔ84Tyr147Ala or UNGΔ84Asn204Asp, were mixed with 200 ng ds- or ss-[^{3}H]cytosine-labelled DNA (150mCi/mmol), or ds- or ss-[^{3}H]thymine-labelled DNA (100mCi/mmol) in 10 mM NaCl, 20 mM Tris-HCl (pH 7.5), 1 mM EDTA, 1 mM dithiothreitol and 0.5 mg/ml bovine serum albumin (final concentrations) in 20 μl separate reactions. The final concentrations of the [^{3}H]cytosine-DNA (C) and [^{3}H]thymine-DNA (T) substrates were 6.5 μM and 10 μM, respectively. Release of radioactivity as a function of time was measured at 37°C. These conditions are later referred to as standard conditions. Substrate synthesis and processing of samples for scintillation counting were as described in Krokan and Wittwer (1981) Nucl. Acids Res., 9: 2599-2613. Single-stranded substrate was generated by boiling double-stranded substrate for 10 min, followed by rapid cooling on ice.

Results

Figure 1 demonstrates time-dependent release of acid-soluble radioactivity by homogeneous UNGΔ84Asn204Asp (CDG) from [^{3}H]cytosine-labelled DNA, but not from [^{3}H]thymine-labelled DNA. Conversely, the homogeneous UNGΔ84Tyr147Ala (TDG) mutant releases acid-soluble radiolabelled material from [^{3}H]thymine-labelled DNA, but not from [^{3}H]cytosine-labelled DNA.

Example 2

Analysis of the radioactive excision products by thin layer chromatography

Methods

The analysis was performed using DC-cellulose as the

stationary phase and methanol:HCl:H_2O - 70:20:10 as the mobile phase. Samples were prepared as follows: 1.5 μg enzyme (UNGΔ84, UNGΔ84Tyr147Ala or UNGΔ84Asn204Asp as prepared in Example 1) was incubated with 1 μg [^{3}H]uracil-labelled DNA (500mCi/mmol), [^{3}H]cytosine-labelled DNA (150mCi/mmol) or [^{3}H]thymine-labelled DNA (100mCi/mmol) in separate 50 μl reactions under standard buffer conditions (see Example 1) for 1 hour. Macromolecules in the samples were then ethanol precipitated, the supernatants after centrifugation were collected, ethanol was removed by evaporation and the remaining material was dissolved in 10 μl H_2O. 1μl was spotted on the membrane. After migration the cellulose sheet was cut in strips and radioactivity measured by scintillation counting in Ready Protein scintillation cocktail.

Results

Separation of the acid-soluble radioactive material by thin layer chromatography (Figure 2) demonstrated that the released material was the free bases [^{3}H]cytosine or [^{3}H]thymine. Separation using another mobile phase (butanol:H_2O, 86:14) verified these results (data not shown). In addition, both mutants release [^{3}H]uracil, whereas "wild type" UDG (UNGΔ84) releases [^{3}H]uracil only (Figure 2).

Example 3

Substrate specificity and uracil inhibition and kinetic properties of UDG-mutants

Methods

For measuring release of uracil from double-stranded (U-ds) or single-stranded (U-ss) DNA the various mutant enzymes (prepared as described in Example 1 and by analogous site-directed mutagenesis methods and identical expression and purification methods) were incubated with 200 ng ds- or ss- [^{3}H]dUMP-labelled DNA

(500mCi/mmol, 2μM final concentration) in 20 μl separate reactions for 10 min under standard conditions as described in Example 1. For measuring release of [^{3}H]cytosine or [^{3}H]thymine, assays were performed as described in Example 1 using an incubation time of 10 min. Uracil inhibition was analysed by adding 5 mM uracil (final concentration) to a standard U-ds assay. 0 activity indicates activity below detection limit (10 pmol per mg protein per min) with 100 ng enzyme and 200 ng DNA substrate at standard conditions.

The kinetic parameters were determined using six different substrate concentrations to obtain the K_m and V_{max} values. Duplicate samples were incubated for 20 min using standard assay buffer conditions and substrates as specified. K_m and V_{max} were calculated using the computer program Enzpack, version 3.0 after the method of Wilkinson (1961) Biochem. J., 80: 324-332. K_{cat} was calculated from V_{max} assuming an M_r = 25000.

Results

The results are shown in Tables 1 and 2. From Table 1 it can be seen that only the substitution Tyr147Ala results in an enzyme which specifically excises thymine. Similarly, only the substitution Asn204Asp results in a mutant which excises cytosine. Both mutant enzymes exhibit activity on single or double-stranded DNA and are also able to excise uracil. From Table 2 it can be seen that the turnover numbers of CDG and TDG are lower than for "wild type" release of uracil.

Discussion

These results demonstrate the significance of Asn204 for specific binding of uracil-containing DNA and the significance of Tyr147 side chain ring structure for preventing binding of thymine.

It is somewhat surprising that the novel CDG of the invention still recognizes uracil, considering the unfavourable proximity of the Asp carboxyl side chain and the O4 atom in uracil. However, it should be noted that the other oxygen atom of the Asp204 carboxyl side chain still may form H-bonds with N3 uracil and that Asp145 main chain carbonyl as well as the amide-N of Asp145 and Gln144 also contribute to the specificity. In addition, the UDG activity remaining is very low (0.04-0.16%) compared with "wild type". CDG has a 10-fold increased preference for single stranded substrate, whereas TDG has a decreased preference (Figure 1 and Table 1).

It is evident that the turnover numbers (K_{cat}) of the novel enzymes releasing either cytosine or thymine, as well as residual UDG activities, are very low when compared with release of uracil by "wild type" UDG (Table 2). However, the very high turnover number of UDG appears to be unique among DNA glycosylases and turnover numbers of other DNA glycosylases may be as low, or even lower than those of the engineered glycosylases CDG and TDG. Thus, a recent biochemical characterisation of recombinant N-methylpurine-DNA glycosylase from mouse gave K_{cat} values of 0.8 min^{-1} and 0.2 min^{-1} for excision of 3-methyladenine and 7-methylguanine respectively (Roy *et al.* (1994) Biochemistry, 33: 15131-15140).

The Escherichia coli inducible 3-methyladenine-DNA glycosylase II (AlkA) is a DNA glycosylase that recognizes at least 6 different damaged bases, among these structurally different alkylated purines and pyrimidines. The turnover number for AlkA on the substrate 3-methyladenine-DNA is calculated to be 0.03 min^{-1} (Bjelland *et al.* (1994) J. Biol. Chem., 269: 30489-30495).

The FPG protein (formamido-pyrimidine-DNA glycosylase) also has a rather low turnover number. The K_{cat} value on the imidazole ring-opened form of 7-methylguanine-DNA substrate is calculated to 1.4 min^{-1} (Boiteux *et al.* (1990) J. Biol. Chem., 265: 3916-3922). A low rate of catalysis is also likely for the naturally occurring T(U)/G-mismatch-DNA glycosylase since band shifts can be demonstrated after mixing the enzyme with substrate (Sassanfar & Roberts (1990) J. Mol. Biol., 212: 79-96).

All of these DNA glycosylases recognize at least two different substrates, and in most cases several damaged pyrimidines or purines. Probably the very high turnover number of UDG reflects a high selectivity of substrate binding in a tight fitting active site, allowing rapid catalysis by this specialized enzyme. In contrast, the DNA glycosylases with a broader substrate specificity may bind substrate less accurately, and excise the base, more slowly.

TABLE 1

	pmol excised per min per mg protein						Inhibition 5mM Uracil
Mutant	U-ds	U-ss	C-ds	C-ss	T-ds	T-ss	%
«Wild type»	4.7×10^7	9.5×10^7	0	0	0	0	80
Gln144Leu	3.4×10^4	4.8×10^4	0	0	0	0	25
Asp145Glu	5.5×10^4	8.5×10^4	0	0	0	0	80
Asp145Asn	1.4×10^4	1.1×10^4	0	0	0	0	80
Tyr147Ala	2.2×10^4	2.2×10^4	0	0	1.3×10^3	7.5×10^2	0
Tyr147Phe	3.2×10^7	6.3×10^7	0	0	0	0	50
Ser169Ala	3.1×10^6	5.6×10^6	0	0	0	0	80
Asn204Asp	1.7×10^4	1.6×10^5	3.0×10^2	3.0×10^3	0	0	0
Asn204Gln	1.5×10^6	2.2×10^6	0	0	0	0	70
His268Leu	1.3×10^5	2.6×10^5	0	0	0	0	75

TABLE 2

	Substrate											
	C-ds		C-ss		T-ds		T-ss		U-ds		U-ss	
Mutant	K_m (μM)	K_{cat} (min^{-1})	K_m (μM)	K_{cat} (min^{-1})	K_m (μM)	K_{cat} (min^{-1})	K_m (μM)	K_{cat} (min^{-1})	K_m (μM)	K_{cat} (min^{-1})	K_m (μM)	K_{cat} (min^{-1})
«Wild type»	-	-	-	-	-	-	-	-	0.10	2500	0.06	5150
Tyr147Ala					6.0	0.06	1.4	0.02	3.5	1.0	0.30	0.6
Tyr147Phe	-	-	-	-	-	-	-	-	0.16	1225	0.10	2370
Asn204Asp	35	0.12	5.3	0.39	-	-	-	-	2.4	1.2	2.0	15
Asn204Gln	-	-	-	-	-	-	-	-	0.40	66	0.23	89

Example 4

Effects of TDG and CDG activity on frequency of rifampicin resistant mutations in E. coli *ung*$^+$ strain (NR8051) and E. coli *ung*$^-$ strain (NR8052)

Methods

An overnight culture of E. coli strains NR8051 and NR8052 (both *recA*$^+$, provided by Tomas Kunkel of National Institute of Environmental Health, USA) containing plasmids pTrc99A, pTUNGΔ84, UNGΔ84Tyr147Ala and UNGΔ84Asn204Asp were prepared as described in Example 1 and grown in LB-medium with ampicillin (100 μg/ml) at 30°C. The culture was then diluted 1:20 in fresh medium and cultured for 5 hours at 37°C. Induced culture contained 1 mM IPTG in the LB-medium. To determine the number of rifampicin resistant bacteria, 100 μl of the culture were mixed with 3 ml top agarose and poured on LB plates containing 100 μg/ml rifampicin and incubated overnight at 37°C. Colonies were counted and the number of rifampicin resistant colonies per 10^8 viable cells was calculated.

Results

The results are shown in Table 3. These results

indicate that the expression of UDG does not cause an increase in mutation frequencies (plasmid pTUNGΔ84 compared to parental pTrc99A). In fact, human UDG complements E. coli *ung*$^-$ cells. This is clear from the reduction in mutation frequencies from 4.4 to 1.3 when UDG is present in induced cells. Uninduced cells are also protected as a result of promoter leakage. In contrast, the mutation frequencies of E. coli *ung*$^+$ cells are increased by a factor of 8.6 and 39 when carrying plasmids encoding CDG and TDG, respectively, compared to host cells carrying the parental plasmid pTrc99A. This increases to approximately 8.9 and 94.4 respectively, in induced *ung*$^+$ cells.

Discussion

Single amino acid substitutions transform the highly uracil-selective uracil-DNA glycosylase into less selective DNA glycosylases that attack normal pyrimidines and confer a mutator phenotype upon the cell, presumably because excess numbers of apyrimidinic-sites are formed.

It may seem surprising that propagation of plasmids expressing CDG or TDG activity is at all possible, since they might be expected to kill the host cells. We believe that the relatively low turnover numbers and the low expression in the absence of inducer (IPTG) is sufficient to reduce the number of depyrimidinations to a level that the DNA repair system can cope with. Nevertheless, DNA degradation is detectable even in the absence of inducer and is strongly increased when IPTG is added (data not shown). The survival of Escherichia coli *recA*$^+$ host cells carrying uninduced CDG or TDG-plasmid is equal to that of the parental cell carrying plasmid pTrc99A (data not shown) although mutation frequencies are increased by a factor of 8.6 and 39 for CDG and TDG, respectively as mentioned above.

Induction of CDG or TDG by IPTG reduces survival of the Escherichia coli host cells (in both NR8051 and NR8052) to less than 50% and 10%, respectively, within 5 hrs. Thus, AP-site repair capacity is sufficient for repair of damage caused by expression of CDG or TDG due to "leakage" from the uninduced promoter. However, this repair is apparently not complete, or may be inaccurate, since the frequency of mutations leading to rifampicin resistance is significantly increased by induction with IPTG (Table 3). The activity of TDG in vivo leads to a 10-fold higher mutation frequency in Escherichia coli than the in vivo CDG activity. This probably reflects the fact that TDG has a higher activity on dsDNA than CDG, as demonstrated by in vitro experiments with homogeneous enzyme (Figure 1), and that the K_m value for TDG on dsDNA is much lower than the K_m for CDG on dsDNA (Table 2). We have observed that TDG and CDG are both highly cytotoxic in a *recA*$^-$ background (Escherichia coli DH5α) even without induction (data not shown). It is likely that this cytotoxic effect is due to a lack of SOS-induction in *recA*$^-$ cells. The chemical nature of the SOS-inducing signal, or signals, is not fully known, and some DNA lesions may indirectly activate the SOS response by interfering with DNA replication (Sassanfar & Roberts, 1990, supra). If generation of AP-sites by TDG and CDG directly or indirectly triggers SOS-induction, this would increase cell survival, at the cost of error prone repair and a high yield of mutations. CDG and TDG should be very useful for exploring the biological consequences of AP-sites in DNA.

The new DNA glycosylases that we have engineered are distinctly different from previously known glycosylases. The mismatch-specific thymine-DNA glycosylase previously reported also releases uracil (Sassanfar & Roberts, 1990, supra; Nedderman & Jiricny,

1993, supra), like the thymine-DNA glycosylase we have constructed. However while the naturally occurring thymine-DNA glycosylase has an absolute requirement for a mismatched U or T opposite of a G, the TDG we have engineered recognises T or U from T(U):A matches, as well as from single stranded substrate. A DNA glycosylase recognizing unmodified cytosine had previously not been reported.

The mutator phenotype caused by a single amino acid substitution is intriguing since it changes an enzyme from its normal role in mutation avoidance into a cytotoxic mutator protein. In the case of CDG this change is the result of a single A→G transition, which in vivo could be the result of several different events, such as deamination of A, O4-alkylation of T in the complementary strand, and replication errors. Since this mutation would be dominant, only one allele would need to be mutated to get a new phenotype. It is possible, however, that this mutation would be lethal, or that it would be without serious consequences due to efficient repair of DNA in mammalian cells. Nevertheless, the generation of repair enzymes having a dominant mutator effect that would give the cells a hypermutable phenotype may represent a new principle in mutagenesis.

TABLE 3

	Frequency of rif^R mutations per 10^8 cells			
	NR8051		NR8052	
Plasmid	Uninduced	Induced	Uninduced	Induced
pTRC99A	0.8 ± 0.3	0.9 ± 0.5	4.2 ± 1.2	4.4 ± 1.1
pTUNGΔ84	0.7 ± 0.4	0.8 ± 0.3	0.9 ± 0.2	1.3 ± 0.4
pTUNGΔ84Tyr147Ala	31 ± 8	85 ± 32	13 ± 5	57 ± 9
pTUNGΔ84Asn204Asp	6.9 ± 4.1	8.0 ± 3.4	2.1 ± 0.7	4.1 ± 1.1

Example 5

Effects of TDG activity on the frequency of rifampicin resistant mutations in E. coli strains BW527, and GW2100 (umuC$^-$)

An overnight culture of E. coli strains BW527 (endoIV$^-$) or GW2100 (umuC$^-$) provided by Erling Seeberg, The National Hospital, Oslo, containing plasmids pTrc99A or UNGΔ84Tyr147Ala (TDG) were prepared as described in Example 1 and grown in LB-medium with ampicillin (100 μg/ml) at 30°C. The culture was then diluted 1:20 in fresh medium and cultured as described in Example 4.

Results

The results are shown in Table 4. These results indicate that the expression of UNGΔ84Tyr147Ala (TDG) in E. coli strains BW527 (endoIV$^-$) or GW2100 (umuC$^-$) enhances the mutagenic effect of TDG compared to strains that do not carry these defects in the repair of AP-sites or defect in umuC especially after induction with IPTG. pTrc99A alone did not exert this effect to any significant extent. Even more importantly, the background mutations in these strains are low and the effects of induction with IPTG is high, thus improving the usefulness of UNGΔ84Tyr147Ala (TDG) for mutagenesis when using more optimal strains.

These results are especially surprising in light of previous findings that mutants in *umuC* are generally difficult to mutate by some methods, for example by UV-light or by chemical challenge.

TABLE 4

Effects of TDG-activity on frequency of rifampicin resistant mutations in E. coli strains BW527 and GW2100

	Frequency of rif^R mutations per 10^8 cells			
	BW527		GW2100	
Plasmid	Uninduced	Induced	Uninduced	Induced
pTRC99A	0.06±0.02	0.07±0.03	$0.24x10^{-3}±0.1x10^{-3}$	$6x10^{-3}±5x10^{-3}$
pTUNGΔ84Tyr147Ala	1.20±0.2	240±122	0.65±3829	84±28

Example 6

Isolation and characterisation of a nuclear form of uracil-DNA glycosylase

Materials and Methods

Materials

Mouse embryonic carcinoma cDNA library, human liver cDNA library and NT2 neuronal precursor cell cDNA library were from Stratagene (La Jolla, CA, USA). All libraries were propagated in the Uni-ZAP™XR vector using XL-1 blue as host. [α-^{32}P]dCTP, [^{35}S]methionine, Rediprime random labelling kit and Hybond N+ filters were all from Amersham (UK). All sequencing primers were from MedProbe (Oslo, Norway). Dye terminator cycle sequencing ready reaction kit was from Applied Biosystems (Foster City, CA). The Dynazyme PCR kit was purchased from Finnzymes Oy (Espoo, Finland). TNT *in vitro* transcription/translation rabbit reticulocyte lysate system kit, pGEM-T TA cloning kit, Alter Sites II *in vitro* Mutagenesis System, primers for sequencing from T3 and T7 promoters and T3 RNA polymerase were from Promega (Madison, WI). The plasmid encoding the red-shifted variant of green fluorescent protein (pEGFP-N1) was from Clontech (Palo Alto, CA, USA).

Restriction enzymes were from New England Biolabs Inc. (Beverly, MA, USA).

Screening of cDNA libraries

All libraries were screened as recommended by the manufacturer, using ^{32}P-labelled *UNG40* cDNA (Olsen et al., 1989, EMBO J., 8, p 3121-3125) as probe. Hybridization was carried out at 65°C overnight in 6 x SSC, 5 x Denhardt's solution and 0.1% SDS. Filters were washed in 0.1 x SSC/0.5 % SDS at 65°C and autoradiographed. Three rounds of screening were done. *In vivo* excision of pBluescript phagemids from the Uni-ZAP^{TM}XR vector was performed as recommended by the manufacturer.

Sequence analysis of clones

Sequencing was performed on an Applied Biosystems Model 373A DNA Sequencing System using the Dye terminator cycle sequencing ready reaction kit as recommended by the manufacturer. The sequences were analysed using the Auto Assembler software (Applied Biosystems).

In vitro transcription, uracil-DNA glycosylase assays and transient transfection of HeLa cells for promoter studies

In vitro transcription/translation was performed with the TNT transcription/translation system with [^{35}S]methionine as recommended by the manufacturer, using 200 ng of the expression constructs per 10 µl reaction volume. The mouse UNG1-pBluescript construct was transcribed from the T3 promoter in the pBluescript vector. The insert of mouse UNG2-pBluescript was amplified by the polymerase chain reaction using Dynazyme PCR kit, ligated into the pGEM-T vector and transcribed from the T7 promoter. The human UNG2-pBluescript was transcribed from the T3 promoter after

*Sac*I/*Nhe*I excision of a 79 bp fragment from the polylinker and the 5'-end of cDNA for UNG2. Human *UNG1* cDNA was transcribed from the T7 promoter as previously described (Slupphaug et al., 1995, Biochemistry, 34, p128-138). The samples were run on a 12% denaturing sodium dodecyl sulfate polyacrylamide gel (SDS-PAGE). The gel was dried, autoradiographed overnight and scanned on an LKB Ultroscan XL Enhanced Laser Densitometer. Uracil-DNA glycosylase activity was measured in parallel samples of the *in vitro* transcription/translation assay mixture containing unlabelled amino acids (Slupphaug et al., 1995, supra). A construct containing both promoters (pGL2-ProAB) linked to the luciferase gene was prepared by insertion of a *Pvu*II/*Mlu*I fragment (the enzymes cleave in positions 418 and 1035, respectively) from the promoter region of the *UNG* gene into the *Sma*I-*Mlu*I sites of pGL2-ProB. A promoter II-luciferase construct (pGL2-ProB) and transient transfection with Transfectam (Promega) have been described previously (Haug et al., 1994, FEBS Letters, 353, p 180-184).

Preparation of pUNG-EGFP-N1 fusion constructs and localization studies

UNG15 cDNA, which encodes UNG1, in pGEM7Zf+ (pUNG15), (Slupphaug et al., 1995, supra; Olsen et al., 1989, supra) was digested with *Bcl*I, which cuts at bp 1019 in *UNG15* cDNA, blunted with DNA polymerase I, (Klenow fragment), and ligated to an *Age*I linker prepared from the oligonucleotide 5'-ACCGGTGCC-3' and its complementary copy. The religated pUNG15 containing the *Age*I linker correctly ligated into the *Bcl*I site (verified by sequencing) was digested with *Rsr*II, which cuts at bp 49 in *UNG15* cDNA (Olsen et al., 1989, supra), blunted as above and finally digested with *Age*I. The fragment was then ligated into pEGFP-N1 digested with *Sma*I (blunt) and *Age*I. The construct was

sequenced to verify that the construct was in frame with the ATG of the EGFP-N1 fusion protein. The TGA stop codon of pUNG15 was changed to GGA by site-directed mutagenesis performed according to the procedure provided by the manufacturer using ssDNA prepared with R408 phage. Potential $pUNG1_{GGA}$-EGFP-N1 constructs were screened by digestion with *BclI* (digests only unmutated plasmids) and verified by sequencing. The correct construct was named pUNG1-EGFP-N1. cDNA for UNG2 (this example) in pBluescript was digested with *Nhe*I, which cuts 54 bp upstream of ATG, and *Eco*NI which cleaves the cDNAs in the sequence that is shared by cDNAs for UNG1 and UNG2 (positions 529 and 520, respectively). The resulting fragment of interest (501 bp) was isolated and ligated to the 5155 bp fragment of *Nhe*I/*Eco*NI-digested pUNG1-EGFP-N1 to obtain pUNG2-EGFP-N1. Transient transfections of HeLa cells were done with the $CaPO_4$-method (Profection, Promega) according to the manufacturer's recommendations. Confocal microscopy (BioRad MRC-600) of HeLa cells and staining of mitochondria with mouse anti human mitochondria antibody (MAB 1273, Chemicon) and Texas Red anti-mouse IgG (Vector) were performed as previously described (Nagelhus et al., 1995, Exptl. Cell Res., 220, p 292-297). Examination of HeLa cells transfected with expression plasmids pEGFP-N1, pUNG1-EGFP-N1 or pUNG2-EGFP-N1 was carried out using an excitation wave length of 488 nm and emission wave length >515 nm at 16 hours after transfection

Results

A human NT2 neuronal precursor cell cDNA library and a mouse embryonic carcinoma cDNA library were screened and a new form of human uracil-DNA glycosylase (human UNG2) encoded by the *UNG* gene, as well as the homologous cDNA from mouse (mouse UNG2) was identified. In addition the cDNA for the mouse homolog (encoding

mouse UNG1) of human UNG1 (Olsen et al., 1989, supra) was identified. cDNA for human UNG2 has an ORF encoding 44 N-terminal amino acids not found in human UNG1 whereas cDNA for human UNG1 has an ORF encoding 35 amino acids not found in human UNG2 (Figure 4). The two forms are identical in the rest of the amino acid presequence, which is not required for enzyme activity, as well as in the catalytic domain, altogether 269 identical consecutive amino acids. The sequence of the 269 amino acids common to UNG1 and UNG2, and the corresponding DNA sequence, is identical to amino acid residues 35-304 in Olsen et al., 1989, supra. cDNAs for human UNG2 and its mouse homolog, are apparently as abundant as UNG1 in cDNA libraries from proliferating cells since among 20 cDNA clones that were sequenced 10 were of the UNG2 type and 10 were similar to the previously known UNG1 type. Among 4 mouse cDNAs sequences, 3 were of the UNG2 type and 1 was of the UNG1 type. However, screeing of a human hepatocyte library with *UNG40* cDNA resulted in the isolation of 80 strongly bybridizing clones and sequencing of 14 of these demonstrated that they were all similar to the previously characterized cDNA for UNG1 or the cDNA *UNG40* (Olsen et al., 1989, supra).

Comparison of the human cDNA for UNG2 with the recently published complete human *UNG* sequence (Haug et al., 1996, Genomics, 36, p408-416) revealed the presence of a previously unrecognised exon (exon 1A) located some 650 base pairs upstream of the previously identified exon 1 (hereinafter called exon 1B). A revised organization of the UNG gene is therefore presented in Figure 3. Exon 1B forms the leader sequence and codon 1-104 of the mRNA enbcoding the previously known form UNG1. The mRNA corresponding to the new human cDNA is formed by joining exon 1A (encoding 44 amino acids) into a consensus splice site after codon 35 in exon 1B

after which the two human cDNAs are identical. The open reading frame of human *UNG2* cDNA predicts a protein of 313 amino acids, as compared to 304 amino acids for UNG1. Genomic clones for the mouse homolog of the *UNG* gene have also been isolated and sequenced. This has revealed that the splice sites for exons 3, 4, 5 and 6 in the *UNG* genes from mouse and man are in identical positions. Furthermore, PCR analyses have demonstated that the rest of the mouse gene is structurally similar to the human gene, as expected from the cDNA clones (data not shown).

Figure 4 shows how the alternative forms of mRNA for UNG1 and UNG2 arise as deduced from human cDNAs and the corresponding *UNG* sequences and indicates the presence of a putative nuclear localization signal of 4 basic residues (RKRH) in the N-terminal end of the new cDNA and putative mitochondrial localization signals in cDNA for UNG1. In addition, and now shown here, both human cDNAs contain a putative nuclear localization signal (RKRHH) in the catalytic domain (residues 258-262 in the ORF of cDNA for UNG1). These residues are located at the surface of the enzyme between α-helix 7 and β-strand 4 (Mol et al., 1995, Cell, 80, p869-878).

Figure 5 shows the genomic structure of exons 1A and 1B, as well as the structure of the previously characterized promoter (hereinafter called promoter II), possible elements in the putative promoter upstream of exon 1A (hereinafter called promoter I) and the alternative splice acceptor site (SEQUENCE I.D. No. 7). Promoter I probably starts after the 3'-terminal end of two *Alu*-repeats (position 425) and ends immediately upstream of the start of exon 1A. However, it can not be excluded that the promoter is located upstream of the *Alu*-repeats. This would require the presence of an exon encoding a leader that would be

joined to exon 1A. This is considered unlikely since promoter motifs upstream of the *Alu*-repeats have not been detected and furthermore transcripts of the required size have also not been detected by Northern analyses (data not shown). Furthermore, the cDNA for UNG2 does not contain sequences from this upstream region.

Figure 6 shows an alignment of predicted amino acid presequences of the human and mouse enzymes (SEQUENCE I.D. Nos 2 and 8-10). Note that UNG1 proteins and UNG2 proteins have been aligned separately in the parts of the proteins that are derived from different exons (up to codon 45 in human UNG2). Table 4 shows the % of identical residues in the different forms, using human UNG2 as the reference (100%). The parts of the protein that are not required for catalytic activity are less well conserved than the catalytic domain. Amino acids that have been found to be critical for catalytic activity or formation of the uracil-binding pocket (Mol et al., 1995, supra; Kavli et al., 1996, EMBO J., 15, p3442-3447) or DNA binding are completely conserved in mouse (residues Q144, D145, P146, Y147, F158, S169, N204, S247, H268, S270, L272, S273, Y275 and R276 in UNG1).

To compare the promoter activity of promoter I alone and promoter I and promoter II in combination, promoter-luciferase gene constructs were prepared and transient transfection experiments performed with HeLa cells. These studies verified the promoter activity of promoter II alone (Haug et al., 1994, supra) and further demonstrated that when both promoters are present in the construct, the luciferase activity increased some 50%, indicating that promoter I is also active in HeLa cells, as expected from the abundance of the new cDNA in proliferating cells (Table 5).

Coupled transcription-translation of the two forms of human and mouse cDNA resulted in easily measurable uracil-DNA glycosylase activity for both forms from mouse and man. For calculations of the relative specific activities, the radioactivity released in uracil-DNA glycosylase assays was compared to band intensities on an SDS-PAGE gel from transcription/translation reactions using [^{35}S] methionine (Table 6).

To examine whether human UNG1 and UNG2 were translocated to different subcellular compartments, constructs expressing fusion proteins of the UNG proteins and a red shifted variant of green fluorescent protein (EGFP-N1) were prepared. These were used for transient transfection experiments with HeLa cells. The major advantage of the green fluorescent protein (over the use of antibodies) is that this method relies on the autofluorescence of this protein alone, and thus possible cross reaction of the antibody with epitopes in irrelevant proteins is not a problem. The control (pEGFP-N1) shows that the green fluorescent protein displays a homogeneous staining over the cells (Figure 7A). In contrast, the UNG2-EGFP-N1 fusion protein is exclusively located in the nuclei (Figure 7C) and the UNG1-EGFP-N1 fusion protein (Figure 7D) is mainly, if not exclusively, located in extranuclear spots that have the same appearance as mitochondria stained with Texas red (Figure 7B). These results provide convincing experimental evidence that UNG2 is a nuclear protein and UNG1 a mitochondrial protein.

Table 4. Conservation of amino acids in four homologs of uracil-DNA glycosylase calculated as % identity with human UNG2*

	% identity of domains			
	Variant presequence# (1-44)	Common presequence (45-63)	Catalytic domain (64-313)	Overall identity (1-313)
hUNG1	2	100	100	90
mUNG2	64	75	91	86
mUNG1	2	75	91	79

* The identity is calculated for the domains in UNG2 compared with the corresponding domains in the other forms. # The identity of the presequences of hUNG1 and mUNG1 is 27% with 82% identity overall.

Table 5. Promoter activites in the *UNG* gene*

Promoter - reporter gene construct	Luciferase activity %
pGL2-Basic	0.8±0.4
pGL2-ProB	100±8
pGL2-ProAB	156±4

* The promoter activity of pGL2-ProB (promoter II) was arbitrarily set to 100%. pGL2-Basic is a control lacking promoter.

Table 6. Relative specific activities of different forms of UNG after translation in rabbit reticulocyte lysates*

Protein	dpm*	Area (mm^2)	Activity (dpm/area)
human UNG1	1291	0.054	23907
human UNG2	6360	0.268	23731
human UNG1	921	0.061	15098
mouse UNG2	856	0.051	16784

* Relative specific activites were calculated from measured dpm-values (^{3}H-uracil released in uracil-DNA glycosylase assays) and areas under the curve of scanned bands on SDS-PAGE gels after subtraction of background values of 123 dpm.

Claims:

1.1. A DNA glycosylase capable of releasing cytosine bases from single stranded (ss) DNA and/or double stranded (ds) DNA (cytosine-DNA glycosylase) or thymine bases from both single stranded (ss) DNA and double stranded (ds) DNA or from single stranded (ss) DNA (thymine-DNA glycosylase) or uracil bases from single stranded (ss) DNA and/or double stranded (ds) DNA (uracil-DNA glycosylase), wherein said uracil-DNA glycosylase is encoded by a nucleic acid molecule comprising the sequence (SEQUENCE I.D. Nos 1 and 2):

```
1   CACAGCCACA GCCAGGGCTA GCCTCGCCGG TTCCCGGGTG GCGCGCGTTC GCTGCCTCCT

61  CAGCTCCAGG ATGATCGGCC AGAAGACGCT CTACTCCTTT TTCTCCCCCA GCCCCGCCAG
                 M  I  G  Q  K  T  L  Y  S  F  F  S  P  S  P  A

121 GAAGCGACAC GCCCCCAGCC CCGAGCCGGC CGTCCAGGGG ACCGGCGTGG CTGGGGTGCC
    R  K  R  H  A  P  S  P  E  P  A  V  Q  G  T  G  V  A  G  V

181 TGAGGAAAGC GGAGATGCGG CGGCCATCCC AGCCAAGAAG GCCCCGGCTG GGCAGGAGGA
    P  E  E  S  G  D  A  A  A  I  P  A  K  K  A  P  A  G  Q  E

241 GCCTGGGACG CCGCCCTCCT CGCCGCTGAG TGCCGAGCAG TTGGACCGGA TCCAGAGGAA
    E  P  G  T  P  P  S  S  P  L  S  A  E  Q  L  D  R  I  Q  R

301 CAAGGCCGCG GCCCTGCTCA GACTCGCGGC CCGCAACGTG CCCGTGGGCT TTGGAGAGAG
    N  K  A  A  A  L  L  R  L  A  A  R  N  V  P  V  G  F  G  E

361 CTGGAAGAAG CACCTCAGCG GGGAGTTCGG GAAACCGTAT TTTATCAAGC TAATGGGATT
    S  W  K  K  H  L  S  G  E  F  G  K  P  Y  F  I  K  L  M  G

421 TGTTGCAGAA GAAAGAAAGC ATTACACTGT TTATCCACCC CCACACCAAG TCTTCACCTG
    F  V  A  E  E  R  K  H  Y  T  V  Y  P  P  P  H  Q  V  F  T

481 GACCCAGATG TGTGACATAA AAGATGTGAA GGTTGTCATC CTGGGACAGG ATCCATATCA
    W  T  Q  M  C  D  I  K  D  V  K  V  V  I  L  G  Q  D  P  Y

541 TGGACCTAAT CAAGCTCACG GGCTCTGCTT TAGTGTTCAA AGGCCTGTTC CGCCTCCGCC
    H  G  P  N  Q  A  H  G  L  C  F  S  V  Q  R  P  V  P  P  P

601 CAGTTTGGAG AACATTTATA AAGAGTTGTC TACAGACATA GAGGATTTTG TTCATCCTGG
    P  S  L  E  N  I  Y  K  E  L  S  T  D  I  E  D  F  V  H  P

661 CCATGGAGAT TTATCTGGGT GGGCCAAGCA AGGTGTTCTC CTTCTCAACG CTGTCCTCAC
    G  H  G  D  L  S  G  W  A  K  Q  G  V  L  L  L  N  A  V  L
```

```
721  GGTTCGTGCC CATCAAGCCA ACTCTCATAA GGAGCGAGGC TGGGAGCAGT TCACTGATGC
     T  V  R  A  H  Q  A  N  S  H  K  E  R  G  W  E  Q  F  T  D

781  AGTTGTGTCC TGGCTAAATC AGAACTCGAA TGGCCTTGTT TTCTTGCTCT GGGGCTCTTA
     A  V  V  S  W  L  N  Q  N  S  N  G  L  V  F  L  L  W  G  S

841  TGCTCAGAAG AAGGGCAGTG CCATTGATAG GAAGCGGCAC CATGTACTAC AGACGGCTCA
     Y  A  Q  K  K  G  S  A  I  D  R  K  R  H  H  V  L  Q  T  A

901  TCCCTCCCCT TTGTCAGTGT ATAGAGGGTT CTTTGGATGT AGACACTTTT CAAAGACCAA
     H  P  S  P  L  S  V  Y  R  G  F  F  G  C  R  H  F  S  K  T

961  TGAGCTGCTG CAGAAGTCTG GCAAGAAGCC CATTGACTGG AAGGAGCTGT GATCATCAGC
     N  E  L  L  Q  K  S  G  K  K  P  I  D  W  K  E  L

1021 TGAGGGGTGG CCTTTGAGAA GCTGCTGTTA ACGTATTTGC CAGTTACGAA GTTCCACTGA

1081 AAATTTTCCT ATTAATTCTT AAGTACTCTG CATAAGGGGG AAAAGCTTCC AGAAAGCAGC

1141 CATGAACCAG GCTGTCCAGG AATGGCAGCT GTATCCAACC ACAAACAACA AAGGCTACCC

1201 TTTGACCAAA TGTCTTTCTC TGCAACATGG CTTCGGCCTA AAATATGCAG AAGACAGATG

1261 AGGTCAAATA CTCAGTTGGC TCTCTTTATC TCCCTTGCCT TTATGGTGAA ACAGGGGAGA

1321 TGTGCACCTT TCAGGCACAG CCCTAGTTTG GCGCCTGCTG CTCCTTGGTT TTGCCTGGTT

1381 AGACTTTCAG TGACAGATGT TGGGGTGTTT TTGCTTAGAA AGGTCCCCTT GTCTCAGCCT

1441 TGCAGGGCAG GCATGCCAGT CTCTGCCAGT TCCACTGCCC CCTTGATCTT TGAAGGAGTC

1501 CTCAGGCCCC TCGCAGCATA AGGATGTTTT GCAACTTTCC AGAATCTGGC CCAGAAATTA

1561 GGGCTCAATT TCCTGATTGT AGTAGAGGTT AAGATTGCTG TGAGCTTTAT CAGATAAGAG

1621 ACCGAGAGAA GTAAGCTGGG TCTTGTTATT CCTTGGGTGT TGGTGGAATA AGCAGTGGAA

1681 TTTGAACAAG GAAGAGGAGA AAAGGGAATT TTGTCTTTAT GGGGTGGGGT GATTTTCTCC

1741 TAGGGTTATG TCCAGTTGGG GTTTTTAAGG CAGCACAGAC TGCCAAGTAC TGTTTTTTTT

1801 AACCGACTGA AATCACTTTG GGATATTTTT TCCTGCAACA CTGGAAAGTT TTAGTTTTTT

1861 AAGAAGTACT CATGCAGATA TATATATATA TATTTTTCCC AGTCCTTTTT TTAAGAGACG

1921 GTCTTTATTG GGTCTGCACC TCCATCCTTG ATCTTGTTAG CAATGCTGTT TTTGCTGTTA

1981 GTCGGGTTAG AGTTGGCTCT ACGCGAGGTT TGTTAATAAA AGTTTGTTAA AAGTTCAAAA

2041 AAAAAAAAAA AAA
```

or a fragment thereof encoding a catalytically active product comprising at least nucleotides 121 to 130 in

addition to the catalytic domain, or a sequence which is degenerate, substantially homologous with or which hybridizes with at least nucleotides 121 to 130 of any such aforesaid sequence.

2. A cytosine-DNA glycosylase (CDG) as claimed in claim 1.

3. A cytosine-DNA glycosylase (CDG) as claimed in claim 1 or 2 wherein said CDG is capable of releasing both cytosine and uracil bases from ssDNA and/or dsDNA.

4. A cytosine-DNA glycosylase (CDG) as claimed in any one of claims 1 to 3 wherein said CDG is derived from UDG.

5. A CDG as claimed in claim 4 wherein Asn at amino acid position 204 in human UDG protein or equivalent residue in other species is substituted or modified.

6. A CDG as claimed in claim 5 wherein said Asn or equivalent residue is replaced with an aspartic acid residue (Asp).

7. A thymine-DNA glycosylase (TDG) as claimed in claim 1.

8. A thymine-DNA glycosylase (TDG) as claimed in claim 1 or 7 wherein said TDG is capable of releasing both thymine and uracil bases from both ssDNA and dsDNA.

9. A thymine-DNA glycosylase (TDG) as claimed in any one of claims 1, 7 or 8 wherein said TDG is capable of releasing thymine bases from A:T DNA pairs.

10. A thymine-DNA glycosylase (TDG) as claimed in any one of claims 1, 7 or 8 wherein said TDG is capable of

releasing thymine bases from single stranded DNA.

11. A thymine-DNA glycosylase (TDG) as claimed in any one of claims 1 or 6 to 9 wherein said TDG is derived from UDG.

12. A TDG as claimed in claim 11 wherein Tyr at amino acid position 147 in human UDG protein or equivalent residue in other species is substituted or modified.

13. A TDG as claimed in claim 12 wherein said Tyr or equivalent residue is replaced with an alanine residue (Ala).

14. A CDG or TDG as claimed in any one of claims 4 to 6 or 11 to 13 wherein said UDG is human.

15. A uracil-DNA glycosylase (UDG) as claimed in claim 1.

16. A uracil-DNA glycosylase (UDG) as claimed in claim 1 or 15 wherein said fragment comprises at least nucleotides 71 to 202 and/or degenerate, substantially homologous and hybridizing sequences are degenerate, substantially homologous with or hybridize with at least nucleotides 71 to 202.

17. A UDG as claimed in claim 16 wherein degenerate, substantially homologous and hybridizing sequences are degenerate, substantially homologous with or hybridize with the entire sequence.

18. Nuclear localization peptides encoded by a nucleic acid molecule comprising the sequence (SEQUENCE I.D. Nos 3 and 4):

ATGATCGGCC AGAAGACGCT CTACTCCTTT TTCTCCCCCA GCCCCGCCAG
M I G Q K T L Y S F F S P S P A

GAAGCGACAC GCCCCCAGCC CCGAGCCGGC CGTCCAGGGG ACCGGCGTGG CTGGGGTGCC
R K R H A P S P E P A V Q G T G V A G V

TGAGGAAAGC GGAGATGCGG CG
P E E S G D A A

or a fragment thereof encoding a functional equivalent or a sequence which is degenerate, substantially homologous with or which hybridizes with any such aforesaid sequence.

19. Nuclear localizing peptides as claimed in claim 18 wherein said peptides include the amino acid sequence RKRH.

20. A DNA glycosylase as claimed in any one of claims 1 to 17 which additionally comprises at least one nuclear localization peptide sequence as defined in claim 18 or 19 or at least one mitochondrial localization peptide sequence encoded by a nucleic acid molecule comprising the sequence (SEQUENCE I.D. Nos 5 and 6):

ATGGGCGTCT TCTGCCTTGG GCCGTGGGGG TTGGGCCGGA AGCTGCGGAC GCCTGGGAAG
M G V F C L G P W G L G R K L R T P G K

GGGCCGCTGC AGCTCTTGAG CCGCCTCTGC GGGGACCACT TGCAG
G P L Q L L S R L C G D H L Q

or a fragment thereof encoding a functional equivalent or a sequence which is degenerate, substantially homologous with or which hybridizes with any such aforesaid sequence.

21. An assay for the identification of DNA glycosylases as defined in any one of claims 1 to 14 in a sample, in which said assay comprises at least the step of assaying for activity in the sample which is capable of excising thymine or cytosine and optionally also uracil from an introduced ssDNA and/or dsDNA

substrate.

22. A nucleic acid molecule comprising a nucleotide sequence which encodes a DNA glycosylase and/or nuclear localizing peptide as defined in any one of claims 1 to 20.

23. An expression vector containing a nucleic acid molecule as defined in claim 22.

24. A transformed or transfected host cell carrying a nucleic acid molecule as defined in claim 22.

25. Use of one or more DNA glycosylases as defined in any one of claims 1 to 17 in in vitro and/or in vivo mutagenesis systems.

26. Use of one or more DNA glycosylases as defined in any one of claims 1 to 14 for removing contaminating DNA prior to PCR amplification.

27. Use of one or more DNA glycosylases as defined in any one of claims 1 to 14 in DNA modification.

28. A method of killing cells, comprising the steps of introducing a DNA glycosylase as defined in any one of claims 1 to 17 into a cell and expressing said DNA glycosylase in the cell to an extent which results in the killing of that cell.

29. A method of performing enzymatic DNA sequencing to determine the position of cytosine and/or thymine bases by treating said DNA with at least one CDG and/or TDG as defined in any one of claims 1 to 14.

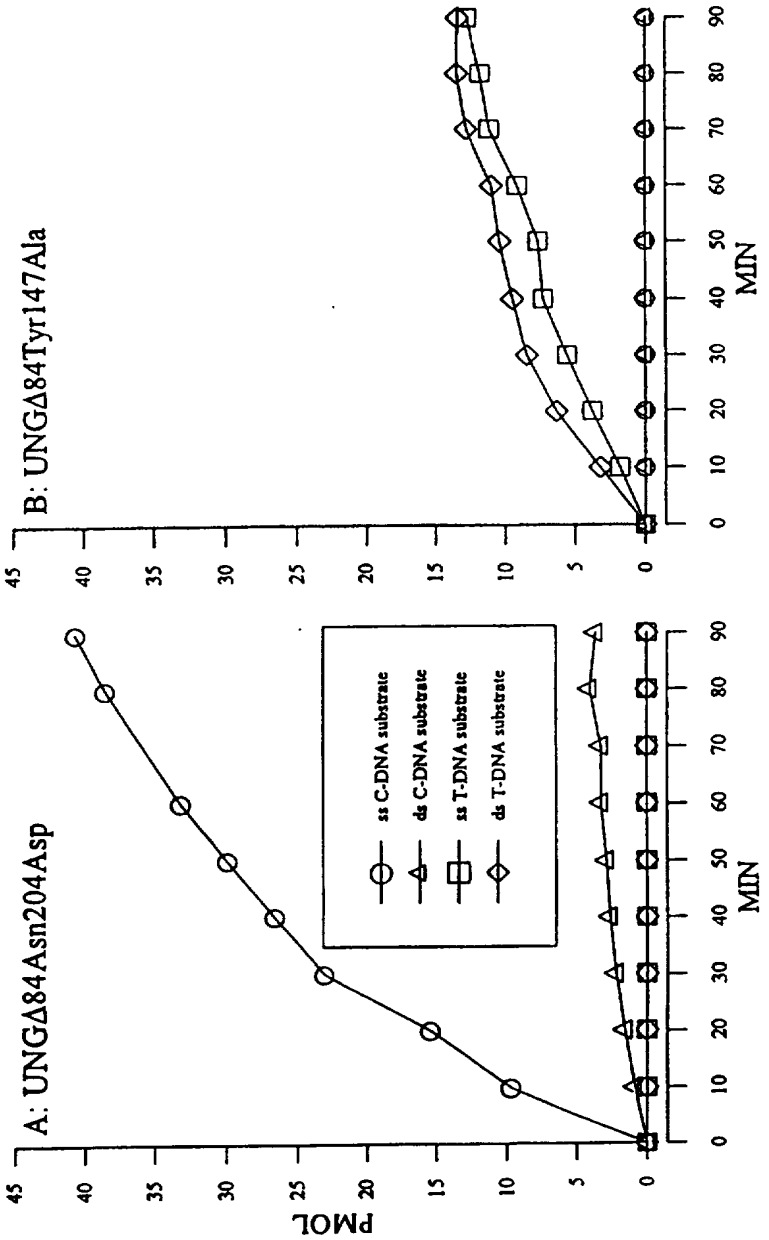

A: UNGΔ84Asn204Asp
B: UNGΔ84Tyr147Ala
PMOL
MIN
ss C-DNA substrate
ds C-DNA substrate
ss T-DNA substrate
ds T-DNA substrate

Figure 1

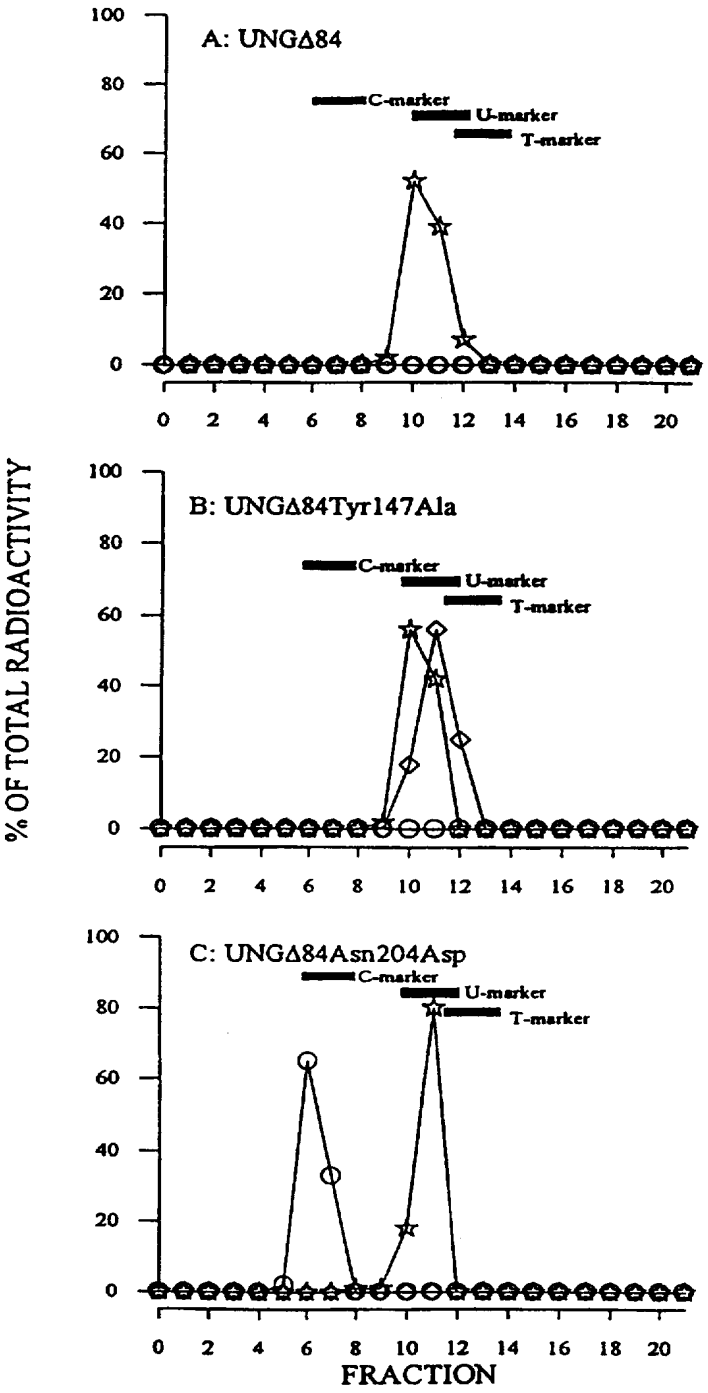

A: UNGΔ84
C-marker
U-marker
T-marker
B: UNGΔ84Tyr147Ala
C-marker
U-marker
T-marker
C: UNGΔ84Asn204Asp
C-marker
U-marker
T-marker
% OF TOTAL RADIOACTIVITY
FRACTION
0
20
40
60
80
100
0 2 4 6 8 10 12 14 16 18 20

Figure 2

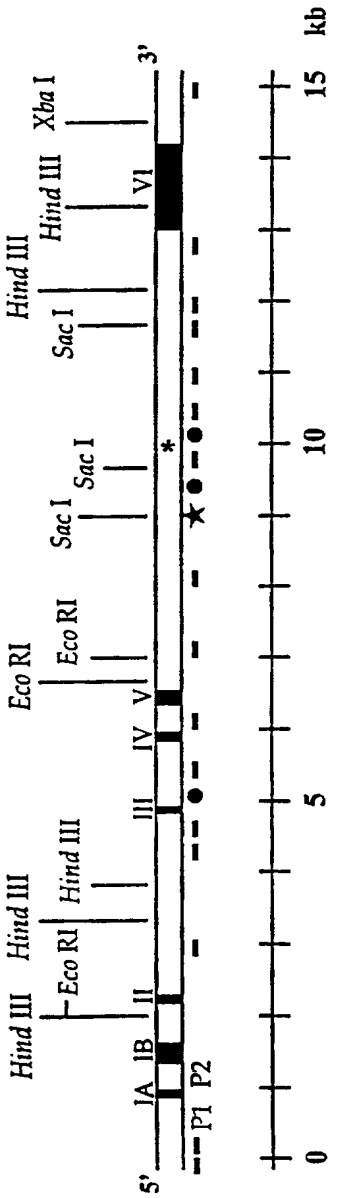

5'
3'
IA
IB
II
III
IV
V
VI
P1
P2
Hind III
Eco RI
Hind III
Hind III
Eco RI
Sac I
Sac I
Sac I
Hind III
Xba I
0
5
10
15
kb

Figure 3

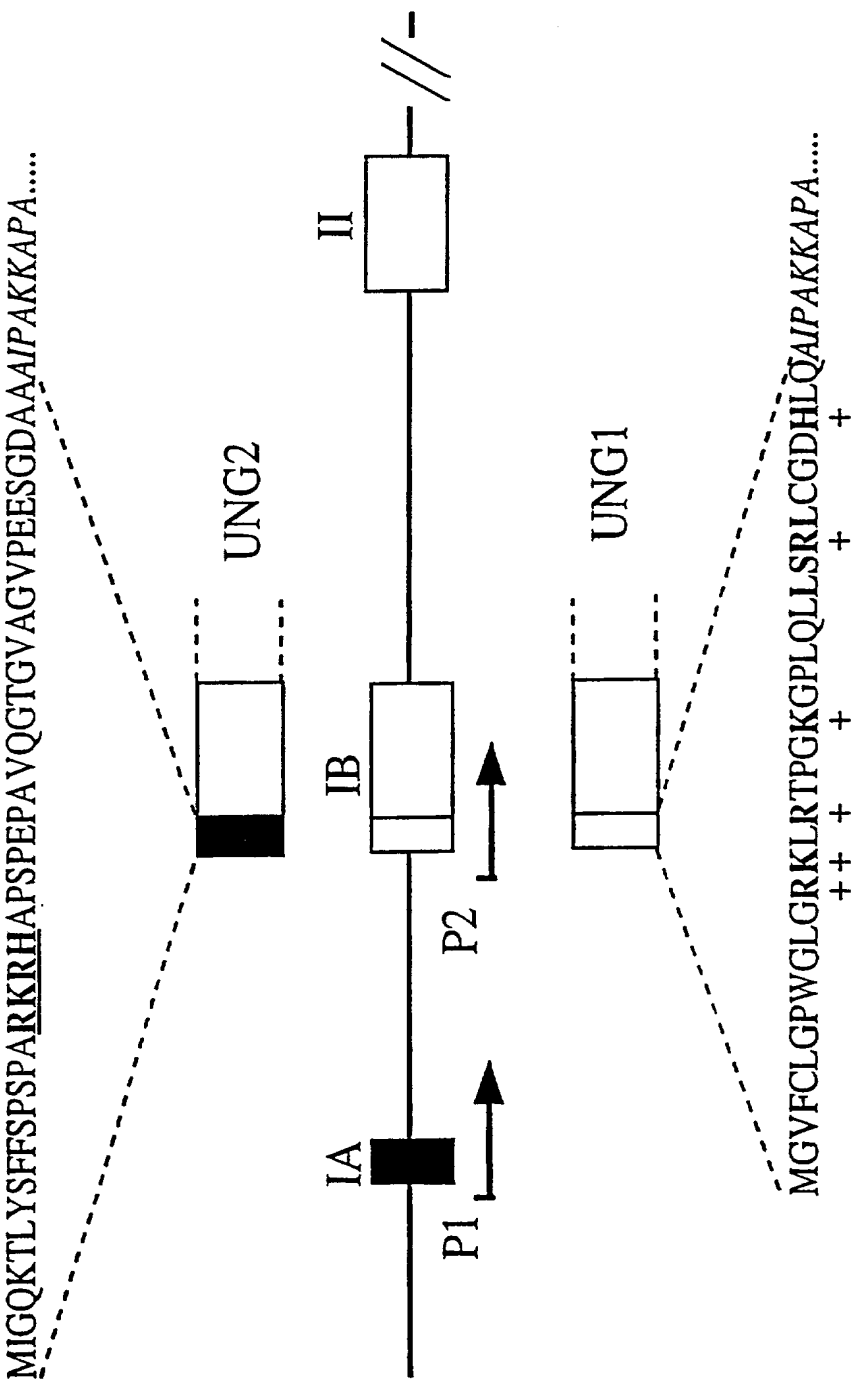

MIGQKTLYSFFSPSPARKRHAPSPEPAVQGTGVAGVPEESGDAAAIPAKKAPA.....
UNG2
II
IB
IA
P2
P1
UNG1
MGVFCLGPWGLGRKLRTPGKGPLQLLSRLCGDHLQAIPAKKAPA.....

Figure 4

```
301  tcaaagctcactacagctcagaccctctggcctcaagcgatcctccagcctgggcctcccaaagcgctaggattacaggcgtgggccaccgcgcctgacc

                                                                                               AP2       C-MYB
401  agtcttctcttcttgcagctgagccttaagagcctgtccaaagagcagaggtgggctgaaggcacaaagcgaatgaaagaataggcccccgggcaccgtt

                    AP2                  AP2                            SP1    SP1/AP2
501  gcacgccccacctcctcccaggggcgttgcactccagccctcccgcacatgcgcactgggccttccaccgccccccgcccccagcaaagccccccgctc

                          CCAAT                Start of exon 1A
601  ggagcatgcgcgggccgcttggcgccaattgctgaccgcCACAGCCACAGCCAGGGCTAGCCTCGCCGGTTCCCGGGTGGCGCGCGTTCGCTGCCTCCTC

               Start codon 1A                                 PEA3
701  AGCTCCAGGATGATCGGCCAGAAGACGCTCTACTCCTTTTTCTCCCCCAGCCCCGCCAGGAAGCGACACGCCCCCAGCCCCGAGCCGGCCGTCCAGGGGA

                          PU.1                                                                PU.1/PEA3
801  CCGGCGTGGCTGGGGTGCCTGAGGAAAGCGGAGATGCGGCGgtgaggcgcggcttgggccggggctaggggggtgaagggggaggaaggcggtgggccccg

                                   AP1                                        C-MYC        REPRESSOR(RPA)
901  cctgacggagggcgtgcaggatcgcgcctctgactcggtaaacccgggctccgctttccaaatagcctccacgtgttcaaaatagccgccgctgtcccccc

     AP2    YY1                 CCAAT                       YY1     CCAAT                  Y1                CCA
1001 atgggccgcccatgctaaagggccagccaatgggaacgcgtctcggggcccatggcgccaatccgcgcgccgcaggccctcctggctcggtgcgctgtccca

     AT AGAGGG AGAGGG  SP1      AGAGGG          E2F           C-MYC               SP1        SP1
1101 atcagaggggagagggggcgggacccagagggaggtttttgccgcgaaaagaccacgtggggacgcggtggggcgggtctggcgggggcggggcacctc

                                                      YY1        Start of exon 1B                   CCAAT
1201 tgtgcagggttcccagtcaccgcgacgctcctcgggaagccatagggcgc CTCCCAGCCCGTCTCCCCGCTCCAGTTTAGAACCTAATTCCCAATTCCC

                                                                      Start codon 1B
1301 GACCGGGCCCAGCCCTGGGCTCTTACTGTCCGCTTTTGCTGGGACCTGTTCCACAAATGGGCGTCTTCTGCCTTGGGCCGTGGGGGTTGGGCCGGAAGCT

                                                                         Alternative splice acceptor site
1401 GCGGACGCCTGGGAAGGGGCCGCTGCAGCTCTTGAGCCGCCTCTGCGGGGACCACTTGCAGGCCATCCCAGCCAAGAAGGCCCCGGCTGGGCAGGAGGAG

1501 CCTGGGACGCCGCCCTCCTCGCCGCTGAGTGCCGAGCAGTTGGACCGGATCCAGAGGAACAAGGCCGCGGCCCTGCTCAGACTCGCGGCCCGCAACGTGC

1601 CCGTGGGCTTTGGAGAGAGCTGGAAGAAGCACCTCAGCGGGGAGTTCGGGAAACCGTATTTTATCAAGgtaaatatggaaatgcaccttccataagggta
```

Figure 5

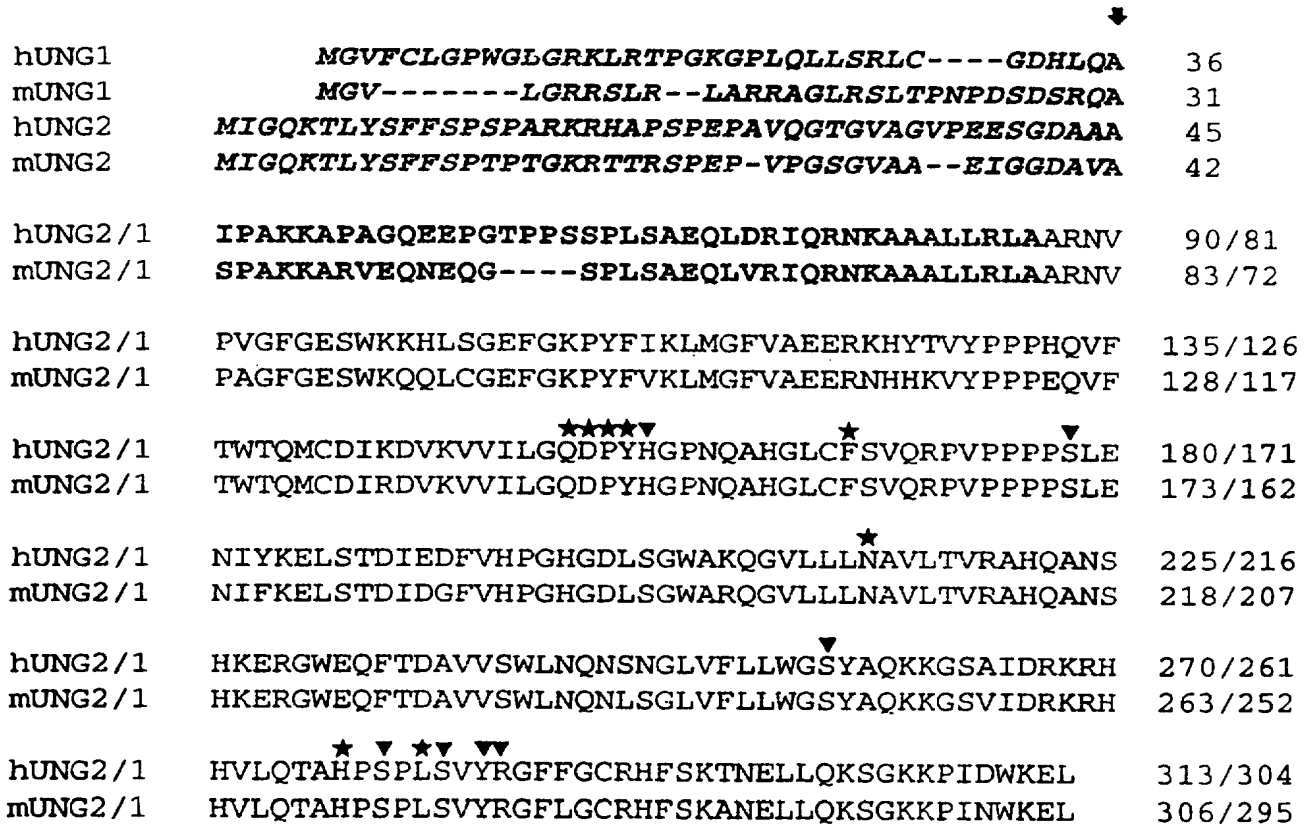

hUNG1 MGVFCLGPWGLGRKLRTPGKGPLQLLSRLC----GDHLQA 36
mUNG1 MGV-------LGRRSLR--LARRAGLRSLTPNPDSDSRQA 31
hUNG2 MIGQKTLYSFFSPSPARKRHAPSPEPAVQGTGVAGVPEESGDAAA 45
mUNG2 MIGQKTLYSFFSPTPTGKRTTRSPEP-VPGSGVAA--EIGGDAVA 42
hUNG2/1 IPAKKAPAGQEEPGTPPSSPLSAEQLDRIQRNKAAALLRLAARNV 90/81
mUNG2/1 SPAKKARVEQNEQG----SPLSAEQLVRIQRNKAAALLRLAARNV 83/72
hUNG2/1 PVGFGESWKKHLSGEFGKPYFIKLMGFVAEERKHYTVYPPPHQVF 135/126
mUNG2/1 PAGFGESWKQQLCGEFGKPYFVKLMGFVAEERNHHKVYPPPEQVF 128/117
hUNG2/1 TWTQMCDIKDVKVVILGQDPYHGPNQAHGLCFSVQRPVPPPPSLE 180/171
mUNG2/1 TWTQMCDIRDVKVVILGQDPYHGPNQAHGLCFSVQRPVPPPPSLE 173/162
hUNG2/1 NIYKELSTDIEDFVHPGHGDLSGWAKQGVLLLNAVLTVRAHQANS 225/216
mUNG2/1 NIFKELSTDIDGFVHPGHGDLSGWARQGVLLLNAVLTVRAHQANS 218/207
hUNG2/1 HKERGWEQFTDAVVSWLNQNSNGLVFLLWGSYAQKKGSAIDRKRH 270/261
mUNG2/1 HKERGWEQFTDAVVSWLNQNLSGLVFLLWGSYAQKKGSVIDRKRH 263/252
hUNG2/1 HVLQTAHPSPLSVYRGFFGCRHFSKTNELLQKSGKKPIDWKEL 313/304
mUNG2/1 HVLQTAHPSPLSVYRGFLGCRHFSKANELLQKSGKKPINWKEL 306/295

Figure 6

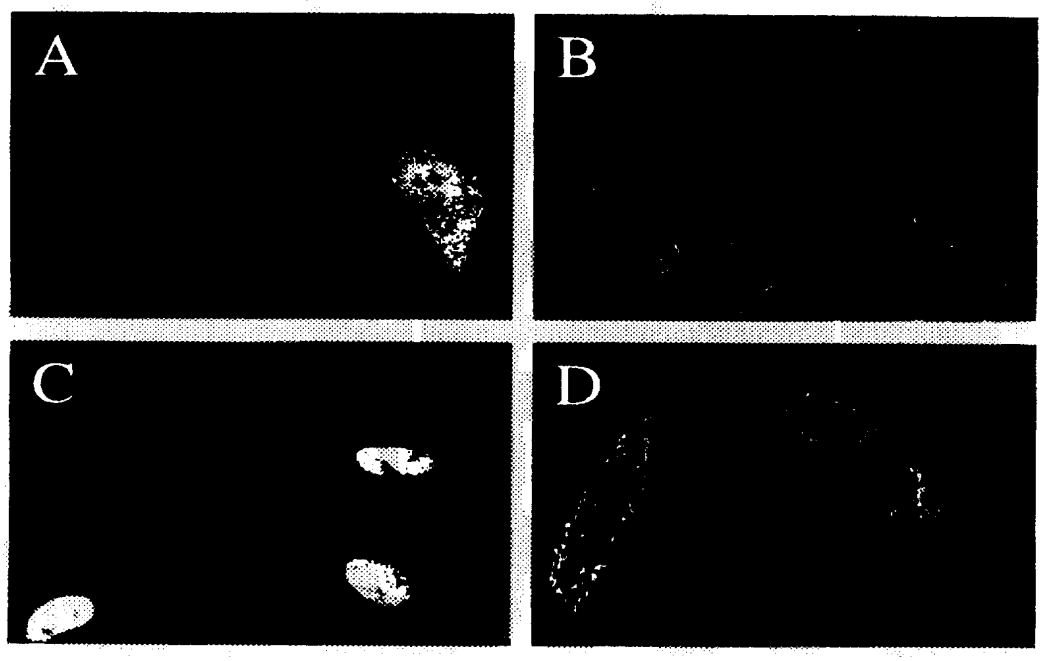


Figure 7

Electronic Acknowledgement Receipt

EFS ID:	46340787
Application Number:	17381888
International Application Number:	
Confirmation Number:	6188
Title of Invention:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ
First Named Inventor/Applicant Name:	Douglas A. Amorese
Customer Number:	21710
Filer:	Thomas C. Meyers/Kelley Warren
Filer Authorized By:	Thomas C. Meyers
Attorney Docket Number:	NUGN-001/04US 34079/550
Receipt Date:	04-AUG-2022
Filing Date:	21-JUL-2021
Time Stamp:	21:53:16
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Foreign Reference	WO9640998A1.pdf	2346149 18a2a719d46fe930ef0e809f52f2f6fd23c05314	no	57

Warnings:

Information:					
2	Foreign Reference	WO9712061A1.pdf	2541433 daaab019b1b097e67c2488a5ba819c200bc329dc	no	55
Warnings:					
Information:					
3	Foreign Reference	WO9725416A2.pdf	2647771 87de3cfde126644762a7ff54af0588e3460ecfef	no	62
Warnings:					
Information:					
4	Foreign Reference	WO1998044151.pdf	4042909 aa4fcc7842419e0a480b6be8de7032ba821a9b35	no	109
Warnings:					
Information:					
5	Foreign Reference	WO9806736A1.pdf	2432965 92b4b07651653ae9be5010b10457812249abf973	no	53
Warnings:					
Information:					
6	Foreign Reference	WO9838296A1.pdf	3314452 a362ecf8e9a5e7be4a8b9d234a40a6b871e576d4	no	79
Warnings:					
Information:					
7	Foreign Reference	WO9910540A1.pdf	3705085 cb4aa9eea0957b7c5caa066cb597cffe1587c93e	no	64
Warnings:					
Information:					
8	Foreign Reference	WO9911819A1.pdf	2035634 b7463eb33b027fe97297c9585c3486a756f52055	no	37
Warnings:					
Information:					

9	Foreign Reference	WO9942618A1.pdf	4183347 c1732d1a1033e906fe3b928ceec8f104b7d89f52	no	80
Warnings:					
Information:					
10	Non Patent Literature	Amos-2000-AmJHumGenet.pdf	69058 42318ac339bacb4eddaf7d78be51c022582de5f3	no	4
Warnings:					
Information:					
11	Non Patent Literature	Bellos-2014-NuclAcRes.pdf	1102622 8d2911668cbc8c03eb87834a02715c932070fe93	no	9
Warnings:					
Information:					
12	Non Patent Literature	Benson_DA_et_al_Nucl_Acids_Res_vol_41_D36-D42_2013.pdf	235735 dc570b9c2b3286e0651700b14d9489ea3811ea98	no	10
Warnings:					
Information:					
13	Non Patent Literature	Blomquist-2013-PLoSOne.pdf	2580571 8a047dee4dd84853727f8a73153b4305d6b4ba2d	no	14
Warnings:					
Information:					
14	Non Patent Literature	Bodi-2013-JBiomolTech.pdf	3237469 38826959ed9ceaadb77d282e32b4e01ca8660044	no	14
Warnings:					
Information:					
15	Non Patent Literature	Church-1988-Science.pdf	1714280 8950bc62db9f08b20c1121fe3e53a15732c18dac	no	5
Warnings:					
Information:					

16	Non Patent Literature	Colbert-2001-PlantPhysiol.pdf	1498887 ef78c14820b03a45293e1ee6d4253b0c07d337ca	no	5
Warnings:					
Information:					
17	Non Patent Literature	Collard-2005-Euphytica.pdf	1221586 2f4a0fbdc94fbf20ccfea6aa6914c57a41751fb6	no	28
Warnings:					
Information:					
18	Non Patent Literature	Eminaga-2013-CurrProtMolBio.pdf	955188 5e8e244687a1ad5c377ae98222f92434f157dc74	no	14
Warnings:					
Information:					
19	Non Patent Literature	FakhraiRad-2002-HumMut.pdf	274998 8c04872482a0132ce2724767275ddf29c7018f98	no	7
Warnings:					
Information:					
20	Non Patent Literature	Grothues_D_et_al_Nucl_Acids_Res_vol_21_pp_1321-1322_1993.pdf	1567172 702766bc2a55cb2e7e06c01e6f67ba942cd1bf16	no	2
Warnings:					
Information:					
21	Non Patent Literature	Hajibabaei-2005-PhilTransRoySocB.pdf	224082 ea85fbb5dcf5d2622dc1e1a110c4bb005419c648	no	9
Warnings:					
Information:					
22	Non Patent Literature	34079_550US_-_Illumina-2008.pdf	543458 9535500997a9ecd186617c2d6d1a6b3774c76a39	no	4
Warnings:					
Information:					

23	Non Patent Literature	Anonymous_TruSeq_RNA_and_DNA_Sample_Preparation_Kits_v2_1-15_Illumina_27_April_2011_20.pdf	396471 e9a6ff79d75793cfccd41a782ee12a3881ee9077	no	4
Warnings:					
Information:					
24	Non Patent Literature	IonTotal-2012-LifeTech.pdf	5078664 67745029de161a1ca44c54b497c72e9864e06ed3	no	82
Warnings:					
Information:					
25	Non Patent Literature	Jiang-2015-NuclAcRes.pdf	2149760 45b6533eda47dbff65cae1e3e4dfa1445234efa4	no	12
Warnings:					
Information:					
26	Non Patent Literature	Krumm-2012-GenomeRes.pdf	775190 a241d391263dff286d9b19803e6156c50a387d36	no	8
Warnings:					
Information:					
27	Non Patent Literature	Lai-2004-GenomeRes.pdf	158504 3e5e6d77fce28151a5c22e5ca1a5204175812b12	no	7
Warnings:					
Information:					
28	Non Patent Literature	Langmead_et_al_Ultragast_and_memory-efficient_alignment_of_short_DNA_sequences_to_the_hum.pdf	754226 14a3bf94660ae9fcb957126a3406e234e9817533	no	10
Warnings:					
Information:					
29	Non Patent Literature	Li-2012-Bioinformatics.pdf	4147152 0c930a536b583f62e2705b8ff2cc08cec8d8be45	no	7
Warnings:					
Information:					

30	Non Patent Literature	Lindstrom-2004-JClinMicrob.pdf	1411298 718ccab42a410e698897debca8b92c6e5a4478a0	no	8
Warnings:					
Information:					
31	Non Patent Literature	Liu-2008-BMCGen.pdf	1243308 b83f8548cd6ac0af3fff292202a0d0fbc705c5eb	no	17
Warnings:					
Information:					
32	Non Patent Literature	Ma-2015-CurrProcHumGen.pdf	1030895 139ebc07cb98eb369eef5f3a0a89aec2d53decb6	no	10
Warnings:					
Information:					
33	Non Patent Literature	Machine-translation-CN105890722A.pdf	2764565 e77c93082b243bcad432076c3aca93194b817564	no	4
Warnings:					
Information:					
34	Non Patent Literature	Margulies_et_al_Genorne_sequencing_in_open_microfabricated_high_density_picoliter_reactor.pdf	395179 1351ef06d128965ac99bdc5473b3f0efdac3b0b5	no	6
Warnings:					
Information:					
35	Non Patent Literature	McCloskey-2007-BioGen.pdf	2664759 1fcf5c873fec5a5406673389b3c656038d2a67ed	no	7
Warnings:					
Information:					
36	Non Patent Literature	Nucl_Acids_Res-2004-Miner-e135.pdf	223972 07cb77de071ce9386df47eb89a0e074d44b66841	no	4
Warnings:					
Information:					

37	Non Patent Literature	Myers_R_Protocol_for_Creating_Multiplexed_miRNA_Libraries_for_Use_in_Illumina_Sequencin.pdf	284752 1704091f97a79007c89517dcf356c816087fb1a8	no	15
Warnings:					
Information:					
38	Non Patent Literature	NuGEN-2014-Ovation.pdf	425063 7ac50befa1f8dd3cc8596ea1408fe0e4088f8855	no	45
Warnings:					
Information:					
39	Non Patent Literature	Plagnol-2012-Bioinformatics.pdf	428034 4cece82824c17cb57434501b31c2096608afe27a	no	8
Warnings:					
Information:					
40	Non Patent Literature	Qiu-2003-PlantPhysiol.pdf	99516 dcd99db61500187117a4a13e8eb590078f6c2fbe	no	7
Warnings:					
Information:					
41	Non Patent Literature	Querfurth-2012-Biotechniques.pdf	744473 b8f6f24d8043312fa98c5ccab309ef74d946d6df	no	6
Warnings:					
Information:					
42	Non Patent Literature	Ronaghi-2001-GenomeRes.pdf	589788 c93b6f67500b80d674d7437a56fe7f8698e43d71	no	10
Warnings:					
Information:					
43	Non Patent Literature	Sathirapongsasuti-2011-Bioinformatics.pdf	4739786 49fd685452dc08c201ed4f921334ebe2652bc795	no	7
Warnings:					
Information:					

44	Non Patent Literature	Schiemer-2011.pdf	100206 e552f37667496 4e2e6dc9a94461744ac522e5ca6	no	5
Warnings:					
Information:					
45	Non Patent Literature	Shapero-2001-GenomeRes.pdf	541354 c4e7ed1f8a4405d12a5cc8add9f7752ed66a36ec	no	10
Warnings:					
Information:					
46	Non Patent Literature	Shendure-2005-Science.pdf	820219 872ca46920fbebdd246052b8f15134f935df3b08	no	5
Warnings:					
Information:					
47	Non Patent Literature	Soni_and_Meller_Progress_toward_ultrafast_DNA_sequencing_using_solid-state_nanopores_200.pdf	269236 64e6782c2ffc868e82a4012b203666ae66c96c16	no	6
Warnings:					
Information:					
48	Non Patent Literature	Sood-2006-Methods.pdf	4627937 e364ce5246393c3af2b00a2d367a47f158a90d20	no	8
Warnings:					
Information:					
49	Non Patent Literature	Staroscik-2004-URIGenomics.pdf	88213 5ee963687cab49d6004a4b4e9a0d3604f2375565	no	1
Warnings:					
Information:					
50	Non Patent Literature	Stratagene-1988-Catalog.pdf	1315599 2414bb48bbf486e4979f9b85dd98bf3963ec859b	no	2
Warnings:					
Information:					

51	Non Patent Literature	Till-2003-GenomeRes.pdf	1206368 580f09cf1e64cf5c02d581a254abd82497798a68	no	8
Warnings:					
Information:					
52	Non Patent Literature	Trapnell_at_al_Transcript_assembly_and_quantification_by_RNA-Seq_reveals_unannotated_tran.pdf	298748 f73e96e906a2158b53fbb3bd4f4bd105ddb1fb85	no	13
Warnings:					
Information:					
53	Non Patent Literature	Trapnell_at_al_Differential_analysis_of_gene_regulation_at_transcript_resolution_with_RNA.pdf	994974 64c91effb477f9c3ea9c3cc9ef6d0105975a0e1c	no	9
Warnings:					
Information:					
54	Non Patent Literature	Unemo-2004-JClinMicrob.pdf	481362 628cd67ddd4f9e6aa62f45308284bbc596a2fa4f	no	9
Warnings:					
Information:					
55	Non Patent Literature	34079_550US_-_Vasil-2012.pdf	1773566 9bcf1553f2eecdea85984d8ed0fa6dd2ad23ae77	no	21
Warnings:					
Information:					
56	Non Patent Literature	Vigal-2002-GenSelEvol.pdf	327389 201abc77be6f4ea46e211469f3fff73b35213371	no	31
Warnings:					
Information:					
57	Non Patent Literature	Walker_et_al_Strand_displacement_amplification--an_isothermal_in_vitro_DNA_amplificatio.pdf	1429029 251aaa86ead034ce972c6296872a80a5d47769df	no	6
Warnings:					
Information:					

58	Non Patent Literature	Westin-2000-NatBiotech.pdf	376091 af4f36c0f5e824ab95109021c12bbdf28945e0b2	no	6
Warnings:					
Information:					
59	Non Patent Literature	Wienholds-2004-MethCellBiol.pdf	865416 69e3210ac32328d39be6ee470322a4775fe95e64	no	22
Warnings:					
Information:					
60	Non Patent Literature	Wolford-2000-HumGenet.pdf	178877 01e762f809ac5144dcef3eba21ed4015a76fdf16	no	5
Warnings:					
Information:					
		Total Files Size (in bytes):	88674790		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Espacenet

Bibliographic data: WO9640998 (A1) — 1996-12-19

PCR-BASED cDNA SUBTRACTIVE CLONING METHOD

Inventor(s): WANG XUN; DUVICK JONATHAN P; BRIGGS STEVEN P ± (WANG, XUN, ; DUVICK, JONATHAN, P, ; BRIGGS, STEVEN, P)

Applicant(s): PIONEER HI BRED INT [US] ± (PIONEER HI-BRED INTERNATIONAL, INC)

Classification: **- international:** ***C07K14/415; C12N15/10; C12Q1/68;*** (IPC1-7): C12Q1/68
- cooperative: **C07K14/415**; **C12N15/1096**; **C12Q1/6809** more

Application number: WO1996US08582 19960605 Global Dossier

Priority number (s): US19950481687 19950607

Also published as: AU6251996 (A) AU700952 (B2) CA2223104 (A1) EP0832294 (A1) US5837468 (A) more

Abstract of WO9640998 (A1)

A process is described in which PCR-based cDNA libraries with anchored ends are made, a PCR--based cDNA subtracted (PCSUB) library is produced therefrom, and the ends of a cDNA clone are isolated, via PCR, from an anchored-ends library.

IDT Exhibit 1020.0497

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68

A1

(11) International Publication Number: **WO 96/40998**

(43) International Publication Date: 19 December 1996 (19.12.96)

(21) International Application Number: PCT/US96/08582

(22) International Filing Date: 5 June 1996 (05.06.96)

(30) Priority Data:
08/481,687 7 June 1995 (07.06.95) US

(71) Applicant: PIONEER HI-BRED INTERNATIONAL, INC. [US/US]; 700 Capital Square, 400 Locust Street, Des Moines, IA 50309 (US).

(72) Inventors: WANG, Xun; Apartment L119, 7800 Iltis Drive, Urbandale, IA 50322 (US). DUVICK, Jonathan, P.; 1707 38th Street, Des Moines, IA 50310 (US). BRIGGS, Steven, P.; 2131 Willowmere Drive, Des Moines, IA 50321 (US).

(74) Agents: BENT, Stephen, A. et al.; Foley & Lardner, Suite 500, 3000 K Street, N.W., Washington, DC 20007-5109 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: PCR-BASED cDNA SUBTRACTIVE CLONING METHOD

(57) Abstract

A process is described in which PCR-based cDNA libraries with anchored ends are made, a PCR--based cDNA subtracted (PCSUB) library is produced therefrom, and the ends of a cDNA clone are isolated, via PCR, from an anchored-ends library.

IDT Exhibit 1020.0498

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

IDT Exhibit 1020.0499

PCR-Based cDNA Subtractive Cloning Method

BACKGROUND OF THE INVENTION

Polymerase chain reaction (PCR) technology is employed in a growing variety of ways, including preparation of cDNA's and constructing cDNA libraries. An early use of PCR to generate a cDNA library was reported by Belyavsky *et al.*, *Nucleic Acids Res.* 17: 2919-32 (1989).

The Belyavsky method utilized oligo (dT) as a primer for reverse transcriptase reaction, followed by poly (dG) tailing via the action of terminal deoxynucleotidyl transferase (TdT). The resulting dG-tailed cDNAs were subsequently amplified with poly (dT) and poly (dC) primers. The cDNA pool thus obtained was cloned into a vector for subsequent cDNA screening.

Since an oligo (dT) primer can anneal at any position of the poly(A) tail of a (+) strand of cDNA, and an oligo (dC) primer can anneal at any position of the poly(G) tail of a (-) strand of cDNA, the amplified cDNAs generated by the Belyavsky method often have varying lengths. Accordingly, these products cannot be analyzed directly, and instead require subcloning and screening of a cDNA library, a time-consuming technique. Furthermore, the use of primers containing homopolymers on the 3' end typically yields a high background of non-specific product.

A technique for rapid amplification of cDNA ends (RACE) was described by M.A. Frohman and his colleagues. See Frohman *et al.*, *Proc. Nat'l Acad. Sci. USA* 85: 8998-9002 (1988), and Frohman, PCR PROTOCOLS, A GUIDE TO METHODS AND APPLICATIONS 28-38 (Academic Press 1990). The RACE protocol produces specific cDNAs by using PCR to amplify the region between a single point on a transcript and the 3' or the 5' ends. One requires knowledge of the sequence of an internal portion of the transcript, however, in order to design a primer for use in conjunction with either the polyT or polyG primers to

amplify the ends. This protocol therefore yields specific cDNAs products only, not whole libraries.

A modification to the RACE protocol introduced by Borson *et al.*, *PCR Methods and Applications* 2: 144-48 (1992), entails the use of a "lock-docking oligo (dT)." The locking mechanism involves extending the poly dT primer, by either one nucleotide (A, C or G) or by two nucleotides (also A, C or G) and yet one more of the four possible nucleotides, at the 3'-end of the primer. This "locks" the primer to the beginning of the poly dT tail, either the natural dT or a poly dT tail attached to the first strand cDNA 3'-end, by use of TdT, resulting in the synthesis of cDNA's of discrete lengths. Subcloning and screening of subclone library is not necessary before analysis, which can speed up the inquiry. Like the RACE protocol, however, Borson's protocol uses a gene-specific internal primer and, hence, produces only specific cDNAs, not whole libraries.

Approaches are described in the literature to identify mRNA expressed differentially, either in only some cell types, or at certain times of a biological process, or during infection by a parasite or a virus, etc. Those studies generally employ subtractive hybridization to reveal the differentially expressed mRNA(s). Liang and colleagues have used the anchored-end technique to look for specific differences in mRNA populations. Liang *et al.*, *Nucleic Acids Res.* 21: 3269-75 (1993). The Liang method, called "differential display," employs a decanucleotide of arbitrary sequence as a primer for PCR, internal to the mRNA, and a polyTMN primer on the 3'-end of mRNAs; "M" in this context is randomly G,C or A, but N is chosen as one of the four possible nucleotides. When such sets of primers are employed, patterns of mRNAs can be visualized, upon polyacrylamide gel electrophoresis of the PCR product, and the comparison of such patterns produced by mRNAs from two sources reveal the differentially expressed mRNAs.

The differential display method can identify individual, differently expressed mRNA's, but cannot constitute a complete library of such mRNA's. As a further consequence of having one primer of an arbitrary sequence, and therefore probably not having an exact match, low copy number mRNAs may not be picked up by this method. Finally, the cDNA candidates identified would still require recovery from the gel and subcloning, if the individual cDNA is desired for further analysis.

Lisitsyn *et al.*, *Science* 259: 946-51 (1993), have described a representational differences analysis (RDA) method which uses subtractive hybridization and PCR technology to define the differences between two genomes. Like other subtractive hybridization protocols, in RDA there are defined two sets of DNAs, the "tester" DNA and the "driver" DNA. According to the RDA protocol, the DNA of the two genomes to be compared are digested by restriction endonucleases, and a dephoshorylated double-stranded oligonucleotide adapter is ligated. After denaturation and hybridization of driver and tester DNA, oligonucleotides from the adaptors covalently linked to tester DNA were used to amplify unique DNA sequences of tester library. The adapters are partially double-stranded DNAs made by partially complementary oligos, where the single-stranded sequence at one end of the double stranded adapter is complementary to the single-strand tail of the digested genomic DNA. The combined use of (i) restriction enzyme- digested DNA as PCR substrate and (ii) the preferential amplification of shorter substrates results in a population of fairly short, amplified DNA molecules. The adapters then are removed by cleavage with the restriction enzymes used originally to digest the DNA. To the tester DNA, new adapters with novel sequences are ligated, the tester and driver DNA are mixed, the DNA strands are separated by heating ("melting"), and the DNA's are cooled to allow for reannealing. PCR is performed with primers complementary to the adapters on tester DNA, thereby

amplifying only target DNA, *i.e.*, only DNA unique to the tester DNA. By restriction enzyme digestion of the adapters from the amplified DNA and ligation of additional, novel adapters, followed by PCR, the target DNA is amplified to become the dominant fraction.

The RDA procedure does not use any physical method of separation between the tester and driver DNA which, if used, would allow enhanced purification of target DNA. The method is used only to identify differences between genomes and was not used to identify differential cDNA expression.

In view of the limitations of the RACE and RDA methodologies, it would be very useful to have one method to provide full-length, anchored-ended cDNA libraries, for creation of differentially expressed cDNA libraries, and to use PCR to screen such libraries for the ends of specific mRNAs.

SUMMARY OF THE INVENTION

It therefore is an object of the present invention to provide a PCR-based method for generating a full-length cDNA library with anchored ends. The method would use lock-docking oligos as PCR primers, one primer, polyTV (SEQ ID NOS:13-15) locking over the polyA tail of eukaryotic mRNA and producing the first strand synthesis, and a second primer, polyGH (SEQ ID NOS:16-18) that would lock onto polyC tail added by TdT to the newly synthesized strand. This would contrast with the methods of Belyevsky and of Borson in that (a) discrete sized PCR products would result which would not necessarily require further subcloning/screening, (b) full-length cDNA's would be produced and (c) cDNA libraries would be produced as opposed to specific cDNA clones.

It is a further object of the present invention to generate a PCR-based cDNA subtractive (PCSUB) library. This would be accomplished by generating in the first place two cDNA libraries with anchored ends, one of tester DNA and one of driver DNA. The two libraries would undergo subtractive hybridization and

amplification, to some extent similar to the RDA method, but with significant differences. In addition to using dephoshorylated adaptors which prevents amplification of driver DNA, we prepared a biotin-tagged driver library by use of biotin labeled dCTP during PCR. This would allow for a physical separation (using streptavidin-coated beads) of driver and of driver/tester hybrid cDNA from the desired and amplified target cDNA, thus enhanced relative amplification of target cDNA. Having a way to remove the driver cDNA also allows for use of a higher ratio of driver/tester cDNA, and therefore more stringent subtraction of cDNA sequences which are not unique to the target cDNA. More importantly, the PCSUB method, unlike RDA, would result in a library representing differentially expressed mRNAs.

It is yet a further object of the present invention to utilize PCR and sequence information derived from cDNA clones from the PCSUB library in order to screen the cDNA anchored end library for the ends of specific cDNA sequences. This approach would employ primers that are complementary to internal sequences, in conjunction with polyTV (SEQ ID NOS:13-15) or polyGH (SEQ ID NOS:16-18) or equivalent primers which comprise restriction enzyme recognition sequences at their respective 5'-ends, to "fish out" from the library the ends of specific mRNA's.

In accomplishing these and other objectives, there has been provided, in accordance with one aspect of the present invention, a method based on PCR for generating a cDNA library with anchored ends, comprising the steps of

(A) providing an RNA preparation that comprises polyadenylated mRNA from a biological sample, the polyadenylated mRNA having a polyA portion and a non-polyA portion;

(B) preparing an oligonucleotide polyTV primer that anneals to the polyA portion of the polyadenylated mRNA, anchoring the polyTV primer such that a reverse transcriptase

reaction can start within one nucleotide from the 3′ end of the non-polyA portion;

(C) using the polyTV primer to generate by PCR from the RNA preparation a DNA strand that is complementary to the polyadenylated mRNA;

(D) using terminal deoxynucleotidyl transferase to add a polynucleotide tail at the 3′ end of the DNA strand, whereby the DNA strand has a first portion that is complementary to the polyadenylated mRNA and a tail portion;

(E) preparing an oligonucleotide polyGH primer that anneals to the polynucleotide tail attached in step (D), anchoring the polyGH primer such that a reverse transcriptase reaction can start one nucleotide downstream from the 5′ end of the first portion of the DNA strand;

and then

(F) using the polyGH primer and the polyTV primer to amplify the DNA strand by PCR.

In a preferred embodiment, the polynucleotide tail in step (D) is a polyC tail.

In accordance with a second aspect of the present invention, a method is provided for generating a cDNA library with anchored ends, as described above, wherein each of the polyGH primer and the polyTV primer contains a recognition site for a restriction endonuclease.

According to yet a third aspect of the present invention, there has been provided a method of comparing two cDNA libraries to identify cDNA that is unique to one of the libraries, comprising the steps of

(A) preparing a first cDNA library and a second cDNA library with anchored ends, wherein the first cDNA library contains the unique cDNA and is prepared according to the aforementioned second aspect of the invention, and wherein the second cDNA library is similarly produced and incorporates biotin-labelled deoxynucleotides, and

(B) digesting the first cDNA library with a restriction enzyme for which a recognition sequence is found on the polyGH and polyTV primers, to produce, at the ends of cDNAs in the first library, single-stranded DNA tails,

(C) preparing multiple sets of double-stranded adapter DNA molecules, each set comprising a first and a second oligonucleotide of such sequence that:

(1) the 3′ end of the first oligonucleotide complements the 5′ end of the second oligonucleotide of the same set

and

(2) the 5′ end of the first oligonucleotide complements the single-strand DNA produced by a restriction enzyme digest at the anchored ends of the first DNA library,

(D) manipulating the libraries by mixing them; subjecting them to DNA-melting conditions; allowing for reannealing of DNA strands in the libraries; purifying the unique cDNA by trapping on streptavidin beads DNA that incorporates the biotin-labeled deoxynucleotides, whereby a cDNA fraction enriched for the unique cDNA is obtained; producing blunt-ended DNA by filling-in the unique cDNA ends with Klenow enzyme reactions; and subject the cDNA fraction to PCR amplification of the unique DNA via oligonucleotide primers that comprise an adapter molecule used in step (D),

(E) digesting the unique cDNA with the restriction enzyme and ligating an adapter set from the multiple sets to the anchored ends of the first library,

and then

(F) repeating steps (D) and (E), each time ligating a new set of adapter DNA molecules from the multiple sets to the cDNA fraction, until the unique cDNA is essentially free of non-unique cDNA from the first library.

Still another aspect of the present invention comprehends a method of isolating the cDNA ends of a unique cDNA from an anchored library produced pursuant to to either the second or the third aspects detailed above. This method comprises the steps of (A) providing a set of PCR primers that hybridize, respectively, to a sequence internal to the unique cDNA and (B) using the primers with the polyTV or the polyGH primers in PCR reactions to produce two ends of the cDNA, wherein the primers contain a restriction enzyme cleavage site.

Other objects, features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and the specific examples, while indicating embodiments of the invention, are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from this detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 depicts a general scheme using PCR to generate a cDNA library with anchored ends. As depicted, a polyTV primer anchors on the beginning of the polyA tail of the mRNA. After synthesis of the first strand (SEQ ID NOS:13-15, respectively) by reverse transcription, a polyC tail is added to the 3'-end, by use of TdT. In the last step, PCR is performed employing polyGH and polyTV to amplify the first strand and produce a cDNA library with anchored ends (SEQ ID NOS:16-18, 21-23, 24-26 and 13-15, respectively). The polyGH and polyTV primers are used here as illustrative primers.

Figure 2 details the making of a PCSUB library. The first few steps shown entail the making of two libraries, employing "tester" RNA and "driver" RNA as substrates. The libraries are constructed similarly to the description of Figure 1. But the restriction site implicit in the 5'-end sequence of the polyTV and polyGH primers of Figure 1 are indicated here as a BamHI site;

the primers therefore are denoted "BamT17V" and "BamG13H," respectively. BamH1 is illustrative only, and other restriction enzyme recognition sequence(s) are possible. Also, an important addition to the scheme of Figure 1 is the use of biotin-dCTP to "tag" the PCR product. In the next step the tester cDNA library is digested with the restriction enzyme(s) for which recognition sites are built in the 5'-end of the anchor, here BamHI. An adaptor DNA then is ligated to the tester cDNA library, followed by "substraction" of target cDNA from the two libraries. This is accomplished by mixing the DNA of the tester and driver libraries, the melting of the DNA molecules into simple stranded DNA, preferably by heating, and the reannealing of complementary strands by a process of slow cooling. Finally, the biotin-labeled cDNA, both from the driver library and from the tester library which "found" a complementary strand to driver library cDNA, are removed by passing through and trapping on a slurry of streptavidin beads. The ends of the DNA are made double-stranded, preferably by Taq polymerase reaction, and the mixture enhanced for target cDNA is amplified using primer(s) complementary to the adapter molecule(s). The process of producing BamHI ends, ligating adapter sets, subtracting the target cDNA from a mixture of tester and driver cDNAs, and amplifying by PCR for the target cDNA is repeated as many times as deemed necessary to get essentially pure target cDNA. Since the sequence of the adapter sets used each time is different, target cDNA is preferentially amplified each time away from remaining process of tester cDNA.

Figure 3 portrays the isolation of cDNA ends from an anchored library (ICEFAL). The preparation of a cDNA anchored library is done as illustrated in Figure 1. Next, gene-specific PCR primer 1, GSP1 (SEQ ID NO:19), is used together with primer BamG13H to produce a clone of the 5'-end of the desired gene, and gene-specific primer 2, GSP2 (SEQ ID NO: 20), is used together with primer BamT17V to produce the 3'-end of the clone.

Figure 4 shows the preparation of tester RNA and of driver RNA for the PCSUB described in Example 1. Here the target cDNA are derived from plant transcripts induced by infection with *Fusarium moniliforme*. Therefore the tester RNA is extracted from *F. monifilorme* infected embryos and driver RNA extracted from embryo RNA + *F. monifilorme* RNA.

Figure 5 illustrates the effects of $Mg+^2$ concentration on PCR under conditions employed to amplify the tester (+) and driver (-) cDNA libraries. Top panel shows effect on total cDNA production, as visualized on EtBr stained agarose gel. Mid-panel shows a Southern analysis, using as probe MPI, a gene known to be induced by Fusarium infection. See Cordero *et al.*, *The Plant J.* 6: 141-50 (1994). Bottom panel shows similar analysis with ACT, an actin gene which is not expected to be induced by infection.

Figure 6 indicates the effect of each of two rounds of subtractive hybridization between tester and driver cDNA libraries. The left panel is probed with maize protease inhibitor, MPI, a gene expected to be induced by Fusarium infection. The right panel is probed by actin, ACT, a gene expected to be expressed regardless of the infection process.

Figure 7 depicts Northern Analysis of RNA from corn embryo, either induced by infection with the fungus *F. moniliforme* (+) or from uninfected embryos (-). Three clones, C-11-3, G-12-3 and G-4-5 of Example 1 (SEQ ID NO:27, 29 and 39, respectively, were fished out from the PCSUB library and were used as probes.

Figure 8 lists the nucleic acid sequence of C-11-3 (SEQ ID NO:27), one of the *F. moniliforme* induced, plant cDNA clones from the PCSUB. An observed open reading frame (SEQ ID NO:28) is indicated.

Figure 9 lists the nucleic acid sequence of G-4-5 (SEQ ID NO:29), one of the *F. moniliforme* induced, plant cDNA clones from the PCSUB library. The longest observed reading frame is underlined (SEQ ID NO:31).

Figure 10 compares the amino acid sequence deduced from the nucleic acid sequence of G-12-3 (SEQ ID NO:39), a cDNA clone from the PCSUB library, with P450 proteins from other sources (SEQ ID NOS:33-38, respectively). The organism of origin for the other P450 proteins is indicated in the bottom panel.

Figure 11 illustrates the isolation of the 5'-end of cDNA clone G-12-3 (SEQ ID NO:39) by using the ICEFAL technique. Panel A depicts schematically how primers GSP1 (SEQ ID NO:19) and BamG13H (SEQ ID NOS:16-18) are used on a cDNA library with anchored ends to amplify the 5'-end of a clone. Panel 2 shows a Southern Blot of the PCR products using the primer pairs indicated. The gel was 1.2% agarose. The probe was clone G-12-3 (SEQ ID NO:39).

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention provides a PCR-based method of creating a full-length cDNA library with anchored ends. The invention assumes that "good quality" mRNA is obtained, either as a polyadenylated fraction, or as total cellular RNA. An RNA fraction which is further enriched for mRNA containing polyA at its 3'-end is preferred, since it provides more substrate for the PCR reaction. The phrase "good quality" RNA denotes full-length, non-degraded RNA. A variety of methods for obtaining RNA, and methods to assess its quality, are known to those versed in the art, and some of those methods are described hereinafter. In accordance with the present invention, a primer for the first-strand cDNA synthesis takes advantage of the polyA tract located at the end of eukaryotic mRNA; hence, the primer consists in part of a polyT chain. But the polyT primer can typically hybridize at any point on the polyA tail of the mRNA. Primer extension by a reverse transcriptase would thus create a first strand of varying lengths, and lead eventually to a situation where one has to screen through various versions of what is basically the same cDNA, to isolate the cDNA of interest.

To eliminate this problem, the primer contains one non-A nucleotide (C, G or T) at its 3'-end. Such a primer would "lock" at the 3'-end of a mRNA, since the C, G or T nucleotide would need to hybridize to a nucleotide other than the adenine nucleotides of the polyA tail. According to the present invention, the primer described above is called "polyTV (SEQ ID NOS:13-15)" where "V" denotes for A, C or G.

The first strand is synthesized by extension of the polyTV primer, by addition of deoxynucleotides, by a reverse transcriptase enzyme. The first strand would be purified by standard methods and a polyC tail would be added to the 3'-end of the cDNA by TdT. A polyGH (SEQ ID NOS:16-18) primer is synthesized which, by analogy to the discussion of the polyTV primer, would lock onto the 5'-end of the cDNA. Here H stands for A,C or T. PCR reactions using the polyTV and polyGH primers would amplify the cDNA library, which then, either could be used directly for further experiments, as described hereinafter, or could be cloned into a plasmid vector. This scenario is illustrated in Figure 1.

To a person skilled in the art, many variations on this theme are readily apparent. By way of example, but not limited to said examples, the polyTV and polyGH primers could have restriction enzyme recognition site(s) built in near the 5'-end. The restriction enzyme recognition site(s) can be the same, or different to allow for unidirectional cloning. The length of the T or G tracts can be varied. TdT can add to the 3'-end of the first strand a nucleotide other than cytosine, with corresponding complementary changes in the sequence of the primer. The PCR conditions can be optimized for the specific primers used. The polymerase enzyme used can be one of a number of polymerases used for PCR. The polymerase can incorporate labelled or modified nucleotides. In all events, a cDNA library thus produced contains full-length cDNAs, anchored at both ends by

known sequences, herein referred to as an "anchored-end cDNA library."

Figure 1 illustrates one embodiment of the invention. Here the polyTV primer has a BamHI site near the 5'-end. The polyT tract is 17 nucleotides long. This specific primer is referred to herein elsewhere as BamT17V (SEQ ID NOS:13-15). The polyGH primer also has a BamHI site near its 5'-end and has a stretch of 13 guanines. The primer is called "BamG13H" (SEQ ID NOS:16-18) elsewhere in this description.

The present invention also provides for a way to make a PCR-based cDNA subtractive (PCSUB) library. The principle is that two pools of RNA serve as substrates for separate cDNA libraries. One pool, the tester RNA, contains some mRNA(s) which are differentially expressed when compared with the mRNA of the other pool, the driver RNA. Two separate cDNA libraries are made; a tester cDNA library and a driver cDNA library. The driver cDNA library serves the purpose of subtracting from the tester cDNA library all the cDNAs which they have in common. This is accomplished by mixing aliquots of the two libraries, with the driver cDNA in large excess, say 100x, allowing for the separation of the DNA strands and reannealing, usually by heating followed by slow cooling. cDNA common to both libraries will form hybrid double stranded molecules, and, if the driver cDNA was tagged in such a manner as to allow its removal, hybrid DNA molecules also will be removed. What cDNA remains is highly enriched for target cDNA, *i.e.*, the product of the differentially expressed RNAs. The remaining cDNA is amplified by PCR. In practice, this is best accomplished if some short stretch of double-stranded DNA molecule, called an "adapter" elsewhere in this description, is attached only to the tester cDNA molecules that remain after the subtractive hybridization step described above. One possible way to accomplish this would require that the tester cDNA library only, prior to the mixing of the libraries for the subtractive hybridization step, is

digested with the restriction enzyme that has a recognition site built in on the 5'-end of the polyTV and polyGH tails. The restriction digest, by virtue of an appropriate choice of restriction enzyme/restriction site design, will produce a cDNA with single-stranded tails.

The adapter molecule comprises two partially complementary oligonucleotide sequences, so that a single-stranded tail protrudes that is complementary to the single-stranded DNA on the ends of the tester cDNA. The adapter is ligated to the tester cDNA, usually by T4DNA ligase. One of the oligos which comprised the adapter then serves as a primer for PCR, thus amplifying the tester DNA only. For an efficient removal of tester cDNA which is not the target cDNA, multiple rounds of subtractive hybridization followed by amplification of remaining tester cDNA may be required. For this purpose the aforementioned cDNA which is enriched for target DNA again is cleaved with a restriction enzyme for which there is a recognition site on the polyTV and polyGH primers. Subtractive hybridization relative to excess driver cDNA then is carried out, using an adapter molecule of a sequence that is different from that of the adapter(s) previously employed.

Three such adapter sets are listed in the following table (SEQ ID NOS:1-12, respectively), described by Lisitsyn (Lisitsyn *et al.*, *Science* 259: 946-51 (1993). The listed adapters are a subset of such adapters mentioned by Lisitsyn *et al.*, *supra*.

Table - Adapter Sets

adapters	name	sequence
1st round	R Bam 24	5'-AGCACTCTCCAGCCTCTCACCGAG-3'
	R Bam 12	5'-GATCCTCGGTGA-3'
	after pairing	5'-AGCACTCTCCAGCCTCTCACCGAG-3' 3'-AGTGGCTCCTAG-5'
2nd round	J Bam 24	5'-ACCGACGTCGACTATCCATGAACG-3'
	J Bam 12	5'-GATCCGTTCATG-3'
	after pairing	5'-ACCGACGTCGACTATCCATGAACG-3' 3'-GTACTTGCCTAG-5'
3rd round	N Bam 24	5'-AGGCAACTGTGCTATCCGAGGGAG-3'
	N Bam 12	5'-GATCCTCCCTCG-3'
	after pairing	5'-AGGCAACTGTGCTATCCGAGGGAG-3' 3'-GCTCCCTCCTAG-5'

The qualities of the initial tester and driver anchored ends library, and the degree of enrichment achieved after each round of subtractive hybridization/amplification, should be checked by any of the various techniques conventionally employed for this purpose. One such approach uses two cDNA clones, one for a gene known to be differentially expressed and the other for a known gene that would be represented in both the tester and driver RNA pools and cDNA libraries. These genes could be used to probe Northern blots and/or dot blots of the starting mRNAs, the cDNA libraries, and of the material after rounds of subtractive hybridization/amplification to determine (a) the quality of starting RNA (expect unique, full-length bands), (b) the quality of the cDNA libraries, and (c) that the differentially expressed mRNA is present in larger amounts after subtractive hybridization but (d) the non-differentially expressed gene is under-represented after subtractive hybridization/amplification.

There are variations, readily apparent to one skilled in the art, to the techniques described above for the making of the PCSUB library. For example, the primers used to anchor the 5'-end of the tester cDNA library need not be the same as the respective primers for the driver cDNA library. This would eliminate the need to place adapters on the material after the first round of subtractive hybridization/amplification. The polyTV and polyGH, or equivalent primers, used to make the tester and driver libraries may have multiple cloning sites (MCS) near their 5'-ends. The tester library probably contains internal to the sequence of some of its members the restriction site used for adapter molecule attachment to the ends of the cDNA. Therefore, any one isolated clone may not be full-length. If the initial tester library had an MCS at its ends, one could use the library in order to fish out the ends of the desired clone.

Other variations in PCR technique, in the choice of polymerase enzyme employed, in the methods applied to clean up the PCR product, and in the method used to remove the biotin tagged cDNA at the end of the subtractive hybridization step, *inter alia*, also are within the scope of the present invention. An embodiment of the preparation of a PCSUB library is illustrated in Figure 2 and used in experiments described hereinafter. According to this embodiment, please refer to Figure 2, the tester and driver cDNA libraries are made with primers BamT17V and BamG13H (SEQ ID NOS:13-15 and 16-18, respectively). The driver cDNA library is tagged by using biotin-dCTP. After the subtractive hybridization step, the biotin labeled cDNA molecules, now a mixture of driver cDNA library and of hybrid tester/driver derived cDNA molecules, are removed. The removal of the biotin labeled DNA is accomplished by trapping the biotin labeled DNA molecules on streptavidin-paramagnetic particles which next are captured on a magnetic strand (Promega Corporation, Madison WI). According to the embodiment, adapter sets are used to facilitate selective

amplification as described above. The specific sets employed are those enumerated in the aforementioned table, used in the order which they appear there (the first set first, etc).

It is another object of the present invention to present a method of isolation of cDNA ends from an anchored library (ICEFAL). The anchored ends cDNA library is prepared as described above and recapitulated in Figure 3. The isolation of the cDNA ends requires knowledge of the nucleic acid sequence over some internal position of the cDNA desired. Such knowledge can come from any source, not limited to the following examples: knowing the sequences of a peptide fragment, guessing the sequence by analogy with a well-preserved section of the equivalent gene from another organism, or, in line with our claims, from having isolated and analyzed clones from our PCSUB library. According to the present invention, two primers are made based on the known internal sequence, each complementary to a different strand of the cDNA. The primer that can be extended to the 5'-end of the cDNA is called "gene-specific primer 1" (GSP1) (SEQ ID NO:19), and is used in conjunction with polyGH (BamG13H (SEQ ID NOS:16-18) in Figure 3) to amplify the 5'-end of the clone. The primer that can be extended to the 3'-end of the cDNA clone is called gene-specific primer 2 (GSP2) (SEQ ID NO:20), and in conjunction with primer polyTV (Bam17TV (SEQ ID NOS:13-15, respectively) in Figure 3), is used to amplify the 3'end. All primers are designed with restriction site(s) near the 5'-end, which allows for subsequent cloning of the cDNA ends (see Figure 3).

The following commentary illustrates the present invention by reference to a series of experiments. The goal of the experiments was to isolate and then analyze plant embryo genes that were induced by infection with the fungus *Fusarium moniliforme*.

Fungal infection and total RNA preparation

Maize seeds (Pioneer Hi-Bred Int'l inbred line HT1) were used as the experimental material. *Fusarium moniliforme* isolate MO33 was obtained from moldy corn ears at Johnston, Iowa.

Fungal infection of germinating embryos was performed as reported by Casacuberta *et al.*, *Plant Molec. Biol.* 16: 527-36 (1991). Seeds were washed with sterile water for 3 minutes followed by a wash with full strength commercial bleach (5.25% sodium hypochloride) at room temperature for 10 minutes. Then the seeds were washed three times with sterile water. After the seeds were imbibed for 4 hours, embryos were dissected from these seeds under sterile conditions. Dissected embryos were germinated on wet filter paper in the dark at 28 °C for 24 hours and then inoculated with the conidial suspension of *Fusarium moniliforme* by adding 50 μl (approximately 1000-3000 spores/ml) to each embryo. Inoculated seeds and sterile control seeds were allowed to continue germination for two more days in the dark at 28 °C.

Figure 4 depicts the preparation of tester and driver RNA pools, some infected embryos (tester RNA) or a mixture of infected embryos and fungus. Total RNA was isolated from infected and non-infected geminating embryos using TriReagent (Molecular Research Center, Inc. Cincinnati, OH).

Preparation of cDNA pools with anchored ends

For first strand cDNA synthesis 0.5 μg poly (A)+ RNA was combined with 2 pmoles of BamT17V (mixture of three oligo nucleotides of CGCGGATCCGTTTTTTTTTTTTTTTTA, CGCGGATCCGTTTTTTTTTTTTTTTG and CGCGGATCCGTTTTTTTTTTTTTTTC (SEQ ID NOS:13-15, respectively) at equal molar ratio), and DEPC-treated water in a final volume of 11 μl. Mixture was heated at 70 °C for 10 minutes and then chilled on ice for one minute. After addition of 4 μl of 5X first strand cDNA synthesis buffer (Gibco BRL, Gaithersburg, MD), 1 μl of 10 mM dNTP, 2 μl of 0.1 M DTT and 1 μl of

placental RNase inhibitor (Promega Corporation, Madison, WI), the mixture was incubated at 42 °C for 2 minutes prior to the addition of 1 μl of SuperScript (Gibco BRL, Gaithersburg, MD). The reaction mixture was further incubated at 42 °C for 30 minutes. After the reaction 2 units of *E. coli* RNase H were added and the mixture was equilibrated at 55 °C for 10 minutes. Primers, unincorporated dNTPs, salts and proteins were removed from first strand cDNAs using GlassMAX™ Spin Cartridges (Gibco BRL) according to the manufacturer's suggestions except the final wash was performed with 400 μl of cold 80% ethanol. First strand of cDNAs were eluted with 150 μl of water (HPLC grade).

The first strand of cDNA then was tailed with oligo-dC using TdT (Gibco BRL). The reaction mixture contains 36 μl of first strand cDNA, 10 μl of 5X TdT reaction buffer (Gibco BRL), 2 μl of 10 mM dCTP, 2 μl of TdT (10 units/μl). The mixture was incubated at 37 °C for 20 minutes followed by a incubation at 70 °C for 10 minutes. dC-tailed first strand cDNAs (dC-cDNAs) were purified with GlassMAX™ Spin Cartridges system as described above and eluted with 100 μl water (HPLC grade).

dC-cDNAs were amplified using Taq polymerase with primers BamT17V and BamG13H, where primer BamG13H is a mixture of three oligo nucleotides of TAAGGATCCTGGGGGGGGGGGGA, TAAGGATCCTGGGGGGGGGGGGGT and TAAGGATCCTGGGGGGGGGGGGGC (SEQ ID NOS:16-18, respectively) at equal molar ratio. The reaction contained 2 μl of dC-cDNA in a final volume of 50 μl 1X amplification buffer [20 mM Tris-HCl (pH 7.3), 50 mM KCl, 3.0 mM $MgCl^2$, 1 μM of each primer, 0.4 mM dNTP]. The mixture was overlaid with 60 μl light mineral oil and placed into a 96-well format MJ Thermal Cycler. The reaction mixture was heated at 94 °C for 4 minutes then the temperature was held at 72 °C during addition of 1 unit *Taq* polymerase (Boehringer Mannheim Corporation). PCR conditions were as follows: 25 cycles

of 94 °C for 30 seconds, 60 °C for 30 seconds, and 72 °C for 5 minutes. After the final cycle the mixture was further incubated at 72 °C for additional 10 minutes. The amplified "anchored" cDNAs (SEQ ID NOS:16-18, 21-23, 24-26 and 13-15, respectively) (Figure 1) were purified with GeneClean (Bio 101, Vista, CA).

Subtractive cloning of cDNAs from genes that are induced upon infection by *Fusarium moniliforme* in germinating embryos of maize

Tester RNA were isolated from 72 *Fusarium* infected geminating embryos. To prepare driver RNA, 36 embryos were homogenized in liquid nitrogen and inoculated with *Fusarium moniliforme*. The resulting materials were combined with 72 non-infected geminating embryos to generate a driver RNA pool. For driver dC-cDNA amplification, a final 16 μM biotin-14-dCTP (Gibco BRL, Gaithersburg, MD) was included to label driver cDNAs. Both driver and tester cDNAs were amplified with BamT17V and BamG13H primers as described in the previous section.

Subtraction was modified from Lisitsyn *et al.*, *supra*. For each cycle of subtraction, 4-10 μg of tester cDNAs or the subtractive cDNAs (from previous cycle of subtraction) were digested with BamH I (100-200 units) at 37 °C for four hours. Then the proteins were removed by phenol and phenol/chloroform extraction. DNA was precipitated with ethanol. 2 μg of the resulting DNA was ligated to an adapter in a final volume of 60 μl: 16.7 mM 24-mer oligo, 16.7 mM 12-mer oligo 1X T4 DNA ligase buffer (New England BioLabs, Inc.). The mixture was heated at 55 oC for 10 minutes and then slowly cooled down to 10 °C (about 1 hour). After addition of 800 units T4 DNA ligase (New England BioLabs, Inc.) the reaction mixture was incubated at 16 °C for 20 hours. The three pairs of adapters used in successive rounds of subtraction were as described in the table above.

For tester cDNA and driver cDNA hybridization 0.4 μg of the resulting tester cDNA was mixed with 40 μg of biotin labeled driver cDNA and the DNA mixture was

precipitated with ethanol. After centrifugation DNA pellet was washed twice with 70% ethanol and air dried. The pellet was resuspended in 4 μl of EE x 3 buffer (30 mM EPPS from Sigma, pH 8.0 at 20 °C, 3 mM EDTA) by vortexing for 2-5 minutes. The mixture was overlaid with 35 μl mineral oil and then heated at 98 °C for 4 minutes during addition of 1 μl 5M NaCl, followed by a 67 °C incubation for 20-40 hours. After the hybridization 500 μl of water (HPLC grade) and 150 μl chloroform were added to the DNA drop. The aqueous phase was transferred to a new tube.

To remove biotin labeled-DNA and the DNAs associated with them, 3 mg Streptavidin-Paramagnetic Particles in 0.5 ml 1 X SSC (Promega Corporation) was added to the driver-tester cDNA hybridization solution and mixture was incubated at room temperature for 20 minutes. Then the Streptavidin-Paramagnetic Particles were captured with a Magnetic Stand (Promega Corporation). After a brief centrifugation the DNA in the supernatant (subtractive cDNA) was isolated with GeneClean system (Bio 101) in a final volume of 150 μl.

Two microliters of the subtractive cDNAs were filled in and amplified in a final volume of 50 μl [20 mM Tris-HCl (pH 7.3), 50 mM KCl, 3.0 mM $MgCl^2$, 1 μM of the corresponding 24-mer primer (R Bam24, J Bam 24, or N Bam 24), 0.4 mM dNTP]. The mixture was mixed with 2 units of Taq polymerase at 72 °C and incubated further at 72 °C for 20 minutes. The resulting DNAs were amplified at following condition: 23 cycles of 94 °C for 30 seconds, 65 °C for 30 seconds, and 72 °C for 5 minutes. After the final cycle the mixture was further incubated at 72 °C for additional 10 minutes. After digestion with BamH I, the amplified subtractive DNA was either subjected the next round of subtraction (for the first two rounds of subtractive cDNA) or cloned into BamH I digested, dephosphoralated pBlueScript vector (for the third round of subtractive cDNA).

PCR conditions for amplification of "anchored" library were optimized. The results indicated that Mg^{2+} concentration was crucial for success of amplification of such cDNA library (Figure 5). As demonstrated in Figure 5, 2.5mM Mg^{2+} seems to be the best concentration.

In order to evaluate the performance of this technique, maize proteinase inhibitor (MPI) cDNA and maize actin (ACT) cDNA fragments were used as probes to analyze subtractive cDNA pools at the end of each subtraction cycle. An analysis of the results indicated that actin cDNA, a cDNA presents equally in both libraries, was removed at the end of the second round procedure. MPI is a cDNA from a gene that was activated upon infection of maize germinating embryos by *Fusarium moniliforme* (Cordero *et al.*, *supra*), and was enriched about 25- to 50-fold (Figure 6).

Pursuant to the above-described procedure, cDNA clones designated C-11-3, G-4-5 and G-12-3 (SEQ ID NOS:27, 29 and 39, respectively) were obtained to exemplify genes that are activated upon *F. moniliforme* infection of germinating maize embryos. Northern blot analysis using these cDNAs as pro bes indicated that the genes are activated during infection (Figure 7). Sequence analysis of these clones indicated that they are novel cDNA's. DNA database search result suggested that G-12-3 (SEQ ID NO:39) may code for a cytochrome P450 protein (Figure 10). Clone C-11-3 (SEQ ID NO:27) (Figure 8) and G-4-5 (SEQ ID NO:29) (Figure 9) did not match up any known sequences in database.

Figure 11 shows how the sequence information on clone G-12-3 was applied to the ICEFAL technique, resulting in the isolation of the 5'-end of the gene. The primers used were GSP1 (SEQ ID NO:19) and GSP2 (SEQ ID NO:20), which have the sequences 5'-CCGCTCTTACTCCGTTCAGTCTTG-3' and 5'-CCATTCCCTTCAATCACCCATTTC-3', respectively.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: PIONEER HI-BRED INTERNATIONAL, INC.
(B) STREET: 700 CAPITAL SQUARE, 400 LOCUST STREET
(C) CITY: DES MOINES
(D) STATE: IOWA
(E) COUNTRY: UNITED STATES OF AMERICA
(F) POSTAL CODE: 50309

(ii) TITLE OF INVENTION: PCR-BASED CDNA SUBTRACTIVE CLONING METHOD

(iii) NUMBER OF SEQUENCES: 39

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Foley & Lardner
(B) STREET: 3000 K Street, N.W., Suite 500
(C) CITY: Washington
(D) STATE: D.C.
(E) COUNTRY: USA
(F) ZIP: 20007-5109

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: NOT YET ASSIGNED
(B) FILING DATE:

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/481,687
(B) FILING DATE: 07-JUN-1995

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: BENT, Stephen A.
(B) REGISTRATION NUMBER: 29,768
(C) REFERENCE/DOCKET NUMBER: 33229/432/PIHI

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (202)672-5300
(B) TELEFAX: (202)672-5399
(C) TELEX: 904136

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

AGCACTCTCC AGCCTCTCAC CGAG 24

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GATCCTCGGT GA 12

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

AGCACTCTCC AGCCTCTCAC CGAG 24

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GATCCTCGGT GA 12

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ACCGACGTCG ACTATCCATG AACG 24

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

IDT Exhibit 1020.0523

(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GATCCGTTCA TG 12

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ACCGACGTCG ACTATCCATG AACG 24

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GATCCGTTCA TG 12

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

AGGCAACTGT GCTATCCGAG GGAG 24

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GATCCTCCCT CG 12

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

AGGCAACTGT GCTATCCGAG GGAG 24

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GATCCTCCCT CG 12

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CGCGGATCCG TTTTTTTTTT TTTTTTTA 28

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CGCGGATCCG TTTTTTTTTT TTTTTTTG 28

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CGCGGATCCG TTTTTTTTTT TTTTTTTC 28

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

TAAGGATCCT GGGGGGGGGG GGGA 24

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TAAGGATCCT GGGGGGGGGG GGGT 24

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

TAAGGATCCT GGGGGGGGGG GGGC 24

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CCGCTCTTAC TCCGTTCAGT CTTG 24

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CCATTCCCTT CAATCACCCA TTTC 24

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GCGCCTAGGC AAAAAAAAAA AAAAAAAT 28

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GCGCCTAGGC AAAAAAAAAA AAAAAAAG 28

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GCGCCTAGGC AAAAAAAAAA AAAAAAAC 28

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

ATTCCTAGGA CCCCCCCCCC CCCT 24

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

ATTCCTAGGA CCCCCCCCCC CCCA 24

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

ATTCCTAGGA CCCCCCCCCC CCCG 24

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 486 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION: 68..364

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 68..364

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

```
GGATCCTGGG GGGGGGGGGA CGAACTCTCT CTATACTCTC CCATCAATCC TTAAATTATC      60

ACGCATT ATG CGA ACT GTT GCA GTA CTC GCT CTC TTT GCC CAA CTG GCG       109
        Met Arg Thr Val Ala Val Leu Ala Leu Phe Ala Gln Leu Ala
          1               5                  10

ACG TGC GCC ATA TTC AAC ATC ACA GGA TCG TGC GCC GAC AGC GAA AAC       157
Thr Cys Ala Ile Phe Asn Ile Thr Gly Ser Cys Ala Asp Ser Glu Asn
 15                  20                  25                  30

GGC CCT GTT TGC GTC ATT ACG AAG AGT GTA GTT AAC CCA GCT ACA GTT       205
Gly Pro Val Cys Val Ile Thr Lys Ser Val Val Asn Pro Ala Thr Val
                35                  40                  45

TGC AAC GGG AAG GCT GAG GCG TAT GCA GGA GAC GGG AAT CAA TGG CAT       253
Cys Asn Gly Lys Ala Glu Ala Tyr Ala Gly Asp Gly Asn Gln Trp His
            50                  55                  60

GAC GGG CTG TAC TGG AAT TGG TTC CCC TTG CAC TTA TGT TTG GCG ATG       301
Asp Gly Leu Tyr Trp Asn Trp Phe Pro Leu His Leu Cys Leu Ala Met
        65                  70                  75

CTA GAC GTT CTT CCT CAA CAT CAA ACT GCG AAG ACA CTG AAT TCG CTT       349
Leu Asp Val Leu Pro Gln His Gln Thr Ala Lys Thr Leu Asn Ser Leu
    80                  85                  90

TCG GAC CTT GGT ATA TAATCAGCTG CAGGTCCTGG CCTACTCCCT GTTCAACTAA       404
Ser Asp Leu Gly Ile
 95

AAGCAACATT GCTTTCGTTT GTCTTCCTGT ATTACCATCA ATCAGAATTA ACATACTCAT      464

CTCTTAAAAA AAAAAAAAAA AA                                              486
```

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 99 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

```
Met Arg Thr Val Ala Val Leu Ala Leu Phe Ala Gln Leu Ala Thr Cys
 1               5                  10                  15
```

```
Ala Ile Phe Asn Ile Thr Gly Ser Cys Ala Asp Ser Glu Asn Gly Pro
            20                  25                  30

Val Cys Val Ile Thr Lys Ser Val Val Asn Pro Ala Thr Val Cys Asn
        35                  40                  45

Gly Lys Ala Glu Ala Tyr Ala Gly Asp Gly Asn Gln Trp His Asp Gly
    50                  55                  60

Leu Tyr Trp Asn Trp Phe Pro Leu His Leu Cys Leu Ala Met Leu Asp
65                  70                  75                  80

Val Leu Pro Gln His Gln Thr Ala Lys Thr Leu Asn Ser Leu Ser Asp
                85                  90                  95

Leu Gly Ile
```

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 471 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: join(1..111, 115..255, 259..270, 274..321, 325..342, 346..402, 406..471)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

```
GTG GAT CCT TCG ACG ACT ACC GCA TGT ACA TCC GCC GCA AGG GGC CTC        48
Val Asp Pro Ser Thr Thr Thr Ala Cys Thr Ser Ala Ala Arg Gly Leu
  1               5                  10                  15

GCG GGA AGA GCC AGG TCG ACT CCC TCA AGG TCG CCG ACG CCG ACG GCA        96
Ala Gly Arg Ala Arg Ser Thr Pro Ser Arg Ser Pro Thr Pro Thr Ala
            20                  25                  30

GAC AGT GCT ACT AGC TAG TAT ATA CCT AGC CAG CCT GCT GCC GAT CGA       144
Asp Ser Ala Thr Ser     Tyr Ile Pro Ser Gln Pro Ala Ala Asp Arg
        35                  40                  45

GAT TGT TTG TAT GTG TGG TGT GTG CAT GCA TTT GCC CAC ACT GAC CAC       192
Asp Cys Leu Tyr Val Trp Cys Val His Ala Phe Ala His Thr Asp His
    50                  55                  60

TGT CCA CAT GTA CGC CGC CAG CTG CCG GCC CTA AAT AAA ACC ATG CAT       240
Cys Pro His Val Arg Arg Gln Leu Pro Ala Leu Asn Lys Thr Met His
65                  70                  75

AGA TTA GCT AGC TTA TGA TTA ATC AAG TCT TAG CAG CTA GAG AGT GCT       288
Arg Leu Ala Ser Leu     Leu Ile Lys Ser     Gln Leu Glu Ser Ala
80                  85                  90

TTG GGT TGG GAC TCT CTC ATA GGA GGG NAT GCT TGA TCG ATC CGA TCA       336
Leu Gly Trp Asp Ser Leu Ile Gly Gly Xaa Ala     Ser Ile Arg Ser
    95                  100                 105

TCA ATT TGA AAC ACC CTG CTA GGT TGT GCA NCT CCG CCG TCC AAN CCA       384
Ser Ile     Asn Thr Leu Leu Gly Cys Ala Xaa Pro Pro Ser Xaa Pro
    110                 115                 120
```

```
CAA AGG GGN GAN GTC AAN TGA AGG GTG AGA NAA CGT CAA NAA CGA AGC        432
Gln Arg Xaa Xaa Val Xaa     Arg Val Arg Xaa Arg Gln Xaa Arg Ser
    125                     130                 135

NAG CTA GTT CCC NTT ATT NGG GTG GTT CTC AAA AAA AAA                    471
Xaa Leu Val Pro Xaa Ile Xaa Val Val Leu Lys Lys Lys
    140                 145                 150
```

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 151 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

```
Val Asp Pro Ser Thr Thr Thr Ala Cys Thr Ser Ala Ala Arg Gly Leu
 1               5                  10                  15

Ala Gly Arg Ala Arg Ser Thr Pro Ser Arg Ser Pro Thr Pro Thr Ala
            20                  25                  30

Asp Ser Ala Thr Ser Tyr Ile Pro Ser Gln Pro Ala Ala Asp Arg Asp
        35                  40                  45

Cys Leu Tyr Val Trp Cys Val His Ala Phe Ala His Thr Asp His Cys
    50                  55                  60

Pro His Val Arg Arg Gln Leu Pro Ala Leu Asn Lys Thr Met His Arg
65                  70                  75                  80

Leu Ala Ser Leu Leu Ile Lys Ser Gln Leu Glu Ser Ala Leu Gly Trp
                85                  90                  95

Asp Ser Leu Ile Gly Gly Xaa Ala Ser Ile Arg Ser Ser Ile Asn Thr
            100                 105                 110

Leu Leu Gly Cys Ala Xaa Pro Pro Ser Xaa Pro Gln Arg Xaa Xaa Val
        115                 120                 125

Xaa Arg Val Arg Xaa Arg Gln Xaa Arg Ser Xaa Leu Val Pro Xaa Ile
    130                 135                 140

Xaa Val Val Leu Lys Lys Lys
145                 150
```

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 148 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

```
Trp Ile Leu Arg Arg Leu Pro His Val His Pro Pro Gln Gly Ala Ser
1               5                  10                  15
```

Arg Glu Glu Pro Gly Arg Leu Pro Gln Gly Arg Arg Arg Arg Gln
20 25 30

Thr Val Leu Leu Ala Ser Ile Tyr Leu Ala Ser Leu Leu Pro Ile Glu
35 40 45

Ile Val Cys Met Cys Gly Val Cys Met His Leu Pro Thr Leu Thr Thr
50 55 60

Val His Met Tyr Ala Ala Ser Cys Arg Pro Ile Lys Pro Cys Ile Asp
65 70 75 80

Leu Ala Tyr Asp Ser Ser Leu Ser Ser Arg Val Leu Trp Val Gly Thr
85 90 95

Leu Ser Glu Gly Met Leu Asp Arg Ser Asp His Gln Phe Glu Thr Pro
100 105 110

Cys Val Val Xaa Leu Arg Arg Pro Xaa His Lys Gly Xaa Xaa Ser Xaa
115 120 125

Glu Gly Xaa Asn Val Xaa Asn Glu Ala Ser Phe Pro Leu Xaa Gly Trp
130 135 140

Phe Ser Lys Lys
145

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 150 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

Gly Ser Phe Asp Asp Tyr Arg Met Tyr Ile Arg Arg Lys Gly Pro Arg
1 5 10 15

Gly Lys Ser Gln Val Asp Ser Leu Lys Val Ala Asp Ala Asp Gly Arg
20 25 30

Gln Cys Tyr Leu Val Tyr Thr Pro Ala Cys Cys Arg Ser Arg Leu Phe
35 40 45

Val Cys Val Val Cys Ala Cys Ile Cys Pro His Pro Leu Ser Thr Cys
50 55 60

Thr Pro Pro Ala Ala Gly Pro Lys Asn His Ala Ile Ser Leu Met Ile
65 70 75 80

Asn Gln Val Leu Ala Ala Arg Glu Cys Phe Gly Leu Gly Leu Ser His
85 90 95

Arg Arg Xaa Cys Leu Ile Asp Pro Ile Ile Asn Leu Lys His Pro Ala
100 105 110

Arg Leu Cys Xaa Ser Ala Val Gln Xaa Thr Lys Gly Xaa Xaa Gln Xaa
115 120 125

Lys Gly Glu Xaa Thr Ser Xaa Thr Lys Xaa Ala Ser Ser Xaa Tyr Xaa
130 135 140

Gly Gly Ser Gln Lys Lys
145 150

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 83 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

Glu Arg Tyr Asn Pro Gln Arg Trp Leu Asp Ile Arg Gly Ser Gly Arg
1 5 10 15

Asn Phe His His Val Pro Phe Gly Phe Gly Met Arg Gln Cys Leu Gly
20 25 30

Arg Arg Leu Ala Glu Val Glu Met Leu Leu Leu Leu His His Val Leu
35 40 45

Lys His Phe Leu Val Glu Thr Leu Thr Gln Glu Asp Ile Lys Met Val
50 55 60

Tyr Ser Phe Ile Leu Arg Pro Gly Thr Ser Pro Leu Leu Thr Phe Arg
65 70 75 80

Ala Ile Asn

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 83 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

Glu Arg Tyr Asn Pro Gln Arg Trp Leu Asp Ile Arg Gly Ser Gly Arg
1 5 10 15

Asn Phe His His Val Pro Phe Gly Phe Gly Met Arg Gln Cys Leu Gly
20 25 30

Arg Arg Leu Ala Glu Ala Glu Met Leu Leu Leu Leu His His Val Leu
35 40 45

Lys His Phe Leu Val Glu Thr Leu Thr Gln Glu Asp Ile Lys Met Val
50 55 60

Tyr Ser Phe Ile Leu Arg Pro Gly Thr Ser Pro Leu Leu Thr Phe Arg
65 70 75 80

Ala Ile Asn

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 86 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

```
Glu Lys Phe Asp Pro Gly His Phe Leu Asn Ala Asn Gly Thr Phe Arg
1               5                  10                  15

Lys Ser Asn Tyr Phe Met Pro Phe Ser Ala Gly Lys Arg Ile Cys Ala
            20                  25                  30

Gly Glu Gly Leu Ala Arg Met Glu Leu Phe Leu Phe Leu Thr Ser Ile
        35                  40                  45

Leu Gln Asn Phe Ser Leu Lys Pro Val Lys Asp Arg Lys Asp Ile Asp
    50                  55                  60

Ile Ser Pro Ile Val Thr Ser Ala Ala Asn Ile Pro Arg Pro Tyr Glu
65                  70                  75                  80

Val Ser Phe Ile Pro Arg
                85
```

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 86 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

```
Glu Lys Phe Asp Pro Gly His Phe Leu Asn Ala Asn Gly Thr Phe Arg
1               5                  10                  15

Arg Ser Asp Tyr Phe Met Pro Phe Ser Ala Gly Lys Arg Ile Cys Ala
            20                  25                  30

Gly Glu Gly Leu Ala Arg Met Glu Ile Phe Leu Phe Leu Thr Ser Ile
        35                  40                  45

Leu Gln Asn Phe Ser Leu Lys Pro Val Lys Asp Arg Lys Asp Ile Asp
    50                  55                  60

Ile Ser Pro Ile Ile Thr Ser Leu Ala Asn Met Pro Arg Pro Tyr Glu
65                  70                  75                  80

Val Ser Phe Ile Pro Arg
                85
```

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 86 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

Gln Asp Phe Asn Pro Gln His Phe Leu Asn Glu Lys Gly Gln Phe Lys
1 5 10 15

Lys Ser Asp Ala Phe Val Pro Phe Ser Ile Gly Lys Arg Asn Cys Phe
20 25 30

Gly Glu Gly Leu Ala Arg Met Glu Leu Phe Leu Phe Phe Thr Thr Val
35 40 45

Met Gln Asn Phe Arg Leu Lys Ser Ser Gln Ser Pro Lys Asp Ile Asp
50 55 60

Val Ser Pro Lys His Val Gly Phe Ala Thr Ile Pro Arg Asn Tyr Thr
65 70 75 80

Met Ser Phe Leu Pro Arg
85

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 87 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

Glu Thr Phe Lys Pro Glu His Phe Leu Asn Glu Asn Gly Lys Phe Lys
1 5 10 15

Tyr Ser Asp Tyr Phe Lys Ala Phe Ser Ala Gly Lys Arg Val Cys Val
20 25 30

Gly Glu Gly Leu Ala Arg Met Glu Leu Phe Leu Leu Leu Ser Ala Ile
35 40 45

Leu Gln His Phe Asn Leu Lys Ser Leu Val Asp Pro Lys Asp Ile Asp
50 55 60

Leu Ser Pro Val Thr Ile Gly Phe Gly Ser Ile Pro Arg Glu Phe Val
65 70 75 80

Ile Cys Val Ile Pro Arg Ser
85

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 87 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

Glu Lys Phe Ile Pro Glu Arg Trp Leu Asn Glu Thr Pro Glu Met Lys
1 5 10 15

Ser Ala Leu Thr Pro Phe Ser Leu Gly Lys Arg Asn Cys Ile Gly Gln
20 25 30

Asn Leu Ala Trp Gln Glu Leu Tyr Trp Ala Val Asn Glu Val Met Arg
35 40 45

Ser Gly Ser Arg Phe Arg Val Ala Glu Glu Met Lys Asp Trp Glu Met
50 55 60

Glu Met Glu Asp Arg Phe Asn Ile Ala Pro Arg Gly Arg Arg Leu Met
65 70 75 80

Leu Thr Ala Ser Gln Val Asn
85

What Is Claimed Is:

1. A method based on PCR for generating a cDNA library with anchored ends, comprising the steps of

(A) providing an RNA preparation that comprises polyadenylated mRNA from a biological sample, said polyadenylated mRNA having a polyA portion and a non-polyA portion;

(B) preparing a oligonucleotide polyTV primer SEQ ID NOS:13-15 that anneals to said polyA portion of said polyadenylated mRNA, anchoring said polyTV primer (SEQ ID NOS:13-15) such that a reverse transcriptase reaction can start within one nucleotide from the 3′ end of said non-polyA portion;

(C) using said polyTV primer (SEQ ID NOS:13-15)to generate by PCR from said RNA preparation DNA strands that are complementary to said polyadenylated mRNA;

(D) using terminal deoxynucleotidyl transferase to add a polynucleotide tail at the 3′ end of said DNA strands, whereby said DNA strands have a first portion that is complementary to said polyadenylated mRNA and a tail portion;

(E) preparing an oligonucleotide polyGH primer (SEQ ID NOS:16-18) that anneals to said polynucleotide tail attached in step (D), anchoring said polyGH primer (SEQ ID NOS:16-18) such that a reverse transcriptase reaction can start one nucleotide downstream from the 5′ end of said first portion of said DNA strands;

and then

(F) using said polyGH primer (SEQ ID NOS: 16-18) and said polyTV primer (SEQ ID NOS:13-15) to amplify said DNA strands by PCR, whereby a cdNA library with anchored ends is generated (SEQ ID NOS:16-18, 22-23, 24-26 and 13-15, respectively),

2. A method according to claim 1, wherein said polynucleotide tail in step (D) is a polyC tail.

3. A method of generating a cDNA library with anchored ends as claimed in claim 1, wherein each of said polyGH primer and said polyTV primer contains a recognition site for a restriction endonuclease.

4. A method of comparing two cDNA libraries to identify cDNA that is unique to one of said libraries, comprising the steps of

(A) preparing a first cDNA library and a second cDNA library with anchored ends, wherein said first cDNA library contains said unique cDNA and is prepared according to claim 3, and wherein said second cDNA library is produced according to claim 3 and incorporates biotin-labelled deoxynucleotides, and

(B) digesting said first cDNA library with a restriction enzyme for which a recognition sequence is found on said polyGH and polyTV primers, to produce, at the ends of cDNAs in said first library, single-stranded DNA tails,

(C) preparing multiple sets of double-stranded adapter DNA molecules, each set comprising a first and a second oligonucleotide of such sequence that:

(1) the 3′ end of the first oligonucleotide complements the 5′ end of the second oligonucleotide of the same set

and

(2) the 5′ end of said first oligonucleotide complements the single-strand DNA produced by a restriction enzyme digest at the anchored ends of the first DNA library,

(D) manipulating said libraries by mixing them; subjecting them to DNA-melting conditions; allowing for reannealing of DNA strands in said libraries; purifying said unique cDNA by trapping on streptavidin beads DNA that incorporates said biotin-labeled deoxynucleotides, whereby a cDNA fraction enriched for said unique cDNA is obtained;

producing blunt-ended DNA by filling-in said unique cDNA ends with Klenow enzyme reactions; and subject said cDNA fraction to PCR amplification of said unique DNA via oligonucleotide primers that comprise an adapter molecule used in step (D),

(E) digesting said unique cDNA with said restriction enzyme and ligating an adapter set from said multiple sets to said anchored ends of said first library,

and then

(F) repeating steps (D) and (E), each time ligating a new set of adapter DNA molecules from said multiple sets to said cDNA fraction, until said unique cDNA is essentially free of non-unique cDNA from said first library.

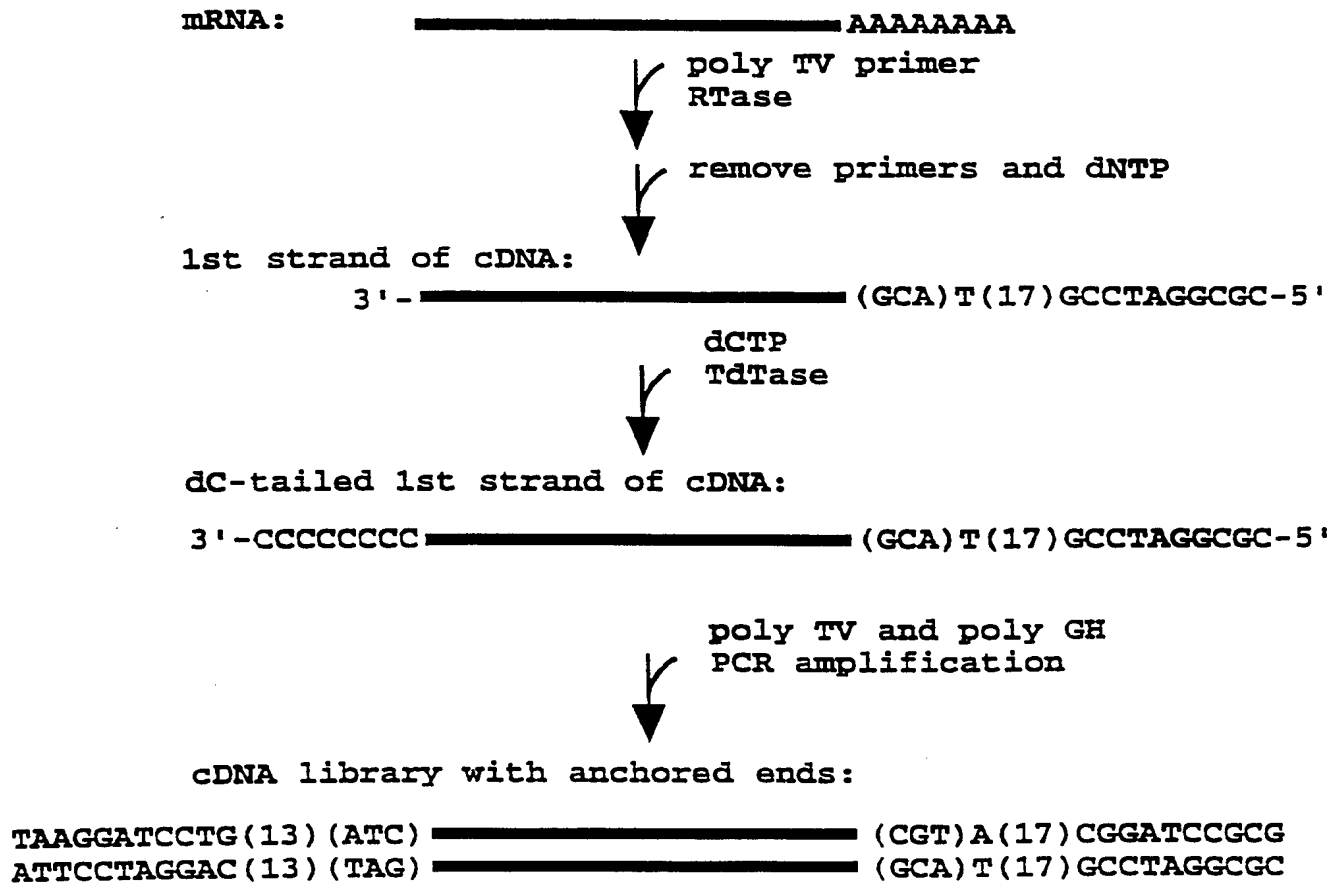
5. A method of isolating the cDNA ends of a unique cDNA from an anchored library produced according to claim 2, comprising the steps of (A) providing a set of PCR primers that hybridize, respectively, to a sequence internal to said unique cDNA and (B) using said primers with said polyTV or said polyGH primers in PCR reactions to produce two ends of said cDNA, wherein said primers contain a restriction enzyme cleavage site.

6. A method of generating a PCR product from a cDNA molecule of an anchored library produced according to claim 1, comprising the steps of (A) providing a PCR primer that hybridizes to a sequence internal to said molecule and (B) using said primer with said polyTV primer (SEQ ID NOS:13-15) or said polyGH primer (SEQ ID NOS:16-18), as appropriate for PCR reactions, to generate a PCR product that corresponds to an end of said cDNA molecule.

7. A method according to claim 6, wherein said primers comprise restriction enzyme cleavage sites.

Figure 1

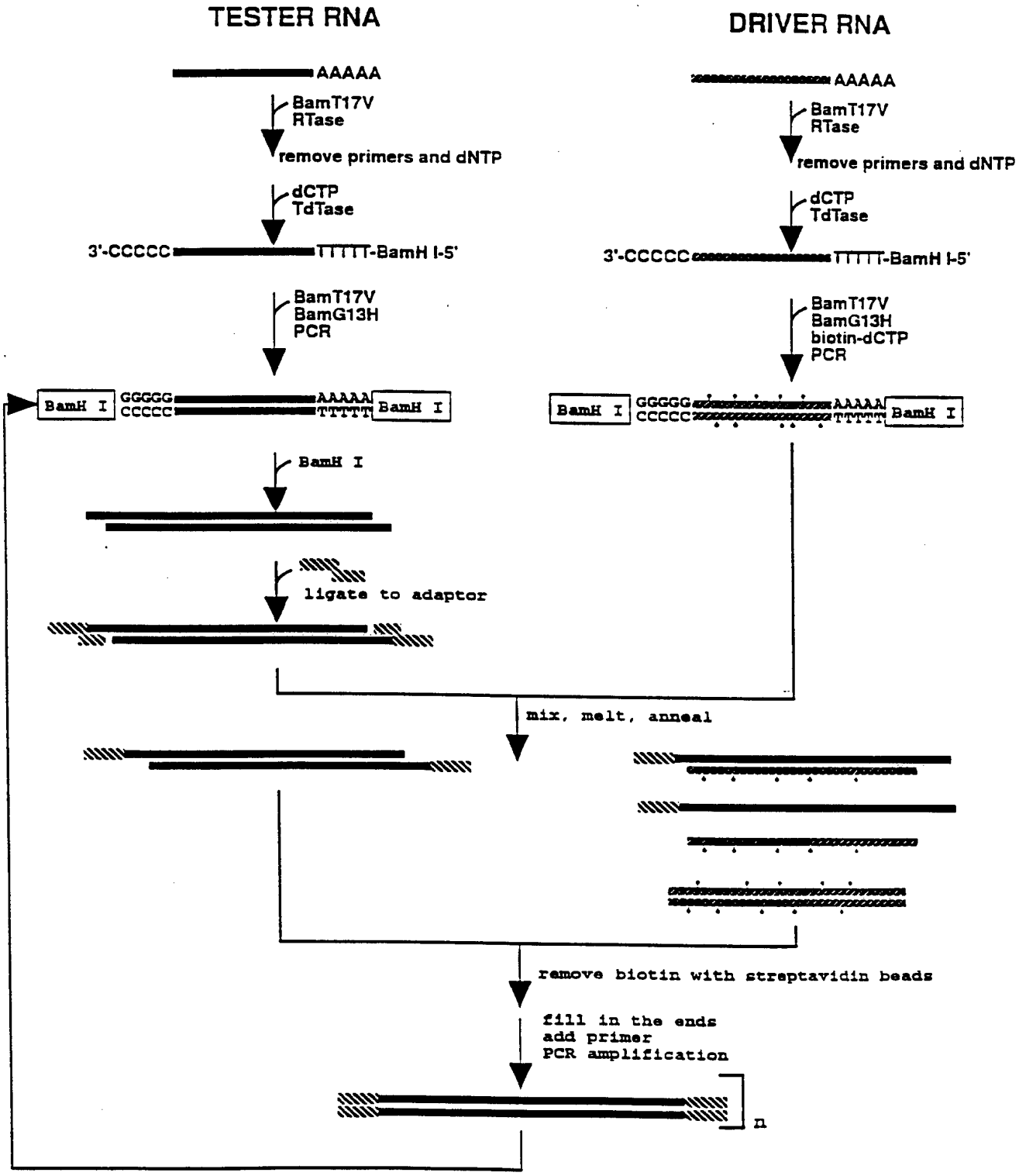
A simple PCR based method for generation of a cDNA library with anchored ends



IDT Exhibit 1020.0540

Figure 2

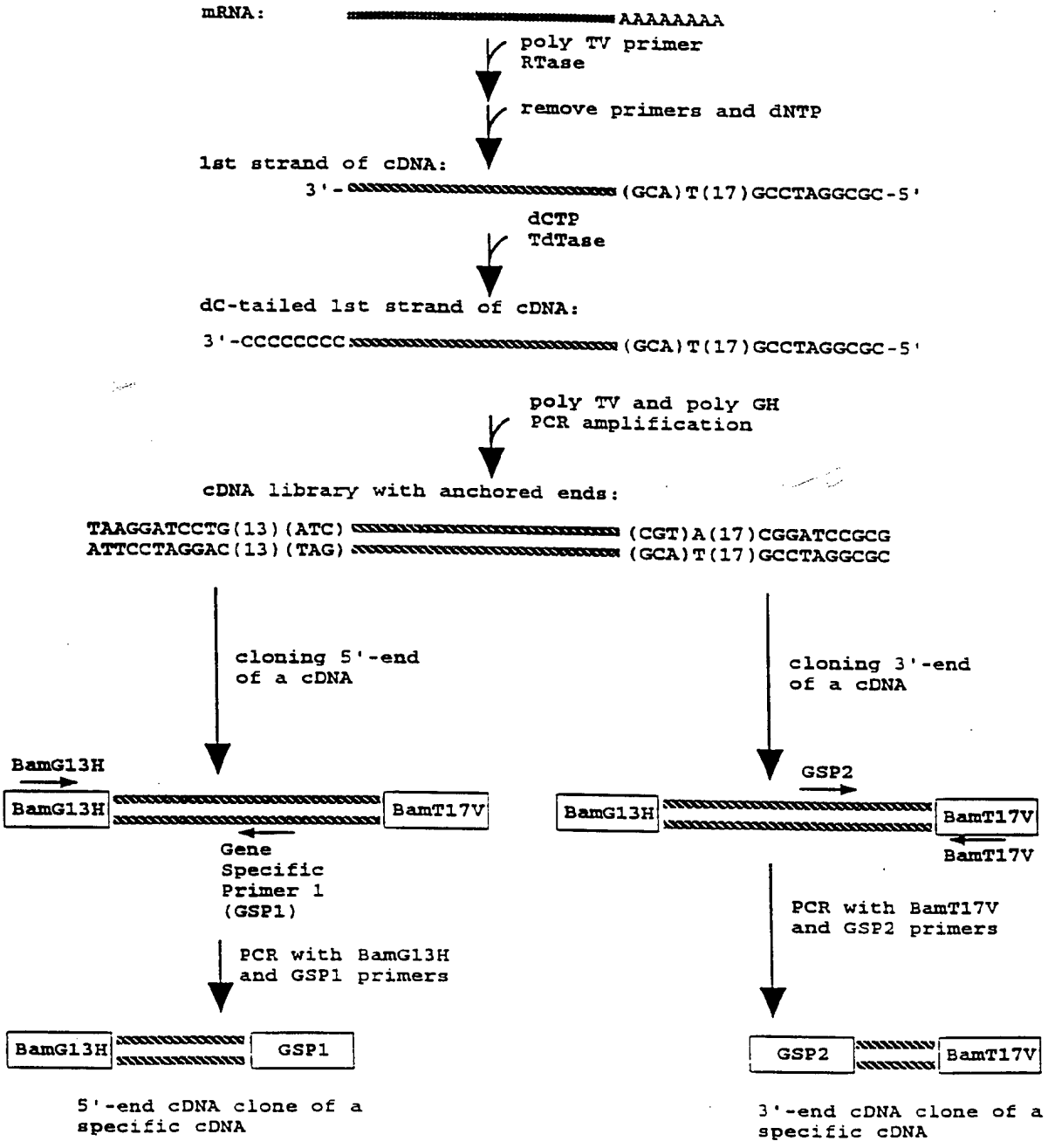
PCR Based cDNA Subtractive (PCSUB) Library



IDT Exhibit 1020.0541

Figure 3

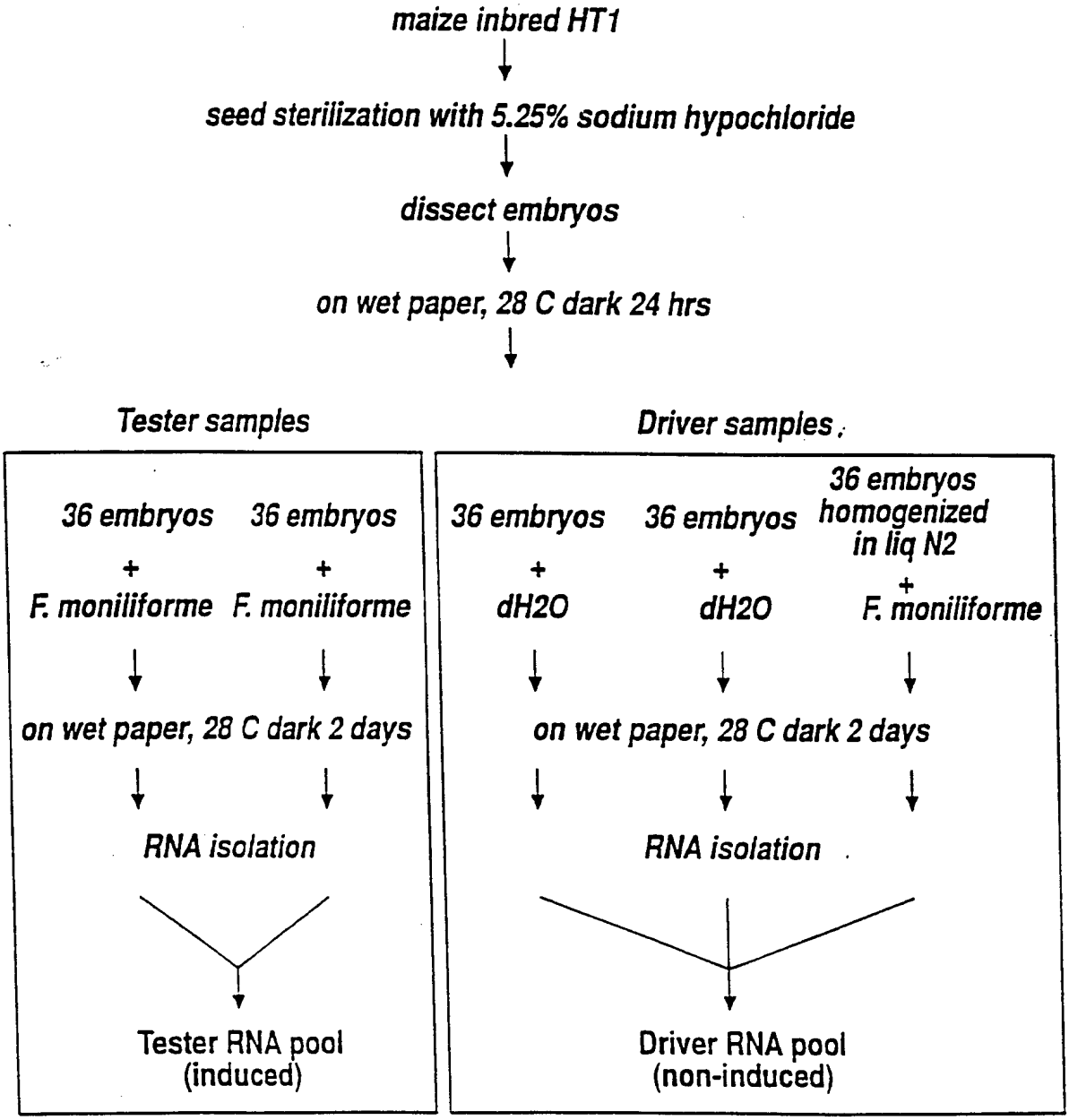
Isolation of cDNA ends from an "anchored" library (ICEFAL)



IDT Exhibit 1020.0542

Figure 4

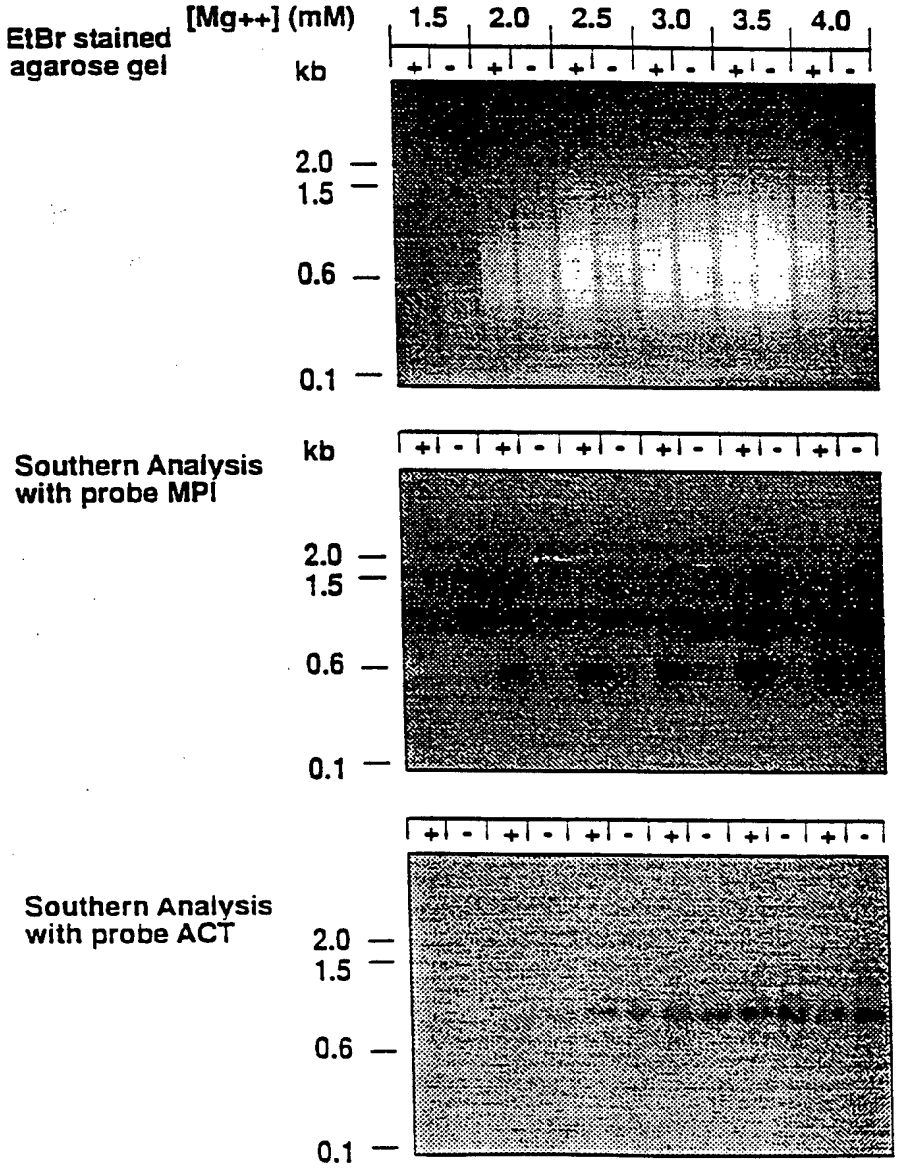
Tester and Driver RNA Preparation



IDT Exhibit 1020.0543

Figure 5

Tester (+) and driver (-) dC-cDNAs were amplified with BamT17V and BamG13H primers at different magnesium concentration. The resulting DNA was analyzed by gel electrophoresis and Southern analysis. Southern results indicated [Mg2+] > 2.5mM was preferable for amplification of both MPI and ACT



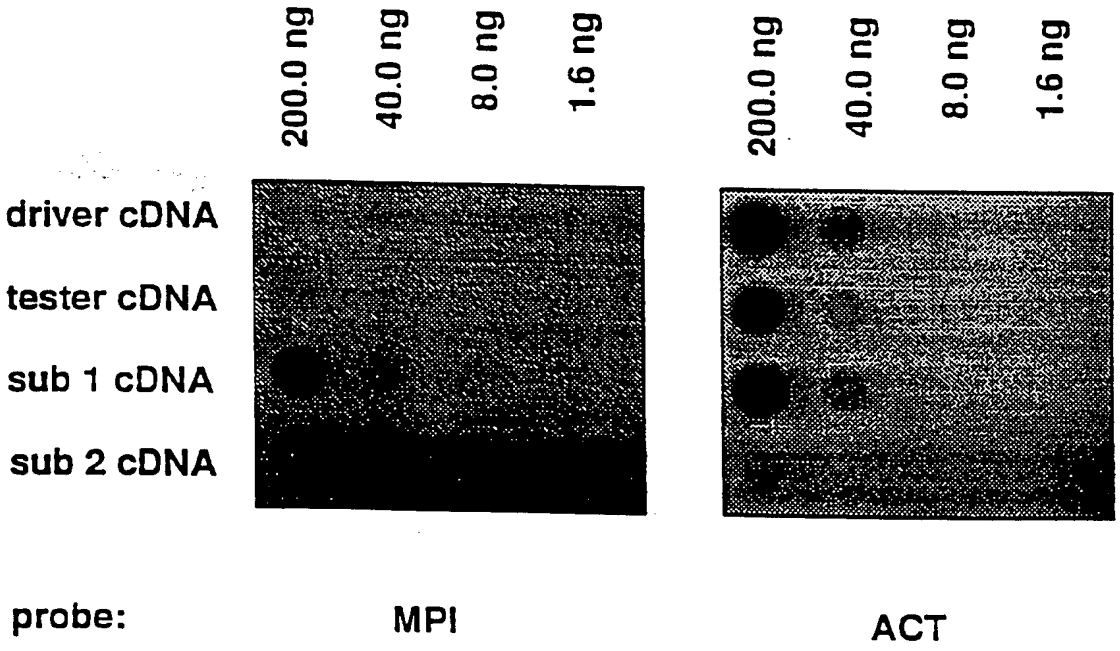
IDT Exhibit 1020.0544

Figure 6

DNA dot-blot analysis with probes MPI and ACT.

sub 1 cDNA = subtractive cDNA obtained after the first cycle of subtraction

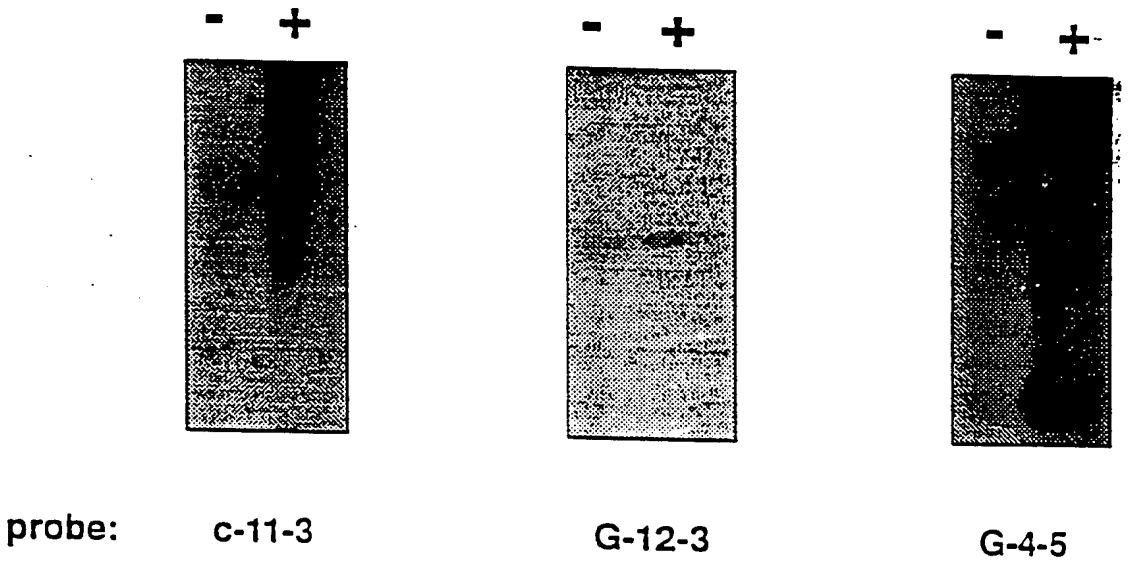
sub 2 cDNA = subtractive cDNA obtained after the second cycle of subtraction



IDT Exhibit 1020.0545

Figure 7

Northern Analysis probed with cDNAs from the Subtractive cDNA library



IDT Exhibit 1020.0546

Figure 8

Nucleic acid sequence of clone C-11-3 and the deduced amino acid sequence in a putative open reading frame

```
  1 GGATCCTGGGGGGGGGGGGACGAACTCTCTCTATACTCTCCCATCAATCCTTAAATTATC 60

 61 ACGCATTATGCGAACTGTTGCAGTACTCGCTCTCTTTGCCCAACTGGCGACGTGCGCCAT 120
          M  R  T  V  A  V  L  A  L  F  A  Q  L  A  T  C  A  I  -

121 ATTCAACATCACAGGATCGTGCGCCGACAGCGAAAACGGCCCTGTTTGCGTCATTACGAA 180
     F  N  I  T  G  S  C  A  D  S  E  N  G  P  V  C  V  I  T  K  -

181 GAGTGTAGTTAACCCAGCTACAGTTTGCAACGGGAAGGCTGAGGCGTATGCAGGAGACGG 240
     S  V  V  N  P  A  T  V  C  N  G  K  A  E  A  Y  A  G  D  G  -

241 GAATCAATGGCATGACGGGCTGTACTGGAATTGGTTCCCCTTGCACTTATGTTTGGCGAT 300
     N  Q  W  H  D  G  L  Y  W  N  W  F  P  L  H  L  C  L  A  M  -

301 GCTAGACGTTCTTCCTCAACATCAAACTGCGAAGACACTGAATTCGCTTTCGGACCTTGG 360
     L  D  V  L  P  Q  H  Q  T  A  K  T  L  N  S  L  S  D  L  G  -

361 TATATAATCAGCTGCAGGTCCTGGCCTACTCCCTGTTCAACTAAAAGCAACATTGCTTTC 420
     I  *

421 GTTTGTCTTCCTGTATTACCATCAATCAGAATTAACATACTCATCTCTTAAAAAAAAAAA 480

481 AAAAAA
```


IDT Exhibit 1020.0547

Figure 9

Nucleic acid sequence of clone G-4-5 and the deduced amino acid sequence in three reading frames. A putative open reading frame was underlined

```
      91  GTGGATCCTTCGACGACTACCGCATGTACATCCGCCGCAAGGGGCCTCGCGGGAAGAGCC 150

a        V  D  P  S  T  T  T  A  C  T  S  A  A  R  G  L  A  G  R  A  -
b         W  I  L  R  R  L  P  H  V  H  P  P  Q  G  A  S  R  E  E  P  -
c          G  S  F  D  D  Y  R  M  Y  I  R  R  K  G  P  R  G  K  S  Q  -

     151  AGGTCGACTCCCTCAAGGTCGCCGACGCCGACGGCAGACAGTGCTACTAGCTAGTATATA 210

a        R  S  T  P  S  R  S  P  T  P  T  A  D  S  A  T  S  *  Y  I  -
b         G  R  L  P  Q  G  R  R  R  R  Q  T  V  L  L  A  S  I  Y  -
c          V  D  S  L  K  V  A  D  A  D  G  R  Q  C  Y  *  L  V  Y  T  -

     211  CCTAGCCAGCCTGCTGCCGATCGAGATTGTTTGTATGTGTGGTGTGTGCATGCATTTGCC 270

a        P  S  Q  P  A  A  D  R  D  C  L  Y  V  W  C  V  H  A  F  A  -
b         L  A  S  L  L  P  I  E  I  V  C  M  C  G  V  C  M  H  L  P  -
c          *  P  A  C  C  R  S  R  L  F  V  C  V  V  C  A  C  I  C  P  -

     271  CACACTGACCACTGTCCACATGTACGCCGCCAGCTGCCGGCCCTAAATAAAACCATGCAT 330

a        H  T  D  H  C  P  H  V  R  R  Q  L  P  A  L  N  K  T  M  H  -
b         T  L  T  T  V  H  M  Y  A  A  S  C  R  P  *  I  K  P  C  I  -
c          H  *  P  L  S  T  C  T  P  P  A  A  G  P  K  *  N  H  A  *  -

     331  AGATTAGCTAGCTTATGATTAATCAAGTCTTAGCAGCTAGAGAGTGCTTTGGGTTGGGAC 390

a        R  L  A  S  L  *  L  I  K  S  *  Q  L  E  S  A  L  G  W  D  -
b         D  *  L  A  Y  D  *  S  S  L  S  S  *  R  V  L  W  V  G  T  -
c          I  S  *  L  M  I  N  Q  V  L  A  A  R  E  C  F  G  L  G  L  -

     391  TCTCTCATAGGAGGGNATGCTTGATCGATCCGATCATCAATTTGAAACACCCTGCTAGGT 450

a        S  L  I  G  G  ?  A  *  S  I  R  S  S  I  *  N  T  L  L  G  -
b         L  S  *  E  G  M  L  D  R  S  D  H  Q  F  E  T  P  C  *  V  -
c          S  H  R  R  ?  C  L  I  D  P  I  I  N  L  K  H  P  A  R  L  -

     451  TGTGCANCTCCGCCGTCCAANCCACAAAGGGGNGANGTCAANTGAAGGGTGAGANAACGT 510

a        C  A  ?  P  P  S  ?  P  Q  R  G  ?  V  ?  *  R  V  R  ?  R  -
b         V  ?  L  R  R  P  ?  H  K  G  ?  ?  S  ?  E  G  *  ?  N  V  -
c          C  ?  S  A  V  Q  ?  T  K  G  ?  ?  Q  ?  K  G  E  ?  T  S  -

     511  CAANAACGAAGCNAGCTAGTTCCCNTTATTNGGGTGGTTCTCAAAAAAAAA 561

a        Q  ?  R  S  ?  L  V  P  ?  I  ?  V  V  L  K  K  K  -
b         ?  N  E  A  S  *  F  P  L  ?  G  W  F  S  K  K  -
c          ?  T  K  ?  A  S  S  ?  Y  ?  G  G  S  Q  K  K  -
```


IDT Exhibit 1020.0548

Figure 10

Comparison of the amino acid sequence deduced from the nucleotide sequence of clone G-12-3 with the sequences of known cytochrome P450 proteins. Asterisk indicates the identical amino acids shared among these sequences

```
          451                                                    500
B34181    ERYNPQRWLD IRGSGRNFHH .VPFGFGMRQ CLGRRLAEVE MLLLLHHVLK
A37088    ERYNPQRWLD IRGSGRNFHH .VPFGFGMRQ CLGRRLAEAE MLLLLHHVLK
A31418    EKFDPGHFLN ANGTFRKSNY FMPFSAGKRI CAGEGLARME LFLFLTSILQ
A24814    EKFDPGHFLN ANGTFRRSDY FMPFSAGKRI CAGEGLARME IFLFLTSILQ
O4hua6    QDFNPQHFLN EKGQFKKSDA FVPFSIGKRN CFGEGLARME LFLFFTTVMQ
S19657    ETFKPEHFLN ENGKFKYSDY FKAFSAGKRV CVGEGLARME LFLLLSAILQ
G-12-3    EKFIPERWLN ETPEMKSA.. LTPFSLGKRN CIGQNLAWQE LYWAVNEVMR
               *   *                *  *  *  * *  **  *

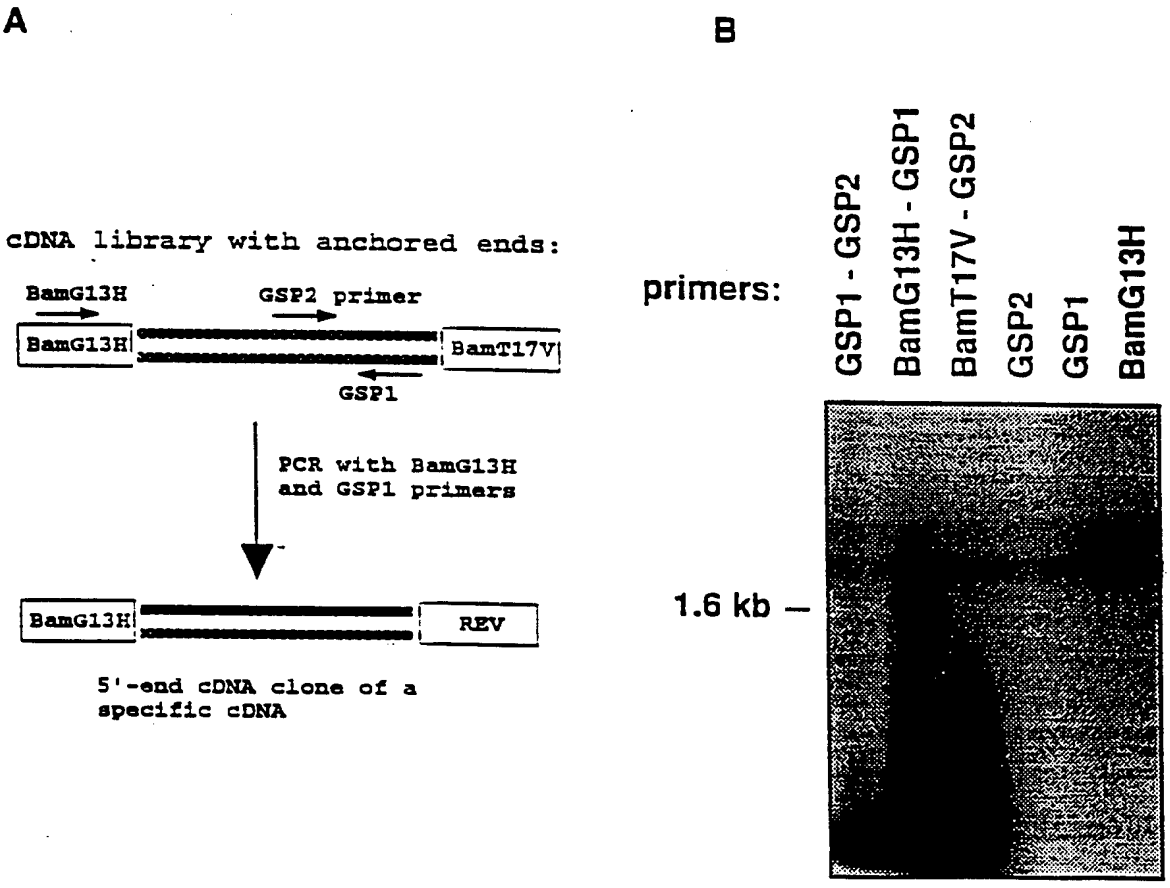
          501                                                541
B34181    HF..LVETLT QEDIKMVYSF ILRPGTSPLL TFRAIN.... .
A37088    HF..LVETLT QEDIKMVYSF ILRPGTSPLL TFRAIN.... .
A31418    NFS.LKPVKD RKDIDISPIV TSAANIPRPY EVSFIPR... .
A24814    NFS.LKPVKD RKDIDISPII TSLANMPRPY EVSFIPR... .
O4hua6    NFR.LKSSQS PKDIDVSPKH VGFATIPRNY TMSFLPR... .
S19657    HFN.LKSLVD PKDIDLSPVT IGFGSIPREF VICVIPRS.. .
G-12-3    SGSRFRVAEE MKDWEME.ME DRFNIAPRGR RLMLTASQVN
                       *
```

B34181: human cytochrome P450 11B1
A37088: human cytochrome P450 11B2
A31418: chicken cytochrome P450 2F2
A24814: chicken cytochrome P450 phenobarbital-inducible
O4hua6: human cytochrome P450 2A6
S19657: mouse cytochrome P450 2E1
G-12-3: maize putative cytochrome P450

IDT Exhibit 1020.0549

Figure 11

Isolation of the 5'-end of cDNA G-12-3 clone by ICEFAL technique. PCR with GSP1 and BamG13H primers resulted in an amplification of a 1.6kb DNA fragment from tester cDNA library. Single primer controls did not result in amplification of this fragment.



IDT Exhibit 1020.0550

INTERNATIONAL SEARCH REPORT

Inte onal Application No
PCT/US 96/08582

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12Q1/68

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12Q

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	NUCLEIC ACIDS RESEARCH, vol. 17, no. 8, 1989, pages 2919-2932, XP000604679 BELYAVSKY ET AL.: "PCR-based cDNA library construction: general cDNA libraries at the level of a few cells" cited in the application see page 2919 - page 2921; figure 1 ---	1
A	PCR METHODS AND APPLICATIONS, vol. 2, no. 1, August 1992, COLD SPRING HARBOR, NY, pages 144-148, XP000196443 BORSON ET AL.: "A lock-docking oligo(dT) primer for 5' and 3' RACE PCR" cited in the application see page 144, left-hand column - page 145, left-hand column --- -/--	1

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
7 October 1996	23 / 10 / 96

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016	Ceder, O

Form PCT/ISA/210 (second sheet) (July 1992)

IDT Exhibit 1020.0551

INTERNATIONAL SEARCH REPORT

Inte onal Application No
PCT/US 96/08582

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JOURNAL OF BIOORGANIC CHEMISTRY, vol. 20, no. 6, 1 January 1994, pages 386-388, XP000575090 LUK'YANOV ET AL: "HIGHLY EFFICIENT SUBTRACTIVE HYBRIDIZATION OF CDNA" see the whole document ---	1
A	WO,A,89 12695 (GENELABS INC) 28 December 1989 see page 5 - page 6 ---	4
A	SCIENCE, vol. 259, no. 5097, 12 February 1993, WASHINGTON, DC, pages 946-951, XP000196441 LISITSYN ET AL.: "Cloning the differences between two complex genomes" cited in the application see page 946 - page 947 -----	4

1

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

IDT Exhibit 1020.0552

INTERNATIONAL SEARCH REPORT

Int ional Application No
PCT/US 96/08582

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO-A-8912695	28-12-89	AT-T-	125873	15-08-95
		AU-B-	620628	20-02-92
		AU-A-	3852389	12-01-90
		CA-A-	1335070	04-04-95
		DE-D-	68923704	07-09-95
		DE-T-	68923704	18-01-96
		EP-A-	0419571	03-04-91
		JP-T-	4501504	19-03-92

IDT Exhibit 1020.0553

Espacenet

Bibliographic data: WO9910540 (A1) — 1999-03-04

DNA METHYLTRANSFERASE GENOTYPING

Inventor(s): LOPEZ OSVALDO J [US]; NELSON R MICHAEL [US] ± (LOPEZ, OSVALDO, J, ; NELSON, R., MICHAEL)

Applicant(s): LOPEZ OSVALDO J [US]; NELSON R MICHAEL [US] ± (LOPEZ, OSVALDO, J, ; NELSON, R., MICHAEL)

Classification: **- international:** ***C12Q1/68;*** (IPC1-7): C12Q1/68
- cooperative: **C12Q1/6827**; **C12Q1/683**; **C12Q1/6858**; **C12Q1/6883**; C12Q2600/156 more

Application number: WO1998US17859 19980828 Global Dossier

Priority number (s): US19970057068P 19970829

Also published as: AR017023 (A1) AU9660698 (A) EP1009865 (A1) US6514698 (B1)

Abstract of WO9910540 (A1)

DNA Methyltransferases can be utilized in methods for quickly and accurately: determining variations, mutations or polymorphisms in DNA sequences; identifying specific alleles in single copy genes; creating genomic fingerprints; creating DNA Paints; and generating ordered maps.

IDT Exhibit 1020.0554

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12Q 1/68

A1

(11) International Publication Number: **WO 99/10540**

(43) International Publication Date: 4 March 1999 (04.03.99)

(21) International Application Number: PCT/US98/17859

(22) International Filing Date: 28 August 1998 (28.08.98)

(30) Priority Data:
60/057,068 29 August 1997 (29.08.97) US

(71)(72) Applicants and Inventors: LOPEZ, Osvaldo, J. [AR/US]; 1921 Pinedale Avenue, Lincoln, NE 68506 (US). NELSON, R., Michael [US/US]; Megabase Research Products, Suite 110, 2820 North 48th Street, Lincoln, NE 68504 (US).

(74) Agent: NEBEL, Heidi, Sease; Zarley, McKee, Thomte, Voorhees & Sease, Suite 3200, 801 Grand Avenue, Des Moines, IA 50309–2721 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: DNA METHYLTRANSFERASE GENOTYPING

(57) Abstract

DNA Methyltransferases can be utilized in methods for quickly and accurately: determining variations, mutations or polymorphisms in DNA sequences; identifying specific alleles in single copy genes; creating genomic fingerprints; creating DNA Paints; and generating ordered maps.

IDT Exhibit 1020.0555

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

IDT Exhibit 1020.0556

TITLE: DNA METHYLTRANSFERASE GENOTYPING

GRANT REFERENCE

Work for this invention was funded at least in part by a grant from United States Department of Agriculture, Small Business Innovation Research Program, Grant Proposal No. 97-03376; and United States National Institutes of Health, Grant No. 1-R43-GM56595-01/120PW. The Government may have certain rights in this invention.

FIELD OF THE INVENTION

This invention relates to methods for DNA genotyping and gene mapping whereby nucleotide variations, mutations and polymorphisms are quickly and accurately detected by the enzymatic methylation of DNA using sequence-specific DNA methyltransferases. In particular, a novel method is described in which DNA can be "painted" at specific sites using radiometric or immunochemical detection. This novel methodology has been identified by the term "DNA Paint". DNA methyltransferases can be used alone, or in combination with restriction endonucleases, to (a) diagnose diseases of plants, animals, or humans; (b) map genes; or (c) detect genetic mutations or polymorphisms.

BACKGROUND OF THE INVENTION

Starting in the early 1970's with the emergence of the recombinant DNA methods and rapid DNA sequence technologies that followed soon thereafter, scientists have relied upon restriction endonucleases to study the genetic makeup of plants and animals. The medical field has shown great advancements through the use of this molecular genetic revolution through the identification of nucleic acid variations, mutations, and polymorphisms within genes and over entire genomes. In addition, the ability to identify and recombine specific alleles for genes of interest allow researchers to create improved livestock or plant varieties much more quickly than traditional breeding methods have allowed.

The discovery of sequence-specific DNA cutting enzymes, restriction endonucleases, made the development of recombinant DNA technologies possible. Over 2,000 different sequence-specific endonucleases have been characterized, of which about 200 are available. R.J. Roberts & D. Macelis, "REBASE--Restriction enzymes and methylases", Nucleic Acids Res., 24:223-235 (1996), incorporated herein by reference. These restriction endonucleases have been widely used in RFLP mapping, DNA

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0557

fingerprinting, gene mapping, to detect mutations responsible for heritable human diseases and polymorphisms associated with traits of interest in animal and plant breeding programs, and to diagnose infectious disease agents or viral, bacterial, or fungal origin.

Gene mapping, DNA fingerprinting, and RFLP technologies routinely exploit the variations and mutations between the genomes of different individuals or species or varietal populations. When a point mutation, insertion, or deletion, alters the DNA sequence within a genome, the restriction endonuclease enzyme can detect these changes in nucleic acid sequence if the change creates or destroys a restriction endonuclease recognition sequence. Restriction endonucleases are sequence-specific DNA-cutting enzymes which recognize a specific nucleic acid sequence pattern and cleave DNA strands at specific locations. The restriction enzymes most often used in gene mapping and DNA fingerprinting technologies are Type II bacterial restriction enzymes, most of which recognize 4 to 6 base pair recognition sequences. Therefore, on average, most DNA restriction fragments are 4^4 to 4^6 (256 to 4096) base pairs (bp) long. When a mutation or variation occurs within one of these DNA sequence recognition sites, those restriction enzymes which cleave at that recognition sequence will no longer be able to cut at that site if the site is destroyed. Conversely, if a DNA sequence recognition site is created, the restriction enzyme for that site will be able to cleave the DNA where it could not previously. It is this variation in the presence and absence of restriction enzyme recognition sites within the genomes of different individuals that allows for DNA mapping and genetic fingerprinting using restriction enzymes. (Botstein et al., "Construction of a genetic linkage map in man using restriction fragment length polymorphisms", Am. J. Human Genetics, 32:314-331 (1980); R. White and J.J. Lalouel, "Chromosome Mapping with DNA Markers", Scientific American, 258:40-48 (1988).

While the advent of the restriction enzymes has revolutionized molecular genetics, it has not come without its inherent weaknesses. In order to visualize the changes in restriction endonuclease target sites (and to therefore define a restriction isotype or an allelic fingerprint), the DNA fragments that result from restriction endonuclease cleavage must be separated by size. In short, if DNA is cut into pieces then the resulting pieces must be separated. Differences in fragment size create different morphological patterns, hence polymorphisms, for each individual or trait of interest. These polymorphic patterns (DNA fingerprints) are best observed when run out by electrophoresis on agarose or polyacrylamide gels.

IDT Exhibit 1020.0558

Another limitation using bacterial restriction enzymes in RFLP mapping or genotyping is that most mutations or polymorphic sites in DNA do not occur within these relatively rare (4 to 6 bp) endonuclease recognition sequences. The present invention overcomes this limitation through the use of sequence-specific DNA Methyltransferases (MTases) which can recognize short 2 to 4 bp sequence recognition sites as well as longer 4 to 8 bp sequence recognition sites. In particular, 75% of point mutations responsible for all known heritable human diseases occur at CG sites. D. Cooper & H. Youssoufin, "The CpG dinucleotide and human genetic disease", Human Genetics 78:151-165 (1988), incorporated herein by reference. Several DNA MTases recognize 2 bp CG sites, such as M.SspMQI, M.SssI, and M.DmtI of mouse (Okano, M., et al, "Cloning and Characterization of a Family of Novel Mammalian DNA (Cytosine-5) Methyltransferases", Nature Genetics, Vol. 19, July 1998, 219-220), human, Arabidopsis, and algal virus DNA MTases (Nelson et al., "DNA Methyltransferases and Site-specific Endonucleases Encoded by *Chlorella* Viruses", in: G.P. Jost & H.P. Salusz HP, eds., DNA Methylation, Birkauser & Basel, pp. 186-211 (1993)). In contrast, no known restriction endonucleases recognize CG or any other dinucleotide sequences.

The present invention uses DNA MTases rather than restriction enzymes to overcome many of the above-mentioned problems. Each restriction endonuclease has a companion sequence-specific DNA methyltransferase (MTase) which has the same DNA recognition site. W. Arber & S. Linn, "DNA Modification and Restriction", Ann. Rev. Biochem. 38:467-500 (1969), incorporated herein by reference. Several hundred of these DNA MTase specificities are known (McClelland et al., "Effect of site-specific modification on restriction endonucleases and DNA MTases", Nucleic Acids Res. 22:3640-3659 (1994), incorporated herein by reference) and more are being discovered each year. In addition, as mentioned above, there are several DNA MTases which recognize 2 to 4 bp DNA sequences, such as the CG dinucleotide sequence, whereas there are no known restriction endonucleases with such short recognition sequences. As described in this patent application for the first time, it is possible to use methyltransferases rather than restriction enzymes to detect genetic polymorphisms and mutations at 2 to 8 bp sites within genomic DNA.

The present invention can utilize PCR amplification and sequence-specific DNA methylation to detect the presence or absence of specific DNA methyltransferase recognition sites.

IDT Exhibit 1020.0559

DNA methyltransferases (MTases) catalyze the transfer of methyl groups from S-Adenosylmethionine (SAM) to specific sites in double-stranded DNA, yielding methylated DNA and
S-Adenosylhomocysteine (SAH).

$$\text{DNA} + \text{SAM} \xrightarrow{\text{DNA MTase}} \text{methyl-DNA} + \text{SAH}$$

If radioactive ^{3}H-methyl-SAM is used as a substrate, then the number of methyl groups incorporated into DNA can be measured by trichloroacetic acid (TCA) precipitation of ^{3}H-methyl-DNA, followed by liquid scintillation counting.

$$\text{DNA} + {}^3\text{H-methyl-SAM} \xrightarrow{\text{DNA MTase}} {}^3\text{H-methyl -DNA} + \text{SAH}$$

This reaction is usually termed a "SAM-dependent DNA methyltransferase reaction". However, it might be better termed a "DNA-dependent SAM methyl transfer reaction", since the ^{3}H-methyl groups are incorporated into DNA only if a sequence-specific MTase recognition site is present. If one or more DNA MTase sequence recognition sites are present, then the number of such sites can be measured quantitatively based on a linear increase in tritium counts per minute (cpm) of radiolabeled ^{3}H-methyl DNA.

Sequence-specific DNA MTase enzymes have been used only sparingly in megabase mapping experiments. M. Nelson & M. McClelland, "The use of DNA MTase/Endonuclease enzyme combinations for megabase mapping of chromosomes by pulsed field gel electrophoresis", Methods in Enzymology 216:279-303 (1992), incorporated herein by reference. These megabase mapping experiments utilize sequential multi-step DNA methylation and restriction enzyme cleavage reactions to create chromosome fragments in the size range from 50 to 2000 kilobases (kb) that are subsequently analyzed using pulse field gel electrophoresis (PFGE). See D.C. Schwartz & C.R. Cantor, 37 Cell 67 (1984); Gardiner et al., 12 Somatic Cell. Mol. Genet. 185 (1986); Cantor et al., 17 Ann. Rev. Biophys. Chem. 287 (1988). These experiments describe how MTases can be used to increase the apparent specificity of restriction enzymes. However, they do not describe how MTases can be used to "paint" DNA at specific 2 to 8 bp sites using either radioactive (^{3}H-methyl-DNA) or immunochemical detection. Furthermore, the techniques of the above cited references are normally

IDT Exhibit 1020.0560

carried out *in situ* on unsheared chromosomes embedded in agarose plugs and it has proven necessary to require: (1) that the purity of the MTases and the restriction endonucleases be critically controlled, (2) that the number of steps be kept at a minimum, and (3) that the reaction conditions be defined which are compatible with PFGE separation. This technique includes the inherent drawbacks of PFGE such as the time and effort required to prepare high molecular weight substrates and to run pulsed-field gels. The present invention overcomes these problems by creating smaller fragments of genomic DNA or utilizing PCR-type techniques, thus eliminating the need for agarose embedded *in situ* reactions or completely eliminating the need for electrophoresis. Further, the present invention allows for much finer mapping -- down to the nucleotide level -- and exploitation of this information for such uses as diagnostic testing in clinical settings.

In general, DNA MTases have been used neither to map genes nor to detect mutant alleles. The present invention shows how DNA MTases can be used to radio- or immuno-chemically "paint" DNA in order to identify polymorphic sites or mutations. Using DNA MTases, one can determine genotypes with significant improvements in speed, accuracy and specificity over RFLP mapping and other methods utilizing restriction endonucleases. In short, the methods of the present DNA MTase invention are extremely fast, do not require gel electrophoresis or nucleic acid hybridization, and result in simple quantitative and/or qualitative assays. As a result, DNA MTase genotyping of the present invention lends itself to automation. Rapid automated PCR-based MTase genotyping may be especially useful in high throughput diagnostic laboratories or in situations where fast, easy, reproducible results are required, such as clinical laboratory or hospital settings.

The present invention does not require that DNA fragments be separated on agarose gels. Southern transfers, nucleic acid hybridization or radioactive labeling are not required, although these are possible uses of DNA MTases in certain embodiments of the present invention. Most importantly, since MTases do not cut the sugar-phosphate backbone of DNA, there are no DNA fragments that need to be separated. The present invention overcomes the lengthy and laborious dependence on gels and Southern transfers. The present invention allows for quick and quantitative analysis without the need to separate by fragment size using gels or time-consuming Southern transfer techniques. Certain embodiments of the present invention do not require the use of restriction enzymes at all. This independence from endonucleases allows the

IDT Exhibit 1020.0561

diagnostician or researcher to avoid erroneous results due to incomplete restriction enzyme digestions. The present invention also allows for the use of very short PCR products (30-100 bp). Since the PCR reaction cycle times are shortened to a few seconds, the concentrations of amplified DNA are increased, resulting in improved signal-to-noise. Thus the present inventive methods utilizing DNA MTases for genotyping are faster, more accurate, and easier to perform than the methods known to those skilled in the art utilizing bacterial restriction enzymes.

It is an object of the present invention to provide a method of genotyping DNA utilizing a DNA methyltransferase in conjunction with a PCR amplicon so that a quick and accurate determination of a mutation or nucleic acid variation can be determined.

A second object of the invention is to provide a method utilizing a DNA methyltransferase to genotype DNA whereby the method of detection relies on immunochemical (non-radioactive) detection, so that radioactive label is not required.

A third object of the invention is to provide a method of genotyping DNA utilizing a DNA methyltransferase without the need for agarose gel fractionation of DNA or Southern transfer.

A fourth object of the invention is to provide a method of ordered genetic maps of PCR-amplified DNA utilizing biotinylated primers and combinations of DNA methyltransferases and endonucleases. In other words, not only can the presence of MTase sites be determined, but their positions relative to the 5' biotinylated end can be located.

A fifth object of the invention is to provide a method of DNA genotyping utilizing DNA methyltransferases whereby the determination of mutation or variation can be accomplished in a single-container reaction with no need for removal of reaction precursors, thus allowing for an even faster turnaround time.

A sixth objective of the invention is to provide a method utilizing DNA methyltransferases to identify individual alleles associated with disease traits or traits of economic importance and which could further be automated for fast, accurate and economic results.

A seventh objective is to provide a genotyping method which relies upon detecting the presence or absence of DNA MTase recognition sites, so that detection of specific DNA regions or alleles can be automated.

Additional objects and advantages of the invention will be set forth in part in the description that follows, and in part will be obvious from the description, or may be

IDT Exhibit 1020.0562

learned by the practice of the invention. The objects and advantages of the invention will be obtained by means of the instrumentalities and combinations, particularly pointed out in the appended claims.

SUMMARY OF THE INVENTION

To achieve the objects and in accordance with the purpose of the invention, as embodied and broadly described herein, the present invention provides a method for exploiting DNA MTases' ability to methylate DNA at specific nucleic acid sequences. The sequence-specific DNA MTases can be used to detect the presence of specific recognition sites in nucleic acid molecules. Known or unknown sequences can be genotyped using DNA MTases. For example, MTases can be used to detect genetic mutations or polymorphisms or simply to gain DNA sequence information of unknown genes. The ability, or inability, of these DNA MTases to methylate at the sites of mutation or polymorphism are then detected by means which can include immunochemical or radiometric detection methods. Upon detection of the sites of methylation, DNA sequence information can be compiled, analyzed and referenced.

Numerous immunochemical detection methods are known to those of ordinary skill in the art and are readily available through commercial sources such as Fisher Scientific, Pittsburgh, PA; Boehringher Mannheim Corp., Indianapolis, IN; and Vector Laboratories, Burlingame, CA. These immunodetection methods rely upon haptens such as biotin, digoxigenin, or those detected by antibodies or lectins conjugated with enzymes such as alkaline phosphatase, horseradish peroxidase, or β-Galactosidase. See, Christopher Kessler, "The digoxigenin:anti-digoxigenin (DIG) technology -- a survey on the concept and realization of a novel bioanlytial indicator system", Molecular and Cellular Probes, 5:161-205 (1991). Immunochemical detection methods can include, among others, labeling systems which utilize chemiluminescence substrates, chromogenic substrates, rhodamine, and fluorochrome-labels such as fluorescein isothiocyanate and tetramethylrhodamine. Numerous radiometric detection methods are also known to those of ordinary skill in the art and readily available through commercial sources such as Amersham Lifesciences, Arlington Heights, IL and New England Nuclear, Boston, MA. However, the use of sequence-specific DNA MTases *in vitro*, followed by immunochemical detection of methylated DNA has not been previously described. Furthermore, none of these earlier methods allow DNA to be internally labeled at defined 2 to 8 bp sites.

IDT Exhibit 1020.0563

The methods of this invention include as preferred embodiments in which these DNA MTases are utilized: (1) a radiometric ^{3}H-methyl incorporation assay, theoretically also $^{3}H_{3}C^{11}$-, position-labeled methyl assay; (2) a non-radioactive immunochemical detection; (3) a combination of sequence-specific DNA methylation, restriction enzyme digestion, and Southern blotting ("DNA Paint"); and (4) an anchored PCR format in which biotinylated primers are used in combination with DNA MTases and endonucleases. Importantly, this last format allows ordered maps to be rapidly constructed without the need for either electrophoresis or nucleic acid hybridization.

In one embodiment, the invention comprises a method for genotyping an unmethylated DNA fragment produced by PCR amplification. Using PCR primers that flank the region of interest, enough DNA is amplified to conduct a series of sequence-specific DNA MTase and/or restriction endonuclease reactions. The design of PCR primers is known to those of skill in the art and are described in Mullis et al., The Polymerase Chain Reaction, Burkhäuser Publishers, Boston, MA (1994), the disclosure of which is incorporated by reference. Amplified DNA is treated with selected DNA methyltransferases and radioactively-labeled-methyl-SAM. The radioactively labeled methyl group will thus be incorporated into a specific 2 to 8 bp site in the DNA. If appropriate, sufficient DNA may be put into reaction tubes to allow for digestion by selected restriction enzymes. The amount of radioactively labeled methyl DNA will be measured using radiometric assay methods well known to those of ordinary skill in the art. This technique will identify a DNA MTase genotype for each recognition sequence site within the PCR generated DNA fragment.

In a second embodiment of the invention, the genotyping of genes, or alleles of genes, associated with a disease condition, or a trait of economic interest in livestock or plants, can be determined utilizing DNA methyltransferases. DNA is first amplified using PCR primers that flank the gene, or segment of the genome, of interest. The amplified DNA is methylated with a radioactively labeled methyl group at a specific site known to vary from one allele to another. A sequence-specific DNA MTase can then be used to detect a defined genetic variation or mutation. Methylation of the DNA may be accomplished either at the same time as the amplification of the PCR fragment using thermostable MTases or following the amplification of the PCR fragment of interest if no thermostable methyltransferase is available for the site of interest. The presence or absence of the variation or mutation in question is determined by radiometric assay. If

IDT Exhibit 1020.0564

adequate signal-to-noise is achieved, then the wild type (a/a) state, heterozygous (a/b) state, or homozygous (b/b) state can be detected for that site.

In a third embodiment, the presence or absence (and the quantification of multiple sites) can be determined utilizing immunochemical detection rather than radioactive detection of methylated DNA. Detection of methylated DNA can be accomplished utilizing the above methods or other methods, as will be evident from this disclosure, by enzymatically incorporating non-radioactive methyl groups utilizing DNA methyltransferase. The resulting DNA can be spotted on to a fixed surface such as nitrocellulose or nylon membranes and adhered through routine methods known in the art. The DNA can also be separated by size on an agarose gel and transferred onto membranes using the Southern transfer technique. The membranes can then blocked to prevent background signal due to non-specific noise. In one example of this embodiment, a nylon membrane containing methylated DNA fragments are UV-irradiated and then incubated in the presence of anti-^{6}-methyladenine rabbit immune serum in buffered nonfat milk blocking solution. The filters are washed and treated with a solution containing alkaline-phosphatase conjugated goat anti-rabbit IgG or antiserum. After washing, NBT/NCIP substrate for alkaline-phosphatase is added. Methylated ^{6m}A DNA is detected as blue spots or bands on a white nylon filter background.

In a fourth embodiment (sequence-specific "DNA Paint" Format), a combination of sequence-specific DNA methylation, restriction digestion, and Southern blotting technique allows methylated DNA fragments to be selectively stained. Such "methylase painting" of DNA fragments can be accomplished by digesting DNA with restriction endonucleases, and methylating the resultant DNA fragments with DNA methylases that have a <u>different</u> sequence specificity site. In general, DNA methylation is conducted prior to endonuclease digestion. However, the order can usually be reversed. The resultant fragments are then separated by size utilizing an agarose gel. After electrophoresis, the DNA is transferred from the gel to a membrane (nylon or nitrocellulose) using Southern transfer technique or other transfer methods known to those of ordinary skill in the art. The membrane-bound methylated fragment patterns are detected using radiometric techniques (if radioactively-labeled-methyl groups were incorporated) or immunochemical detection methods. This embodiment results in a DNA membrane or image exhibiting a unique pattern, or "DNA Painting", for each genome or DNA fragment of interest. Simply put in binary code, any particular DNA

IDT Exhibit 1020.0565

fragment is either radioactively or immunochemically labeled (1) or else it is not labeled (0). The presence or absence of methylated 2 to 8 bp DNA MTase recognition sites appear as colored bands or spots.

In a fifth embodiment, the invention comprises a method whereby the PCR fragment is anchored to a solid matrix, for example, a magnetic bead or Streptavidin SPA bead, so that radioactively labeled methyl groups can be counted using appropriate radiometric assays.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a graphic representation of a single reaction mixture for thermocycling wherein DNA is amplified using polymerase chain reaction (PCR) and simultaneously methylation is carried out using a thermostable DNA methyltransferase.

Figure 2 shows a graphic representation of a single reaction mixture for thermocycling wherein DNA is amplified using strand displacement amplification and methylated using DNA methyltransferase, simultaneously.

Figures 3A and 3B are double-stranded DNA sequences from the human platelet antigen glycoprotein (GPIIIa) gene. Figure 3A shows the 362 bp fragment containing the mutated M.CviRI recognition sequence. Figure 3B shows the 63 bp fragment containing both the wild type M.SssI recognition sequence and the mutated CviRI sequence. Primers used on opposite sides of the GPIIIa CG to CA mutation are also indicated as shown.

Figure 4A is a graphic representation of a partial digest band of homozygous Yuk $^{b/b}$ compared to complete restriction digestion of wild type, heterozygous or homozygous Yuk alleles.

Figure 4B is a photographic representation of an agarose gel containing a partial digest of Yuk $^{b/b}$ compared to Yuk $^{a/a}$, Yuk $^{a/b}$ and Yuk $^{b/b}$ fragments generated by the PCR amplified DNA MTase genotype method. The far left lane is uncut DNA, the next lanes are genotype a/a, a/b, and bb. The far right lane represents a partial b/b digestion that could be mistaken for an a/b genotype.

Figure 5 is a diagram showing the actual sequence of two alleles, one from HSV-1 and the other from HSV-2. This sequence shows how it is possible to PCR amplify and map an amplicon using a combination of MTases and restriction enzymes.

IDT Exhibit 1020.0566

Figure 6 is a photograph of a nylon membrane "dot blot" showing alkaline-phosphatase immunostaining of ^{6m}A-methylated DNA using rabbit anti-^{6m}A immunoglobin followed by alkaline-phosphatase goat anti-rabbit antibodies and developed with BCIP substrate. Various amounts of ^{6m}A-methylated Chlorella virus NY2A DNA, ^{6m}A-methylated Chlorella virus PBCVI DNA or unmethylated bacteriophage T7 DNA were spotted onto the nylon sheet (left to right: 10 ng, 1 ng, 100 pg, 10 pg, 1 pg, 100 fg, 10 fg, 1 fg of DNA). The DNA was fixed to the nylon by UV light treatment and then the methylated DNA was immunochemically detected. This figure shows clearly that ^{6m}A methylated NY2A (37% ^{6m}A) and PBCVI (3% ^{6m}A) DNAs are specifically immunostained; in contrast unmethylated phage T7 DNA does not stain using anti-^{6m}A antibodies.

Figures 7A and 7B show an early "purple DNA Paint" experiment.

Figure 7A is a photograph of an ethidium bromide stained agarose gel showing lambda dam$^+$ and dam$^-$ phage DNA. The lambda dam- DNA is naturally methyl-deficient (negative control; lanes 5-8), and the lambda dam$^+$ DNA is naturally methylated (lanes 1-4). Lanes 2 and 6 are digested with EcoRI, lanes 3 and 7 are digested with BamHI, and lanes 4 and 8 are digested with HindIII.

Figure 7B is a photograph of the same gel after transfer of DNA to a nylon membrane. DNA is painted purple after immunochemical detection using rabbit anti-^{6m}A-antibodies and alkaline phosphatase-conjugated goat-anti-rabbit IgG. Methylated DNA fragments appear as dark purple/black bands on a light purple background. As shown in lanes 2, 3 and 4, only those bands containing ^{6m}A-methylated G ^{6m}ATC sites are immunostained. This figure shows that methylated DNA fragments can be detected after electrophoresis and transfer to a nylon filter. Only those fragments with at least 6 to 8 ^{6m}A sites can be detected. Improved sensitivity was subsequently achieved, as shown in Figures 9 to 12.

Figures 8A and 8B show the immunochemical detection of *in vitro* methylated DNAs using rabbit anti-^{6m}A antibody and immunoenzymatic staining. Figure 8A is a photograph of the ethidium bromide stained gel and Figure 8B is a photograph of the anti ^{6m}A immunophosphatase stained nylon membrane that corresponds to the gel in Figure 8A.

Figures 9A and 9B show a sequence-specific "purple DNA Paint" experiment using phage T7 DNA methylated in vitro using M.dam (G^{6m}ATC) MTase.

IDT Exhibit 1020.0567

Figure 9A is a photograph of an ethidium bromide stained agarose gel containing phage T7 DNA. The unmethylated T7 DNA (negative control), G^{6m}ATC-methylated DNA (T7 M.dam) and TCG^{6m}A-methylated DNA (T7 M.Taq I) were digested using DpnII, TaqI, SpeI or NruI endonucleases(lanes A-L, respectively).

Figure 9B is a photograph of a nylon membrane following transfer of the gel in Figure 9A. Figure 9B further shows the immunochemical detection of the ^{6m}A methylated DNA. This figure shows that an *in vitro* sequence-specifically methylated DNA fragment can be detected after electrophoresis and transfer to a nylon filter.

Figure 10A, 10B and 10C show a sequence-specific "purple DNA Paint" experiment of an *in vitro* methylated amplicon from human ß-globin gene.

Figure 10A is a photograph of of an ethidium-bromide stained agarose gel showing a 536 bp amplicon from human ß-globin gene. The amplicon was digested with Nco I (C↓CATGG; lanes 2 and 6), Ava II (G↓GWCC; lanes 3 and 7), or Dpn II (↓GATC; lanes 4 and 8).

Figure 10B is a photograph of an anti- ^{6m}A immunostained nylon membrane following transfer of the gel is shown in Figure 10A. The amplicon was not methylated (lanes 1-4), or else methylated at G^{6m}ATC sites using M.dam MTase (lanes 5-8). Figure 10B shows sequence-specific immunophosphatase/BCIP immunochemical detection of DNA fragments containing G^{6m}ATC sites.

Figure 10C is a map of the human ß-globin 536 bp amplicon showing the expected sizes of methylated DNA fragments.

Figures 11A, 11B and 11C show a sequence-specific "red DNA Paint" of a 1.85 Kb PCR amplicon from the human ß-globin gene. The sensitivity of immunostaining is sufficient to detect a single M.EcoRI MTase site in less than 1 µg of PCR amplified DNA.

Figure 11A is a photograph of an ethidium-bromide stained agarose gel showing a 1850 bp amplicon from human ß-globin gene. The amplicon was digested with EcoRI (G↓AATTC; lanes 3 and 7), BsrGI (T↓GTACA; lanes 4 and 8), BamHI (G↓GATCC; lanes 5 and 9).

Figure 11B is a photograph of a ^{6m}A-methylated DNA fixed to a nylon membrane following transfer of the gel in Figure 11A. The amplicon was not methylated (lanes 2-5) or else methylated *in vitro* by M EcoRI MTase (GA^{6m}ATTC; lanes 6-9). Figure 11B further shows the sensitivity of immunochemical detection of the GA^{6m}ATTC-methylated DNA using rabbit anti^{6m}A antibodies followed by alkaline-

IDT Exhibit 1020.0568

phosphatase conjugated goat anti-rabbit antibodies and fast red substrate (Pierce, Rockford, IL).

Figure 11C is a map of the 1.85 Kb ß-globin amplicon.

Figures 12A, 12B and 12C show a sequence-specific "red DNA Paint" of a PCR amplified product specifically methylated at ^{4m}C, ^{5m}C or ^{6m}A. The sensitivity of the immunostaining is sufficient to detect a single M.EcoRI MTase site (GA^{6m}ATTC) in 1 µg of amplified DNA.

Figure 12A is a photograph of an ethidium-bromide stained agarose gel showing a 536 bp amplicon from human ß-globin gene. The amplicon was methylated with M·HaeIII MTase (GGC^{5m}C; lanes 6-8), M.dam MTase (G^{6m}ATC; lanes 9-11), M·MvaI MTase (C^{4m}CWGG; lanes 12-14), or not methylated (lanes 1-5). The amplicon was digested with NcoI (C↓CATGG; lanes 2, 7, 10 and 13), DpnII (↓GATC; lanes 3 and 11), BstNI (CC↓WGG; lanes 4 and 14), HaeIII (GG↓CC; lanes 5 and 8). Note that amplicons in lanes 4, 8 and 11 are not digested because cleavage by BstNI, HaeIII, and DpnII is blocked by MTases, M.MvaI, M.HaeIII and M.dam, respectively.

Figure 12B further shows the immunochemical detection of the ^{6m}A methylated DNA using rabbit anti^{6m}A antibody followed by alkaline-phosphatase anti-rabbit secondary antibody and fast red substrate (Pierce, Rockford, IL).

Figure 12C is a map of the ß-globin 536 bp amplicon showing the expected sizes of methylated DNA fragments.

DETAILED DESCRIPTION OF THE INVENTION

Reference will now be made in detail to the presently referred embodiments of the invention, which together with the following examples, serve to explain the principles of the invention.

The DNA of bacteria that encode type II restriction endonucleases are protected from endonuclease catalyzed auto degradation by corresponding isospecific methylases. Specific DNA methyltransferases "...recognize the same nucleotide sequences as their cognate endonucleases and transfer a methyl group from S-adenosyl methionine to specific adenine (forming N^6-methyladenine) or cytosine (forming 5-methylcytosine) residues within the site." Maxine Singer and Paul Berg, Genes and Genomes 253 (University Science Books 1991). As a result, the restriction endonuclease cannot cleave at that modified (methylated) recognition site. In addition, many bacterial species contain enzymes, the *dam* and *dcm* methylases, which do not have a cognate restriction

IDT Exhibit 1020.0569

endonuclease and yet will methylate DNA at well-defined recognition sequences. Roger L.P. Adams, "Review Article: DNA Methylation--The Effect of Minor Bases on DNA-Protein Interactions", 265 Biochem. J. 309, 309-10 (1990). As new DNA methyltransferase genes are isolated and sequenced, predicted amino acid sequences specific to each type of methylase are identified as conserved regions. Id.

DNA methyltransferases are available through commercial suppliers such as New England Biolabs, Beverly, MA; MBI Fermentas, Flamborough, ON, Canada; and Megabase Research Products, Lincoln, NE. In addition, the Type II DNA methylases (S-Adenosylmethionine: DNA Methyltransferases) have been purified from a number of different sources and purification procedures have been published and are therefore well known to those of ordinary skill in the art. See Michael Nelson and Michael McClelland, "Purification and Assay of Type II DNA Methylases", 155 Methods in Enzymology 332, 332-341 (1987). Extensive lists of known DNA methyltransferases have been published such as that in Table III of M. McClelland et al., "Effect of Site-Specific Modification on Restriction Endonucleases and DNA MTases", 22 Nucleic Acids Research 3640-3659 (1994). See also R.J. Roberts and D. Macelis, "REBASE -- Restriction Enzymes and Methylases", 22 Nucleic Acids Research 3626-3639 (1994); Michael Nelson et al., "DNA Methyltransferases and DNA Site-Specific Endonucleases Encoded by Chlorella Viruses", (Jost, J.P., and Sulsz H.P., eds., DNA Methylation, Birkauser, Basel) 186-211 (1993); and Geoffrey G. Wilson & Noreen E. Murray, "Restriction and Modification Systems", 25 Ann. Rev. Genetics 585-627 (1992).

DNA methyltransferases can be selected for use in the present invention based on known sequences of regions of DNA from genes or intergenic regions of interest. DNA methyltransferases can also be selected randomly, in much the same way that restriction endonucleases are randomly chosen for mapping and genotyping experiments. When determining the genotype of a specific region of DNA or gene of interest, a DNA methyltransferase is chosen that will methylate a specific recognition sequence site within that DNA sequence. If the sequence is not known, then the restriction endonucleases (R) and their cognate DNA methylases (M) can be utilized to determine both the presence or absence of their respective R or M recognition sites and the order in which they occur. These and other embodiments will be more fully detailed in the disclosure that follows.

The present invention relates to the use of DNA MTases to detect genetic mutations or polymorphisms at specific DNA sites. A single nucleotide point mutation

IDT Exhibit 1020.0570

or polymorphism can create or destroy the sequence-specific recognition site of a DNA MTase and its companion restriction endonuclease. The ability of a DNA MTase to methylate at the site of a mutation or polymorphism is detected by means which can include immunochemical detection or radiometric detection methods. Once the sites of DNA methylation are detected, sequence data can be compiled, analyzed, and referenced.

The invention also relates to the use of DNA methyltransferases for genotyping specific alleles within genes of interest, genotyping or DNA paint analysis of genomes or specific regions within genomes, and creation of ordered maps for specific regions within genomes. As used here, the term "genotype" means the particular combination of sequence-specific sites present or absent in individual genome or fragment of DNA from an individual genome. The methods of the present invention can be used to determine not only those same DNA methylase sequence-specific recognition sites within a DNA sample, but also those sites associated with the companion restriction endonuclease. DNA MTases also can be used in embodiments whereby point mutations responsible for heritable diseases or traits can be detected, including those which occur at CG sites. This ability to recognize CG sites is unique to DNA methyltransferases and provides an additional improvement over the use of restriction endonucleases alone in genotyping DNA. Three fourths of known mutations responsible for heritable human diseases occur at CG sites (D. Cooper & H. Youssoufin, "The CpG dinucleotide and human genetic disease", Human Genetics, 78:151-165 (1988)). As shown in Table 3, such frequent CG to TG mutations can be detected using M.SssI MTase, which methylates DNA specifically at hypermutable 2 bp CG sites.

In radiometric ^{3}H-methyl-incorporation assays, DNA methyltransferase genotyping is carried out using a sequence-specific methyltransferase to transfer ^{3}H-methyl groups from ^{3}H-methyl-SAM to DNA. The incorporation of this radioactively labeled methyl group is monitored by trichloroacetic acid (TCA) precipitation followed by liquid scintillation counting. If several different sequence-specific DNA methylation reactions are carried out, then the pattern of ^{3}H-methyl incorporation defines the DNA MTase genotype.

A DNA MTase genotype is a "fingerprint" of DNA based upon the presence or absence of MTase recognition sites. Similarly, a restriction modification (R-M) genotype is a DNA fingerprint in which both restriction (R) and methylase (M) sites are monitored. R-M genotyping of 5' biotinylated allows ordered maps to be rapidly

IDT Exhibit 1020.0571

constructed. As a result, alleles or MTase isotypes can be expressed as binary code words. For example, the presence or absence of a single restriction or modification enzyme recognition site can be expressed in binary code as "1" for the presence of the site or "0" for the absence of the site. When looking at the inheritance of an allele associated with a disease state, the resulting DNA can be expressed in ternary code as "0" for homozygous wild type alleles, "1" for heterozygous wild type/disease alleles, or "2" homozygous for the alleles associated with the expression of the disease or trait of interest. This conversion of genotypes into an easily readable digital code allows for larger scale, automated screening procedures to be utilized. Moreover, in the case of radioactive MTase genotyping or combination R-M genotyping, the quantitative assay allows higher order codes to be employed.

Radiometric ^{3}H-methyl incorporation can be utilized to genotype entire genomes or smaller DNA fragments of varying lengths. For the analysis of entire genomes, DNA MTases can be used in a method similar to RFLP analysis. In general, RFLP analysis is a technique well known to those skilled in the art. See for example, Erlich, U.S. Patent No. 4,582,788 issued April 15, 1986 and Gasella, U.S. Patent No. 4,666,828 issued May 19, 1987, Frosserd, U.S. Patent No. 4,772,549 issued September 20, 1988, and Frossert, U.S. Patent No. 4,861,708 issued August 29, 1989, all of which are disclosed for example and are incorporated herein by reference. Broadly speaking, the RFLP technique involves obtaining the DNA to be studied, digesting the DNA with restriction endonucleases, separating the resulting fragments, and then detecting the fragments of altered electrophoretic mobility associated with those genes of interest.

In the present invention, DNA fragments are detected not by using electrophoretic markers associated with genes, but by utilizing sequence-specific DNA methyltransferases which incorporate radioactively or immunochemically labeled methyl groups into defined 2 to 8 bp DNA sequence-recognition sites.

In the simplest format, DNA is enzymatically methylated; and then the presence or absence of 2 to 8 bp MTase sites is detected either radiometrically or immunochemically. The presence (1) or absence (0) of MTase recognition sites is a "DNA MTase genotype". However more complex "R-M genotypes," based upon combination, of restriction (R) or modification (M) enzymes are possible. For example, DNA can be methylated at specific sites and then digested by a restriction enzymes with a different sequence specificity. After gel electrophoresis and Southern blotting to nylon filters, only those fragments containing sequence-specific radioactively labeled DNA

IDT Exhibit 1020.0572

methyl groups are detected via autoradiograms. A variation of this method can be performed by first digesting with a sequence-specific restriction endonuclease followed by incorporation of the radioactively labeled methyl group utilizing a sequence-specific methyltransferase which recognizes a different recognition sequence than the restriction enzyme used.

Another embodiment of the present invention utilizes amplification of desired DNA fragments through the use of standard techniques, such as polymerase chain reaction (PCR). In general, the PCR technique is well known to those skilled in the art and the following references are disclosed for example. The PCR technique is described in Mullis et al., U.S. Patent No. 4,683,195, issued July 28, 1987, Mullis, U.S. Patent No. 4,683,202 issued July 28, 1987, Mullis et al., U.S. Patent No. 4,800,159 issued January 24, 1989, Delfand et al., U.S. Patent No. 4,889,818 issued December 26, 1989, and Columbus et al., U.S. Patent No. 4,902,624 issued February 20, 1990, all of which are incorporated herein by reference.

In this embodiment DNA MTase genotyping is carried out using two sequential reactions. First, DNA is amplified using PCR(Mullis et al. 1994):

$$\text{(1)DNA + primers + dNTP's} \xrightarrow{\text{Taq Polymerase}} (\text{DNA})_n + \text{dNTP's}$$

Second, amplified DNA is methylated using ^{3}H-methyl-SAM and a sequence-specific DNA MTase:

$$\text{(2)DNA + } ^3\text{H-methyl-SAM} \xrightarrow{\text{DNA MTase}} ^3\text{H-methyl-DNA + SAH}$$

This PCR/MTase assay requires neither gel electrophoresis nor hybridization. It is therefore a very fast method for detecting polymorphic sites in DNA. Furthermore, if thermostable polymerase and thermostable MTase enzymes, such as Taq Polymerase and M.TaqI, are used then the DNA amplification and methylation reactions can be coupled in a single reaction mixture. The two major advantages of the coupled PCR/MTase assay are its speed and simplicity. DNA, primers, DNTP's, ^{3}H-methyl-SAM, Taq Polymerase, and a thermostable MTase are placed in a single reaction mixture for thermal cycling. (See Figure 1) After several cycles of thermal denaturation, annealing, and elongation, the DNA fragments are both amplified and ^{3}H-methylated at defined sequences. ^{3}H-methyl-DNA can then be TCA precipitated, spotted

IDT Exhibit 1020.0573

onto a filter paper, washed, and counted. These counts can be expressed as raw data (cpm) or mathematically as either a binary code or a ternary code as described above.

As an alternative to PCR amplification by thermocycling, it is possible to methylate DNA amplified isothermally at 37°C by strand displacement amplification (SDA) (Walker et al., 1992). In SDA, the template DNA is mixed with 4 primers, dCTP, dTTP, dGTP, and α-S-dATP at 94°C. Following denaturation at 94°C, HincII restriction endonuclease and Klenow exo- Polymerase are added whereby α-S-substituted DNA is repeatedly nicked and amplification occurs isothermally at 37°C by strand displacement.

It is also possible to couple SDA amplification and DNA methylation in a single reaction mixture by a modification of the above procedure. In this case, a DNA MTase and ^{3}H-SAM could be added at the same time as the HincII restriction endonuclease and the Klenow exo-Polymerase. This would allow for methylation to occur in the same reaction mix as the amplification of the DNA. (See Figure 2) A large number of sequence-specific DNA MTases are suitable for isothermal SDA/MTase genotyping because thermal stable enzymes are not required.

In both the PCR/MTase assay and the SDA/MTase assay, the presence or absence of a sequence-specific mutation or variation can be detected by the presence or absence of radioactively labeled methyl groups following TCA precipitation, spotting onto filter paper, washing, and scintillation counting.

In a more preferred embodiment of this method, short (30-100 bp) PCR products are used. In RFLP mapping of amplified DNA, the size of the amplicon is usually constrained by the need for fragments that are separable by gel electrophoresis and can be detected by ethidium bromide staining. DNA fragments are typically several hundred base pairs long. However, in DNA MTase genotyping, improved speed and sensitivity are possible when short (30-100 bp) amplicons are methylated.

The method is faster when amplification of short PCR or SDA products is used. For example, if a central 20 bp polymorphic DNA sequence is amplified using two 20-mer primers, then the size of the resulting amplicon is 60 bp (20 bp from each primer plus 20 bp of amplified DNA). At an elongation rate of about 50 nucleotides per second at 72°C for Taq Polymerase, only 1 to 2 seconds are needed to replicate the 60 bp amplicon during each cycle of amplification. In contrast, typical PCR thermal cycling products require elongation times of approximately 1 minute or more per cycle.

IDT Exhibit 1020.0574

In addition, because the molecular weight of the amplified DNA product is smaller, its molar concentration is proportionately higher. For example, 1 μg of a 200 bp DNA fragment in a 25 μl reaction has a molar concentration of approximately 3 μM, but 1 μg of a 60 bp fragment in the same volume has a concentration of approximately 9 μM. If each of these DNA fragments has a single MTase recognition site, then the number of methyl groups incorporated into this shorter 60 bp fragment will be three-fold greater. In short, the use of short PCR products allows for faster amplification times and improved sensitivity as evidenced through stronger signals due to higher molar concentration.

Another embodiment of the present invention is the use of non-radioactive immunochemical detection of methylated DNA. Prior to the present invention, the use of immunochemical detection following *in vitro* DNA methylation was unknown. Immunochemical detection methods are well known to those of skill in the art and the creation of antibodies, including monoclonal antibodies, to modified nucleotides has been the subject of many publications. See, e.g. D.C. Eichler & D.G. Glitz, "Nucleotide-Specific Antibodies as Potential Blocking Agents in the Structural Analysis of Nucleic Acids", 335 BBA 303-317 (1974); C.W. Achwal and H.S. Chandra, "A sensitive immunochemical method for detecting 5mC in DNA fragments", FEB Letters, 150:469-472 (1982); B.F. Erlanger & S.M. Beiser, "Antibodies to methylated bases in nucleic acids", Proc. Nat. Acad. Sciences, USA, 52:68 (1964); Brooks et al., "An improved method for the purification of IgG monoclonal antibodies from culture supernatants", Journal of Immunological Methods, 155:129-32 (1992); Sano et al., "Detection of heavy methylation in human repetitive DNA subsets by a monoclonal antibody against 5-methylcytosine", Biochemica et Biophysica Acta., 951:157-65 (1988). A variety of non-radioactive formats are currently available including the use of infra-red tags, fluorescent dyes, chemiluminescence, and alkaline-phosphatase conjugated antibodies. A more preferred embodiment utilizes specific antibodies which can be used to immunochemically detect ^{6m}A-or ^{4m}C-methylated DNA. Although there are several references which teach the creation of antibodies to ^{5m}C, (See e.g. C.W. Achwal & H.S. Chandra, "A Sensitive Immunochemical Method for Detecting 5mC in DNA Fragments", Vol. 150 No. 2, FEBS Letters 469-472; C.W. Achwal et al., "Estimation of the Amount of 5-Methylcytosine in *Drosophila Melanogaster* DNA by Amplified ELISA and Photoacoustic Spectroscopy", Vol. 3 No. 2 EMBO Journal, 263-266 (1984); J.A. Vilpo et al., "Radioimmunoassay of 5-Methyl-2'-Deoxycytidine: A Method for the Quantification

IDT Exhibit 1020.0575

of DNA Methylation", 75 Journal of Immunological Methods, 241-246 (1984)), it was a surprising result to find poor immuno-reactivity in anti-^{5m}C antibodies in the present invention due to poor signal-to-noise ratio. In addition, the creation of ^{5m}C antiserum in rabbits has proven to be very difficult. However, this technical problem may be due to a tolerance-type lack of response in the rabbits after injection of the ^{5m}C conjugates. Therefore, in some examples of the present invention, the use of anti-^{6m}A- or anti-^{4m}C-antiserum was preferred. One skilled in the art could appreciate that anti-^{5m}C antibodies suitable for immunostaining of ^{5m}C methylated DNA are also possible.

In one embodiment utilizing immunochemical detection methylated DNA fragments can be adhered to a solid matrix. For example, the methylated DNA fragments can be run out on an agarose gel and adhered to a solid nylon membrane using Southern blotting methods. Another example includes the use of methylated DNA fragments where one of the primers used to create the methylated amplicon was biotinylated and this resulting biotinylated fragment is captured on a Streptavidin-coated ELISA plate (Boehringer). The presence of the methylated nucleoside can then be detected utilizing anti-^{6m}A or anti-^{4m}C immune serum followed by an anti-anti-^{6m} antiserum IgG or anti-anti-^{4m} immune serum which has been conjugated with either alkaline phosphatase or peroxidase. Colorimetric assays can then be conducted. In the above examples, NBT/BCIP substrate for alkaline-phosphatase would be added to the nylon membrane whereby blue bands, or in the case of dot blots--blue spots, on the white background of the nylon filter would be observed. In the case of a Streptavidin coated ELISA plate, peroxidase-conjugated goat anti-rabbit IgG would be detected through the addition of ABTS® (Boehringer) whereby the resultant color would be measured at 450 nm using an ELISA plate reader.

Yet another embodiment of the present invention utilizes the use of DNA methylation and restriction endonuclease digestion followed by immunochemical detection or "DNA paint". In the "DNA paint" format, DNA is first methylated at specific sites and then digested by restriction enzymes which possess a different sequence specificity. Conversely, the DNA can be first restriction endonuclease digested followed by methylation at different sequence-specific sites. After digested/methylated DNA undergoes gel electrophoresis and Southern blotting to nylon filters, only those fragments containing sequence-specific DNA methyl groups are immunochemically detected as described above.

IDT Exhibit 1020.0576

Yet another embodiment of the invention utilizes an anchored PCR format in which biotinylated primers in combination with DNA MTases and restriction endonucleases are used. One advantage of using a biotinylated anchored PCR format is that generation of ordered maps can be created by using a solid matrix such as Streptavidin-coated magnetic beads (MGP, Lincoln Park, NY). In this method, the DNA is PCR amplified using a 5' biotinylated forward primer and non-biotinylated reverse primer. The resulting amplicons are then methylated using sequence-specific DNA methyltransferases and ^{3}H-methyl-SAM. If appropriate buffer conditions are employed, the DNA fragment can be enzymatically ^{3}H-methylated and digested with sequence-specific endonucleases in the same reaction mix. Each amplicon is anchored to Streptavidin-coated magnetic beads via the biotin present in the forward primer. If a restriction enzyme cuts between the ^{3}H-methyl site and the biotin anchored 5' end then the ^{3}H-methyl group is no longer attached to the bead. The ^{3}H-methyl groups in the remaining DNA are measured by TCA precipitation and liquid scintillation counting as described above. Ordered maps are constructed for restriction enzymes (R) and MTases (M) using an (R+1) x (M+1) coding. By looking at the results in matrix format with restriction endonucleases along one axis and methyltransferases along the opposite axis, one can determine the order of the restriction and modification sites as oriented by the 5' biotinylated forward primer.

Another embodiment of this anchored PCR format utilizes SPA (scintillation proximity assay) beads. N. Bosworth & P. Towers, "Scintillation Proximity Assay", Nature 341:167-168 (1989), incorporated herein by reference. This method is even faster than the magnetic bead detection of ^{3}H-methyl-DNA because removal of ^{3}H-methyl-SAM or other sample processing steps are not required prior to the detection of the H^{3}C-labeled DNA. In SPA, reaction products labeled with low energy beta emitters such as ^{3}H are captured on the surface of Streptavidin-coated fluoromicrospheres (SPA beads). Since beta particles from tritium decay travel only about 1 micrometer in water, unincorporated ^{3}H-labeled precursor molecules are, on the average, too far away from the SPA beads to excite fluorescence. However, biotinylated ^{3}H-methyl labeled DNA reaction products are tightly bound to Streptavidin SPA beads and cause the scintillation beads to emit light. SPA assays are extremely fast because no sample processing or separation of unincorporated radiolabeled precursors is required. In an SPA format, ^{3}H-methyl-SAM does not need to be removed. Time consuming, labor intensive steps such as TCA precipitation,

IDT Exhibit 1020.0577

filter processing, and the addition/disposal of scintillation fluid are unnecessary. If biotinylated primers are used for PCR (or SDA) amplification, followed by sequence-specific methylation using ^{3}H-methyl-SAM, then biotinylated ^{3}H-methyl-DNA can be captured on Streptavidin beads and counted directly as an aqueous suspension. Using coupled DNA amplification/methylation reactions and SPA detection, it is possible to diagnose biologically and medically important pathogens or traits in less than thirty minutes. If combinations of restriction enzymes (R) and methylases (M) are employed in an anchored Streptavidin-SPA bead format, then ordered maps of R-M sites in 5' biotinyulated DNA can be rapidly constructed. Binary (0,≥1) or ternary (0,1, ≥1) signal encoding of bound ^{3}H-methyl-DNA signals is readily decoded using an (R+1) x (M+1) array.

It is to be understood that one of skill in the art can utilize the knowledge of DNA methyltransferases and restriction endonucleases and the methylation and restriction sites of those MTases and endonucleases in conjunction with the teachings of the present invention to analyze and/or exploit the DNA genotype of any sequence of DNA underlying different methyltransferase/endonuclease/genome or gene combinations or methyltransferase/genome or gene combinations. It is also to be understood that the application of the teachings of the present invention to a specific problem or environment will be within the capabilities of one having ordinary skill in the art in light of the teachings contained herein. It is to be further understood that all citations to articles, patent applications, patents, etc., herein are hereby expressly incorporated by reference. The examples of the products and processes of the present invention appear in the following examples.

EXAMPLE 1
DNA MTase genotyping of pBluescript DNA.

A preliminary experiment was carried out to test the feasibility of DNA MTase genotyping using pBluescript plasmid (Stratagene) in a simple model system. T3 and T7 DNA sequencing primers (Stratagene) were used to amplify 10 ng of the polylinker region of pBluescript SK+ DNA using Taq Polymerase. About 1 μg of PCR-amplified plasmid DNA was untreated or else digested by EcoRI, BamHI, or PstI endonucleases (10 units, 2 hours). These DNA's were then phenol extracted, alcohol precipitated in 1.5 ml Eppendorf tubes and resuspended in 30 μl of 15 mM Tris-HCl, pH7.5, 5 mM EDTA, 100 μg/ml BSA, approximately 2 mM ^{3}H-methyl-SAM (New England Nuclear, approximately 85 Ci/mmole). 5 units of M.EcoRI or M.BamHI was added and

IDT Exhibit 1020.0578

methylation was carried out for one hour at 23°C. The amount of ^{3}H-methyl-DNA was measured by TCA precipitation and liquid scintillation counting. M. Nelson & M. McClelland, "Purification and Assay of Type II DNA Methylases", Methods in Enzymology 155:332-341 (1987), incorporated herein by reference.

This initial feasibility experiment shows that the presence or absence of DNA MTase sites in PCR-amplified DNA can be determined by ^{3}H-methyl incorporation assay.

TABLE 1
DNA MTase Genotyping of 191 Base Pair PCR Amplified Polylinker Region of pBlueScript DNA

DNA Pre-Digested by	TCA Precipitable ^{3}H-Methyl-DNA Counts per Minute		
	No MTase	M.EcoRI	M.BamHI
Undigested	574	85,043	69,719
EcoRI	750	770	1306
BamHI	661	32,830	999
PstI	836	27,674	817

Secondly, the results of this DNA MTase genotyping experiment conclusively show that the specificity of DNA MTase genotyping is the same as that of RFLP mapping: DNA sites precut by EcoRI or BamHI restriction endonucleases cannot be methylated by M.EcoRI or M.BamHI MTases, respectively. (See Table 1). Third, the signal-to-noise performance using a ^{3}H-methyl-DNA incorporation assay is excellent. Using M.EcoRI (GA^{6m}ATTC) or M.BamHI (GGAT^{4m}CC), sequence-specific incorporation of 70,000 to 85,000 cpm of ^{3}H-methyl groups into 1 μg of PCR amplified 191 base paired DNA is observed against a background of approximately 600 to 800 counts per minute (cpm). For approximately 200 base pair amplicons, as little as 1 μg of PCR amplified DNA could be easily detected. Even greater sensitivity can be achieved if shorter amplicons are used. Finally in some cases, predigestion by an endonuclease interferes with methylation by a DNA MTase that has a different sequence specificity. For example, pBluescript DNA that has been digested by EcoRI is methylated poorly by M.BamHI. The proximity of "frayed" EcoRI DNA ends to M.BamHI MTase sites interferes with methylase catalysis. This steric effect allows the relative locations of certain DNA MTase sites to be identified and can be used to confirm the authenticity of PCR products.

IDT Exhibit 1020.0579

EXAMPLE 2
DNA MTase Genotyping of a Heritable Human Disease

MTase genotyping technique can be used to rapidly diagnose heritable human diseases, such as neonatal alloimmune thrombocytopenia (NAT), a childhood blood disease which affects ~1:800 babies.

NAT is a blood-borne maternal-fetal immune disease that affects approximately 40% of babies in neonatal intensive care units. In NAT, the mother produces antibodies against her child's blood platelet antigens. Affected infants are born with bruises and are prone to hemorrhages. If left untreated, about one-fourth of these babies die of cranial hemorrhages or are neurologically impaired (Castle et al, 1986). However, if properly diagnosed, NAT infants can be treated by transfusion or washed (antibody-free) blood platelets from their mothers (Abnor et al., 1969).

The mother of an NAT child is deficient in platelet glycoprotein antigens of the GPIb/GPII/GPIIIa complex; but the baby inherits genes for immunoreactive platelet antigens from its father. Four alleles of human platelet antigen (HPA) are known: HPA-1 (Zw), HPA-2 (Ko), HPA-3 (Bak), or HPA-4 (Yuk).

In mutant HPA-4 ($Yuk^{b/b}$) homozygotes the GPIIIa glycoprotein gene is mutated from TGCG at position 526 to TGCA. Like most mutations responsible for heritable human diseases (Cooper & Youssoufian, 1988), this mutation is due to $^{5m}CG \rightarrow TG$ deamination on the lower DNA strand. The loss of CG (M.SssI) site from Yuk^a DNA corresponds to a gain of a TGCA (M.Cvi RI site)in Yuk^b DNA. As a result, wild type Yuk (Yuk^a) and mutant (Yuk^b) alleles can be genotyped by PCR amplification and DNA methylation using M.SssI and/or M.CviRI. The two step DNA amplification/methylation method requires less than two hours.

DNA was isolated from blood samples of wild type ($Yuk^{a/a}$) and homozygous ($Yuk^{b/b}$) patients containing mutations in the GPIIIa glycoprotein gene. These DNA's had previously been typed by restriction isotyping (RFLP analysis) using restriction enzyme CviRI (TG↓CA). PCR amplification was carried out using 5'-GGTGGGAGAAGAAGATAAAAAG-3' and 3'-CCAGGTTCAGGACCAAGGAT-5' primers. A 362 base pair DNA fragment spanning the Yuk allele of the GPIIIa glycoprotein gene was amplified using 30 cycles of Taq Polymerase denaturation/annealing/elongation. (See Figure 3A). About 2 μg of PCR-amplified Yuk DNA was alcohol precipitated and resuspended in 30 μl of TE buffer PH 8.0. 5 μl (about 0.25 μg) of this amplified $Yuk^{a/a}$ or $Yuk^{b/b}$ DNA was added to a 30 μl reaction containing 50 mM Tris-acetate, PH 7.5, 5 mM

IDT Exhibit 1020.0580

EDTA, 100 μg/ml BSA, and approximately 2 μM of ^{3}H-methyl-SAM (New England Nuclear, approximately 85 Ci/mmole). 5 units of either M.CviRI (TGC^{6m}A) or M.HaeIII (GG^{5m}CC) MTase was added, and DNA methylation was carried out for 2 hours at 23°C. TCA precipitation and liquid scintillation counting were carried out as described by Nelson & McClelland (1987) (See Table 2).

TABLE 2
DNA MTase Genotyping of 362 Base Pair PCR Amplified Yuk$^{a/a}$ or Yuk$^{b/b}$ Alleles of Neonatal Alloimmune Thrombocytopenia patients

DNA	TCA-Precipitable ^{3}H-Methyl- DNA Counts per Minute		
	No MTase	M.CviRI	M.HaeIII
pUC19 (Control)	615	118,417	26,156
Yuk$^{a/a}$	472	506	36,442
Yuk$^{b/b}$	501	12,241	34,217

DNA MTase genotyping can clearly be used to genotype heritable diseases. This experiment shows that the speed, accuracy, and sensitivity of MTase genotyping are suitable for diagnosing many traits of medical or economic interest. In the present example, PCR/MTase genotyping was able to detect a point mutation in a single copy gene responsible for a heritable human disease. The presence of a TGCA site and PCR-amplified Yuk$^{b/b}$ DNA is easily distinguished from its absence in Yuk$^{a/a}$ DNA by M.CviRI MTase genotyping.

In addition, the signal-to-noise performance is excellent. Using approximately 0.25 μg of 362 bp amplified DNA from human blood, 12,241 cpm of ^{3}H-methyl-DNA is observed in Yuk$^{b/b}$ DNA which contains one TGCA site. Only 506 cpm of ^{3}H-methyl-DNA are observed in Yuk$^{a/a}$ DNA, which has no TGCA sites. If shorter (60-70 base pair) PCR amplicons are employed, then even greater (5-6 fold greater) sensitivity should be possible. Under "short PCR" conditions, it should be possible to easily discriminate Yuk$^{a/a}$, Yuk$^{b/b}$, and heterozygous Yuk$^{a/b}$ ^{3}H-methyl-DNA's as signals of approximately 500 cpm, approximately 60,000 counts per minute, or approximately 30,000 counts per minute, respectively, would result. For example, using PCR primers 5'-GAACCTGGGTACCAAGCTTGGC-3' and 3'-GTGGTCATTGGACTCCTAACCG-5'; the 61 bp fragment spanning the M.CviRI and M.SssI sites can be amplified (See Figure 3B).

Finally, the success of MTase genotyping of NAT patients is of particular interest because the use of short PCR conditions and/or the use of both M.CviRI and M.SssI should allow the diagnostician to determine whether the allele composition is homozygous Yuk$^{a/a}$, homozygous Yuk$^{b/b}$, or heterozygous Yuk$^{a/b}$. The method currently

IDT Exhibit 1020.0581

being used in NAT diagnosis utilizes restriction endonuclease enzymes to digest fragments for fingerprinting using gels and/or Southern blots. In addition to being slow, a severe drawback results from this use of restriction endonucleases alone when incomplete digestion of a homozygous $Yuk^{b/b}$ DNA gives the appearance of a $Yuk^{a/b}$ carrier. (See Figures 4A and 4B) When a patient is determined to be $Yuk^{a/b}$, no therapy or platelet transfusion is given. This can have tragic results. If an affected $Yuk^{b/b}$ infant does not receive a proper therapy, then it may die. As mentioned above, affected infants are prone to hemorrhages and can die of cranial hemorrhaging or are neurologically impaired. However, with the use of the present inventive method a clearcut genotype can rapidly and accurately be obtained. Using quantitative PCR/MTase genotyping, the different levels of ^{3}H-methyl incorporation into DNA by M.CviRI(TGC^{6m}A) or M.SssI (^{5m}CG) can be encoded using a ternary (0,1,2) code. For example, using approximately 1 μg of amplified human DNA (a/a=0, a/b=1, and b/b=2), the difference between a signal level of 1 and 2 is tens of thousands of ^{3}H-methyl cpm. By ternary encoding, an algebraic distance of two can be achieved between a/a, a/b, and b/b genotypes (See Table 3). Systematic error detection is possible, so that misdiagnosis of a/b heterozygote carriers for b/b homozygotes is much less likely. R.W. Hamming, Coding and Information Theory, 2nd Ed., Prentice-Hall, Englewood Cliffs, NJ (1986), incorporated herein by reference.

TABLE 3
Ternary (0, 1, 2) Signal Encoding of DNA MTase Genotypes
DNA MTase Genotype

DNA	No MTase Control	M.SssI (^{5m}CG)	M.CviRI (TGC^{6m}A)
Wild type $Yuk^{a/a}$	0	2	0
Heterozygous $Yuk^{a/b}$	0	1	1
Homozygous $Yuk^{b/b}$	0	0	2

Hypothetical cpm ranges:
0:^{3}H-methyl cpm ≤800
1:800<cpm<10,000 cpm
2:cpm>10,000

EXAMPLE 3
Radiometric Detection and Ordered Maps in <30 Minutes

In classic RFLP mapping (Bottstein et al., 1980), DNA fragments must be long enough to contain polymorphic 4 to 6 bp restriction sites and to be stained by ethidium bromide after gel electrophoresis. However, since MTases do not cut the sugar-

IDT Exhibit 1020.0582

phosphate backbone of DNA, in DNA MTase genotyping, there are no DNA fragments to be separated. Therefore, gel electrophoresis and ethidium bromide staining are unnecessary. Improved speed and sensitivity are possible when relatively short (30 to 100 bp) amplicons are methylated: the shorter the better. In simple terms, if the molecular weight of a DNA fragment is smaller, then its molar concentration is higher. For example, 1 μg of a 300 bp DNA fragment in a 25 μl reaction volume has a concentration of ~2 μM, but 1 μg of a 100 bp fragment in the same reaction volume has a concentration of ~6 μM. If each of these DNA fragments has a single MTase recognition site, then the ^{3}H-methyl incorporation into the shorter 100 bp fragment will be three-fold greater.

As an added benefit, amplification of short PCR products is also faster because thermocycle times can be reduced to a few seconds per cycle. This time can be further reduced by using a hot-air thermocycler and a rapid sample preparation protocol. Sequence-specific DNA 3H-methylation can be conveniently detected in a homogenous solid phase Scintillation Proximity Assay (SPA; Bosworth and Towers, 1989) using a biotinylated PCR primer and Streptavidin-coated SPA beads (Amersham, Arlington Heights, IL). Radiometric DNA MTase Genotyping (DMG) has been used diagnose viral pathogens and to detect genetic polymorphisms in less than 30 minutes.

Moreover, if combinations of DNA MTases (M) and restriction endonucleases (R) are employed in an anchor-dependent SPA bead detection format, then ordered physical maps can be constructed in a few minutes without the need for electrophoresis. Radiometric DNA MTase Genotyping requires three steps: (1) PCR-amplification of DNA, (2) 3H-methylation using a sequence-specific MTase, (3) capture of ^{3}H-methyl-DNA on SPA beads. The presence (binary 1) or absence (binary 0) of any particular DNA MTase recognition site is easily determined in a few minutes. When a "universal buffer" such as KGB is employed, all three steps can be carried out sequentially in the same reaction buffer. (McClelland et al., "KGB: A single buffer for all restriction endonucleases", Nucleic Acid Research, 16(1):364 (1988).) This simple 1-2-3 method requires only 20 to 30 minutes to perform manually; and probably less than 15 minutes in automated analysis. For example, as shown in Table 4 and Figure 5, ordered maps of Herpes Simplex Virus DNA from HSV-1 and HSV-2 can be constructed in about 30 minutes using a 5' biotinylated primers, various combinations of restriction enzymes and MTases, and rapid Streptavidin-SPA bead detection.

IDT Exhibit 1020.0583

A 103 b.p. amplicon obtained by PCR amplification of 2 ng of HSV-1 or HSV-2 DNA with a 5' biotinylated reverse primer was ^{3}H-methylated using 10 U M.MspI (CCGG), M.TaqI (TCGA), or M.HhaI (GCGC) MTases in the presence of ~2 μM ^{3}H-methyl-SAM. As noted, DNA was also digested using 10 U restriction endonucleases MspI (C↓CGG), RsaI (GT↓AC), or BstUI (CG↓CG). The resulting 5' biotinylated ^{3}H-methylated/restriction digested DNA was mixed with Streptavidin-SPA beads (Amersham), suspended in ~1 ml phosphate-buffered saline and counted for bound ^{3}H-methyl-DNA cpm using a Packard scintillation counter (See Table 4).

TABLE 4

Ordered Maps* of Short PCR-Amplified HSV-1 and HSV-2 DNAs Using Combinations of DNA MTases and Restriction Endonucleases

DNA source	DNA MTase	Restriction Endonuclease	^{3}H-methyl DNA cpm	Binary signal (> 10,000 cpm = 1, ≤ 10,000 cpm = 0)
HSV-1	None	None	425	0
HSV-2	None	None	225	0
HSV-1	M. Taq I	None	9,056	0
HSV-2	M. Taq I	None	128,236	1
HSV-1	M. Taq I	Rsa I	1,667	0
HSV-2	M. Taq I	Rsa I	5,550	0
HSV-1	M. Taq I	Msp I	1,667	0
HSV-2	M. Taq I	Msp I	9,319	0
HSV-1	M. Hha I	None	8,963	0
HSV-2	M. Hha I	None	246,516	1
HSV-1	M. Hha I	Rsa I	3,565	0
HSV-2	M. Hha I	Rsa I	122,134	1
HSV-1	M. Hha I	Msp I	1,061	0
HSV-2	M. Hha I	Msp I	5,243	0
HSV-1	M. Msp I	None	166,300	1
HSV-2	M. Msp I	None	115,020	1
HSV-1	M. Msp I	BstU I	2,866	0
HSV-2	M. MSP I	BstU I	1,005	0

* See Figure 5

By measuring at the ^{3}H cpm (presence of MTase site >10,000 ^{3}H-methyl-DNA, absence of MTase site ≤ 10,000 ^{3}H-methyl DNA cpm), relative locations of restriction enzyme sites and methylation sites relation to the 5' biotinylated end bound to the SPA Bead can be determined. For example, a comparison of the cpm for M.TaqI ^{3}H-methyl labeled HSV-1 and HSV-2 when uncut and cut with RsaI shows the HSV-1 does not

IDT Exhibit 1020.0584

have a M.TaqI (TCGA) site and that RsaI restriction enzyme site (GTAC) is between the M.TaqI site and the 3' SPA-Bead of HSV-2. Note the immediate proximity of the M.HhaI site and the RsaI restriction enzyme site reduces the efficiency of the RsaI digest and only cuts about half the time. By comparing the results in Table 4 to the DNA sequence found in Figure 5, it is clear that methyltransferase ordered mapping is a quick, accurate and useful procedure. It can also be seen that monitoring the absence or presence of a DNA site by counting ^{3}H-methyl-DNA bound to a solid phase bead is a very useful and versitile tool for quick screening of samples and is conducive to large scale analysis of multiple samples.

EXAMPLE 4
Non-Radioactive Immunochemical Detection of Methylated DNA

DNA MTase genotyping can also be carried out in a non-radioactive format. In particular, specific antibodies have been used to immunochemically detect enzymatically methylated ^{6m}A-labeled DNA. Similar immunochemical detection procedures could also be used to detect ^{5m}C or ^{4m}C methylated DNA.

Rabbit antibodies to ^{6m}A were prepared by the method of B.F. Erlinger & S.M. Beiser, "Antibodies to methylated bases in nucleic acids", Proc. Nat. Acad. Sciences USA 52:68 (1964) as modified by B.D. Stollar, "The Experimental Induction of Antibodies to Nucleic Acids", Methods in Enzymology 70:70-85 (1980), both incorporated herein by reference. Periodate-oxidized ^{6m}A ribonucleosides were coupled to hemocyanin at pH 9.5, reduced using 0.1 M $NaBH_4$, and dialyzed against 0.1 M $NaHCO_3$ (pH 8.5) and then against distilled H_2O. The resulting ^{5m}C- and ^{6m}A- hemocyanin conjugates were injected into rabbits. About 30 ml of blood was collected from hyper-immunized animals, and clarified in anti-^{5m}C and anti-^{6m}A sera were frozen at -20°C.

The specificity of rabbit anti-^{6m}A serum for ^{6m}A residues in UV-irradiated DNA bound to nylon sheets was tested by immunochemical staining of in vivo methylated (or control unmethylated) DNA's, using alkaline-phosphatase conjugated goat anti-rabbit IgG.

Various dilutions (0/1 ng to 100 ng) of *in vivo* ^{6m}A-methylated Chlorella virus NY-s1 (37% ^{6m}A), phage λ*dam*$^{+}$, (1.5% ^{6m}A) and plasmid pBR322 *dam*$^{+}$ DNAs (1.5% ^{6m}A) were spotted onto nylon Hypobond™ membranes and uv-crosslinked prior to immunochemical staining. The membrane was blocked for one hour using 10 ml of (100 mM Tris, pH 7.6, 100 mM NaCl, 10% nonfat milk). 100 µl of 1/500 diluted rabbit anti-

IDT Exhibit 1020.0585

^{6m}A antisera was then added in fresh blocking solution, followed by incubation for 30 minutes at 23°C. After four 5 minute washes in blocking buffer, 100 µl of alkaline-phosphatase conjugated goat anti-rabbit secondary antibody (Vector Labs) was added, followed by incubation for 20 minutes at 23°C. After four 5 minute washes to remove excess enzyme-antibody conjugate, BCIP substrate (KPL Labs) was added and the reaction was developed for 5 minutes at 23°C; and then stopped by placing the filter in distilled H_20. (Results from this initial immunochemical staining experiment are not shown.)

Two conclusions were drawn from this experiment: First, ^{6m}A methylated DNA can be detected immunoenzymatically as dark spots on a white or light purple filter background using a 1/500 diluted anti-^{6m}A immune serum. As little as 1 nanogram of heavily methylated NYS1 DNA (37% ^{6m}A) can be detected. When sequence-specific DNA methylation occurs at G^{6m}ATC (dam$^+$) sites (1.5% ^{6m}A), as little as 10 ng of PBR322 (dam$^+$) DNA or 20 ng of λ(dam$^+$) DNA can be detected. Second, the background binding of anti-^{6m}A serum to methyl deficient dam- DNA is barely detectable. Phage λ(dam$^-$) and plasmid PBR322 (dam$^-$) DNA's are essentially not stained using the anti ^{6m}A antibody. This light background staining of 200 ng of λ(dam$^-$) and PBR322 (dam$^-$) DNA's can probably be eliminated by purification of anti-^{6m}A immunoglobins or with the use of monoclonal antibodies.

In vivo methylated Chorella virus DNA was obtained from NY2A (37% ^{6m}A) and PBCVI (3%^{6m}A) strains as described in Nelson et al., 1993. T7 bacterial phage DNA (unmethylated) was obtained from Sigma Chemicals, St. Louis, Missouri. Various dilutions of methylated or unmethylated DNA was spotted onto nylon Hypobond™ filters and UV irradiated for two minutes. The membranes were incubated in blocking solution (100 mM Tris pH 8' 200 mM NaCl and 1 mM EDTA containing 10% nonfat milk) overnight at 4°C. The membrane was washed for one hour at room temperature in 10 mls of the 1:500 dilution of anti-^{6m}A rabbit immune serum (Megabase Research Products, Lincoln, Nebraska) in the above blocking solution. Filters were washed 3 times with 100 mM Tris pH 8, 200 mM NaCl, 100 mM EDTA, 0.5% Tween 20, and then 3 times with the same solution without detergent. A 1/500 dilution of alkaline-phosphatase conjugated goat anti-rabbit IgG antiserum (Vector Labs, Burlingame, CA) was used as a secondary antibody. After washing as described above, NBT/BCIP substrate for alkaline-phosphatase (KPL Laboratories, Gaithersburg, MD) was added.

IDT Exhibit 1020.0586

Methylated DNA is detected as purple spots on white nylon filter background. Figure 6 is a photograph showing the immuno staining results. The first row of DNA is the NY2A (NY), the second row is the PBCVI(cvid), the third row is the T7 bacterial phage (T7). Columns 1-8 are DNA in the amounts of 10 ng, 1 ng, 100 pg, 10 pg, 1 pg, 100 fg, 10 fg, and 1 fg, respectively. The results of this experiment show that ^{6m}A-methylated DNA can be specifically detected using 1/500 diluted anti-^{6m}A immune serum. As little as 1 picogram of methylated NY2A DNA (37% ^{6m}A) and 10 picogram of PBCVI (3% ^{6m}A) can be detected. In addition, the results show no background binding of anti-^{6m}A serum to unmethylated DNA is detectable using this method. Finally, the results show the anti-^{6m}A serum can detect below the limit of detection of ethidium bromide: as little as 1 picogram of ^{6m}A methylated DNA is detectable.

EXAMPLE 5
Restriction Digestion Followed by Immunochemical Detection

A more sophisticated immunochemical detection scheme for staining DNA fragments which are internally methylated between the 5' and 3' ends relies on Southern blotting. DNA is first methylated at specific sites and then digested by restriction enzymes which have a different sequence specificity. After gel electrophoresis and Southern blotting to nylon filters, only those fragments containing sequence-specific DNA methylation groups are immunochemically stained. In simple terms, DNA restriction fragments from endonuclease #1 can be "painted" at specific sites using MTase #2.

DNA is methylated using ^{6m}A-specific DNA MTases in (50 mM Tris PH7.5, 5mM EDTA, 100 μM SAM). *In vitro* methylated DNA is then digested using one or more restriction endonucleases. The DNA is separated on a 1% agarose gel (Agarose GTG, FMC, Rockland, ME) and stained with ethidium bromide and Southern blotted onto a nylon membrane which is UV crosslinked prior to immunochemical staining. The blotted DNA is later immunochemically detected with a rabbit anti-^{6m}A antisera and a goat anti-rabbit antiserum (Megabase, Lincoln, Nebraska) as described in Example 4.

In Figure 7A and 7B a preliminary experiment is shown in which lambda dam+ or dam- DNA was digested with restriction endonucleases. One microgram of lambda DNA was digested for 2 hours at 37°C with 15 units of EcoRI, BamHI or HindIII restriction endonucleases in a 20 μl volume. The reaction mix was mixed with dye-EDTA stop buffer and electrophoresed for 4 hours at 60 Volts in a 1% agarose gel.

IDT Exhibit 1020.0587

Figure 7A is a photograph of the gel used for transfer showing the λ(dam+) DNA in lanes 1-4 undigested, EcoRI digested, BamHI digested and HindIII digested respectively. Lanes 5-8 of Figure 7A show the λ(dam-) DNA undigested, EcoRI digested, BamHI digested and HindIII digested respectively. The DNA was Southern blotted to a nylon Hybond™ membrane using techniques known to those of skill in the art and UV irradiated for two minutes. Methylated DNA fragments are detected as purple bands on white nylon filter background.

Figure 7B is a photograph of the nylon membrane after immunochemical detection. It can be seen that BamHI or HindIII digested DNA fragments which were subsequently M.EcoRI (GA^{6m}ATTC) methylated can be seen after transfer to nylon membranes and immunochemically detected. These results show that ^{6m}A sites in DNA which has been enzymatically methylated *in* vitro can be stained using anti ^{6m}A antibodies, whereas unmethylated DNA is nonstained. In short, Example 5 shows sequence-specific MTase painting of DNA at 6 bp long GA^{6m}ATTC sites.

EXAMPLE 6
Sequence-Specific DNA Methylation and Restriction Digestion Followed by Immunochemical Detection (DNA Paint)

This example is an early sequence specific "purple DNA Paint" experiment showing the specificity of methylation by first methylating with ^{6m}A- and ^{5m}C- specific methyltransferases, followed by immunochemical detection using rabbit anti-^{6m}A-antibodies and alkaline phosphatase-conjugated goat anti-rabbit antibodies. (See Figures 8A and 8B.)

Methylation was conducted on lambda dam- phage DNA using M.dam MTase (G^{6m}ATC, lane D-F), M.EcoRI MTase (G^{6m}AATTC, lanes G-I), M.SssI MTase (^{5m}CG, lanes J-K), and unmethylated dam- phage lambda (lanes A-C). The DNA was then digested with restriction enzymes MboI (↓GATC, lanes A, D, G, J), EcoRI (G↓AATTC, lanes B, E, H, K) or HpaII (C↓CGG, lanes C, F, I) and run out on an agarose gel stained with ethidium bromide (see Figure 8A). As can be seen in lanes D and H, the digestion is blocked due to the methylation at the sequence-specific restriction site.

The gel-separated DNA was then transferred to a nylon membrane by Southern blot. The immunochemical detection was performed as described in Example 4. Methylated DNA fragments stain as dark purple bands on a light purple background. Lanes D-I all contain methylated ^{6m}A sites and Lanes J and K contain methylated ^{5m}C sites. Figure 8B is a photograph of the immunostained membrane and shows that only

IDT Exhibit 1020.0588

those fragments containing ^{6m}A-methylated DNA stained purple. While Lanes D-I all contained ^{6m}A-methylated DNA, only those fragments with at least 6 to 8 ^{6m}A sites were detected (lanes D and E).

Figures 9A and 9B also show a sequence-specific "purple DNA Paint" experiment using phage T7 DNA and M.dam (G^{6m}ATC) MTase. DNA was either unmethylated (lanes A, B, C, D) or was G^{6m}ATC-methylated with M.dam MTase (lanes E, F, G, H) or TCG^{6m}A--methylated with M.TaqI MTase (lanes I, J, K, L). The DNAs were then digested using DpnII (↓GATC)(lanes A, E, I), TaqI (T↓CGA)(lanes B, F, J), SpeI (A↓CTAGT) (lanes C, G, K), or NruI (TCG↓CGA) (lanes D, H, L).

As can be seen in Figure 9A, lanes A, B, C, D are cut to completion. Lane E shows the G^{6m}ATC-methylated DNA did not cut using DpnII(↓GATC) and lane J shows the TCG^{6m}A-methylated DNA did not cut using TaqI(T↓CGA). Figure 9B shows the nylon membrane following transfer of the gel in Figure 9A.

Figure 9B is a photograph of the membrane following immunodetection of the ^{6m}A-methylated DNA. There are no bands in the unmethylated lanes (A, B, C, D). While lanes E through L all show that a sequence-specifically methylated DNA fragment can be detected, lanes I through L appear to give much stronger signal suggesting either a greater number of M.TaqI MTase sites in T7 DNA.

From these figures, the following conclusions can be drawn: (1) It is possible to detect DNA which has been methylated *in vitro* using Dam (G^{6m}ATC) MTase. (2) ^{6m}A methylated DNA is specifically stained using anti ^{6m}A antibodies, whereas unmethylated DNA and ^{5m}C methylated DNA is unstained.

With modest improvements in sensitivity, it should be possible to detect single MTase sites in restriction fragments from approximately 1-2 μg of phage λ DNA. For example, using more DNA, partially purifying anti-^{6m}A IgG's, or sandwich immuno staining protocols (biotinylated goat anti-rabbit antibody and avidin conjugated alkaline phosphatase) should increase sensitivity at least ten-fold.

EXAMPLE 7

Example 7 shows such improved sensitivity. Refinements of our DNA MTase genotyping procedure make it possible to immunochemically "paint" single copy human genes at defined 2 to 6 bp sites using PCR amplification, sequence-specific methylation and immunodetection of ^{6m}A-labeled DNA. One ^{6m}A-labeled MTase recognition site in a 536 bp fragment of the single copy human beta globin gene can be detected by improved immunostaining.

IDT Exhibit 1020.0589

Figures 10A, 10B, and 10C show a sequence-specific (purple DNA Paint) experiment of an *in vitro* methylated amplicon from human β-globin gene. The 536 bp amplicon is first PCR amplified and then methylated at G^{6m}ATC sites using M.dam MTase. Amplified methylated DNA is then digested with 10 units of restriction endonucleases NcoI(C↓CATGG), AvaII(G↓G(A/T)CC), DpnII(↓GATC) or is left uncut. Figure 10A shows a 1.7% agarose gel containing the following DNA treatments:

Lane 1; unmethylated, uncut

Lane 2; unmethylated, NcoI

Lane 3; unmethylated, AvaII

Lane 4; unmethylated, DpnII

Lane 5; methylated, uncut

Lane 6; methylated, NcoI

Lane 7; Methylated, AvaII

Lane 8; Methylated, DpnII

Lane 9; BioRad Molecular Weight marker 50-2000 bp.

Note that DpnII(↓GATC) cleavage is completely blocked by M.dam MTase (G^{6m}ATC), whereas unmethylated DNA is cut by DpnII.

After 1.7% agarose gel electrophoresis, the gel was Southern blotted to a nylon membrane and uv-crosslinked prior to immunochemical detection (MegaPurple™).

Figure 10B shows a photograph of the nylon membrane following sequence-specific immunophosphatase/BCIP immunochemical detection of DNA fragments containing G^{6m}ATC sites. By comparing fragment sizes (see map shown in Figure 10C) with a G^{6m}ATC-methylation site it is clear that the 536 and 405 bp fragments are easily detectable. However, the 161 bp fragment of the AvaII digest is barely visible. There is no non-specific background labeling of unmethylated DNA. Therefore, Example 7 demonstrates sequence-specific immunochemical painting of a single methylated G^{6m}ATC site in a 536 bp fragment of a single copy human gene.

EXAMPLE 8

This example demonstrates it is possible to paint methylated DNA fragments larger than 1 kb from a single human gene. Immunochemical detection of *in vitro* GA^{6m}ATCC methylated PCR amplicon from human β-globin gene after EcoRI, BsrGI or BamHI digestion and Southern blotting to a nylon membrane yields "red DNA Paint" when fast red substrate (Pierce, Rockford, IL) is employed for immunostaining.

IDT Exhibit 1020.0590

A 1850 bp fragment of untreated or else human β-globin gene is PCR amplified and then either methylated using M.EcoRI MTase (GA^{6m}ATCC) or left untreated. Restriction enzyme digests are then conducted on 20 µl aliquot with 10 units of EcoRI(G↓AATTC), BsrGI(T↓GTACA), or BamHI(G↓GATCC) endonuclease. Note that EcoRI cleavage is blocked by M.EcoRI MTase methylation; but unmethylated DNA is cut by EcoRI. Methylated or unmethylated DNA restriction fragments are then separated on a 1.7% agarose gel in the following order (see Figure 11A):

Lane 1; dam+, Molecular Marker, 100bp DNA Labeling (New England Biolabs)
Lane 2; unmethylated, uncut
Lane 3; unmethylated, EcoRI
Lane 4; unmethylated, BsrGI
Lane 5; unmethylated, BamHI
Lane 6; methylated, uncut
Lane 7; methylated, EcoRI
Lane 8; methylated, BsrGI
Lane 9; methylated, BamHI

After 1.7% agarose gel electrophoresis, the gel was Southern blotted to a nylon membrane and uv-crosslinked prior to immunochemical detection (MegaRed™). The immunochemical detection of the ^{6m}A methylated DNA using rabbit anti^{6m}A methylated antibodies followed by alkaline-phosphatase anti-rabbit antibodies and fast red substrate is shown in Figure 11B. Comparison of the "red DNA Paint" results with the map of Figure 11C shows the fragments containing the GA^{6m}ATTC-sequence-specific methylated site were "painted" red by the immunochemical staining method. In other words, DNA fragments as large as 1.8 kb from single copy human genes can be specifically painted at GA^{6m}ATTC sites; and they can be painted red.

EXAMPLE 9

This example shows the precise specificity of the rabbit antibodies, which only recognize 6-methyladenine (^{6m}A). DNA which has been enzymatically methylated at ^{4m}C or ^{5m}C sites is not immunochemically stained. Immunochemical detection of *in vitro* M.dam methylated PCR amplicon from human β-globin gene after NcoI, DpnII, BstNI or HaeIII digestion and Southern blotting to a nylon membrane selectively identifies those fragments containing G^{6m}ATC sites. Only ^{6m}A methylated DNA is "painted red".

IDT Exhibit 1020.0591

A 536 bp fragment of human β-globin gene was PCR amplified using methods described previously. A portion of this amplified DNA was methylated using M.HaeIII MTase (GGC^{5m}C), M.dam MTase (G^{6m}ATC), or M.MvaI MTase (C^{4m}CWGG). Portions of each of these methylated DNAs, as well as unmethylated DNA were then restriction endonuclease digested with 10 units of NcoI(C↓CATGG); DpnII(↓GATC); BstNI(CC↓WGG) or HaeIII(GG↓CC). These DNA samples were then size separated on a 1.7% agarose gel in the following order:

Lane 1; unmethylated; uncut
Lane 2; unmethylated; NcoI
Lane 3; unmethylated; DpnII
Lane 4; unmethylated; BstNI
Lane 5; unmethylated; HaeIII
Lane 6; M.HaeIII MTase; uncut
Lane 7; M.HaeII MTase; NcoI
Lane 8; M.HaeII MTase; HaeIII
Lane 9; M.dam MTase; uncut
Lane 10; M.dam MTase; NcoI
Lane 11; M.dam MTase; DpnII
Lane 12; M.MvaI MTase; uncut
Lane 13; M.MvaI MTase; NcoI
Lane 14; M.MvaI MTase; BstNI
Lane 15; Molecular Weight Marker; 50-2000 bp (BioRad)

Note that DpnII cleavage is blocked by M.dam MTase (G^{6m}ACT), HaeIII cleavage is blocked by M.HaeIII MTase (GGC^{5m}C), and BstNI cleavage is blocked by M.MvaI MTase (C^{4m}CWGG). (See Figure 12A).

After 1.7% agarose gel electrophoresis, the gel was Southern blotted to a nylon membrane and uv-crosslinked prior to immunochemical detection (MegaRed™). (See Figure 12B).

Figure 12C shows a map of the β-globin 536 bp amplicon exhibiting the expected sizes of methylated DNA fragments. By comparing Figure 12B and Figure 12C it is clear that all unmethylated lanes (1-5 and 15) are not labeled by "red DNA Paint" specific for ^{6m}A methylated DNA. Further, the ^{5m}C methylated DNA lanes (6-8) and the ^{4m}C methylated DNA lanes (12-14) are also not "painted red". However, the 536 bp DNA amplicon of M.dam MTase methylated DNA of lane 9, the 405 bp fragment of NcoI

IDT Exhibit 1020.0592

digested M.dam MTase methylated DNA of lane 10, and the uncut 536 mp DNA amplicon of DpnII digested M.dam MTase methylated DNA of lane 11 are all painted red by the "red DNA Paint" reaction of the present invention. In this example, the immunochemical detection of the ^{6m}A methylated DNA is accomplished using rabbit anti-^{6m}A IgG antibody followed by alkaline phosphatase anti-rabbit IgG and Fast Red™ substrate (Pierce, Rockford, IL).

In short, Example 9 shows that DNA which is unmethylated, ^{4m}C-methylated, or ^{5m}C-methylated is not stained by anti ^{6m}A antibody. In contrast, DNA which has been methylated at G^{6m}ATC sites in vitro by M.dam MTase is immunostained red.

IDT Exhibit 1020.0593

What is claimed is:

1. A method for identifying DNA genotype comprising: methylating DNA utilizing a DNA methyltransferase specific for a sequence recognition site; assaying for the presence of the methylation at the sequence recognition site; and determining the allele composition at the sequence recognition site.

2. The method of claim 1 wherein the means of assaying for the presence of methylation is immunochemical or radiometric.

3. A method for identifying a DNA genotype comprising: amplifying a DNA fragment using a Polymerase and primers flanking a DNA methyltransferase recognition sequence wherein an allele contains the recognition sequence and another allele does not contain the recognition sequence; methylating the DNA fragment utilizing a DNA methyltransferase that recognizes the recognition site and radioactively-labeled-methyl-SAM in appropriate methylation buffer; assaying for the presence of radioactively-labeled-methylated DNA; and determining allele composition at the recognition sequence site.

4. The method of claim 1 wherein the radioactively-labeled methyl-SAM is ^{3}H-methyl-SAM.

5. The method of claim 3 wherein the primers are selected to create an amplified DNA fragment capable of identifying a DNA genotype.

6. The method of claim 3 wherein the primers are selected to create an amplified DNA fragment in the range of about 30-100 base pairs.

7. The method of claim 3 wherein the DNA genotype is determined qualitatively as a positive for homozygous alleles a or heterozygous for alleles a/b and negative for alleles b/b.

8. The method of claim 3 wherein the DNA genotype is determined quantitatively as a range of counts per minute for alleles a/a, a/b and b/b.

9. The method of claim 3 wherein the DNA fragment is amplified using polymerase chain reaction technique.

IDT Exhibit 1020.0594

10. The method of claim 3 wherein the DNA fragment is amplified using strand displacement amplification.

11. A method for identifying a DNA genotype comprising: amplifying and methylating a DNA fragment using an appropriate buffer solution and a thermostable DNA methyltransferase, radioactively-labeled-methyl-SAM, Polymerase and primers flanking the thermostable DNA methyltransferase recognition sequence wherein an allele contains the recognition sequence and another allele does not contain the recognition sequence; assaying for the presence of radioactively-labeled-methylated DNA; and determining allele composition at the recognition sequence site.

12. The method of claim 11 wherein the radioactively-labeled-methyl-SAM is ^{3}H-methyl-SAM.

13. The method of claim 11 wherein the primers are selected to create an amplified DNA fragment capable of identifying a DNA genotype.

14. The method of claim 11 wherein the primers are selected to create an amplified DNA fragment in the range of about 30-70 base pairs.

15. The method of claim 11 wherein the DNA genotype is determined qualitatively as a positive for homozygous alleles a or heterozygous for alleles a/b and negative for alleles b/b.

16. The method of claim 11 wherein the DNA genotype is determined quantitatively as a range of counts per minute for alleles a/a, a/b and b/b.

17. The method of claim 11 wherein the DNA fragment is amplified using polymerase chain reaction technique.

18. The method of claim 11 wherein the DNA fragment is amplified using strand displacement amplification.

19. The method of claim 3 wherein one of the primers is biotinylated.

20. The method of claim 19 wherein the resulting DNA fragment is captured on a streptavidin-coated fluoromicrosphere.

21. The method of claim 11 wherein one of the primers is biotinylated.

IDT Exhibit 1020.0595

22. The method of claim 21 wherein the resulting DNA fragment is captured on a streptavidin-coated fluoromicrosphere.

23. A method for genotyping DNA comprising: digesting the DNA with a restriction endonuclease; methylating the DNA utilizing radioactively-labeled-methyl-SAM and a DNA methyltransferase specific for a recognition sequence different than the recognition sequence of the restriction endonuclease; separating the digested, methylated DNA on an agarose gel; transferring the separated, digested, methylated DNA onto a membrane; and identifying a banding pattern specific for the restriction endonuclease/DNA methyltransferase/DNA utilizing a radiometric assay.

24. The method of claim 23 wherein the radioactively-labeled-methyl-SAM is ^{3}H-methyl-SAM.

25. A method for genotyping DNA comprising: digesting the DNA with a restriction endonuclease; methylating the DNA utilizing methyl-SAM and a DNA methyltransferase specific for a recognition sequence different than the recognition sequence of the restriction endonuclease; separating the digested, methylated DNA on an agarose gel; transferring the separated, digested, methylated DNA onto a membrane; and identifying a banding pattern specific for the restriction endonuclease/DNA methyltransferase/DNA utilizing an immunochemical technique.

26. The method of claim 25 wherein the immunochemical technique utilizes an antibody to the methylated DNA and a secondary antibody to the antibody to the methylated DNA conjugated to an enzymatic staining element whereby a substrate is then added to the antibody-secondary antibody-enzyme conjugate to form a detectable color.

27. The method of claim 26 wherein the antibody is rabbit anti-^{6m}A antisera.

28. The method of claim 26 wherein the secondary antibody is alkaline-phosphatase conjugated goat anti-rabbit antibody.

29. The method of claim 26 wherein the antibody is rabbit anti-^{5m}C antisera.

30. The method of claim 26 wherein the antibody is a monoclonal antibody selected from the group consisting of anti-^{6m}A antibody, anti-^{5m}C antibody, and anti-^{4m}C antibody.

IDT Exhibit 1020.0596

31. The method of claim 26 wherein the secondary antibody is horseradish-peroxidase-conjugated goat anti-rabbit antibody.

32. A method for identifying a DNA genotype comprising: amplifying a DNA fragment using a Polymerase and primers flanking a DNA methyltransferase recognition sequence wherein an allele contains the recognition sequence and another allele does not contain the recognition sequence; methylating the DNA utilizing a DNA methyltransferase that recognizes the sequence recognition site and methyl-SAM in an appropriate methylation buffer; removing unincorporated methyl-groups; dotting dilutions of DNA onto membrane; and determining allele composition utilizing an immunochemical technique.

33. The method of claim 32 wherein the immunochemical technique utilizes an antibody to the methylated DNA and a secondary antibody to the antibody to the methylated DNA conjugated to an enzymatic staining element whereby a substrate is then added to the antibody-secondary antibody-enzyme conjugate to form a detectable color.

34. The method of claim 33 wherein the antibody is rabbit anti-^{6m}A antisera.

35. The method of claim 33 wherein the secondary antibody is alkaline-phosphatase conjugated goat anti-rabbit antibody.

36. The method of claim 33 wherein the antibody is rabbit anti-^{5m}C antisera.

37. The method of claim 33 wherein the secondary antibody is horseradish-peroxidase-conjugated goat anti-rabbit antibody.

38. The method of claim 32 wherein the primers are selected to create an amplified DNA fragment less than or equal to about 200 base pairs.

39. The method of claim 32 wherein the primers are selected to create an amplified DNA fragment in the range of about 30-100 base pairs.

40. The method of claim 32 wherein the DNA genotype is determined qualitatively as a positive for homozygous alleles a or heterozygous for alleles a/b and negative for alleles b/b.

IDT Exhibit 1020.0597

41. The method of claim 32 wherein the DNA genotype is determined quantitatively as a range of detectable color for alleles a/a, a/b and b/b.

42. The method of claim 40 wherein the DNA genotype is determined utilizing binary code.

43. The method of claim 41 wherein the DNA genotype is determined utilizing ternary code.

44. The method of claim 32 wherein the DNA fragment is amplified using polymerase chain reaction technique.

45. The method of claim 32 wherein the DNA fragment is amplified using strand displacement amplification.

46. A method for generating ordered maps from a DNA fragment comprising: amplifying a DNA fragment using a Polymerase and primers flanking a region of DNA to be mapped wherein one of the primers is biotinylated; digesting the DNA fragments with a series of specific restriction endonucleases; methylating each restriction endonuclease digested DNA fragment utilizing ^{3}H-methyl-SAM and a series DNA methyltransferases wherein one of the DNA methyltransferases recognizes the same recognition sequence as the restriction endonuclease and the other DNA methyltransferases recognize different recognition sequences than the restriction endonuclease; anchoring each amplicon to a streptavidin-coated magnetic bead via the biotin present in the biotinylated primer; removing unanchored DNA; treating the bead-anchored DNA with Proteinase K to remove DNA from bead; removing the bead; assaying for the presence of ^{3}H-methylated DNA; and creating an ordered map based on presence or absence of the ^{3}H-methylated DNA.

47. The method of claim 46 wherein the DNA fragment is amplified using polymerase chain reaction technique.

48. The method of claim 46 wherein the DNA fragment is amplified using strand displacement amplification.

49. A kit for analyzing a DNA sequence comprising: a DNA methyltransferase specific for a sequence recognition site; a means for assaying for the presence of the

IDT Exhibit 1020.0598

methylation at the sequence recognition site; and a means for determining the allele composition at the sequence recognition site.

50. The kit of claim 49 wherein the means for assaying for the presence of the methylation is radiometric or immunochemical.

51. The kit of claim 49 wherein the means for determining the allele composition is a pattern of dots or bands, or a range of radioactive counts, each associated with a specific composition of allele or alleles.

52. A kit for analyzing a DNA sequence comprising: a DNA methyltransferase specific for a sequence recognition site; a DNA restriction endonuclease specific for a sequence recognition site; a means for assaying for the presence of the methylation at the methylation sequence recognition site; and a means for determining the allele composition at the methylation and endonuclease sequence recognition sites.

53. The kit of claim 52 wherein the means for assaying for the presence of the methylation is radiometric or immunochemical.

54. The kit of claim 52 wherein the means for determining the allele composition is a pattern of dots or bands, or a range of radioactive counts, each associated with a specific composition of allele or alleles.

55. A kit for analyzing a DNA sequence comprising: a set of primers flanking the sequence of interest; a DNA methyltransferase specific for a sequence recognition site; a DNA restriction endonuclease specific for a sequence recognition site; a means for assaying for the presence of the methylation at the methylation sequence recognition site; and a means for determining the allele composition at the methylation and endonuclease sequence recognition sites.

56. The kit of claim 55 wherein the means for assaying for the presence of the methylation is radiometric or immunochemical.

57. The kit of claim 55 wherein the means for determining the allele composition is a pattern of dots or bands, or a range of radioactive counts, each associated with a specific composition of allele or alleles.

58. A polymorphic site identified by the kit of claim 49.

IDT Exhibit 1020.0599

59. A polymorphic site identified by the kit of claim 52.

60. A polymorphic site identified by the kit of claim 55.

61. A polymorphic site identified by the use of the method of claim 3.

62. A polymorphic site identified by the use of the method of claim 11.

63. A polymorphic site identified by the use of the method of claim 23.

64. A polymorphic site identified by the use of the method of claim 25.

65. A polymorphic site identified by the use of the method of claim 32.

IDT Exhibit 1020.0600

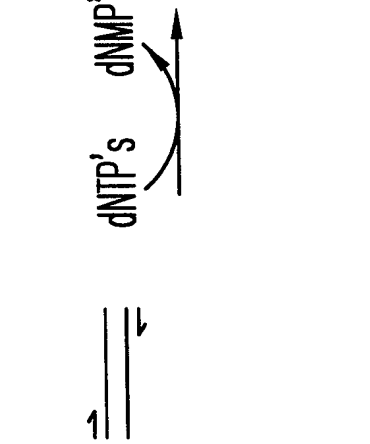

dNTP's
dNMP's
(——) n
Amplified DNA
^{3}H-methyl
SAM
SAH
Thermostable
DNA MTase
C*H$_3$
(——) n
C*H$_3$
^{3}H-methyl
Amplified DNA

Fig. 1

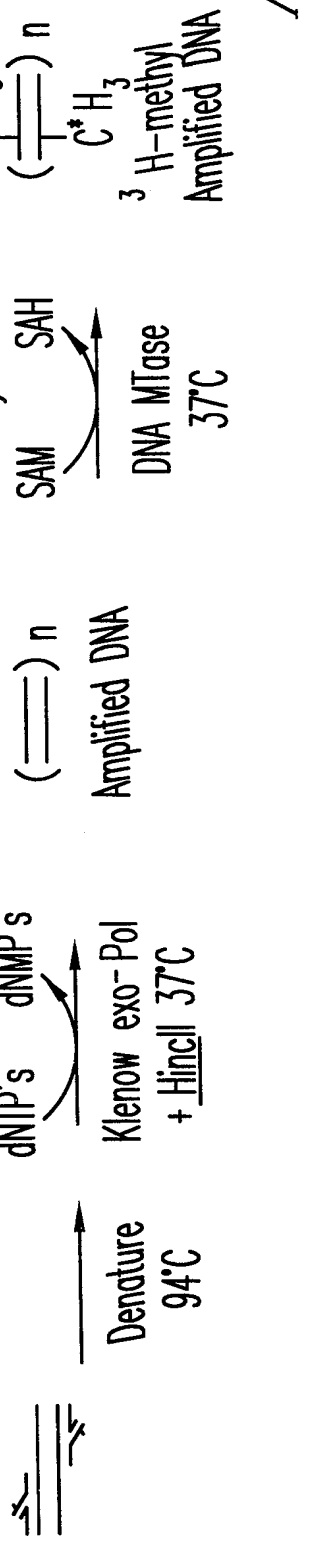

Denature
94°C
dNTP's
dNMP's
Klenow exo- Pol
+ HincII 37°C
(——) n
Amplified DNA
^{3}H-methyl
SAM
SAH
DNA MTase
37°C
C*H$_3$
(——) n
C*H$_3$
^{3}H-methyl
Amplified DNA

Fig. 2

GGTGGGAGAAGAAGATAAAAAG →

Yuk[a] GGTGGGAGAAGAAGATAAAAAGTAACATCTTTTCTGCCTTCCAGATGATTCGAAGA

ATTTCTCCATCCAAGTGCGGCAGGTGGAGGATTACCCTGTGGACATCTACTACTT

GATGGACCTGTCTTACTCCATGCCGGATGATCTGTGGAGCATCCAGAACCTGGGT

TGC**A** M.CviRI =Yuk[b]

ACCAAGCTGGCCACCCAGATGC**G**AAAGCTCACCAGTAACCTGCGGATTGGCTTCG

GGGCATTTGTGGACAAGCCTGTGTCACCATACATGTATATCTCCCCACCAGAGGC

CCTCGAAAACCCCTGCTATGAGTAAGTCCCTCCTCCAGACGCCAGGACAGCATCC

TTTGCCCAGGAAGGTCCAAGTCCTGGTTCCTA

← CCAGGTTCAGGACCAAGGAT

Fig. 3A

$TGC^{6m}A$ M. CviRI = Yuk^b

^{5m}CG M.Sss I = Yuk^a

22-mer primer
GAACCTGGGTACCAAGCTTGGC-->

GPIIIa DNA
GAACCTGGGTACCAAGCTTGGCCACCCAGATGCGAAGCTCACCAGTAACCTGCGGATTGGC
CTTGGACCCATGGTTCGAACCGGTGGGTCTACGCTTCGAGTGGTCATTGGACGCCTAACCG

<--GTGGTCATTGGACTCCTAACCG
22-mer primer

Fig. 3B

Complete Restriction Digestion
Wild-type.
Heterozygote
Homozygote
a/a
a/b
b/b
Partial Digestion
Homozygote
b/b
Partial digest band leads to misdiagnosis of homozygote DNA
Electrophoresis
M1
M2
M3

Fig. 4A

A B C D E F

Fig. 4B

```
                    m.MvaI   RsaI
HSV-1 CCAAGCTGACGGACATCTACAAGGTCCCCCTGGACGGGTACCGGCCGCATG
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
HSV-2 CCAAGCTGACGGAGATCTACAAGGTCCCGCTCGACGGGTACGGGCGCATG
                        m.TaqI  RsaI  HhaI

          m.MspI      BstUI
HSV-1 AACGGCCGGGGCGTGTTTCGCGTGTGGGACATAGGCCAGAGCCACTTCCAGAA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
HSV-2 AACGGCCGGGGTGTGTTCCGCGTGTGGGACATCGGCCAGAGCCACTTCCAGAA
          m.MspI      BstUI
```

Fig. 5

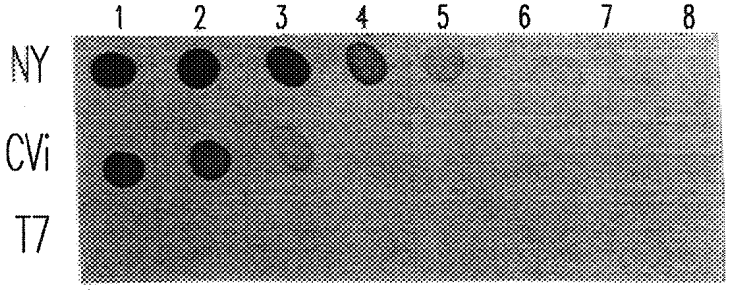


Fig. 6

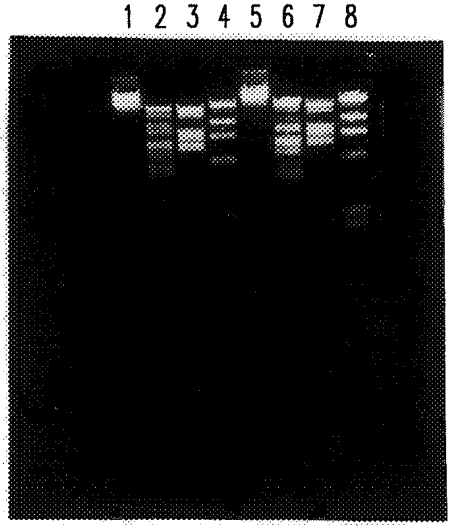


Fig. 7A

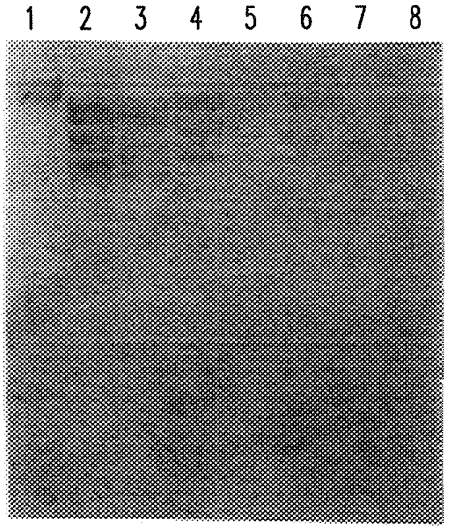


Fig. 7B

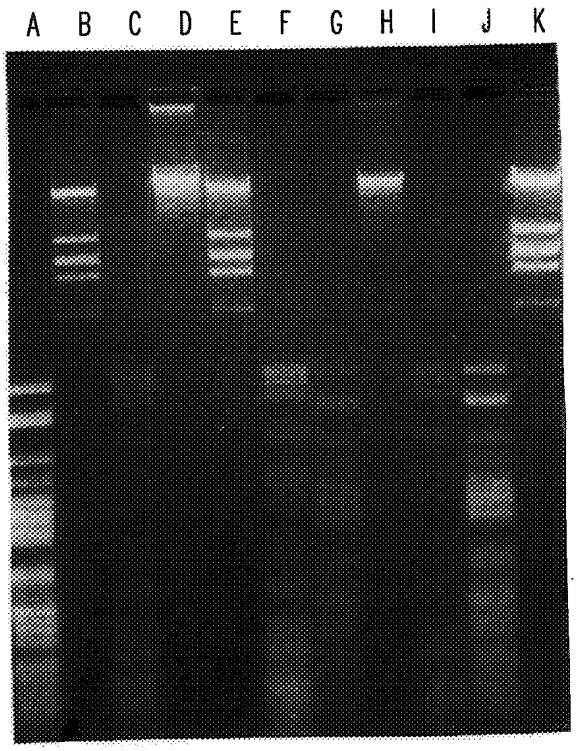


Fig. 8A

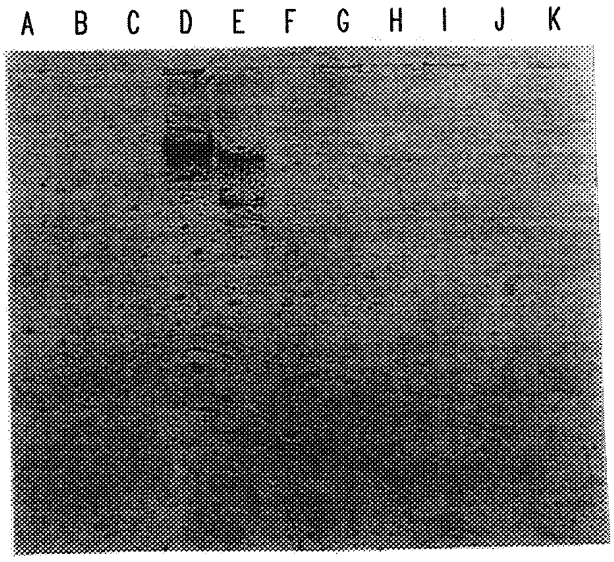


Fig. 8B



Fig. 9A

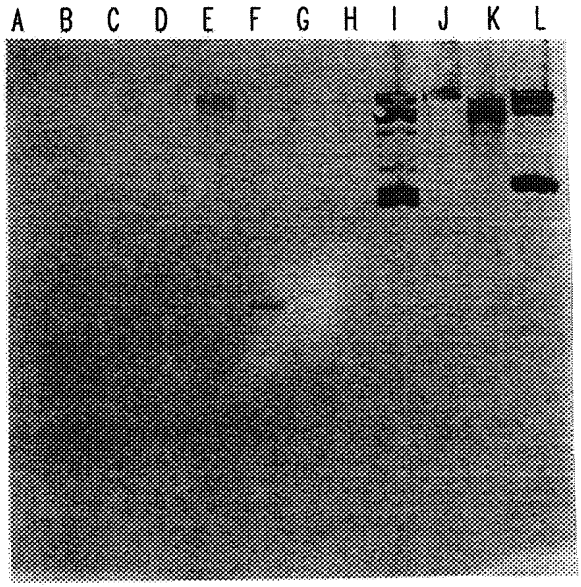


Fig. 9B

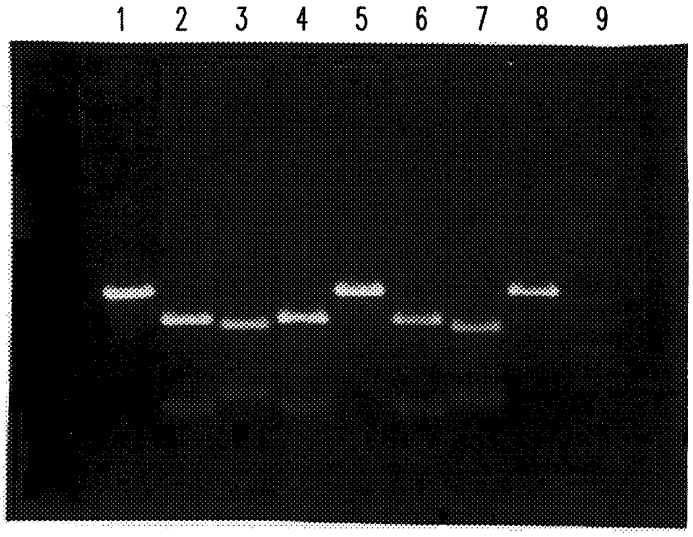


Fig. 10A

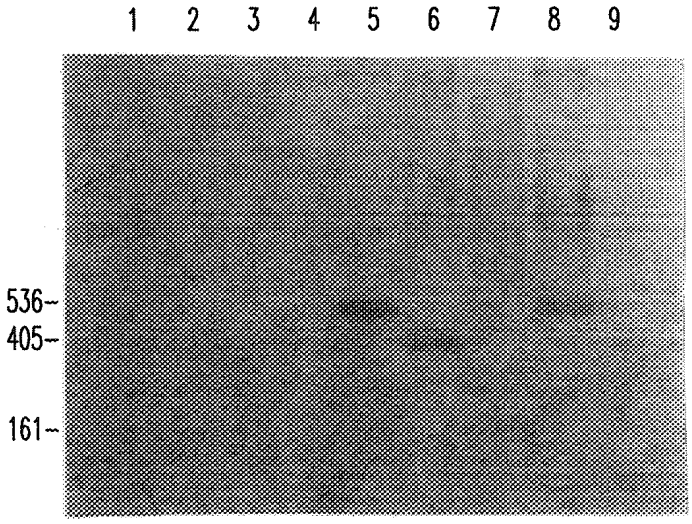


Fig. 10B

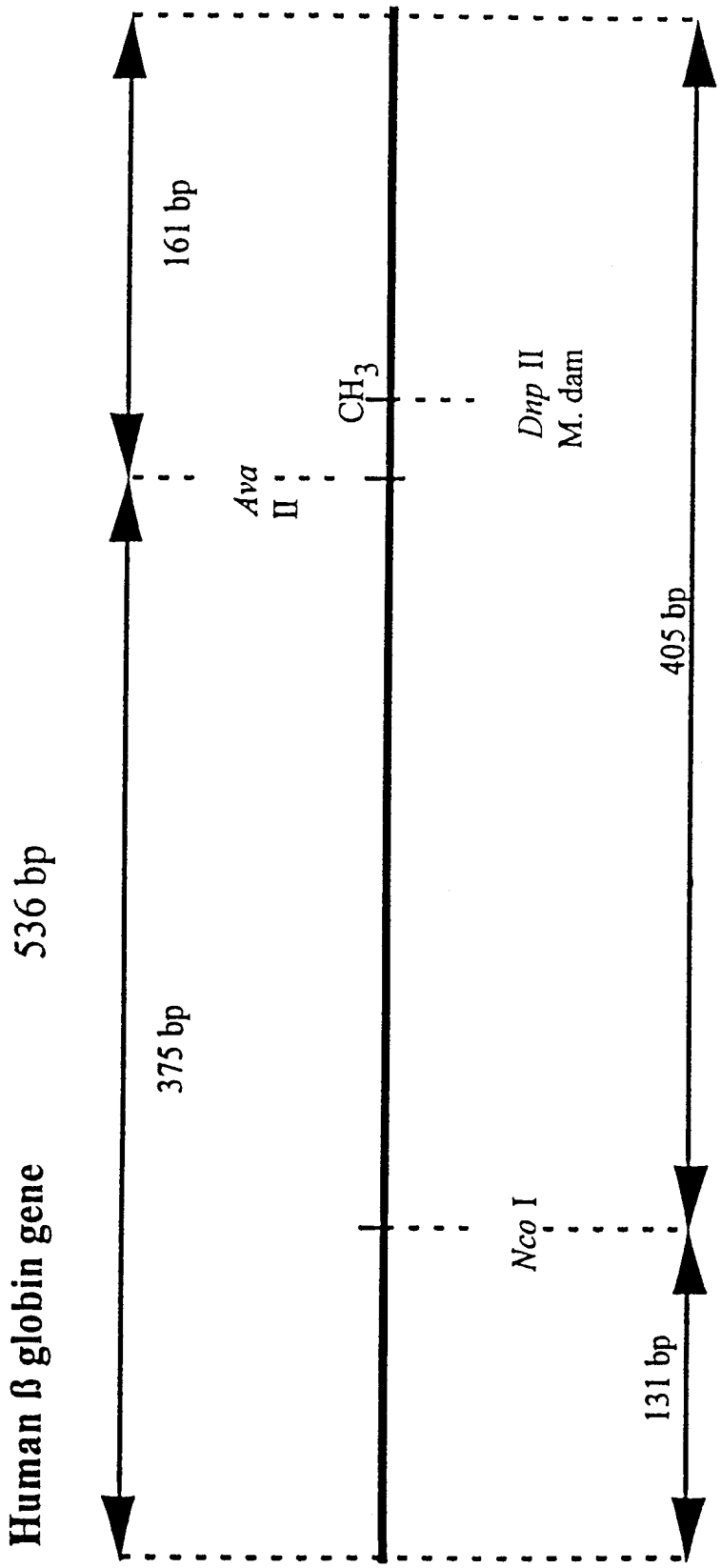

Human ß globin gene
536 bp
161 bp
375 bp
405 bp
131 bp
Ava II
CH3
Dnp II
M. dam
Nco I

Fig. 10C

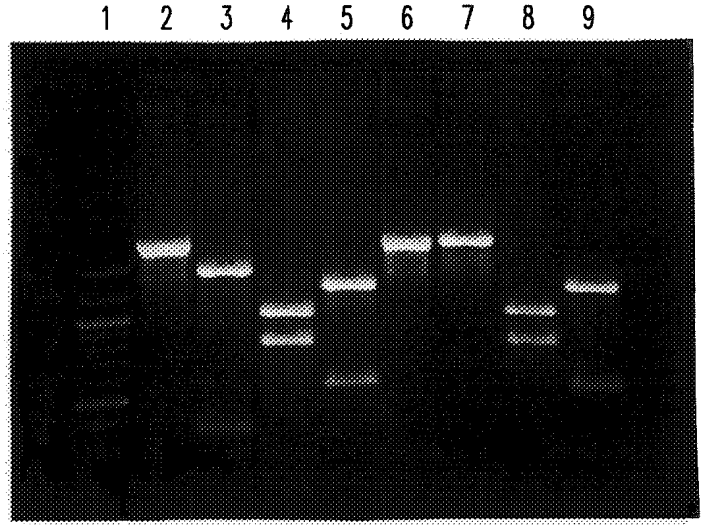


Fig. 11A

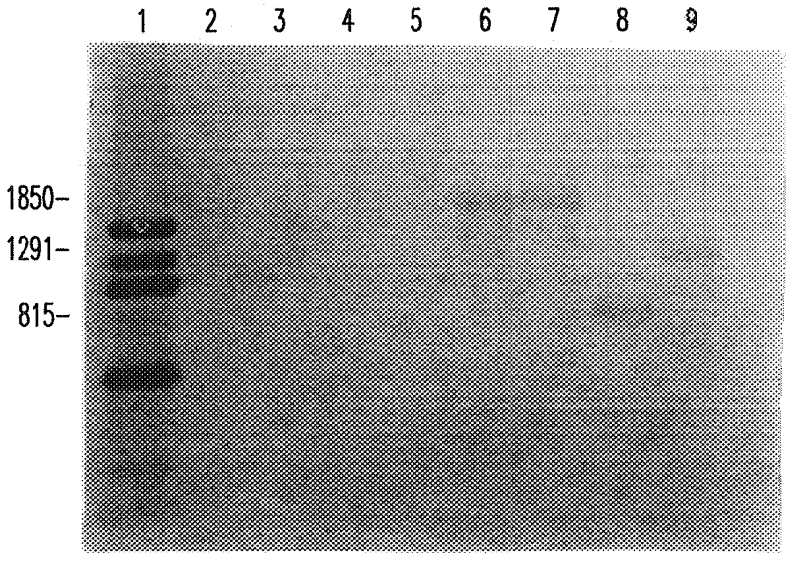


Fig. 11B

IDT Exhibit 1020.0611

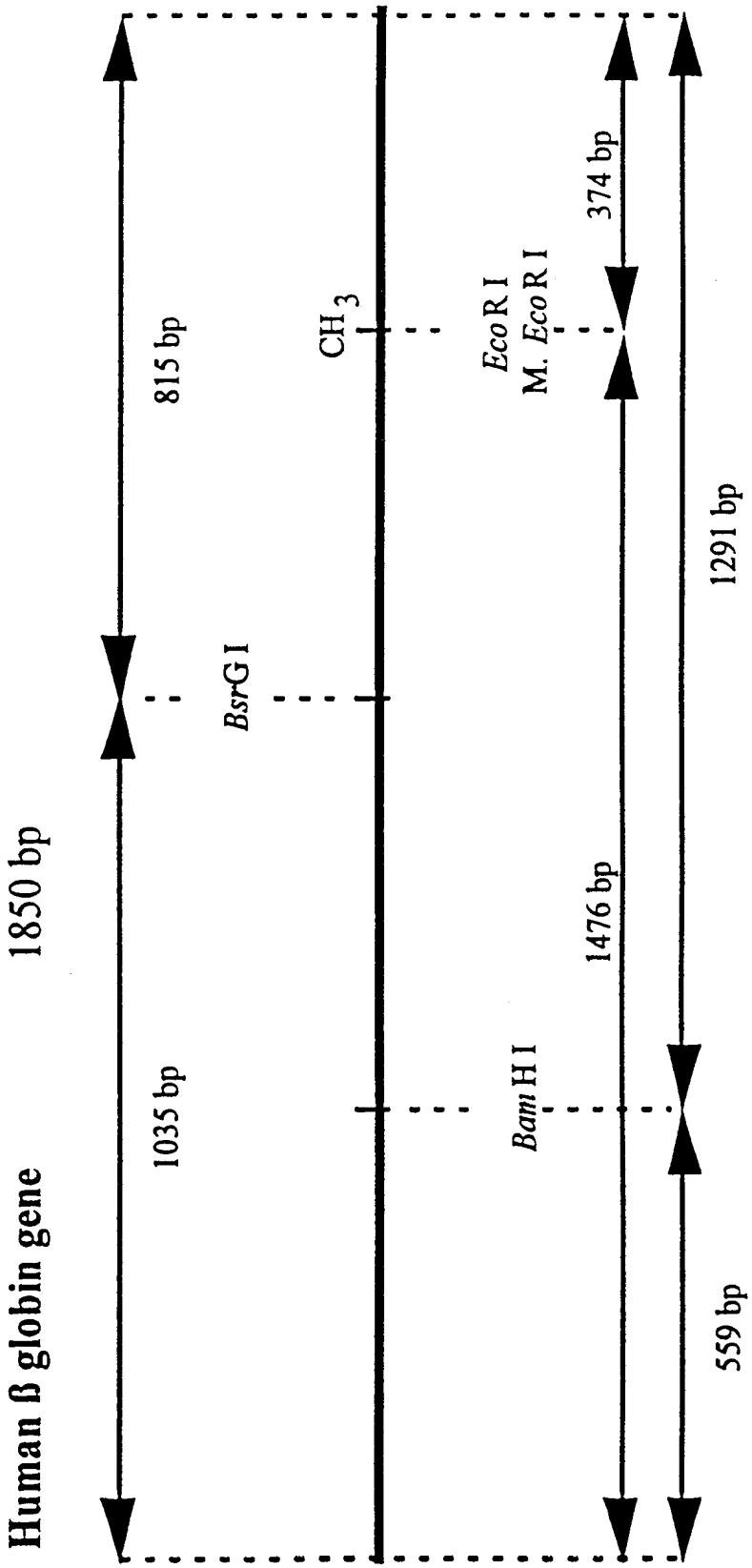

Human ß globin gene
1850 bp
1035 bp
815 bp
BsrG I
CH3
Eco R I
M. Eco R I
374 bp
1476 bp
1291 bp
Bam H I
559 bp

Fig. 11C

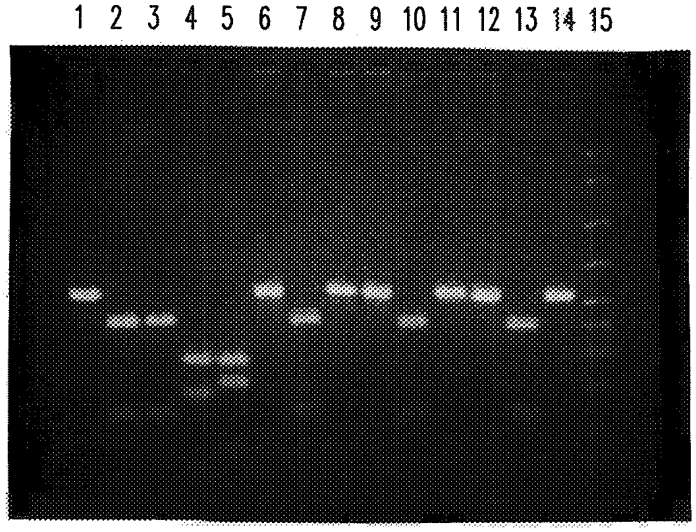


Fig. 12A

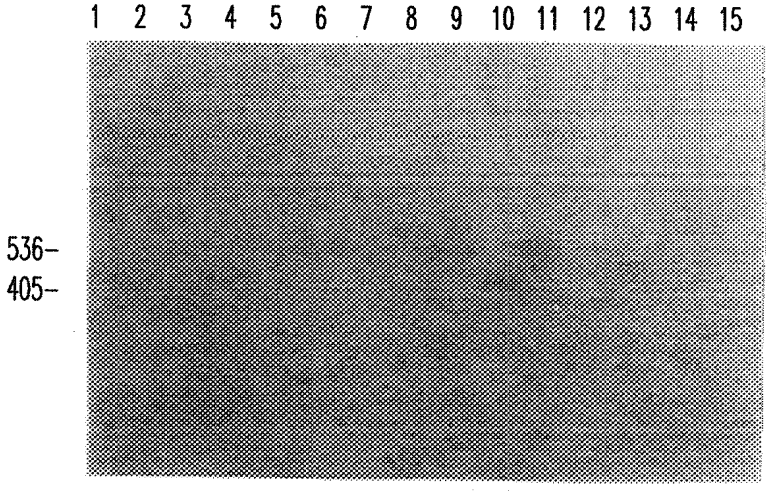


Fig. 12B

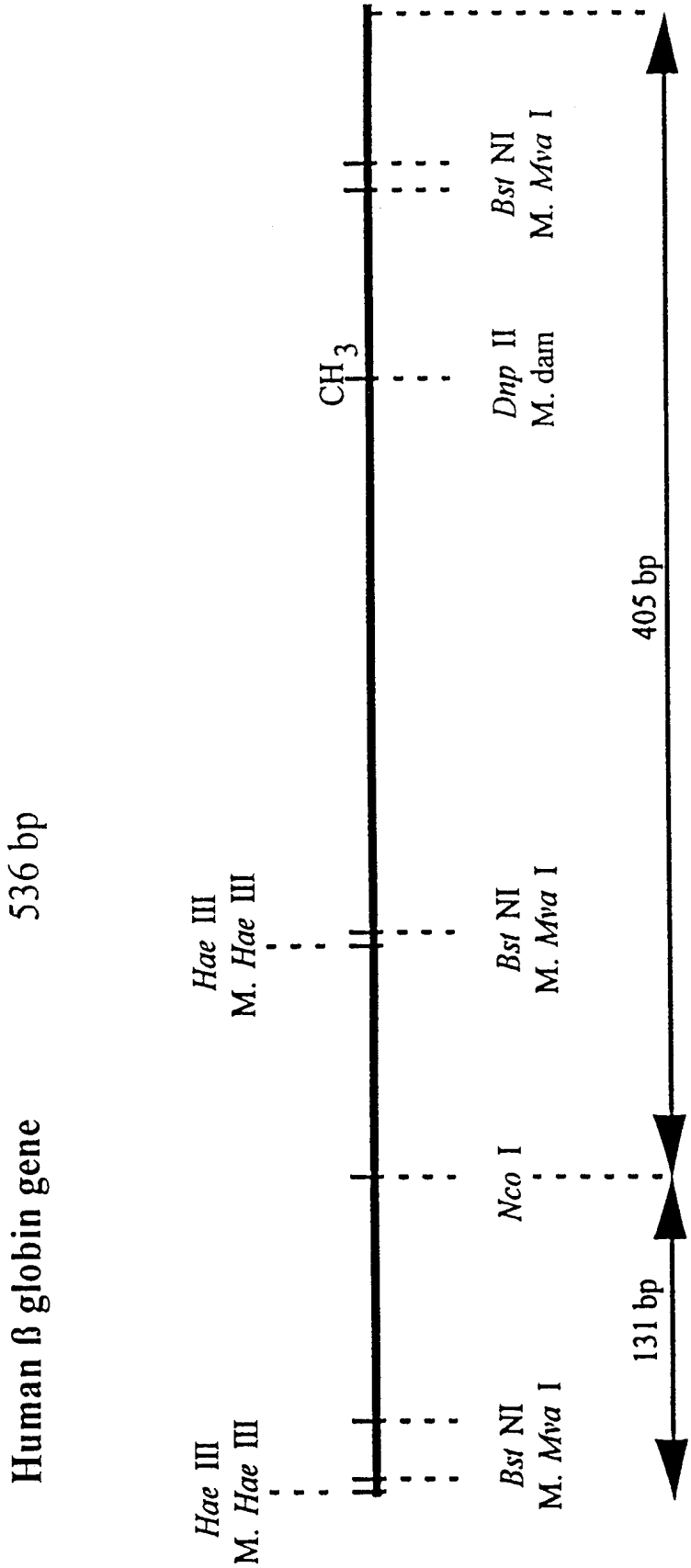

Human ß globin gene
536 bp
Hae III
M. Hae III
Hae III
M. Hae III
Bst NI
M. Mva I
Nco I
Bst NI
M. Mva I
Bst NI
M. Mva I
Dnp II
M. dam
CH3
Bst NI
M. Mva I
405 bp
131 bp

Fig. 12C

INTERNATIONAL SEARCH REPORT

Inte onal Application No
PCT/US 98/17859

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12Q1/68

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12Q

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	NELSON M ET AL: "Use of DNA methytransferase /endonuclease enzyme combinations for megabase mapping of chromosomes " METHODS IN ENZYMOLOGY, vol. 216, 1992, pages 279-303, XP002091913 cited in the application see the whole document ---	1,11, 23-25, 32,46, 52-55
A	WO 96 27024 A (AVITECH DIAGNOSTICS INC) 6 September 1996 see the whole document ---	1-3,9, 11, 23-25, 32,52-55

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
2 February 1999	16/02/1999

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Osborne, H

1

Form PCT/ISA/210 (second sheet) (July 1992)

IDT Exhibit 1020.0615

INTERNATIONAL SEARCH REPORT

Inte onal Application No
PCT/US 98/17859

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	WO 90 05195 A (DNA PLANT TECHN CORP) 17 May 1990 see the whole document	1,11, 23-25, 32,46, 52-55
A	WO 92 20825 A (BAYLOR COLLEGE MEDICINE ;UNIV EMORY MED (US)) 26 November 1992 see claims 27-31	1,11, 23-25, 32,46, 52-55
A	WO 95 25176 A (RAPAPORT ERICH ;YISSUM RES DEV CO (IL); BEN NERIAH YINON (IL); ZIO) 21 September 1995 see claims 1-6	1,11, 23-25, 32,46, 52-55
A	WO 90 12891 A (GENMAP INC) 1 November 1990 see page 13, line 18 - page 15, line 32	1,11, 23-25, 32,46, 52-55

1

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

IDT Exhibit 1020.0616

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inter nal Application No
PCT/US 98/17859

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9627024 A	06-09-1996	US 5843649 A	01-12-1998
		AU 5171496 A	18-09-1996
WO 9005195 A	17-05-1990	NONE	
WO 9220825 A	26-11-1992	AU 2185492 A	30-12-1992
WO 9525176 A	21-09-1995	IL 108978 A	22-02-1998
		AU 2071295 A	03-10-1995
WO 9012891 A	01-11-1990	AU 3730689 A	16-11-1990
		EP 0467883 A	29-01-1992

IDT Exhibit 1020.0617

Espacenet

Bibliographic data: WO9911819 (A1) — 1999-03-11

IMPROVED DETECTION OF MUTATIONS IN NUCLEIC ACIDS BY CHEMICAL CLEAVAGE

Inventor(s): BLOCH WILL ± (BLOCH, WILL)

Applicant(s): PERKIN ELMER CORP [US] ± (THE PERKIN-ELMER CORPORATION)

Classification: **- international:** ***C12N15/09; C12Q1/68; G01N33/50;*** (IPC1-7): C12Q1/68
- cooperative: **C12Q1/6827** more

Application number: WO1998US16385 19980806 Global Dossier

Priority number (s): US19970919385 19970828

Also published as: AU731069 (B2) AU8772798 (A) CA2268368 (A1) EP0991779 (A1) JP2002511767 (A) more

Abstract of WO9911819 (A1)

Improved methods, compositions and kits for detecting mutations in nucleic acids are disclosed, such comprise annealing a piece of control nucleic acid without mutations to a piece of test nucleic acid very similar in sequence to the control nucleic acid but possibly containing mutations, treating this mixture with potassium permanganate or hydroxylamine to remove mismatched bases from the duplex nucleic acid, treating the resulting nucleic acid with any of a class of diamines, triamines, and tetraamines analogous to 1,2-ethylenediamine to cleave abasic sites, and then analyzing the chemically treated nucleic acid to determine whether cleavage has occurred and approximately at what position in the nucleic acid any cleavage has occurred. Inclusion of betaine in the base-removal reactions improves their sensitivity and specificity. This series of non-hazardous, non-noxious, chemical treatments can be performed rapidly in a single reaction tube with no or minimal intervening separation steps; and the cleavage product can be prepared for the analysis step without removing the chemical reagents used to effect cleavage.

IDT Exhibit 1020.0618

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68

A1

(11) International Publication Number: **WO 99/11819**

(43) International Publication Date: 11 March 1999 (11.03.99)

(21) International Application Number: PCT/US98/16385

(22) International Filing Date: 6 August 1998 (06.08.98)

(30) Priority Data:
08/919,385 28 August 1997 (28.08.97) US

(71) Applicant: THE PERKIN–ELMER CORPORATION [US/US]; 850 Lincoln Centre Drive, Foster City, CA 94404 (US).

(72) Inventor: BLOCH, Will; 607A N. Claremont Street, San Mateo, CA 94401 (US).

(74) Agent: GROSSMAN, Paul, D.; The Perkin–Elmer Corporation, 850 Lincoln Centre Drive, Foster City, CA 94404 (US).

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: IMPROVED DETECTION OF MUTATIONS IN NUCLEIC ACIDS BY CHEMICAL CLEAVAGE

(57) Abstract

Improved methods, compositions and kits for detecting mutations in nucleic acids are disclosed, such comprise annealing a piece of control nucleic acid without mutations to a piece of test nucleic acid very similar in sequence to the control nucleic acid but possibly containing mutations, treating this mixture with potassium permanganate or hydroxylamine to remove mismatched bases from the duplex nucleic acid, treating the resulting nucleic acid with any of a class of diamines, triamines, and tetraamines analogous to 1,2–ethylenediamine to cleave abasic sites, and then analyzing the chemically treated nucleic acid to determine whether cleavage has occurred and approximately at what position in the nucleic acid any cleavage has occurred. Inclusion of betaine in the base–removal reactions improves their sensitivity and specificity. This series of non–hazardous, non–noxious, chemical treatments can be performed rapidly in a single reaction tube with no or minimal intervening separation steps; and the cleavage product can be prepared for the analysis step without removing the chemical reagents used to effect cleavage.

IDT Exhibit 1020.0619

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

IDT Exhibit 1020.0620

IMPROVED DETECTION OF MUTATIONS IN NUCLEIC ACIDS BY CHEMICAL CLEAVAGE

FIELD OF THE INVENTION

This invention relates to the detection of mutations in nucleic acids. As such, it serves the fields of genetics, molecular biology, and medicine, being especially useful as an economical, first-stage, clinical diagnostic screen for mutations causing or predisposing for disease.

REFERENCES

Baskaran et al., *Genome Research*, 6: 633-638 (1996)
Bloch, European Patent Application 0 507 591 A2 (1992)
Carmody and Vary, *BioTechniques*, 15:692-699 (1993)
Cotton et al., U.S. Pat.No.5,217,863 (1993)
Dianzani et al., *American Journal of Human Genetics*, 48: 423-424 (1991)
Dieffenbach & Dveksler, *PCR Primer: A Laboratory Manual*, CSHL Press (1995)
Gogos et al., *Nucleic Acids Research*, 18: 6807-6814 (1990)
Maxam and Gilbert, *Methods in Enzymology*, 65, 499-560 (1980)
McHugh and Knowland, Nucleic Acids Research 23: 1664-1670 (1995)
Melchior and von Hippel, *Proceedings of the National Academy of Sciences U.S.A*, 70: 292-302 (1973)
Meo et al.,WO95/07361 (1995)
Mytelka and Chamberlin, *Nucleic Acids Research*, 24(14):2774-2781 (1996)
Papp et al., *Molecular Diagnosis*, 1: 59-64 (1996)
Ramus and Cotton, *BioTechniques* , 21, 216-220 (1996)
Rees et al., *Biochemistry*, 32:, 137-144 (1993)
Rubin and Schmid, *Nucleic Acids Research*, 8:4613-4619 (1980)
Shapiro et al., *Biochemistry*, 9: 3233-3241 (1969)
Weissensteiner and Lanchbury, *BioTechniques*, 21: 1102-1108 (1996)

BACKGROUND

Understanding the variation, or polymorphism, in a species's genome serves multiple scientific and practical goals. Examples include the study of evolutionary relatedness, the study of the genetic relatedness and geographical dispersion of populations within a species, the study of disease mechanism and predisposition, genetic linkage mapping in genome study and animal and plant breeding, the forensic identification and differentiation of individuals, the clinical diagnosis and prognosis of inherited diseases

IDT Exhibit 1020.0621

and cancer, and the clinical diagnosis and prognosis of drug resistance in infectious microorganisms.

A variety of methods have been described for discovering mutations, a process sometimes called "mutation scanning." Representative methods include denaturing gradient gel electrophoresis, constant denaturant gel electrophoresis, temporal temperature gradient gel electrophoresis, single-stranded conformational polymorphism analysis, denaturing HPLC, direct heteroduplex electrophoretic analysis, protein truncation test, immobilized mismatch binding protein assay, cleavage fragment length polymorphism analysis, enzymatic mismatch scanning, chemical cleavage of mismatch analysis (CCM), and complete sequence analysis (e.g., Sanger sequencing). As presently practiced, each of these methods suffers from one or more disadvantages which limits its value for cost-effective, high-throughput, reliable, mutation scanning, including low clinical specificity (a significant false-positive rate); low clinical sensitivity (a significant false-negative rate); low sensitivity to mutations present in only a small fraction of the nucleic acid (with a high normal background); limitation to relatively short sequences; low information content; low throughput; low potential for automation; high labor requirement; long turnaround time; dependence on expensive, labile, or proprietary reagents; or dependence on toxic or noxious reagents.

Of the mutation scanning methods listed above, CCM is one of the most attractive, especially for clinical applications requiring specificity, sensitivity, high information content, and low cost. However, conventional CCM methods (Cotton; Meo) suffer from two major weaknesses: (1) the use of toxic and noxious reagents and (2) the need for multiple, labor-intensive, hard-to-automate, technically demanding separation steps to remove the CCM reagents from the nucleic acid sample.

Conventional CCM entails the application of Maxam-Gilbert sequencing chemistry (Maxam and Gilbert) to the reactive mismatched bases generated when largely identical normal and mutant ("parental") sequences are mixed, denatured, and re-annealed to create a mixture of parental (perfectly matched) duplexes and heteroduplexes, where the heteroduplexes, containing one strand from each of the parental sequences, are mismatched at any mutant positions. There are two common ways to create heteroduplexes. (1) Nucleic acid samples purified from control and test individuals or tissues can be mixed, denatured by heating or adding alkali, and renatured by cooling or adding acid to return the pH approximately to neutrality. This method works whether the test sample is homozygous or heterozygous for the mutation. (2) Alternatively, as long as the test sample is likely to contain a mixture of normal and mutant sequences (e.g., because it is heterozygous for any mutation or because it is purified from a mixture of

IDT Exhibit 1020.0622

normal and mutant cells, as in many cancers), it can be denatured and renatured to form heteroduplex without adding control nucleic acid (Dianzani). The nucleic acid in question is almost always entirely DNA, but CCM is distinguished from some of the other mutation scanning methods in that one or both nucleic acid samples could be RNA. The parental nucleic acids could, in principle, be single-stranded, as long as the mixture contained the complementary strands needed to form a duplex molecule; but in practice the parental molecules are almost always duplex DNA, generated by PCR amplification of a sequence from genomic DNA or from cDNA generated from RNA by a reverse transcription reaction.

In conventional CCM, chemical treatment of potential heteroduplex to cleave at mismatched sites occurs in two stages. First, in a base-removal step, the nucleic acid is treated either (a) with moderately concentrated (16 mM) aqueous osmium tetroxide (OsO_4) in fairly concentrated (0.3 M) pyridine under conditions favoring destruction of mismatched thymines while minimally affecting matched base pairs and mismatched purines and cytosines, or (b) with concentrated (about 4 M) aqueous hydroxylamine (NH_2OH) in concentrated (about 2 M) diethylammonium chloride under conditions favoring destruction of mismatched cytosines, substantially sparing matched base pairs and mismatched purines and thymines. The two base-removal reactions are run separately, in parallel, in order to assure that all mutations are detected and to supply some information about the chemical nature of the mutation. Following the base-removal reaction, the base-removal reagent is separated from the nucleic acid, typically by ethanol precipitation. Next, in a cleavage reaction, concentrated (1 M) aqueous piperidine is incubated with the nucleic acid and incubated at about 90°C for about 20 minutes to cleave any abasic sites generated by the base-removal reaction. Because the base-removal reaction usually removes only pyrimidines, cleavage commonly generates nicked duplex nucleic acid rather than double-stranded breaks. Then ethanol precipitation is used again, to eliminate the cleavage reagent, before dissolving the nucleic acid in a solvent suitable for subsequent analysis. Typically, analysis of the cleaved heteroduplex is performed by denaturing gel electrophoresis, in a high-resolution sequencing gel capable of detecting differences the size of the parental strands and any cleavage fragments to within one or a few nucleotides. Also typically, the nucleic acid strands are tagged at the 5' end with either radioactivity or fluorescent dyes, in a way which allows approximate or exact location of the cleavage point in the nucleic acid sequence. For example, one PCR primer is 5'-tagged with a fluorescein dye, the other primer is tagged with a rhodamine dye, and the cleavage product is analyzed on a multicolor, automated DNA sequencer (Meo). Any heteroduplex molecule with only one mismatched position can be cleaved on only the strand(s) containing a mismatched pyrimidine, but every heteroduplex population generated from duplex parental molecules must contain two, "complementary", mismatched heteroduplex

IDT Exhibit 1020.0623

molecules. Therefore, several different fluorescent electropherograms can be obtained for base-substitution mutations, depending on whether the mismatch pair is A-A and T-T, C-C and G-G, A-G and T-C, or A-C, and G-T, and on whether OsO_4 or NH_2OH was used for mismatched pyrimidine removal. With two-dye fluorescent analysis, the position and chemical nature of the mutation can be determined from the two patterns obtained with the two base-removal reagents.

Informative though the conventional CCM method is, it is unsuitable for economical, high-throughput, clinical mutation screening. OsO_4 is volatile, toxic, and noxious; pyridine is volatile and noxious; NH_2OH is volatile, putatively carcinogenic, and mildly noxious; and piperidine is volatile and noxious. This toxicology is aggravated by the high concentrations needed for effectiveness. Furthermore, the several nucleic acid precipitation and washing steps increase operator exposure to the undesirable reagents, require considerable operator skill and judgment to get good recoveries, and are difficult to automate. Finally, hot aqueous piperidine occasionally causes some background cleavage at sites where no base removal has occurred.

One published effort to reduce the operational complexity of conventional CCM effected mismatched cytosine and thymine removal serially on one sample rather than in parallel on two samples, but could not avoid the inclusion of a separation step after each chemical step (Ramus and Cotton). It was reported to be important to remove cytosine before thymine.

Gogos et al. showed that dilute (0.1 mM) potassium permanganate ($KMnO_4$, a non-noxious and effectively non-toxic reagent) could replace OsO_4/pyridine for mismatched thymine removal from heteroduplexes, with much improved sensitivity and specificity if the reaction occurred in 3 M tetramethylammonium chloride (TMAC) (Gogos). They also showed that 0.5 to 4 M tetraethylammonium chloride (TEAC) (2 M preferred) increased the sensitivity and specificity of mismatched cytidine removal by NH_2OH. They reduced the need for demanding separation steps by adsorbing the DNA to an ion-exchange paper, performing the chemical treatments on paper-bound DNA, and changing chemical conditions simply by shifting the paper from one reagent solution to another with an intervening water wash. Unfortunately, the commercial supplier of the ion-exchange paper (Amersham) since has discontinued its sale. Furthermore, the piperidine treatment removed the cleaved DNA from the paper, so that time-consuming lyophilization was needed to remove the concentrated piperidine from the DNA before electrophoretic analysis. In addition, Gogos *et al.* stated that their improved reagents were more specific (and therefore successful) for immobilized DNA than for DNA treated in

IDT Exhibit 1020.0624

solution. Another limitation of this work was that almost all experiments studied the cleavage of labeled oligonucleotide probes interrogating only rather short (approximately 30 nt) sequences: unrealistic models for the PCR-generated polynucleotide samples needed for cost-effective clinical mutation scanning.

In a totally different application not dependent on base-removal reagents (chemical detection of photoproducts in UV-damaged DNA), McHugh and Knowland showed that 1,2-ethylenediamine and two related compounds (piperazine and N,N'-dimethylethylenediamine) can replace piperidine for cleaving abasic sites in duplex DNA. These relatively non-toxic reagents generate cleavage products with subtle structural differences yielding observable electrophoretic mobility differences, but they have several striking advantages over piperidine which could greatly benefit mutation detection. (1) They are effective at much lower concentrations than piperidine (as low as 20 mM for ethylenediamine). (2) They cause much less background cleavage of perfectly matched duplex DNA than does piperidine.

The TMAC found by Gogos to improve $KMnO_4$ sensitivity and specificity may be crucial to this advance because of the documented binding to and stabilization of duplex DNA by tetramethylammonium ion (Shapiro; Melchior and von Hippel). The necessarily high concentration of TMAC, a strong electrolyte, should interfere with direct electrophoretic analysis because it increases conductivity of the electrophoresis sample by over an order of magnitude. It may also block the electrostatic interaction of ethylenediamine and its analogs with duplex DNA which was theorized by McHugh and Knowland to be responsible for their unusual cleavage efficiency.

Betaine, a zwitterionic analog of tetramethylammonium ion which has no net charge at pH values near and above neutrality, appears to affect duplex DNA stability much the same way that tetramethylammonium ion does (Rees). The benefit of high betaine concentrations already has been applied to the sequencing and PCR amplification of difficult, usually GC-rich, nucleic acid targets, where this additive improves the interaction of the DNA polymerase with its substrate (in some way not yet completely understood) without increasing the ionic strength to levels which would inhibit the enzyme (Papp; Mytelka and Chamberlin; Baskaran; Weissensteiner and Lanchbury).

The operational simplicity of the present invention derives significantly from finding chemical conditions where both pyrimidines (and even purines) are removed from mismatched positions while remaining substantially untouched in correctly matched duplex DNA. Uracil can be completely substituted for thymine in PCR with some reduction of amplification efficiency; radical reduction in the sensitivity of hybridization-based

IDT Exhibit 1020.0625

detection of dU-containing amplicon implies that replacement of T with dU significantly changes the conformation or stability of AT-rich sequences of duplex DNA (Carmody and Vary). It has been noted in sequencing applications, which examine single-stranded DNA, that although thymine resists hydroxylamine attack at all pH values, uracil can be removed by NH_2OH at high pH (Rubin and Schmid). This observation raises the question whether under appropriate conditions NH_2OH can function as a single reagent for removal of both mismatched C and U from duplex DNA. The technical challenge in improving mutation scanning by CCM is to optimize and streamline the use of chemistries developed to analyze single-stranded DNA in an application which preserves the structure and chemical inertia of perfectly matched regions of duplex DNA.

SUMMARY

The present invention is directed towards the discovery of improved methods, reagents and kits for performing mutation scanning by chemical cleavage of mismatched heteroduplex nucleic acid.

It is an object of the invention to minimize the number of steps required to perform a CCM process, in particular, to avoid separation steps.

It is another object of the invention to reduce the toxicity of reagents used in a CCM process.

It is yet another object of the invention to reduce the level of nonspecific cleavage in a cleavage reaction step of a CCM process.

It is another object of the invention to improve the speed, automatability and reliability of a CCM process.

It is yet another object of the invention to perform all chemical steps in only one or two reaction vessels with minimal liquid transfer of the nucleic acid.

A first aspect of the invention entails the use of potassium permanganate in the base-removal step of CCM, at a concentration and under conditions of time, temperature, and solvent composition effective to cause removal of mismatched thymines and cytosines in duplex nucleic acid substantially without removal of correctly matched bases. In a first embodiment, this aspect employs a base-removal solvent containing a tetraalkylammonium salt in the concentration range of approximately 1M to approximately 5M. The preferred alkyl groups of the tetraalkylammonium cation are methyl and ethyl. Many

IDT Exhibit 1020.0626

tetraalkyammonium salts are effective; the anion of the salt may be taken from the group consisting of chloride, fluoride, hydrogen sulfate, sulfate, nitrate, silicate, and tetrafluoroborate. In this embodiment, this reaction product is substantially desalted before the cleavage step, for example, by adding a 10- to 100-fold excess volume of water and subjecting the diluted mixture to diafiltration until approximately the original volume is obtained. To the desalted mixture is added a cleavage reagent, taken from the group consisting of 1,2-diamines, at a concentration effective for the cleavage of any abasic sites generated in the base-removal step, substantially without cleavage of normal phosphodiester bonds. After incubation at a pH between about 6 and about 9, at a temperature and for a time effective for cleavage of abasic sites, the cleavage mixture is diluted directly into the reagent used to prepare the sample for fragment size analysis. The group of 1,2-diamines comprises 1,2-ethylenediamine, diethylenetriamine, triethylenetetraamine, piperazine, N,N'-dimethylethylenediamine, 1,2-diaminopropane, 1,2-diamino-2-methylpropane, cis-1,2-diaminocyclohexane, and trans-1,2,-diaminocyclohexane. Because cleavage of abasic sites is much more efficient in single-stranded than in double-stranded nucleic acid, the cleavage reagent preferably is accompanied by a nucleic acid denaturant. Preferred denaturants are urea, formamide, dimethyl sulfoxide, and lactams, such as 2-pyrrolidinone and 1-methyl-2-pyrrolidinone. In this case, the thermal treatment for cleavage should include transient exposure to a temperature high enough to denature any duplex nucleic acid.

In a second, preferred, embodiment of the first aspect of the invention, no tetraalkylammonium salt is used during base removal; instead, the base-removal solvent contains betaine in the concentration range of approximately 1 M to 5 M. Alternatively, neither betaine nor tetraalkylammonium salt may be present during mismatched base removal. In this latter embodiment, the product of the base-removal step enters directly into the cleavage step, following addition of a cleavage reagent taken from the group consisting of 1,2-diamines, at a concentration effective for the cleavage of any abasic sites generated in the base-removal step, substantially without cleavage of normal phosphodiester bonds. After incubation at a pH between about 6 and about 9, at a temperature and for a time effective for cleavage of abasic sites, the cleavage mixture is diluted directly into the reagent used to prepare the sample for fragment size analysis. The cleavage reaction preferably is effected in the presence of nucleic acid denaturant at a concentration which renders the test nucleic acid substantially single-stranded.

IDT Exhibit 1020.0627

A second, less preferred, aspect of the invention entails first the use of potassium permanganate in the base-removal step of CCM, at a concentration and under conditions of time, temperature, and solvent composition effective to cause removal primarily of mismatched thymines in duplex nucleic acid substantially without removal of correctly matched bases or mismatched cytidines, followed by addition to this reaction mixture of hydroxylamine in the presence of substantially no additional salt, at a pH between about 8 and about 10, and at a concentration effective for the removal of mismatched cytidines from duplex nucleic acid substantially without removal of correctly matched bases. In a first embodiment, this aspect employs a base-removal solvent containing a tetraalkylammonium salt in the concentration range of approximately 1 M to approximately 5 M. Many tetraalkylammonium salts are effective; the anion of the salt may be taken from the group consisting of chloride, fluoride, hydrogen sulfate, sulfate, nitrate, silicate, and tetrafluoroborate. In this embodiment, the base-removal product is substantially desalted before the cleavage step, for example, by adding a 10- to 100-fold excess volume of water and subjecting the diluted mixture to diafiltration until approximately the original volume is obtained. To the desalted mixture is added a cleavage reagent, taken from the group consisting of 1,2-diamines, at a concentration effective for the cleavage of any abasic sites generated in the base-removal step, substantially without cleavage of normal phosphodiester bonds. After incubation at a pH between about 6 and about 9, at a temperature and for a time effective for cleavage of abasic sites, the cleavage mixture is diluted directly into the reagent used to prepare the sample for fragment size analysis. The cleavage reaction preferably is effected in the presence of a nucleic acid denaturant at a concentration which renders the test nucleic acid substantially single-stranded.

In a second, preferred, embodiment of the second aspect of the invention, no tetramethylammonium salt is used during base removal; instead, the base-removal solvent contains betaine in the concentration range of approximately 1 M to 5 M. Alternatively, neither betaine nor a tetraalkylammonium salt may be present during mismatched base removal. In this latter embodiment, the product of the base-removal step enters directly into the cleavage step, following addition of a cleavage reagent, taken from the group consisting of 1,2-diamines, at a concentration effective for the cleavage of any abasic sites generated in the base-removal step, substantially without cleavage of normal

IDT Exhibit 1020.0628

phosphodiester bonds. After incubation at a pH between about 6 and about 9, at a temperature and for a time effective for cleavage of abasic sites, the cleavage mixture is diluted directly into the reagent used to prepare the sample for fragment size analysis. The cleavage reaction preferably is effected in the presence of a nucleic acid denaturant at a concentration which renders the test nucleic acid substantially single-stranded.

A third, preferred, aspect of the invention entails the use of dUTP instead of TTP in the PCR used to prepare samples for CCM analysis and the use of hydroxylamine in the base-removal step of CCM, at a concentration and under conditions of time, temperature, and solvent composition (including pH) effective to cause removal of mismatched uracils and cytosines in duplex DNA substantially without removal of correctly matched bases. The product of the base-removal step enters directly into the cleavage step, following addition of a cleavage reagent, taken from the group consisting of 1,2-diamines, at a concentration effective for the cleavage of any abasic sites generated in the base-removal step, substantially without cleavage of normal phosphodiester bonds. After incubation at a pH between about 6 and about 9, at a temperature and for a time effective for cleavage of abasic sites, the cleavage mixture is diluted directly into the reagent used to prepare the sample for fragment size analysis. The cleavage reaction preferably is effected in the presence of a nucleic acid denaturant at a concentration which renders the test nucleic acid substantially single-stranded. In an especially preferred embodiment of this aspect of the invention, the base-removal step of CCM occurs in the presence of betaine in the concentration range of about 1 M to about 5 M. In another preferred embodiment, the same approximate concentration of betaine is present during the PCR used to prepare samples for CCM analysis. A less preferred embodiment of this aspect of the invention employs a base-removal solvent containing a tetraalkylammonium salt in the concentration range of approximately 1M to approximately 5 M, wherein the alkyl group is taken from the group consisting of ethyl and methyl. Many such tetraalkylammonium salts are effective; the anion of the salt may be taken from the group consisting of bromide, chloride, fluoride, hydrogen sulfate, sulfate, nitrate, acetate, formate, hydrogen phthalate, silicate, and tetrafluoroborate. In this embodiment, the base-removal product is substantially desalted before the cleavage step, for example, by adding a 10- to 100-fold excess volume of water and subjecting the diluted mixture to diafiltration until approximately the original volume is obtained.

IDT Exhibit 1020.0629

The invention further comprises methods, compositions, and kits which incorporate and enable the various aspects and embodiments described above.

These and other objects, features, and advantages of the present invention will become better understood with reference to the following description, drawings, and appended claims.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Reference will now be made in detail to the preferred embodiments of the invention. While the invention will be described in conjunction with the preferred embodiments, it will be understood that they are not intended to limit the invention to those embodiments. On the contrary, the invention is intended to cover alternatives, modifications, and equivalents, which may be included within the invention as defined by the appended claims.

Generally, mutation scanning by chemical cleavage of mismatch (CCM) is a four-step process. First, normal and putatively mutated versions of the same general nucleic acid sequence, commonly prepared by PCR, are mixed, denatured, and renatured so that, if mutations are present, heteroduplex molecules will be formed which have mismatched bases at the mutated positions. Second, the heteroduplex nucleic acid is treated with a base-removal reagent which removes bases, commonly the pyrimidines, e.g., thymine and cytosine, from the mismatched positions in the heteroduplexes while sparing correctly matched bases. Such base removal results in an abasic site at that position. Third, the base-removal product is treated with a cleavage reagent which cleaves the heteroduplex at the abasic sites created in the base-removal reaction, preferably under conditions which render the treated heteroduplex substantially single-stranded. Finally, the cleavage product is analyzed, commonly by denaturing electrophoresis, to characterize the cleavage products produced in the cleavage reaction.

The success of various aspects and embodiments of the invention depends in part (1) on the surprising chemical compatibilities (a) of betaine with $KMnO_4$ and NH_2OH removal of mismatched bases from heteroduplex nucleic acid, (b) of betaine with 1,2-diamine cleavage of abasic sites, (c) of $KMnO_4$ with NH_2OH removal of mismatched cytosine, (d) of $KMnO_4$ and hydroxylamine with 1,2-diamine cleavage of abasic sites, (e) of nucleic acid denaturants such as 2-pyrrolidinone with 1,2-diamine cleavage of abasic sites, and (f) of $KMnO_4$, NH_2OH, betaine, and 1,2-diamines with electrophoretic

IDT Exhibit 1020.0630

separation; (2) on the discovery of chemical conditions where mismatched thymine (or uracil) and cytosine are removed from duplex nucleic acid with comparable efficiencies; and (3) on the surprising improvements in cleavage of abasic sites effected by certain members of the group of 1,2-diamines, especially under conditions wherein the nucleic acid has been rendered single-stranded.

I. DEFINITIONS

Unless stated otherwise, the following terms and phrases as used herein are intended to have the following meanings:

"1,2-diamine" means a molecule having the general structure



wherein (1) R_1 and R_8 are H atoms; (2) R_{2-7} can be H atoms, alkyl groups, or aralkyl groups; (3) any single pair of R_{2-7} can be covalently linked to form a ring composed of 5-10 atoms; (4) R_2 and R_7 can be alkyl groups wherein one or more methylene groups have been replaces with NH groups; (5) R_2 and R_7 can be alkyl groups wherein one or more methyl groups have been replaced with NH_2 groups; (7) R_{2-7} can contain one or more carboxyl, carboxamide, or carboxylic ester groups; (8) the atoms in R_2 and R_7 immediately next to the indicated N atoms are either H atoms or C atoms bonded to a total of four other atoms; and (9) said molecule contains no aldehyde groups, ketone groups, or other groups with substantial covalent reactivity toward amino groups at temperatures below about 50C. For the purpose of the present invention, preferred 1,2-diamines are 1,2-ethylenediamine, diethylenetriamine, triethylenetetraamine, piperazine, N,N'-dimethylethylenediamine, 1,2-diaminopropane, 1,2-diamino-2-methylpropane, cis-1,2-diaminocyclohexane, and trans-1,2,-diaminocyclohexane.

"Nucleic acid denaturant" means a non-ionic, water-soluble, organic compound which, when added to an aqueous solution of nucleic acid, promotes the transformation of any duplex or triplex nucleic acid into single-stranded nucleic acid. The solution may be heated to a temperature in the approximate range of 50-100°C to effect this transformation, or may be so treated and then rapidly cooled to approximately 25°C or below. For the purpose of the invention, preferred nucleic acid denaturants are dimethyl

IDT Exhibit 1020.0631

sulfoxide, urea, formamide, or a lactam, applied separately or in combination. Preferred lactams are 2-pyrrolidinone and 1-methyl-2-pyrrolidinone.

II. TEST NUCLEIC ACID

Although the chemical basis of mutation scanning by chemical cleavage of mismatch (CCM) serves equally well for RNA and DNA, in practice the method is preferably applied to a DNA test nucleic acid because RNA is harder to obtain in sufficient quantity in stable form.

Preferably the test nucleic acid component of the nucleic acid sample is obtained through a polymerase chain reaction (PCR) which targets a sequence of about 100 to about 10,000 nucleotides in length. Most practically, the PCR product is no longer than about 3000 nucleotides. The upper limit of test nucleic acid length is dictated by the degree of size resolution of the analysis method used to analyze the cleavage fragments. Generally, longer PCR products ("amplicons") are preferred simply because fewer different PCR reactions are needed to cover a given genetic target thereby reducing the cost of a mutation analysis. The PCR may be applied to genomic DNA or, in the case of RT-PCR, to cDNA generated by applying a reverse transcription reaction to RNA, usually mRNA. For eukaryotic genes, RT-PCR is somewhat preferred over amplification of genomic targets simply because amplicon size can be maximized without concern about exon size. There is some preference to exclude intronic target sequence, except possibly that close to the splice junctions, because the introns are often polymorphic in ways which have no known biological consequence. Such "silent" polymorphisms can act as a background signal in CCM output thereby reducing sensitivity to clinically important mutations which influence the amount and/or structure of mRNA and therefore of expressed protein.

PCR and RT-PCR methods are well known to practitioners of the art of genetic analysis and are therefor in need of little elaboration (Dieffenbach & Dveksler). Preferably, the PCR amplification is optimized such that (1) it generates between about 1 and 100 nM of specific PCR product (enough to see on an ethidium-stained agarose electrophoresis gel using traditional methods), (2) creates negligible misprimed or truncated side products, and (3) results in less primer oligomer than specific PCR product. Hot-Start PCR methods, also well known in the field of mutation detection, often serve to improve amplification specificity and sensitivity, especially when amplification is performed on a small number of target copies, e.g., less than about 1000 molecules, or on long targets. Relatively long annealing and extension times, from 2 to about 6 minutes, also help to minimize truncation product and maximize specific PCR product yield, especially for targets longer than about 1000 nucleotides.

IDT Exhibit 1020.0632

Amplicon is labeled according to the analytical method utilized in the analysis step of the CCM process. Preferred labels include radioactive labels (e.g., ^{32}P), fluorescent dyes, or a binding moiety which is a member of a binding pair, e.g., biotin, antigen, receptor and the like. Preferably, two different spectrally resolvable dyes are attached to the two primers used to create an amplicon. Then, if the cleavage fragments are analyzed in a multi-color fluorescent nucleic acid fragment analyzer, such as an automated DNA sequencer, the association of fragment size with color unambiguously estimates the position of any mutation in the amplicon. More preferably, the same two dye-tagged primers are used to create both the control nucleic acid (if a control is used) and the test nucleic acid. Finally, with this fragment-tagging strategy, it often is practical and desirable (in the interest of cost reduction) to co-amplify several amplicons of resolvable size. As modern multi-color automated DNA sequencers can simultaneously analyze 4 different dyes and soon will be able to process more than 4 dyes at the same time, different amplicons can use different pairs of dyes to allow unambiguous identification of which multiplexed amplicon generated a given set of fragments. However, even if only two dyes are used, this distinction still can be made by correlating the appearance of cleavage fragments with the loss of full-length amplicon in the analytical electropherogram, as long as every amplicon gives a unique signal.

Several labeling strategies may be used to label the nucleic acids of the invention. In a first method, the 5'-end of the primer is labeled, e.g., by performing the amplification reaction using a labeled primer. Alternatively, the amplicon is internally labeled, e.g., by incorporating labeled dNTP's (mixed with un-tagged dNTP's) during the PCR. Although internal labeling has some advantage with respect to signal strength, especially if multiple mutations are expected in an amplicon, it has several weaknesses in most contexts. Because all fragments created by a cleavage are labeled, it is hard to estimate the position of the mutation in the target sequence, complicating second-tier analysis, especially when amplicons over about 500 nt in length are made. In addition, if the internal labels are dyes or binding moieties, there will be a binomial distribution of amplicons bearing different numbers of labels. Because such label molecules affect electrophoretic mobility, such heterogeneous internal labeling broadens peaks, reducing sensitivity and resolution. Finally, internal labeling results in large differences in signal strength between large and small fragments, complicating development of a method which detects all fragments with comparable sensitivity.

In one alternative aspect of the present invention wherein hydroxylamine is used in place of permanganate ion in the base-removal reaction (see below), it is necessary to replace the customary thymines in the nucleic acid sample with uracils. This change,

IDT Exhibit 1020.0633

accomplished by replacing dTTP completely with dUTP in the PCR used to generate the test nucleic acid and any control nucleic acid which is used, may reduce PCR yield by as much as an order of magnitude. Effective remedies to restore PCR efficiency include (a) increasing the dUTP concentration by a factor of 2 to 5, and (b) increasing the annealing or extension times by several minutes. Inclusion of betaine in the PCR, at a concentration of about 1M to 3M, also may counteract the mild inhibition of PCR by dUTP. This change may require a reduction in the annealing or extension temperatures by up to about 5°C in order to restore amplification efficiency.

III. HETERODUPLEX FORMATION

There are two preferred methods suitable for generating the heteroduplex nucleic acid of the present invention. In a first method, the nucleic acid sample is composed of a control nucleic acid and a test nucleic acid from the individual or tissue suspected of carrying mutations. This first approach is required to identify mutations for which the test sample is homozygous or otherwise genetically uniform (as, for example, results from oncogenic loss of heterozygosity). In an alternative method, no control nucleic acid is included and only a target nucleic acid is used in the heteroduplex formation reaction. Thus, in the second method, heteroduplex is formed simply by subjecting the test nucleic acid to denaturation and renaturation, as described above. This second approach suffices if the test sample is heterozygous for the mutation of interest or if the source of the test sample is a tissue, such as a tumor or a precancerous lesion, containing both normal and oncogenically mutated cells.

Whether using the first or second heteroduplex formation method, the general steps of the method are as follows. The nucleic acid sample is denatured briefly to create single-stranded nucleic acid, and it then is allowed to renature to regenerate duplex DNA in which the two strands from the test nucleic acid (and control nucleic acid if present) have reassociated in a random manner. Denaturation/renaturation most conveniently is performed by heating the mixture to a temperature between about 90°C and 100°C, holding the mixture at the higher temperature for only a few seconds (to minimize potential damage to the nucleic acid), and then returning the temperature to a value between about 4 °C and 25 °C. A time- and temperature-programmable thermal cycler, such as is used to control the PCR, is a preferred apparatus for completing the heteroduplex formation operation.

IV. BASE-REMOVAL REACTION

In an important aspect of the present invention, base removal is effected using a cleavage reagent including permanganate ion, e.g., $KMnO_4$. This reaction occurs optimally at a temperature between about 0°C and about 50°C, for times between about 5

IDT Exhibit 1020.0634

min and about 30 min, at final permanganate concentrations of between about 0.1 mM and about 10 mM, buffered at a pH in the range of 6 to 9, e.g., by a mixture of ammonia and ammonium sulfate (above about pH 8.25) or by a phosphate salt (below about pH 8.25). Because of the way that time, temperature, pH, and permanganate ion concentration contribute to reaction rate, there are many possible combinations of these variables with comparable base-removal efficiency. Shorter times, down to about 5 min, are preferable in order to minimize total analysis time, however times substantially shorter than 5 minutes are undesirable because the reaction becomes more difficult to control reproducibly. Higher temperatures are preferred because they offer a simple way to control reaction completeness, i.e., the base-removal reaction can be effectively stopped by returning the temperature to the range of 0 °C to 25 °C. For example, the base-removal reaction can be controlled by adding a small volume, perhaps a few microliters, of a 1 to 10 mM $KMnO_4$ stock solution to a 10 to 100 µl volume of denatured/renatured DNA in PCR buffer at 4°C in a 500 microliter microfuge tube, raising the temperature to 50 °C in a DNA thermal cycler and holding it there for 5 minutes, and then returning the temperature rapidly to 4 °C. Base-removal temperatures above about 50 °C are risky because they might lead to sufficient local denaturation of perfectly matched AT-rich sequences to yield unacceptable false positive results. $KMnO_4$ treatment of PCR reaction mixtures will oxidize residual untreated dNTP's and primers, thereby reducing the amount of $KMnO_4$ available to react with mismatched bases. This side reaction can be minimized by designing the PCR to employ the lowest practical dNTP concentrations, on the order of 10 µM, and in any case below about 100 µM. A pH9 buffer consisting of a mixture of ammonia and ammonium sulfate is especially effective for PCR, instead of the Tris buffer commonly used for PCR.

Optimization of base removal by permanganate ion treatment has several goals: (1) maximize removal of mismatched T's (the easiest target for permanganate ion), (2) maximize removal of mismatched C's (a harder target for permanganate ion), and (3) minimize removal of any nonmismatched nucleotides in the heteroduplex nucleic acid. Certain control reactions are useful for monitoring the performance of the base-removal reaction and identifying optimal reaction conditions: (1) a normal control (preferably one with an AT-rich subdomain), containing no heteroduplex, (2) an originally heterozygous sample carrying a TA to AT transversion, and (3) an originally heterozygous sample containing a CG to GC transversion. [The background reaction also can be monitored at non-mutated positions in the two transversion samples.]

In a preferred embodiment of the present invention, the reagent used to effect the base-removal reaction includes, in addition to permanganate ion, a high concentration, e.g., 1 to 5 M, of either a tetraalkylammonium (preferably tetramethylammonium) salt or betaine (a neutral, zwitterionic, tetramethylammonium analogue). These additives serve to

IDT Exhibit 1020.0635

enhance base-removal efficiency and specificity, and, by stabilizing AT-rich duplex nucleic acid, permit the use of higher base-removal temperatures (for enhanced control, as described above) without increasing false positive signals.

Betaine is the most preferred alternative because it contributes nothing to the ionic strength of the reaction mixture and therefore does not have to be removed from the reaction product prior to conducting the cleavage reaction and subsequent analysis of the processed heteroduplex. Betaine is soluble in water up to about 5 M. Working betaine concentrations as high as 2 to 3 M are attained easily by mixing a 5 M betaine stock solution, possibly containing diluted permanganate ion, with approximately an equal volume of the nucleic acid sample. However, a preferred way to reach a high betaine concentration in the base-removal reaction is to perform the PCR (used to generate the test and control nucleic acids) directly in 2 to 3 M betaine. In addition to facilitating the CCM reaction, this modification of conventional PCR conditions often improves PCR specificity and does not reduce amplification efficiency if the annealing and extension temperatures are reduced from otherwise optimal values by a few Celsius. Modest increases in PCR annealing and extension times are also desired when including betaine in the PCR. Betaine is capable of reducing $KMnO_4$, thereby reducing the effectiveness of the latter reagent for mismatched base removal. Therefor $KMnO_4$ oxidation in the presence of betaine preferably is conducted at a temperature below about 25°C.

If a tetraalkyammonium salt is added to the base-removal reaction instead of betaine, it preferably is not included in the PCR in the high concentration range needed to benefit the base-removal reaction. Therefore the tetraalkylammonium salt is preferably added after completion of the PCR from a concentrated stock solution, which also may serve as the permanganate ion stock solution, typically by diluting the DNA, already in a PCR buffer, approximately 50% with the stock reagent. Preferred tetraalkylammonium salt anions include chloride, fluoride, hydrogen sulfate, sulfate, nitrate, silicate, and tetrafluoroborate (Aldrich Fine Chemicals, Milwaukee, WI).

When tetraalkylammonium salt is used in the base-removal reaction, its concentration must be substantially reduced (by at least 90 to 99%) before subjecting the base-removal product to a cleavage reaction. While ethanol precipitation may be used to remove the tetraalkylammonium salt, in the present invention, alternative desalting methods are preferred which are faster, require less rigorous temperature control, more reliably give high recovery, and require less operator skill. In one preferred desalting method the base-removal reaction product is filtered using a diafilter having a membrane which substantially retains duplex DNA larger than about 50 base pairs and passes smaller molecules. An additional benefit of diafiltration is that the membrane also passes tagged

IDT Exhibit 1020.0636

PCR primers, thereby reducing the low-molecular-weight background in the CCM analysis step. In practice, small-volume diafiltration such as this can be performed by diluting the sample 10- to 100-fold with water and applying relatively low gas pressure to the diluted sample above or vacuum to the space below a diafiltration membrane mounted in a pressure/vacuum manifold (or in a microtiter plate mounted in a vacuum manifold). However, preferably diafiltration is performed by placing the diluted base-removal reaction mixture in a spin-filtration microconcentrator and centrifuging at low speed (less than about 5000 rcf) in an angle-head microcentrifuge for the period of time, commonly less than about 10 minutes, needed to return the volume to a value 50 to 100% as large as the undiluted sample volume. A suitable commercially available device is the disposable Pall Filtron (Northborough, MA) NANOSEP™ 30K Microconcentrator, though the NANOSEP™ 10K device also works. In a preferred technique for minimizing sample-handling steps all CCM chemical processing steps are performed in the sample reservoir of such a centrifugal microconcentrator, as the membrane in the bottom passes negligible fluid without the application of pressure and the sample reservoir cap minimizes evaporative loss.

In an alternative desalting method the base-removal reaction mixture is contacted with a small volume of a beaded mixed-bed ion exchange or ion-retardation resin, such as AG501-X8, Bio-Rex MSZ501(D), or AG11 A8, from Bio-Rad Laboratories (Hercules, CA). The minimum wet capacities of these resins (0.7 to 1.5 meq/mL) are such that on the order of 200 to 500 μL of resin are needed to desalt 100 μL of 3 M tetraalkylammonium salt, and the resin size exclusion limit, about 1000 daltons, assures that analytically important nucleic acid is not captured inside the beads.

In an additional embodiment of the present invention, cytidine removal during the base-removal reaction is facilitated by addition of an effective amount of hydroxylamine to the permanganate base-removal reagent. Comparable hydroxylamine treatment conditions are used, whether betaine or a tetraalkylammonium salt is present. In contrast to conventional hydroxylamine treatment in CCM, the hydroxylamine stock solution is prepared from the free base (commercially available as a 50% aqueous solution) rather than the hydrochloride salt, pH-adjusted with acid to the highest pH, on the order of 8 to 10, consistent with complete avoidance of duplex nucleic acid denaturation. Such an approach minimizes the contribution of hydroxylamine to the ionic strength, so that if hydroxylamine treatment is performed in the presence of betaine instead of a tetraalkylammonium salt, a desalting step is not needed before cleavage and subsequent analysis. The minimum hydroxylamine concentration is used which is effective for mismatched cytidine cleavage. To minimize the required hydroxylamine concentration,

IDT Exhibit 1020.0637

the highest practical reaction temperature, e.g., approximately 50°C, and the longest practical reaction time, e.g., 30 to 60 minutes, are used.

In an alternative aspect of the present invention, base removal is accomplished using hydroxylamine in place of permanganate ion. In this aspect, the nucleic acid sample is prepared with complete replacement of thymine by uracil. The principal condition applied to the hydroxylamine treatment is that the pH must be high enough, preferably above approximately pH 9, to favor ionization of mismatched uracils, yet low enough, preferably below approximately pH 10.5, not to destabilize correctly matched duplex nucleic acid. The inclusion of about 1 M to 3 M betaine, tetramethylammonium salt, or tetraethylammonium salt in the base-removal reaction helps to stabilize duplex nucleic acid, thereby raising the maximum tolerated pH. Should a tetraalkylammonium salt be used, it is removed, e.g. by diafiltration, before cleavage of abasic sites as described above. Betaine is preferred for duplex nucleic acid stabilization, not only because it avoids the need for desalting, but also because its nonionic nature favors the base-removal reaction.

V. CLEAVAGE REACTION

After completion of the base-removal reaction, the heteroduplex nucleic acid is subjected to a cleavage reaction which causes the cleavage of the heteroduplex at any abasic site, i.e., nucleotide positions at which a base has been removed. According to the present invention, the cleavage reaction is effected by incubating the heteroduplex with a 1,2-diamine, e.g., 1,2-ethylenediamine, diethylenetriamine, triethylenetetraamine, piperazine, N,N'-dimethylethylenediamine, 1,2-diaminopropane, 1,2-diamino-2-methylpropane, cis-1,2-diaminocyclohexane, or trans-1,2,-diaminocyclohexane.

Preferably, the cleavage reaction is conducted at a 1,2-diamine concentration in the range of about 10 to 100 mM, at a pH of about 5 to about 9, at a temperature of about 25°C to about 40°C, for an elapsed time between about 10 minutes and one hour. Also preferably, the cleavage reaction contains a nucleic acid denaturant, such as 2-pyrrolidinone or urea, at a concentration in the range of about 1 M to 10 M. The exact combination of conditions depends on which cleavage reagent is used and what combination of reagents and solvents was used in the base-removal reaction.

If the base-removal reagent is not removed by desalting, preferably the temperature of the cleavage reaction is at least 10°C lower than that used in the base-removal reaction. This lower reaction temperature is desirable in order to minimize the activity of any residual base-removal reagent (in order to minimize base removal at non-mismatched sites).

IDT Exhibit 1020.0638

If hydroxylamine is used for base removal and is not removed by a desalting step prior to conducting the cleavage reaction, residual hydroxylamine activity may be reduced by conducting the cleavage reaction at a pH in the 5-6 range and using a cleavage reagent which maintains an unprotonated amino group in that pH range, e.g., diethylenetriamine, triethylenetetraamine, cis-1,2-diaminocyclohexane, or trans-1,2,-diaminocyclohexane. If a tetraalkylammonium salt is used during base removal, then desalting, as described above, is done prior to conducting the cleavage reaction, and the preferred desalting buffer is used instead of water to dilute the sample before desalting. If betaine and hydroxylamine are used during base removal, the need for desalting depends on whether hydroxylamine adds enough ionic strength to interfere with the post-cleavage separation step. If betaine and no hydroxylamine is used during the base-removal reaction, no desalting is done before adding the cleavage reagent to the nucleic acid. Preferably, a cleavage reagent stock solution is prepared which is 10 to 100 times more concentrated than is desired for the actual cleavage reaction. This stock is pH-adjusted so that after dilution into the nucleic acid sample, the desired final pH is attained.

Regardless of the particular chemistry used to effect the base removal and cleavage reactions, subsequent to the cleavage reaction the cleavage product is in a suitable condition to be analyzed under denaturing conditions without further separation steps to remove residual reagents. Preferably, in order to minimize the activity of residual reagents which might create a background signal unrelated to the presence of mutations, the cleavage product is incubated at 0°C - 4°C subsequent to the cleavage reaction and prior to analysis.

VI. ANALYSIS OF CLEAVED HETERODUPLEX

Any method which can identify the presence of and location of heteroduplex cleavage by fragment size or otherwise may be used to analyze the cleavage products of the present invention. One preferred analytical method is denaturing electrophoresis using a "sequencing gel" or in a capillary containing entangled polymer using a detection system including a multicolor fluorescent detector, e.g., the Model 373A (slab gel), Model 377 (slab gel), or Model 310 (capillary) sequencer from Applied Biosystems (Foster City, CA). Analysis is performed according to instructions provided by Applied Biosystems for GeneScan™ fragment analysis. These instructions define the electrophoresis loading reagent (a mixture of a DNA denaturant, a chelator, and a loading dye, usually augmented with fluorescent size-standard fragments) into which the cleavage product is diluted, the methods of preparing and loading the electrophoresis medium, the method of running the electrophoresis with real-time fluorescent detection, and the method of analyzing the

IDT Exhibit 1020.0639

resulting electropherograms to determine which fragment sizes were produced with which dye tag.

Alternatively, cleavage products may be analyzed in only a few minutes by anion-exchange HPLC, as described by the present inventor in European Patent Application No. 0 507 591 A2, (1992). The resolution and size range of this separation can be adjusted to the desired performance level by computer control of the length and steepness of the salt elution gradient. Furthermore, automated fraction collection of fragments, using standard equipment and procedures, can supply material for direct sequence analysis to confirm the exact nature of a mutation.

Regardless of the nature of the mutation, e.g., base substitution, insertion, or deletion, heteroduplex generation results in two heteroduplex molecules and regenerates the two parental homoduplex molecules (normal and mutant), because there are four ways that the two plus strands (normal and mutant) and the two minus strands (normal and mutant) of a denatured mixture of normal and mutant duplex nucleic acid molecules can recombine. At each mutated position, there is a mismatched pyrimidine/purine, pyrimidine/pyrimidine or purine/purine base pair in each heteroduplex molecule. CCM chemistry cleaves primarily, though not always completely, at mismatched pyrimidines. Therefore, because there are two heteroduplex molecules for each mutated position, a total of four single-stranded cleavage fragments are formed. As long as the CCM process uses base-removal chemistry which cleaves at all mismatched pyrimidines and as long as the labeling technique used employs distinguishable labels for each of the two primers responsible for any amplicon, every mutation results in two tagged cleavage fragments in one or two colors (in the case of a base substitution, depending on whether it is a transition or transversion). The relationships between fragment color and size determine unambiguously the approximate position of the mutation in the amplicon. The two tagged peaks in the analytical output (e.g., electropherogram) may not be of similar strength, because base-removal efficiency depends on sequence context.

VII. EXAMPLES

The invention will be further clarified by a consideration of the following prophetic examples, such examples being intended to be purely exemplary of the invention and not to in any way limit its scope.

EXAMPLE 1

DNA is purified by methods well known to molecular biologists, from two immortalized cell lines from the Coriell Human Genetic Mutant Cell Repository (Camden, NJ). Cell line # GM12785 is a compound heterozygote for two cystic fibrosis (CF)

IDT Exhibit 1020.0640

mutations (in the CFTR gene): R347P (a G-to-C transversion at cDNA nucleotide position 1172 in exon 7) and G551D (a G-to-A transition at cDNA nucleotide position 1784 in exon 11). Cell line # GM11282 is also a CF compound heterozygote: mutations 621+1G/T (a G-to-T transversion at the first intronic nucleotide downstream from cDNA nucleotide position 621, just beyond the end of exon 4) and G85E (a G-to-A transition at cDNA nucleotide position 386 in exon 3). Because all four mutations are heterozygous, an approximately equimolar mixture of normal and mutant DNA is created for a mutation whenever primers specific for one of the two mutated exons is used to amplify any one of the DNA samples. Four primer pairs are used in this example, as summarized in Table I.

Table I: CFTR Exonic Primer Pairs

Exon	Strand	Primer Name	Primer Sequence	5" Dye Tag
3	plus	CF001	CCCTGGGATAGAGAGCTGGCT	FAM
	minus	CF002	CCCTGATACATAATGAATGTACAAATGAG	TET
4	plus	CF003	CCCGGAAGAATCATAGCTTCCTATG	FAM
	minus	CF004	CCCAACATGTACGATACAGAATATATGTG	HEX
7	plus	CF007	CCCGAAGGCAGCCTATGTGAG	FAM
	minus	CF008	CCCGTTTGTACAGCCCAGGG	TET
11	plus	CF011	CCCCCTTTCAAATTCAGATTGAG	FAM
	minus	CF012	CCGAATGACATTTACAGCAAATG	HEX

A total of 8 PCR's are performed. Each DNA sample is amplified with each of the four primer pairs of Table I. Table II summarizes the DNA fragments expected to be seen if the denatured/renatured PCR products are subjected to CCM in which cleavage occurs only at mismatched pyrimidines in the heteroduplexes and the cleavage products are analyzed on a multicolor fluorescent automated DNA sequencer.

Table II. Predicted CCM Fluorescent Fragment Patterns

Source DNA	Primers	Parental Fragments	Cleavage Fragments
GM11282	CF001/002	139nt (FAM,TET)	53nt (FAM,TET)
	CF003/004	237nt (FAM,HEX)	183nt(FAM),53nt(HEX)
	CF007/008	189nt (FAM,TET)	none

IDT Exhibit 1020.0641

	CF011/012	206nt (FAM,HEX)	none
GM12785	CF001/002	139nt (FAM,TET)	none
	CF003/004	237nt (FAM,HEX)	none
	CF007/008	189nt (FAM,TET)	149nt(FAM),39nt(TET)
	CF011/012	206nt (FAM,HEX)	76nt (FAM,HEX)

A wax-mediated Hot Start PCR is performed according to the following design. A bottom reaction mixture is prepared, consisting of 40 μM of each of the four standard dNTP's, 0.05% Triton X-100, and 0.1 units/μL of AmpliTaq® DNA polymerase (Perkin Elmer-Applied Biosystems, Foster City, CA). For each PCR, 10 μL of bottom reaction mixture is added to the bottom of a MicroAmp™ reaction tube (PE-Applied Biosystems); one PCR Gem-50™ (PE-Applied Biosystems) is added to each tube, which then is closed and heated to 70°C for 3 min in a GeneAmp™ PCR System 9600 thermal cycler (PE-Applied Biosystems) to layer the wax. When the tube has returned to room temperature, over the wax is placed 10 μL of a top reaction mixture consisting of 40 mM ammonium sulfate adjusted to pH 9.3 with ammonia, 2.7 mM magnesium sulfate, 4 M betaine, 100 genome copies (0.6 ng) of the desired purified human genomic DNA, and two primers (specific for the exon targeted) each at a concentration of 0.5 μM. Amplification is performed in the same thermal cycler by running 3 cycles, at the fastest possible ramp rates, with the following settings: 98°C for 5 seconds, 55°C for 6 minutes, 70°C for 2 minutes; then 26 cycles with the following settings: 94°C for 5 seconds, 55°C for 6 minutes, 70°C for 2 minutes. Then the polymerase is inactivated and heteroduplex is formed by heating at 99.9°C for 20 minutes, followed by rapid cooling to room temperature (fastest possible ramp rate). This step can be programmed to occur immediately after the last PCR cycle rather than as a separate operation. [The polymerase should be inactivated before heteroduplex formation to assure that primer extension does not compete with heteroduplex formation. If the prolonged heating is undesirable, e.g., because the DNA is labile under these conditions, enzyme can be inactivated by adding disodium EDTA to a final concentration of 5 mM between the end of the PCR and a much milder denaturation/renaturation step, such as heating to 95°C for 15 seconds before return to room temperature.]

In order to achieve mismatched thymine and cytidine removal from the heteroduplex formed by the preceding operations, each 20 μL reaction volume is mixed with 5 μL of 5 mM $KMnO_4$ in water and incubated at 50°C for 5 min in a thermal cycler before rapid cooling to room temperature. In order to effect cleavage of abasic sites, each 25 μL reaction volume is (a) mixed with 20 μL of 2-pyrrolidinone and 5 μL of 400 mM 1,2-ethylenediamine which has been adjusted to pH 8.0 with acetic acid, (b) incubated at

IDT Exhibit 1020.0642

99°C for 5s, and (c) is incubated at 30°C for 30 minutes in a thermal cycler before rapid cooling to room temperature. All three of these steps are performed in the same microtube used for PCR and heteroduplex formation.

The cleavage products then are subjected to GeneScan™ analysis in a Model 310 automated DNA sequencer (PE-Applied Biosystems) according to the manufacturer's instructions. FAM-tagged fragments are observed in the blue channel of filter set C; TET-tagged fragments are tracked in the green channel; HEX-tagged fragments are tracked in the yellow channel. The red channel is used for a TAMRA-tagged GeneScan™ size-standard cocktail (PE-Applied Biosystems). The terms, FAM, TET, HEX, and TAMRA, refer to a set of spectrally resolved, fluorescein and rhodamine derivatives developed by PE-Applied Biosystems, attached to the primer 5' ends by chemistries fully described by the manufacturer. The GeneScan™ electropherograms for the eight different PCR products subjected to the CCM process just described show parental and cleavage-product peaks at approximately the sizes indicated in Table II, above, and substantially no other peaks except in the size range below about 30 nt. This latter smattering of very small fragments likely arises from low-level primer inhomogeneity, not the existence of mutations, as can be verified by their presence at comparable strength when homozygous normal samples are examined (i.e., the primer and DNA combinations predicted in Table II to have no cleavage fragments).

EXAMPLE 2

PCR, heteroduplex formation, mismatched base removal, abasic site cleavage, and GeneScan analysis are performed as in Example 1 except for the following changes: (1) betaine is omitted from the PCR top reaction mixture; (2) annealing temperatures in the PCR thermal cycles are increased from 55°C to 59°C; (3) for the first stage of mismatched base removal (thymine removal), the 20 μL of denaturation/renaturation product is mixed with 10 μL of 1.2 mM $KMnO_4$ 6M TMAC and incubated at 50°C for 5 minutes before rapid cooling to room temperature; (4) (a new step) for mismatched cytidine removal, 5 μL of 1M hydroxylamine, adjusted to pH8.5 with acetic acid, are added to the 30 μL of thymine-removal product, which then is incubated at 50°C for 15 minutes before rapid cooling to room temperature; (5) (a new step) the entire product (35 μL) from the cytidine-removal step is mixed with 465 μL water, added to the sample reservoir of a NANOSEP™ 30K microconcentrator (Pall Filtron, Northborough, MA) and centrifuged at room temperature in an angle-head rotor at 4000 rcf until the volume has been reduced to about 10 μL and the total volume is restored to 25 μL with water before addition of 2-pyrrolidinone and ethylenediamine, as above.

IDT Exhibit 1020.0643

Abasic site cleavage and GeneScan analysis are performed as described in Example 1. The results are substantially as described in Example 1.

EXAMPLE 3

PCR, heteroduplex formation, mismatched base removal, abasic site cleavage, and GeneScan analysis are performed as in Example 1 except for the following changes: (1) for the first stage of mismatched base removal (thymine removal), the 20 μL of denaturation/renaturation product are mixed with 5 μL of 2.0 mM $KMnO_4$ and incubated at 50°C for 5 minutes before rapid cooling to room temperature; (2) (a new step) For mismatched cytidine removal, 4 μL of 1M hydroxylamine, adjusted to pH 8.5 with acetic acid, are added to the 25 μL of thymine-removal product, which then is incubated at 50°C for 15 minutes before rapid cooling to room temperature; (3) for abasic site cleavage, 23 μL of 2-pyrrolidinone and 6 μL of 400 mM 1,2-ethylenediamine which has been adjusted to pH 8.0 with acetic acid are added to the 29 μL of cytosine-removal product.

The balance of the abasic site cleavage process and GeneScan analysis are performed as described in Example 1. The results are substantially as described in Example 1.

EXAMPLE 4

PCR, heteroduplex formation, mismatched base removal, abasic site cleavage, and GeneScan analysis are performed as in Example 1 except for the following changes: (1) betaine is omitted from the PCR top reaction mixture; (2) annealing temperatures in the PCR thermal cycles are increased from 55°C to 59°C; (3) For mismatched base removal (thymine and cytidine removal), the 20 μL of denaturation/renaturation product are mixed with 10 μL of 3.0 mM $KMnO_4$ 6M TMAC and incubated at 50°C for 5 minutes before rapid cooling to room temperature; (4) (a new step) the entire product (25 μL) from the mismatched pyrimidine-removal step is mixed with 475 μL water, added to the sample reservoir of a NANOSEP™ 30K microconcentrator (Pall Filtron, Northborough, MA) and centrifuged at room temperature in an angle-head rotor at 4000 rcf until the volume has been reduced to about 10 μL. The total volume is restored to 25 μL with water before addition of 2-pyrrolidinone and ethylenediamine, as above.

Abasic site cleavage and GeneScan analysis are performed as described in Example 1. The results are substantially as described in Example 1.

EXAMPLE 5

PCR, heteroduplex formation, mismatched base removal, abasic site cleavage, and GeneScan analysis are performed as in Example 1 except for the following changes: (1)

IDT Exhibit 1020.0644

dTTP in the PCR bottom reaction mixture is completely replaced by 200 μM dUTP; (2) the duration of the PCR annealing step (at 55°C) is increased from 2 to 4 minutes in all cycles; (3) mismatched base removal by $KMnO_4$ is omitted, and instead, the 20 μL of heteroduplex is mixed with 5 μL of unbuffered 2M hydroxylamine and incubated at 50°C for 30 minutes before rapid cooling to room temperature; (4) the 400mM 1,2-ethylenediamine stock solution is adjusted to pH 5.0 instead of pH 8.0 before addition to the base-removal product. The results are substantially as described in Example 1.

All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

Although only a few embodiments have been described in detail above, those having ordinary skill in the molecular biology art will clearly understand that many modifications are possible in these preferred embodiments without departing from the teachings thereof. All such modifications are intended to be encompassed within the following claims.

IDT Exhibit 1020.0645

I CLAIM:

1. A mutation detection method comprising the steps of:

denaturing a nucleic acid sample, the nucleic acid sample comprising a test nucleic acid;

reannealing the nucleic acid sample to form a heteroduplex nucleic acid;

performing a base-removal reaction by reacting the heteroduplex nucleic acid with permanganate ion;

performing a cleavage reaction by reacting the heteroduplex nucleic acid with 1,2-diamine; and

analyzing the heteroduplex nucleic acid to determine whether and at what location cleavage of the heteroduplex nucleic acid has occurred.

2. The method of claim **1** wherein the nucleic acid sample further includes a control nucleic acid.

3. The method of claim **1** wherein the 1,2-diamine is selected from the group consisting of 1,2-ethylenediamine, diethylenetriamine, triethylenetetraamine, piperazine, N,N'-dimethylethylenediamine, 1,2-diaminopropane, 1,2-diamino-2-methylpropane, cis-1,2-diaminocyclohexane, or trans-1,2,-diaminocyclohexane.

4. The method of claim **3** wherein the 1,2-diamine is selected from the group consisting of diethylenetriamine, triethylamine tetraamine, cis-1,2-diaminocyclohexane, trans-1,2,-diaminocyclohexane, 1,2-diaminopropane, and 1,2-diamino-2-methylpropane.

5. The method of claim **1** wherein the base-removal reaction is performed in the presence of betaine.

6. The method of claim **5** wherein the betaine is present at a concentration of between about 1 M and about 5 M.

7. The method of claim **5** wherein the base-removal reaction further includes reacting the heteroduplex nucleic acid with hydroxylamine.

8. The method of claim **1** wherein the base-removal reaction is performed in the presence of tetramethylammonium salt.

9. The method of claim **8** wherein the tetramethylammonium salt is present at a concentration of between about 1 M and about 5 M.

IDT Exhibit 1020.0646

10. The method of claim **8** wherein the base-removal reaction further includes reacting the heteroduplex nucleic acid with hydroxylamine.

11. The method of claim **8** wherein the tetramethylammonium salt contains an anion selected from the group consisting of bromide, chloride, fluoride, hydrogen sulfate, sulfate, nitrate, acetate, formate, hydrogen phthalate, silicate, and tetrafluoroborate.

12. The method of claim **8** wherein the heteroduplex nucleic acid is desalted subsequent to the base-removal reaction.

13. The method of claim **12** wherein the heteroduplex nucleic acid is desalted by diafiltration.

14. The method of claim **13** wherein the diafiltration is performed using a membrane with an effective size cutoff below about 50 base pairs of duplex DNA.

15. The method of claim **1** wherein the base-removal reaction further includes reacting the heteroduplex nucleic acid with hydroxylamine.

16. The method of claim **1** wherein the step of analyzing the heteroduplex nucleic acid comprises an analytical separation.

17. The method of claim **16** wherein the analytical separation is electrophoresis.

18. The method of claim **17** wherein the electrophoresis takes place in a medium which denatures the nucleic acid.

19. The method of claim **16** wherein the analytical separation is HPLC.

20. The method of claim **1** wherein the heteroduplex is labeled.

21. The method of claim **20** wherein the label is a fluorescent label.

22. A composition comprising:
an aqueous solution of potassium permanganate and betaine;
wherein the aqueous solution is buffered at a pH of between about 5.5 and about 10.0.

IDT Exhibit 1020.0647

23. The composition of claim **22** wherein the potassium permanganate is present at a concentration of between about 0.05 mM and 10 mM.

24. The composition of claim **22** wherein the betaine is present at a concentration of between about 0.5 M and about 5 M.

25. The composition of claim **22** further comprising a heteroduplex nucleic acid.

26. A kit useful for conducting a mutation screening assay comprising:
a base-removal reagent comprising a permanganate salt; and
a cleavage reagent comprising a 1,2-diamine solution.

27. The kit of claim **26** wherein the 1,2-diamine is selected from the group consisting of 1,2-ethylenediamine, diethylenetriamine, triethylenetetraamine, piperazine, N,N'-dimethylethylenediamine, 1,2-diaminopropane, 1,2-diamino-2-methylpropane, cis-1,2-diaminocyclohexane, and trans-1,2,-diaminocyclohexane.

28. The kit of claim **26** further including betaine.

29. The kit of claim **28** wherein the betaine is present at a concentration of between about 1 M and about 5 M.

30. The kit of claim **26** further including a tetramethylammonium salt.

31. The kit of claim **30** wherein the tetramethylammonium salt is present at a concentration of between about 1 M and about 5 M.

32. The kit of claim **30** wherein the tetramethylammonium salt includes an anion selected from the group consisting of chloride, fluoride, hydrogen sulfate, sulfate, nitrate, silicate, and tetrafluoroborate.

33. The kit of claim **26** further including hydroxylamine.

34. The kit of claim **26** further including a control nucleic acid.

35. A mutation detection method comprising the steps of:
denaturing a nucleic acid sample, the nucleic acid sample comprising a test nucleic acid, wherein substantially all thymine nucleotides in the nucleic acid sample have been replaced with uracil nucleotides;

IDT Exhibit 1020.0648

reannealing the nucleic acid sample to form a heteroduplex nucleic acid;

performing a base-removal reaction by reacting the heteroduplex nucleic acid with hydroxylamine;

performing a cleavage reaction by reacting the heteroduplex nucleic acid with 1,2-diamine; and

analyzing the heteroduplex nucleic acid to determine whether and at what location cleavage of the nucleic acid sample has occurred.

36. The method of claim **35** wherein the nucleic acid sample further includes a control nucleic acid.

37. The method of claim **26** wherein the 1,2-diamine is selected from the group consisting of 1,2-ethylenediamine, diethylenetriamine, triethylenetetraamine, piperazine, N,N'-dimethylethylenediamine, 1,2-diaminopropane, 1,2-diamino-2-methylpropane, cis-1,2-diaminocyclohexane, or trans-1,2,-diaminocyclohexane.

38. The method of claim **37** wherein the 1,2-diamine is selected from the group consisting of diethylenetriamine, triethylamine tetraamine, cis-1,2-diaminocyclohexane, trans-1,2,-diaminocyclohexane, 1,2-diaminopropane, and 1,2-diamino-2-methylpropane.

39. The method of claim **26** wherein the step of analyzing the heteroduplex nucleic acid comprises an analytical separation.

40. The method of claim **39** wherein the analytical separation is electrophoresis.

41. The method of claim **40** wherein the electrophoresis takes place in a medium which denatures the nucleic acid.

42. The method of claim **39** wherein the analytical separation is HPLC.

43. The method of claim **26** wherein the heteroduplex is labeled.

44. The method of claim **43** wherein the label is a fluorescent label.

45. A kit useful for conducting a mutation screening assay comprising:
a base-removal reagent comprising hydroxylamine; and
a cleavage reagent comprising a 1,2-diamine.

IDT Exhibit 1020.0649

46. The kit of claim **45** wherein the 1,2-diamine is selected from the group consisting of 1,2-ethylenediamine, diethylenetriamine, triethylenetetraamine, piperazine, N,N'-dimethylethylenediamine, 1,2-diaminopropane, 1,2-diamino-2-methylpropane, cis-1,2-diaminocyclohexane, and trans-1,2,-diaminocyclohexane.

47. The kit of claim **45** further including betaine.

48. The kit of claim **47** wherein the betaine is present at a concentration of between about 1 M and about 5 M.

49. The kit of claim **45** further including a tetramethylammonium salt.

50. The kit of claim **49** wherein the tetramethylammonium salt is present at a concentration of between about 1 M and about 5 M.

51. The kit of claim **49** wherein the tetramethylammonium salt includes an anion selected from the group consisting of chloride, fluoride, hydrogen sulfate, sulfate, nitrate, silicate, and tetrafluoroborate.

52. The kit of claim **45** further including a control nucleic acid.

IDT Exhibit 1020.0650

INTERNATIONAL SEARCH REPORT

Ir. .ational Application No
PCT/US 98/16385

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12Q1/68

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12Q

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	US 4 818 268 A (HOLTZ R B) 4 April 1989 see abstract ---	22 23-25
Y	GOGOS J A ET AL.: "Detection of single base mismatches of thymine and cytosine by potassium permanganate and hydroxylamine in the presence of tetraalkylammonium salts" NUCLEIC ACIDS RESEARCH, vol. 18, no. 23, 1990, pages 6807-6814, XP002086178 cited in the application see the whole document --- -/--	1-3, 7-12, 15-27, 30-37, 39-46, 49-52

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
1 December 1998	14/12/1998

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Knehr, M

2

Form PCT/ISA/210 (second sheet) (July 1992)

IDT Exhibit 1020.0651

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
Y	ROBERTS E ET AL.: "Potassium permanganate and tetraethylammoniumchloride are a safe and effective substitute for osmium tetroxide on solid-phase fluorescent chemical cleavage of mismatch" NUCLEIC ACIDS RESEARCH, vol. 25, no. 16, 1997, pages 3377-3378, XP002086179 see the whole document	1-3, 7-12, 15-18, 20-23, 25-27, 30-37, 39-41, 43-46, 49-52
Y	MCHUGH P J AND KNOWLAND J: "Novel reagents for chemical cleavage at abasic sites and UV photoproducts in DNA" NUCLEIC ACIDS RESEARCH, XP002086180 cited in the application see abstract	1,3,26, 27, 35-37, 45,46
Y	RAMUS S J AND COTTON R G H: "Single-tube chemical cleavage of mismatch: Successive treatment with hydroxylamine and osmium tetroxide" BIOTECHNIQUES, vol. 21, no. 2, 1996, pages 216-220, XP002086181 cited in the application see the whole document	1-3,7, 10,12, 15-18, 20,26, 27, 33-37, 39-41, 43,45, 46,52
Y	EP 0 507 591 A (PERKIN ELMER CORP) 7 October 1992 cited in the application see abstract see page 2, line 38 - line 44 see page 5, line 24 - line 37 see page 10, line 10 - line 40; figure 4; example 2	8,11,19, 30,32, 35,42, 49,51
A	EP 0 329 311 A (MEDICAL RES COUNCIL) 23 August 1989 * see especially page 9, lines 22 - 28 * see the whole document	
A	COTTON R G H ET AL.: "Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES USA, vol. 85, 1988, pages 4397-4401, XP002086182 see the whole document	

-/--

2

IDT Exhibit 1020.0652

INTERNATIONAL SEARCH REPORT

Ir. ational Application No
PCT/US 98/16385

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication,where appropriate. of the relevant passages	Relevant to claim No.
A	COTTON R G H: "DETECTION OF MUTATIONS IN DNA AND RNA BY CHEMICAL CLEAVAGE" 1991 , METHODS IN MOLECULAR BIOLOGY, VOL. 9, PAGE(S) 39 - 49 XP000606219 see the whole document -----	

2

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

IDT Exhibit 1020.0653

INTERNATIONAL SEARCH REPORT

Information on patent family members

In ational Application No
PCT/US 98/16385

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 4818268 A	04-04-1989	CA 1300537 A	12-05-1992
		GB 2210868 A,B	21-06-1989
		GB 2180529 A	01-04-1987
EP 0507591 A	07-10-1992	AT 154132 T	15-06-1997
		AU 645674 B	20-01-1994
		AU 1403492 A	08-10-1992
		CA 2063855 A,C	04-10-1992
		DE 69220132 D	10-07-1997
		DE 69220132 T	13-11-1997
		EP 0691148 A	10-01-1996
		ES 2053416 T	01-08-1994
		GR 93300019 T	28-04-1993
		IL 101356 A	04-08-1996
		IL 115082 A	20-11-1997
		JP 5099909 A	23-04-1993
EP 0329311 A	23-08-1989	JP 2104300 A	17-04-1990
		US 5217863 A	08-06-1993

IDT Exhibit 1020.0654

Espacenet

Bibliographic data: WO9806736 (A1) — 1998-02-19

STABLE COMPOSITIONS FOR NUCLEIC ACID AMPLIFICATION AND SEQUENCING

Inventor(s): RASHTCHIAN AYOUB; SOLUS JOSEPH ± (RASHTCHIAN, AYOUB, ; SOLUS, JOSEPH)

Applicant(s): LIFE TECHNOLOGIES INC [US] ± (LIFE TECHNOLOGIES, INC)

Classification: - international: ***C12N15/09; C12P21/08; C12Q1/68; G01N33/50;*** (IPC1-7): C07H21/04; C12P19/34
- cooperative: **C12N9/1252**; **C12N9/96**; **C12P19/34**; **C12Q1/686**; **C12Q1/6869** more

Application number: WO1997US14266 19970814 Global Dossier

Priority number (s): US19960689815 19960814 ; US19970801720 19970214

Also published as: AU4065397 (A) EP0942917 (A1) EP0942917 (A4) EP0942917 (B1) EP2264045 (A1) more

Abstract of WO9806736 (A1)

This invention is directed to compositions comprising mixtures of reagents, including thermostable enzymes (e.g., thermostable DNA polymerases), buffers, cofactors and other components, suitable for immediate use in nucleic acid amplification or sequencing techniques without dilution or addition of further components. The compositions contain no stabilizing agents (e.g., glycerol or serum albumin) and unexpectedly maintain activity for extended periods of time upon storage at temperature above freezing. These compositions are useful, alone or in the form of kits, for nucleic acid amplification (e.g., by the Polymerase Chain Reaction, as exemplified in the Figure) and sequencing (e.g., by dideoxy or "Sanger" sequencing), or for any procedure utilizing thermostable DNA polymerases in a variety of medical, forensic and agricultural applications. In particular, the compositions and methods are useful for amplifying and sequencing nucleic acid molecules that are larger than about 7 kilobases in size.

IDT Exhibit 1020.0655

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C07H 21/04, C12P 19/34 — A1

(11) International Publication Number: **WO 98/06736**

(43) International Publication Date: 19 February 1998 (19.02.98)

(21) International Application Number: PCT/US97/14266

(22) International Filing Date: 14 August 1997 (14.08.97)

(30) Priority Data:

08/689,815	14 August 1996 (14.08.96)	US
08/801,720	14 February 1997 (14.02.97)	US

(71) Applicant: LIFE TECHNOLOGIES, INC. [US/US]; 9800 Medical Center Drive, Rockville, MD 20850 (US).

(72) Inventors: RASHTCHIAN, Ayoub; 22418 Sweetleaf Lane, Gaithersburg, MD 20882 (US). SOLUS, Joseph; 4 Ivyberry Court, Gaithersburg, MD 20879 (US).

(74) Agents: ESMOND, Robert, W. et al.; Sterne, Kessler, Goldstein & Fox P.L.L.C., Suite 600, 1100 New York Avenue, N.W., Washington, DC 20005-3934 (US).

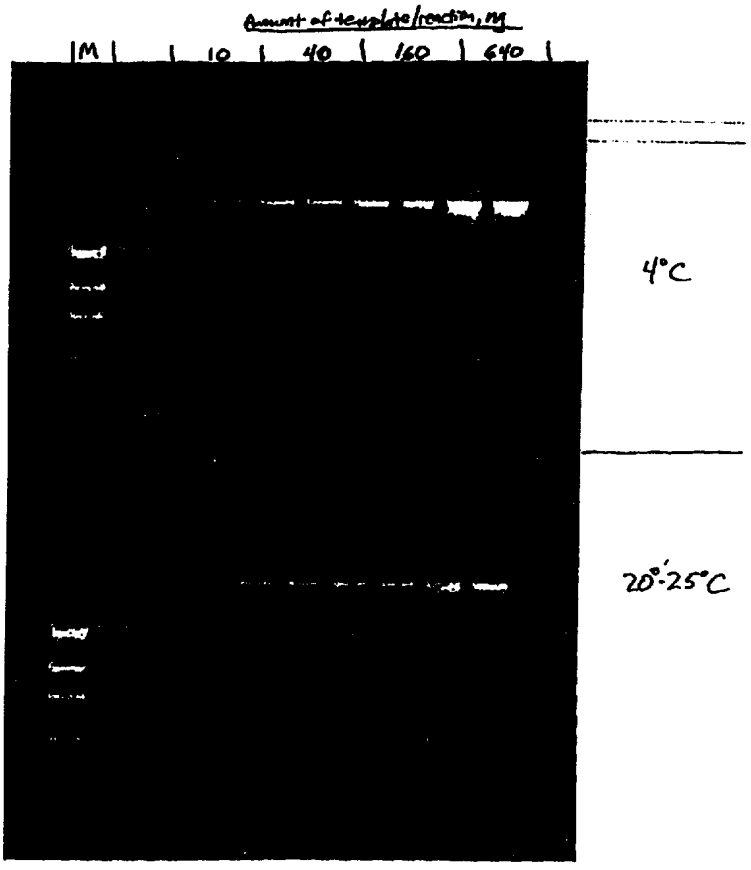
(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: STABLE COMPOSITIONS FOR NUCLEIC ACID AMPLIFICATION AND SEQUENCING

(57) Abstract

This invention is directed to compositions comprising mixtures of reagents, including thermostable enzymes (e.g., thermostable DNA polymerases), buffers, cofactors and other components, suitable for immediate use in nucleic acid amplification or sequencing techniques without dilution or addition of further components. The compositions contain no stabilizing agents (e.g., glycerol or serum albumin) and unexpectedly maintain activity for extended periods of time upon storage at temperature above freezing. These compositions are useful, alone or in the form of kits, for nucleic acid amplification (e.g., by the Polymerase Chain Reaction, as exemplified in the Figure) and sequencing (e.g., by dideoxy or "Sanger" sequencing), or for any procedure utilizing thermostable DNA polymerases in a variety of medical, forensic and agricultural applications. In particular, the compositions and methods are useful for amplifying and sequencing nucleic acid molecules that are larger than about 7 kilobases in size.



IDT Exhibit 1020.0656

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

Stable Compositions for Nucleic Acid Amplification and Sequencing

FIELD OF THE INVENTION

This invention is in the fields of molecular and cellular biology. The invention is particularly directed to reagent compositions for use in techniques whereby nucleic acids (DNA or RNA) are amplified or sequenced, and to methods of amplifying and sequencing long nucleic acid molecules, especially via the polymerase chain reaction (PCR).

BACKGROUND OF THE INVENTION

DNA Polymerases

During growth and reproduction of viruses and cellular organisms, the DNA containing the parental genetic information must be faithfully copied and passed on to the progeny. This highly regulated process of DNA replication is carried out *in vivo* by a complex of enzymes and associated proteins and cofactors (Kornberg, A., *Science 131*:1503-1508, 1959; Hibner, U., and Alberts, B.M., *Nature 285*:300-305, 1980). The primary enzymes taking part in this process are the DNA polymerases, which catalyze the addition of deoxynucleoside triphosphate (dNTP) bases into the newly forming DNA strands. Together with other enzymes (*e.g.*, helicases, ligases and ATPases), the DNA polymerases thus ensure rapid and relatively faithful replication of DNA in preparation for proliferation in prokaryotes, eukaryotes and viruses.

DNA polymerases are also used to manipulate DNA *in vitro* in a variety of molecular genetic techniques. These enzymes have proven useful not only for *in vitro* DNA synthesis, but also for determining the nucleotide sequence (*i.e.*, "sequencing") of DNA fragments or genes. This latter application relies on the fact that, in addition to an activity which adds dNTPs to DNA in the 5' to 3' direction (*i.e.*, "polymerase" activity), many DNA polymerases also possess activities which *remove* dNTPs in the 5' to 3' and/or the 3' to 5' direction (*i.e.*, "exonuclease" activity). DNA polymerases can also be used for sequencing via their incorporation of labelled chain-terminating agents such as dideoxynucleoside triphosphates (*See* U.S. Patent Nos. 4,962,020; 5,173,411; and 5,498,523). Thus the same enzyme, *e.g.*, DNA polymerases I and III from the bacterium *Escherichia coli*, DNA polymerase γ from animal cells or DNA polymerase from bacteriophage T7 (*See* U.S Patent No. 4,795,699), may be used *in vitro* both for DNA synthesis, involving the elongation of the

DNA strand, and for DNA sequencing, involving either the synthesis or the digestion of the DNA strand.

The dual activity of certain DNA polymerases is, however, a drawback for some *in vitro* applications. For example, the *in vitro* synthesis of an intact copy of a DNA fragment by the polymerase activity, an elongation process which proceeds in a 5' to 3' direction along the template DNA strand, is jeopardized by the exonuclease activities which may simultaneously or subsequently degrade the newly formed DNA. To overcome this technical problem, a fragment of *E. coli* DNA polymerase I lacking the 5' to 3' exonuclease activity (named the "Klenow fragment" after its discoverer) is often employed for *in vitro* DNA synthesis. The Klenow fragment provides for *in vitro* DNA synthesis at approximately the same rate as intact *E. coli* DNA polymerase I, but the newly synthesized DNA molecules are less subject to enzymatic degradation and are correspondingly more stable.

Unfortunately, the error rate (*i.e.*, the rate at which incorrect dNTPs are incorporated into the new DNA strand by the enzyme) is somewhat higher for the Klenow fragment than for many other commonly used DNA polymerases, including *E. coli* DNA polymerases I and II and the polymerases from bacteriophages T4 and T7 (Kunkel, T.A., *et al.*, *J. Biol. Chem. 259*:1539-1545, 1984; Tindall, K.R., and Kunkel, T.A., *Biochemistry 27*:6008-6013, 1988; Mattila, P., *et al.*, *Nucl. Acids Res.* 19:4967-4973, 1991). Thus, until recently the rates of synthesis, degradation and error had to be weighed together when choosing which DNA polymerase to use for *in vitro* DNA synthesis.

DNA Sequencing

In general, two techniques have been traditionally used to sequence nucleic acids. In the first method, termed "Maxam and Gilbert sequencing" after its co-developers (Maxam, A.M. and Gilbert, W., *Proc. Natl. Acad. Sci. USA 74*:560-564, 1977), DNA is radiolabeled, divided into four samples and treated with chemicals that selectively destroy specific nucleotide bases in the DNA and cleave the molecule at the sites of damage. By separating the resultant fragments into discrete bands by gel electorphoresis and exposing the gel to X-ray film, the sequence of the original DNA molecule can be read from the film. This technique has been used to determine the sequences of certain complex DNA molecules, including the primate virus SV40 (Fiers, W., *et al.*, *Nature 273*:113-120, 1978; Reddy, V.B., *et al.*, *Science 200*:494-502, 1978) and the bacterial plasmid pBR322 (Sutcliffe, G., *Cold Spring Harbor Symp. Quant. Biol. 43*:77-90, 1979).

An alternative technique for sequencing, named "Sanger sequencing" after its developer (Sanger, F., and Coulson, A.R., *J. Mol. Biol. 94*:444-448, 1975), is more commonly employed. This method uses the DNA-synthesizing activity of DNA polymerases which, when combined with mixtures of reaction-terminating dideoxynucleoside triphosphates (Sanger, F., *et al.*, *Proc. Natl. Acad. Sci. USA 74*:5463-5467, 1977) and a short primer (either of which may be detectably labeled), gives rise to a series of newly synthesized DNA fragments specifically terminated at one of the four dideoxy bases. These fragments are then resolved by gel electrophoresis and the sequence determined as described for Maxam and Gilbert sequencing above. By carrying out four separate reactions (once with each ddNTP), the sequences of even fairly complex DNA molecules may rapidly be determined (Sanger, F., *et al.*, *Nature 265*:678-695, 1977; Barnes, W., *Meth. Enzymol. 152*:538-556, 1987). While Sanger sequencing usually employs *E. coli* or T7 DNA polymerase (U.S. Patent No. 4,795,699), recent modifications of this technique using T7 polymerase mutants allow sequencing to be accomplished using a single sequencing reaction containing all four chain-terminating ddNTPs at different concentrations (U.S. Patent Nos. 4,962,020 and 5,173,411). Further modifications to the technique, to reduce or eliminate the buildup of reaction-poisoning pyrophosphate in the reaction mixtures, have also been described (U.S. Patent No. 5,498,523).

The Polymerase Chain Reaction

Soon after their identification and characterization, it was recognized that the activities of the various enzymes and cofactors involved in DNA synthesis could be exploited *in vitro* to dramatically increase the concentration of, or "amplify," one or more selected DNA sequences. For many medical, diagnostic and forensic applications, amplification of a particular DNA sequence is essential to allow its detection in, or isolation from, a sample in which it is present in very low amounts. More recently, *in vitro* amplification of specific genes has provided powerful and less costly means to facilitate the production of therapeutic proteins by molecular biological techniques, and may have applications in genetic therapy as well.

While a variety of nucleic acid amplification processes has been described, the most commonly employed is the Polymerase Chain Reaction (PCR) technique disclosed in U.S. Patent Nos. 4,683,195 and 4,683,202. In this process, a sample containing the nucleic acid sequence to be amplified (the "target sequence") is first heated to denature or separate the two strands of the nucleic acid. The sample is then cooled and mixed with specific oligonucleotide primers which hybridize to the target sequence. Following this hybridization, DNA polymerase in a buffered aqueous solution

is added to the sample, along with a mixture of the dNTPs that are linked by the polymerase to the replicating nucleic acid strand. After allowing polymerization to proceed to completion, the products are again heat-denatured, subjected to another round of primer hybridization and polymerase replication, and this process repeated any number of times. Since each nucleic acid product of a given cycle of this process serves as a template for production of two new nucleic acid molecules (one from each parent strand), the PCR process results in an exponential increase in the concentration of the target sequence. Thus, in a well-controlled, high-fidelity PCR process, as few as 20 cycles can result in an over one million-fold amplification of the target nucleic acid sequence (*See* U.S. Patent Nos. 4,683,195 and 4,683,202).

Thermostable DNA Polymerases

Overview

Initially, the DNA polymerases of choice for use in DNA sequencing or PCR were *E. coli* DNA polymerase I, the Klenow fragment, or T4 or T7 polymerases owing to their ease of isolation and well-characterized activities. However, the use of these enzymes necessitated their addition prior to the start of each sequencing or PCR cycle, due to their thermolability at the temperatures used to denature the DNA strands in the initial steps of the processes (typically 70° to 95°C) (Saiki, R.K., *et al.*, *Science 230*:1350-1354, 1985; Mullis, K.B., and Faloona, F.A., *Meth. Enzymol. 155*:335-350, 1987; U.S. Patent No. 4,795,699). This need for the addition of fresh enzyme at the beginning of each cycle increased the amount of time required for these processes, and increased the risk of operator error and contamination of the samples during reagent introduction, often leading to undesirable results.

These difficulties were partially overcome by the use of thermostable DNA polymerases in the PCR process (Saiki, R.K., *et al.*, *Science 239*:487-491, 1988). The thermostable DNA polymerase most commonly used in PCR is *Taq* polymerase, isolated from the thermophilic bacterium *Thermus aquaticus* (Saiki *et al.*, 1988, *Id.*; U.S. Patent Nos. 4,889,818 and 4,965,188). *Taq* polymerase functions optimally at temperatures of 70-80°C, and is able to maintain substantial activity upon repeated exposure to temperatures of 92°-95°C as are often used in the initial steps of PCR (Gelfand, D.H., and White, T.J., *in*: *PCR Protocols: A Guide to Methods and Applications*, Innis, M.A., *et al.*, eds., Academic Press, pp. 129-141, 1989; Bej, A.K., and Mahbubani, M.H., *in*: *PCR Technology: Current Innovations*, Griffin, H.G., and Griffin, A.M., eds., CRC Press, pp. 219-237, 1994).

The use of *Taq* polymerase in PCR eliminated the need to add fresh enzyme to the reaction mix prior to each PCR cycle. Instead, a quantity of *Taq* polymerase sufficient to catalyze DNA polymerization over the desired number of cycles can be mixed with the other components prior to the initiation of the first PCR cycle, and the enzyme continues to function throughout the repetitive cycles of increased and decreased temperatures. The use of *Taq* polymerase has also facilitated the automation of the PCR process (Gelfand and White, *Id.*), thereby at once dramatically reducing time constraints and the risks of operator error and sample contamination that are problematic with thermolabile polymerases. Currently, most PCR amplification of nucleic acids for industrial and academic applications is performed using *Taq* polymerase and automated thermal cycling instrumentation.

In addition to *Taq* polymerase, other thermostable polymerases have found similar application in PCR (Bej and Mahbubani, *Id.*). Particularly useful as substitutes for *Taq* polymerase in PCR are polymerases isolated from the thermophilic bacteria *Thermus thermophilus* (*Tth* polymerase), *Thermococcus litoralis* (*Tli* or VENT™ polymerase), *Pyrococcus furiosus* (*Pfu* or DEEPVENT polymerase), *Pyrococcus woosii* (*Pwo* polymerase) and other *Pyrococcus* species, *Bacillus sterothermophilus* (*Bst* polymerase), *Sulfolobus acidocaldarius* (*Sac* polymerase), *Thermoplasma acidophilum* (*Tac* polymerase), *Thermus flavus* (*Tfl/Tub* polymerase), *Thermus ruber* (*Tru* polymerase), *Thermus brockianus* (DYNAZYME™ polymerase), *Thermotoga neapolitana* (*Tne* polymerase; *See* WO 96/10640), *Thermotoga maritima* (*Tma* polymerase; *See* U. S. Patent No. 5,374,553) and other species of the *Thermotoga* genus (*Tsp* polymerase) and *Methanobacterium thermoautotrophicum* (*Mth* polymerase). While each of these polymerases is useful for particular applications (*See* Bej and Mahbubani, *Id.*, p. 222), *Taq* polymerase is still by far the most commonly used polymerase in PCR.

Thermostable polymerases have also found application in DNA sequencing techniques, particularly in automated methods of dideoxy sequencing such as "cycle sequencing." These approaches resemble PCR in most respects except that, in place of dNTPs, automated DNA sequencing uses ddNTPs which allow determination of the sequence of the template DNA as described above. Use of higher denaturation temperatures in automated sequencing also improves sequencing efficiency (*i.e.*, fewer misincorporations occur) and allows the sequencing of templates that are GC-rich or contain significant secondary structure (such as supercoiling).

The use of thermolabile DNA polymerases such as *E. coli* or T7 DNA polymerases in these approaches, however, is subject to the same limitations described above for their use in PCR.

Accordingly, automated methods of DNA sequencing utilizing higher temperatures have increasingly employed thermostable DNA polymerases, the most commonly used of which is, as for PCR, *Taq* polymerase.

Technical Limitations

The use of *Taq* and other thermostable polymerases in sequencing and PCR is not, however, without drawback. For example, the error rate for *Taq* polymerase is substantially higher (*i.e.*, the final product is of "lower fidelity") than that for most of the thermolabile DNA polymerases, including the Klenow fragment of *E. coli* DNA polymerase I (Tindall and Kunkel, *Id.*), averaging about 10^{-4} misincorporations per base pair per cycle. In addition, *Taq* polymerase is only useful for amplifying relatively short stretches of DNA (maximum length on the order of 5-6 kilobases; Barnes, W.M., *Proc. Natl. Acad. Sci. USA 91*:2216-2220, 1994), thus precluding its use in PCR amplification of large genes and whole genomes as is necessary in many current applications.

These technical limitations are apparently related: it has been theorized that the *Taq* polymerase PCR length limitation is due to the low efficiency of elongation of the newly synthesized DNA strands at the sites of incorporation of mismatched bases in the parent strands (Barnes, *Id.*). Further contributing to this difficulty is the absence in *Taq* polymerase of a 3' to 5' exonuclease activity, which in other polymerases acts in a "proofreading" capacity to correct these mismatches and reduce the error rate (Bej and Mahbubani, *Id.*). The 5' to 3' exonuclease activity present in most thermostable DNA polymerases can also degrade the 5' ends of the oligonucleotide primers (also a complication with the 3' to 5' exonuclease activity which can degrade the 3' ends of the primers), yielding undesirable results due to an early termination of the PCR process (*See* WO 92/06200; Barnes, *Id.*).

In sequencing reactions, *Taq* polymerase is subject to a limitation shared by *E. coli* polymerase I and the Klenow fragment. These enzymes each are "discriminatory," meaning that they preferentially incorporate dNTPs over ddNTPs into newly synthesized DNA. Thus, to use *Taq* polymerase in automated sequencing reactions, relatively high concentrations of ddNTPs must be maintained in the reaction mixtures, to kinetically favor ddNTP incorporation by the enzyme. This need for high levels of ddNTPs can be prohibitively expensive, particularly when large DNA fragments are being sequenced.

As another technical limitation, the DNA polymerases have heretofore been maintained in highly concentrated stock solutions in storage buffers containing glycerol, bovine serum albumin

and/or other stabilizing agents and stored at -20°C or lower (*See, e.g.*, WO 92/06188; U.S. Patent No. 5,436,149). The conventional understanding in the field has been that the enzymes would rapidly lose activity at more dilute working concentrations (as do many bioactive proteins) and in solutions without glycerol or other stabilizing agents. Moreover, the solutions of enzymes had to be mixed with dNTPs or ddNTPs, cofactors (such as Mg^{++}) and one or more detergents immediately prior to use in the sequencing or PCR processes, as it was believed that premixture and storage of these solutions would also deleteriously affect their stability. In addition, dNTPs have traditionally been stored at temperatures below -20°C and have also been thought to be unstable if stored otherwise (Maniatis, T., *et al.*, *Molecular Cloning, A Laboratory Manual*, CRC Press, 1992). Together, these limitations have made the use of compositions containing thermostable DNA polymerases in DNA sequencing and PCR more costly and time-consuming than would be desired.

Overcoming Technical Limitations

Several approaches have been undertaken to attempt to surmount these technical difficulties. The results of DNA sequencing methodologies, for example, have been improved by the use of mutant *Taq* enzymes such as Δ*Taq* (WO 92/06188) lacking the 5' to 3' exonuclease activity. Improved sequencing results have also been obtained by including in the reaction mixture an agent such as pyrophosphatase which breaks down the pyrophosphate that can be formed during dideoxy sequencing reactions (*See* U.S. Patent Number 5,498,523). To overcome the discrimination between ddNTPs and dNTPs, some investigators have used T7 DNA polymerase in sequencing, as this enzyme is "nondiscriminatory," meaning that it incorporates ddNTPs at approximately the same rate as dNTPs (*See* U.S. Patent No. 4,795,699). Alternatively, mutants of DNA polymerase from a variety of organisms (*e.g., E. coli*) which are nondiscriminatory have also been described; *see* copending U.S. Patent Application No. 08/525,057 of Deb K. Chatterjee, filed September 8, 1995, entitled "Mutant DNA Polymerases and Use Therof," the disclosure of which is expressly incorporated herein by reference. However, as described above, both *E. coli* and T7 DNA polymerases are thermolabile, so their use in automated sequencing requires addition of fresh enzyme at the beginning of each cycle.

More recently, mutations in *Tne* polymerase from *Thermotoga neapolitana* have been described, which overcome these limitations (WO 96/10640). One of these mutations, in which a phenylalanine residue at amino acid position number 730 in the wildtype protein (SEQ ID NO:1) is replaced with a tyrosine residue, results in a mutant *Tne* polymerase (SEQ ID NO:2) which is both

thermostable *and* nondiscriminatory. This mutant *Tne* polymerase thus provides a solution to both the problems of thermolability and ddNTP discrimination found in other enzymes used in automated DNA sequencing. *See* also the co-pending U.S. Patent Application of A. John Hughes and Deb K. Chatterjee, entitled "Cloned DNA Polymerases from *Thermotoga* and Mutants Therof," filed on even day herewith, which is incorporated by reference herein in its entirety.

In PCR applications, the low fidelity of *Taq*-produced PCR products has been alleviated to some extent by the use of *Pfu* DNA polymerase which contains the proofreading 3' to 5' exonuclease activity lacking in *Taq* polymerase (Lundberg, K.S., *et al.*, *Gene 108*:1-6, 1991). Other thermostable DNA polymerases, including *Tli*/VENT™ (Bej and Mahbubani, *Id.*) and DEEPVENT™ (Flaman, J.-M., *et al.*, *Nucl. Acids Res. 22(15)*:3259-3260, 1994) have also been shown to improve the fidelity of PCR products.

As a means of overcoming this length limitation, mutant enzymes lacking the 5' to 3' exonuclease activity have been prepared, including N-terminal deletion mutants of *Taq* polymerase that are analogous to the Klenow fragment of *E. coli* DNA polymerase I. Several of these mutant enzymes, including Klentaq-1 and Klentaq-278 (Barnes, W.M., *Gene 112*:29-35, 1992; U.S. Patent No. 5,436,149), the *Taq* Stoffel fragment (Lawyer, F.C., *et al.*, *PCR Meth. Appl. 2*:275-287, 1993), and mutants of other thermostable DNA polymerases lacking the 5' to 3' exonuclease activity (*e.g.*, those disclosed in WO 92/06200), have been shown to provide increasingly stable PCR products and primers. However, as these enzymes also lack the 3' to 5' proofreading activity, their use subjects the PCR process to the increased error rates described above. Thus, even with these mutant enzymes, high-fidelity PCR amplification of DNA fragments larger than 5-6 kilobases has proven exceedingly difficult.

By combining specific quantities of several enzymes, however, high-fidelity PCR amplification of large DNA sequences has been achieved. For example, use of a combination of a high concentration of a thermostable DNA polymerase lacking the 3' to 5' exonuclease activity (*e.g.*, Klentaq278) and a low concentration of a thermostable DNA polymerase exhibiting the 3' to 5' exonuclease activity (*e.g.*, *Pfu*/DEEPVENT or *Tli*/VENT™) provides for amplification to high concentrations of DNA sequences of at least 35 kilobases in length with significantly improved fidelity (Barnes, *Id.*; U.S. Patent No. 5,436,149). Similar results have been obtained with mixtures of *Tth* polymerase and low levels of thermostable polymerases from *Pyrococcus*, *Tli* or *Tma* (U.S. Patent No. 5,512,462). Apparently, the low level of 3' to 5' exonuclease activity is sufficient to remove any mismatched bases incorporated by the majority polymerase, but is insufficient to

significantly degrade the primers. While this approach has heretofore been applied to simple DNA sequences such as those from bacteriophage λ, it may prove applicable to larger and more complex sequences as well, including those of the genomes of bacteria, yeast, plants and animals.

Despite these successes in overcoming the ddNTP discrimination, length and fidelity limitations, however, compositions comprising the reagents necessary for DNA sequencing or PCR capable of extended storage above freezing at working concentrations, without stabilizing agents, have not heretofore been reported. Thus, the time and economic constraints to the use of solutions of thermostable enzymes in most applications have yet to be overcome.

BRIEF SUMMARY OF THE INVENTION

The present invention overcomes these temporal and economic limitations of previously available reagent compositions used in nucleic acid amplification and sequencing methods. Specifically, the invention is directed to compositions comprising mixtures of reagents at working concentrations suitable for use with or without dilution and maintaining activity upon storage for an extended time, said mixtures consisting essentially of at least one thermostable enzyme and at least one buffer salt. The invention further provides such compositions for use in nucleic acid amplification further comprising at least one deoxynucleoside triphosphate, magnesium salts and at least one nonionic detergent, wherein the thermostable enzyme is at least one thermostable DNA polymerase. In another embodiment, the invention provides such compositions for use in nucleic acid sequencing further comprising at least one deoxynucleoside triphosphate, at least one dideoxynucleoside triphosphate, magnesium salts and at least one nonionic detergent, wherein the thermostable enzyme is at least one thermostable DNA polymerase. More specifically, the invention is directed to such compositions wherein the thermostable DNA polymerase is *Taq*, *Tne*, *Tma*, *Pfu*, *Pwo* or *Tth* DNA polymerase, or mutants thereof, preferably at concentrations of about 0.1-200 units per milliliter, about 0.1-50 units per milliliter, about 0.1-40 units per milliliter, about 0.1-36 units per milliliter, about 0.1-34 units per milliliter, about 0.1-32 units per milliliter, about 0.1-30 units per milliliter, or about 0.1-20 units per milliliter, and most preferably at concentrations of about 20 units per milliliter. The invention is also directed to such compositions that further comprise VENT or DEEPVENT™ DNA polymerase, preferably at concentrations of about 0.0002-200 units per milliliter, about 0.002-100 units per milliliter, about 0.002-20 units per milliliter, about 0.002-2.0 units per milliliter, about 0.002-1.6 units per milliliter, about 0.002-0.8 units per milliliter, about

0.002-0.4 units per milliliter, or about 0.002-0.2 units per milliliter, most preferably at concentrations of about 0.40 units per milliliter. The invention is further directed to kits for DNA amplification or sequencing, said kits comprising a carrier means, such as a box, carton, tube or the like, having in close confinement therein one or more container means, such as vials, tubes, ampules, bottles or the like, wherein a first container means contains a composition comprising a mixture of reagents at working concentrations suitable for use without dilution and maintaining activity upon storage for extended time, said mixture consisting essentially of at least one thermostable DNA polymerase, buffer salts, magnesium salts and at least one nonionic detergent. The first container means may also contain a mixture of dNTPs (for PCR applications) or ddNTPs (for sequencing applications). Alternatively, the dNTPs or ddNTPs may be included in a second container means also closely confined within the carrier means of the kit.

The invention is also directed to methods of amplifying or sequencing a nucleic acid molecule, comprising contacting the nucleic acid molecule to be amplified or sequenced, which is preferably larger than about 4-8 kilobases in size, more preferably larger than about 5-7 kilobases in size and most preferably larger than about 7 kilobases in size, with the compositions of the invention. The invention also provides nucleic acid molecules amplified by these methods.

The present invention is also directed more generally to compositions containing thermostable proteins or enzymes useful in molecular biology. These compositions comprise mixtures of reagents at working concentrations suitable for use with or without dilution which maintain the enzyme or protein activity upon storage for an extended time. The compositions of this aspect of the invention comprise at least one thermostable enzyme and at least one buffer salt. The thermostable enzymes in these compositions include, but are not limited to, polymerases, restriction enzymes, alkaline phosphatases, reverse transcriptases, ligases, nucleases, pyrophosphatases, DNAses, RNAses, exonucleases, RNAse inhibitors, kinases, topoisomerases, guanylyltransferases and glycosylases (*e.g.*, uracil DNA glycosylase).

The invention is further directed to kits for conducting a procedure associated with the thermostable enzymes or proteins (restriction enzymes, phosphatases, etc.), the kits comprising a container means such as a box, carton, tube and the like, having in close confinement therein one or more container means, such as a vial, tube, ampule, bottle or the like, wherein a first container means contains a composition comprising a mixture of reagents at working concentrations suitable for use with or without dilution. The reagents in the first container means include at least one thermostable enzyme or protein and at least one buffer salt.

The compositions of the invention have unexpectedly been found to maintain enzyme activity for extended periods of time compared to conventional compositions. For example, the present compositions maintain enzyme activity for at least four weeks when stored at ambient temperature (about 20°-25°C), for at least one year at about 4°C and for at least two years at about -20°C.

These stable, ready-to-use reagent compositions are useful for nucleic acid amplification (including the Polymerase Chain Reaction or PCR) and sequencing (including dideoxy sequencing), or for any procedure using thermostable DNA polymerases or other enzymes (restriction enzymes, phosphatases, kinases, etc.) in fields such as medical therapeutics and diagnostics, forensics and agricultural science.

Other features, advantages and applications of the present invention will be apparent to those skilled in the art from the following description of the preferred embodiments thereof, and from the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a photograph of an agarose gel (visualized by ethidium bromide fluorescence under ultraviolet illumination) of PCR amplification of a 1.3 kilobase human genomic DNA fragment from 10 nanograms of template using 15 different reagent compositions (corresponding to samples 10-24 in Table 1) stored as indicated in Table 1, and a freshly made control sample. Lane marked "M" contains markers indicating amount of DNA loaded; bands correspond to (from top to bottom) 100 nanograms, 60 nanograms, 40 nanograms and 20 nanograms of DNA mass markers.

Figure 2 is a photograph of an agarose gel (visualized by ethidium bromide fluorescence under ultraviolet illumination) of PCR amplification of a 4.1 kilobase human genomic DNA fragment using the amounts indicated (in nanograms) of template for each amplification reaction. The samples on the upper portion of the gel were amplified with a nucleic acid amplification composition stored for 10 weeks at 4°C, while those on the lower portion of the gel were amplified with a composition stored for 10 weeks at ambient temperature (about 20-25°C). Leftmost lanes in each portion are markers as indicated in Figure 1.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

Throughout this disclosure, various terms that are generally understood by those of routine skill in the art are used. Certain terms as used herein, however, have specific meanings for the purposes of the present invention. The term "dNTP" (plural "dNTPs") generically refers to the deoxynucleoside triphosphates (*e.g.*, dATP, dCTP, dGTP, dTTP, dUTP, dITP, 7-deaza-dGTP, αdATP, αdTTP, αdGTP and αdCTP), and the term "ddNTP" (plural "ddNTPs") to their dideoxy counterparts, that are incorporated by polymerase enzymes into newly synthesized nucleic acids. The term "unit" as used herein refers to the activity of an enzyme. When referring to a thermostable DNA polymerase, one unit of activity is the amount of enzyme that will incorporate 10 nanomoles of dNTPs into acid-insoluble material (*i.e.*, DNA or RNA) in 30 minutes under standard primed DNA synthesis conditions. "Working concentration" is used herein to mean the concentration of a reagent that is at or near the optimal concentration used in a solution to perform a particular function (such as amplification, sequencing or digestion of nucleic acids). The term "detergent" as used herein refers to a nonionic surfactant such as TRITON X-100®, Nonidet P-40 (NP-40), Tween 20 or Brij 35. The terms "stable" and "stability" as used herein generally mean the retention by an enzyme of at least 70%, preferably at least 80%, and most preferably at least 90%, of the original enzymatic activity (in units) after the enzyme or composition containing the enzyme has been stored for at least four weeks at a temperature of about 20-25°C, at least one year at a temperature of about 4°C or at least 2 years at a temperature of -20°C.

Overview

The present invention provides, in a first preferred embodiment, compositions comprising mixtures of at least one thermostable enzyme (*e.g.*, a thermostable DNA polymerase, restriction enzyme, etc.), at least one buffer salt, and other reagents necessary for carrying out the procedure associated with the enzyme(s) (*e.g.*, deoxynucleoside triphosphates (dNTPs) for amplification of nucleic acids, dNTPs and dideoxynucleoside triphosphates (ddNTPs) for sequencing of nucleic acids, etc.). These compositions contain no stabilizing compounds (*e.g.*, glycerol, serum albumin or gelatin) that have been traditionally included in stock reagent solutions, and exhibit increased stability (measured as maintenance of enzyme activity) even upon storage at temperatures above freezing. Furthermore, the invention provides these reagent compositions in ready-to-use concentrations,

obviating the time-consuming dilution and pre-mixing steps necessary with previously available solutions. Unexpectedly, even at these diluted concentrations the reagent compositions are stable for extended periods of time at temperatures ranging from ambient (about 20-25°C) to about -70°C.

In additional preferred embodiments, the present invention provides these ready-to-use compositions in the form of kits that are suitable for immediate use to carry out the procedure associated with the enzyme(s) (*e.g.*, nucleic acid amplification or sequencing in the case of DNA polymerases). These kits are also stable for extended periods of time at temperatures ranging from ambient (about 20-25°C) to -70°C.

Sources of Reagents

The compositions of the present invention may be formed by mixing the component reagents at the concentrations described below. The components for making the ready-to-use compositions can be obtained from, for example, GIBCO/LTI (Gaithersburg, Maryland).

Thermostable Enzymes

The thermostable enzymes (*e.g.*, DNA polymerases, restriction enzymes, phosphatases, etc.) used in the present invention may be isolated from natural or recombinant sources, by techniques that are well-known in the art (*See* Bej and Mahbubani, *Id.*; WO 92/06200; WO 96/10640), from a variety of thermophilic bacteria that are available commercially (for example, from American Type Culture Collection, Rockville, Maryland) or may be obtained by recombinant DNA techniques (WO 96/10640). Suitable for use as sources of thermostable enzymes or the genes thereof for expression in recombinant systems are the thermophilic bacteria *Thermus thermophilus*, *Thermococcus litoralis*, *Pyrococcus furiosus*, *Pyrococcus woosii* and other species of the *Pyrococcus* genus, *Bacillus sterothermophilus*, *Sulfolobus acidocaldarius*, *Thermoplasma acidophilum*, *Thermus flavus*, *Thermus ruber*, *Thermus brockianus*, *Thermotoga neapolitana*, *Thermotoga maritima* and other species of the *Thermotoga* genus, and *Methanobacterium thermoautotrophicum*, and mutants thereof. It is to be understood, however, that thermostable enzymes from other organisms may also be used in the present invention without departing from the scope or preferred embodiments thereof. As an alternative to isolation, thermostable enzymes (*e.g.*, DNA polymerases) are available commercially from, for example, GIBCO/LTI (Gaithersburg, Maryland), New England Biolabs (Beverly, Massachusetts), Finnzymes Oy (Espoo, Finland) and Perkin Elmer Cetus (Norwalk,

Connecticut). Once obtained, the purified enzymes may be placed into solution at working concentrations and stored according to the methods of the present invention.

dNTPs

The dNTP components of the present compositions serve as the "building blocks" for newly synthesized nucleic acids, being incorporated therein by the action of the polymerases. These dNTPs -- deoxyadenosine triphosphate (dATP), deoxycytosine triphosphate (dCTP), deoxyguanosine triphosphate (dGTP), deoxythymidine triphosphate (dTTP), and for some applications deoxyuridine triphosphate (dUTP) and deoxyinosine triphosphate (dITP), α-thio-dATP and 7-deaza-dGTP -- are available commercially from sources including GIBCO/LTI (Gaithersburg, Maryland), New England Biolabs (Beverly, Massachusetts) and Sigma Chemical Company (Saint Louis, Missouri). The dNTPs may be unlabeled, or they may be detectably labeled by coupling them by methods known in the art with radioisotopes (*e.g.*, ^{3}H, ^{14}C, ^{32}P or ^{35}S), vitamins (*e.g.*, biotin), fluorescent moieties (*e.g.*, fluorescein, rhodamine, Texas Red, or phycoerythrin) or other detection agents. Labeled dNTPs may also be obtained commercially, for example from GIBCO/LTI (Gaithersburg, Maryland) or Sigma Chemical Company (Saint Louis, Missouri). Once obtained, the dNTPs may be placed into solution at working concentrations and stored according to the methods of the present invention.

ddNTPs

The ddNTP components of the present compositions serve as the "terminating agents" in the dideoxy nucleic acid sequencing methodologies, being incorporated into newly synthesized nucleic acids by the action of the polymerases. These ddNTPs -- dideoxyadenosine triphosphate (ddATP), dideoxycytosine triphosphate (ddCTP), dideoxyguanosine triphosphate (ddGTP), dideoxythymidine triphosphate (ddTTP), and for some applications dideoxyuridine triphosphate (ddUTP) and dideoxyinosine triphosphate (ddITP) -- are available commercially from sources including GIBCO/LTI (Gaithersburg, Maryland), New England Biolabs (Beverly, Massachusetts) and Sigma Chemical Company (Saint Louis, Missouri). The ddNTPs may be unlabeled, or they may be detectably labeled by coupling them by methods known in the art with radioisotopes (*e.g.*, ^{3}H, ^{14}C, ^{32}P or ^{35}S), vitamins (*e.g.*, biotin), fluorescent moieties (*e.g.*, fluorescein, rhodamine, Texas Red, or phycoerythrin) or other detection agents. Labeled ddNTPs may also be obtained commercially, for example from GIBCO/LTI (Gaithersburg, Maryland) or Sigma Chemical Company (Saint Louis,

Missouri). Once obtained, the ddNTPs may be placed into solution at working concentrations and stored according to the methods of the present invention.

Buffers/Salts

All buffers and cofactor salts comprising the compositions of the present invention, and concentrated stock solutions thereof, are available from a variety of commercial sources including GIBCO/LTI (Gaithersburg, Maryland) and Sigma Chemical Company (Saint Louis, Missouri). Particularly preferred buffers for use in forming the present compositions are the sulfate, hydrochloride, phosphate or free acid forms of tris-(hydroxymethyl)aminomethane (TRIS®), although alternative buffers of the same approximate ionic strength and pKa as TRIS® may be used with equivalent results. In addition to the buffer salts, cofactor salts such as those of potassium (preferably potassium chloride) and magnesium (preferably magnesium chloride or sulfate) are included in the compositions. Once obtained, the buffers and cofactor salts may be placed into solution at working concentrations and stored according to the methods of the present invention.

Detergents

At least one detergent may be included as a component of the present compositions, to provide for both increased stability and activity of the component enzymes. Nonionic detergents are preferred, to maintain a balanced ionic strength and prevent chelation of cofactors and aggregation or inactivation of proteins. Particularly preferred as detergents are TRITON X-100® , Brij 35, Tween 20 and Nonidet P-40 (NP-40), although other nonionic surfactants and mixtures thereof may also be used in the present compositions. These detergents are available commercially from sources such as Sigma Chemical Company (Saint Louis, Missouri), usually as concentrated aqueous solutions or in powder form. Once obtained, the detergents may be placed into solution at working concentrations and stored according to the methods of the present invention.

Formulating the Reagent Compositions

Once the reagent components have been obtained, they are mixed at working concentrations to form a solution suitable for immediate use with or without dilution or addition of further reagents. The water used in the formulations of the present invention is preferably distilled, deionized and sterile filtered (through a 0.1-0.2 micrometer filter), and is free of contamination by DNase and RNase enzymes. Such water is available commercially, for example from Sigma Chemical Company

(Saint Louis, Missouri), or may be made as needed according to methods well known to those skilled in the art.

Although the components of the present compositions may be admixed in any sequence, it is often preferable to first dissolve the buffer(s) and cofactor salts in water and to adjust the pH of the solution prior to addition of the remaining components. In this way, the pH-sensitive components (particularly the enzymes, ddNTPs and dNTPs) will be less subject to acid- or alkaline-hydrolysis during formulation.

To formulate the buffered salts solution, a buffer salt which is preferably a salt of tris(hydroxymethyl)aminomethane (TRIS®), and most preferably the hydrochloride salt thereof, is combined with a sufficient quantity of water to yield a solution having a TRIS® concentration of 5-150 millimolar, preferably 10-60 millimolar, and most preferably about 20-60 millimolar. To this solution, a salt of magnesium (preferably either the chloride or sulfate salt thereof) may be added to provide a working concentration thereof of 1-10 millimolar, preferably 1.5-5 millimolar, and most preferably about 1.5-2 millimolar. A salt of potassium (most preferably potassium chloride) may also be added to the solution, at a working concentration of 10-100 millimolar and most preferably about 50 millimolar. An ammonium salt, for example ammonium sulfate, may also be added to the mixture, at a working concentration of 2-50 millimolar, preferably 10-30 millimolar and most preferably 18 millimolar. Combinations of ammonium sulfate and potassium chloride (or other salts) may also be used in formulating the compositions of the present invention. A small amount of a salt of ethylenediaminetetraacetate (EDTA) may also be added (preferably about 0.1 millimolar), although inclusion of EDTA does not appear to be essential to the function or stability of the compositions of the present invention. After addition of all buffers and salts, this buffered salt solution is mixed well until all salts are dissolved, and the pH is adjusted using methods known in the art to a pH value of 7.4 to 9.2, preferably 8.0 to 9.0, and most preferably about 8.3 for compositions to be used in amplification or sequencing of nucleotide fragments up to about 5-6 kilobases in size (hereinafter referred to as "standard compositons"), and about 8.9 for compositions to be used for amplification or sequencing of nucleotide fragments larger than about 5-6 kilobases in size (hereinafter referred to as "large sequence compositions").

To the buffered salt solution, the remaining components of the present composition are added. It is well known in the field that the addition of one or more detergents to an aqueous buffer will aid in the subsequent solubilization of added proteins. Accordingly, at least one nonionic detergent such as TRITON X-100® (preferably at a working concentration of 0.1-1%), Brij 35

(preferably at a concentration of 0.01-1% and most preferably of about 0.1%) or Nonidet P-40 (NP-40, preferably as an admixture with a concentration of 0.004-1%, and most preferably in admixture with Tween 20 at a working concentration of 0.1% for standard compositions and 0.02% for large sequence compositions) may be added to the buffer solution. This detergent is preferably added prior to the introduction of the remaining components into the solution, although the detergent may equivalently be added at any step of formulation. Following formulation, the buffered salt solutions may be filtered through a low protein-binding filter unit that is available commercially (for example from Millipore Corporation, Bedford, Massachusetts) and stored until use.

The remaining components are then added to the solution to formulate the compositions of the present invention. At least one thermostable enzyme (*e.g.*, DNA polymerase) is added and the solution is gently mixed (to minimize protein denaturation). For standard DNA amplification (including via PCR) or sequencing of DNA segments up to about 5-6 kilobases in length, any thermostable DNA polymerase (hereinafter the "primary polymerase") may be used in the standard compositions, although *Taq*, *Tne*, *Tma*, VENT™, DEEPVENT™, *Pfu* or *Pwo* polymerases are preferable at a working concentration in the solution of about 0.1-200 units per milliliter, about 0.1-50 units per milliliter, about 0.1-40 units per milliliter, about 0.1-36 units per milliliter, about 0.1-34 units per milliliter, about 0.1-32 units per milliliter, about 0.1-30 units per milliliter, or about 0.1-20 units per milliliter, and most preferably at a working concentration of about 20 units per milliliter. For amplification of DNA segments larger than 5-6 kilobases in length, large sequence compositions should be formulated by adding to the standard compositions a low concentration of one or more additional thermostable DNA polymerases (hereinafter the "secondary polymerase") containing a 3'-5' exonuclease activity. Particularly suited for this application are VENT™, *Pfu*, *Pwo* or *Tne*, and most preferably DEEPVENT™, DNA polymerases. The additional polymerase(s) should be added to the solution in sufficient quantity to give a final working concentration of about 0.0002-200 units per milliliter, about 0.002-100 units per milliliter, about 0.002-20 units per milliliter, about 0.002-2.0 units per milliliter, about 0.002-1.6 units per milliliter, about 0.002-0.8 units per milliliter, about 0.002-0.4 units per milliliter, or about 0.002-0.2 units per milliliter, most preferably at concentrations of about 0.40 units per milliliter.

It has heretofore been thought that the activity ratios of the primary to secondary polymerases should be maintaied at 4:1 - 2000:1 for large sequence amplification (*see* U.S. Patent No. 5,436,149). It has now been discovered, however, that in the compositions of the present invention that activity

ratios of the primary to secondary polymerases of 1:1, 1:2, 1:4, 1:5, 1:8, 1:10, 1:25, 1:50, 1:100, 1:250, 1:500, 1:1000 and 1:2000 may be suitable for amplification of large nucleotide sequences.

For nucleic acid sequencing, the reagent compositions may be used as formulated above. For nucleic acid sequencing by the dideoxy method (*See* U.S. Patent Nos. 4,962,020, 5,173,411 and 5,498,523), however, preferably the mutant *Tne* DNA polymerase shown in SEQ ID NO:2 is added to the reagent compositions. *Tne* polymerase is added to the solution to give a working concentration of about 0.1-10,000 units per milliliter, about 0.1-5000 units per milliliter, about 0.1-2500 units per milliliter, about 0.1-2000 units per milliliter, about 0.1-1500 units per milliliter, about 0.1-1000 units per milliliter, about 0.1-500 units per milliliter, about 0.1-300 units per milliliter, about 0.1-200 units per milliliter, about 0.1-100 units per milliliter, or about 0.1-50 units per milliliter, and most preferably of about 300 units per milliliter.

For dideoxy sequencing, a solution of each ddNTP is also prepared. The base of each solution contains dATP, dCTP, dTTP, 7-deaza-GTP and/or other dNTPs, each at a working concentration of about 10-1000 micromolar, about 10-500 micromolar, about 10-250 micromolar, or about 10-100 micromolar, most preferably at a concentration of about 100 micromolar, in a solution of buffer and chelating salts, for example TRIS®-HCl most preferably at a working concentration of about 10 millimolar (pH about 7.5) and disodium-EDTA most preferably at a concentration of about 0.1 millimolar. To this base, one of the ddNTPs is added to make each of four solutions. Preferably, the sodium or lithium salt of ddATP, ddCTP, ddGTP or ddTTP is added to the solution to give a working concentration of the ddNTP of about 0.5-10 micromolar, about 0.5-8 micromolar, about 0.5-5 micromolar, about 0.5-3 micromolar, about 0.5-2.5 micromolar, or about 0.5-2 micromolar, and most preferably about 2 micromolar. For cycle sequencing applications, the pH of the ddNTP solutions will preferably be about 9.0, and the concentrations of ddNTPs may be lower, preferably about 0.05 to 1.0 micromolar or about 0.05 to 0.8 micromolar, and most preferably about 0.08 to 0.8 micromolar. For some applications, it may be desirable to also incorporate or substitute ddITP, ddUTP, and/or α-thio-dATP into the compositions at approximately the same working concentrations. Thus, four solutions are prepared, each containing one of the four ddNTPs, which are combined with the polymerase compositions of the present invention to carry out the four separate reactions used in dideoxy sequencing. Alternatively, for single-solution sequencing as disclosed in U.S. Patent Nos. 4,962,020 and 5,173,411, the four ddNTPs may be combined into a single solution which is added to the polymerase compositions of the present invention to perform the sequencing reaction.

For nucleic acid amplification, including PCR, dNTP salts are added to the reagent compositions. Preferably, the sodium or lithium salts of dATP, dCTP, dGTP and dTTP are added to the solution to give a working concentration of each dNTP of 10-1000 micromolar, preferably 200-300 micromolar, and most preferably about 200 micromolar. For some applications, it may be desirable to also incorporate or substitute dITP or dUTP into the compositions at the same working concentrations.

To reduce component denaturation, storage of the reagent compositions is preferably in conditions of diminished light, *e.g.*, in amber or otherwise opaque containers or in storage areas with controlled low lighting. The ready-to-use reagent compositions of the present invention are unexpectedly stable at ambient temperature (about 20°-25°C) for about 4-10 weeks, are stable for at least one year upon storage at 4°C, and for at least two years upon storage at -20°C. Surprisingly, storage of the compositions at temperatures below freezing (*e.g.*, -20°C to -70°C), as is traditional with stock solutions of bioactive components, is not necessary to maintain the stability of the compositions of the present invention.

In other preferred embodiments, the compositions of the present invention may be assembled into kits for use in nucleic acid amplification or sequencing. Sequencing kits according to the present invention comprise a carrier means, such as a box, carton, tube or the like, having in close confinement therein one or more container means, such as vials, tubes, ampules, bottles and the like, wherein a first container means contains a stable composition comprising a mixture of reagents, at working concentrations, which are at least one thermostable DNA polymerase, at least one buffer salt, at least one deoxynucleoside triphosphate and at least one dideoxynucleoside triphosphate. The sequencing kits may further comprise additional reagents and compounds necessary for carrying out standard nucleic sequencing protocols, such as pyrophosphatase, agarose or polyacrylamide media for formulating sequencing gels, and other components necessary for detection of sequenced nucleic acids (*See* U.S. Patent Nos. 4,962,020 and 5,498,523, which are directed to methods of DNA sequencing).

Similarly, amplification kits according to the present invention comprise carrier means, such as a box, carton, tube or the like, having in close confinement therein one or more container means, such as vials, tubes, ampules, bottles and the like, wherein a first container means contains a stable composition comprising a mixture of reagents, at working concentrations, which are at least one thermostable DNA polymerase, at least one buffer salt, and at least one deoxynucleoside triphosphate. The amplification kits encompassed by this aspect of the present invention may further

comprise additional reagents and compounds necessary for carrying out standard nucleic amplification protocols (*See* U.S. Patent Nos. 4,683,195 and 4,683,202, which are directed to methods of DNA amplification by PCR).

Use of the Reagent Compositions

The compositions and kits embodied in the present invention will have general utility in any application utilizing nucleic acid sequencing or amplification methodologies. Amplification methods in which the present compositions may be used include PCR (U.S. Patent Nos. 4,683,195 and 4,683,202), Strand Displacement Amplification (SDA; U.S. Patent No. 5,455,166; EP 0 684 315), and Nucleic Acid Sequence-Based Amplification (NASBA; U.S. Patent No. 5,409,818; EP 0 329 822). Nucleic acid sequencing techniques which may employ the present compositions include dideoxy sequencing methods such as those disclosed in U.S. Patent Nos. 4,962,022 and 5,498,523, as well as more complex PCR-based nucleic acid fingerprinting techniques such as Random Amplified Polymorphic DNA (RAPD) analysis (Williams, J.G.K., *et al.*, *Nucl. Acids Res. 18(22)*:6531-6535, 1990), Arbitrarily Primed PCR (AP-PCR; Welsh, J., and McClelland, M., *Nucl. Acids Res. 18(24)*:7213-7218, 1990), DNA Amplification Fingerprinting (DAF; Caetano-Anollés *et al.*, *Bio/Technology 9*:553-557, 1991), microsatellite PCR or Directed Amplification of Minisatellite-region DNA (DAMD; Heath, D.D., *et al.*, *Nucl. Acids Res. 21(24)*: 5782-5785, 1993), and Amplification Fragment Length Polymorphism (AFLP) analysis (EP 0 534 858; Vos, P., *et al.*, *Nucl. Acids Res. 23(21)*:4407-4414, 1995; Lin, J.J., and Kuo, J., *FOCUS 17(2)*:66-70, 1995). In particular, the compositions and kits of the present invention will be useful in the fields of medical therapeutics and diagnostics, forensics, and agricultural and other biological sciences, in any procedure utilizing thermostable DNA polymerases. Furthermore, the methods by which the compositions of the present invention are formulated may be extendable to all thermostable enzymes or mixtures thereof, and may allow the formulation of ready-to-use compositions of a variety of bioactive enzymes or other proteins that demonstrate increased stability upon extended storage at temperatures above freezing.

The compositions and kits of the invention are particularly useful in methods for amplifying and sequencing nucleic acid molecules. Nucleic acid amplification methods according to this aspect of the invention comprise contacting a nucleic acid molecule to be amplified with one or more of the compositions of the invention, thus providing a population of amplified copies of the nucleic acid molecule. Nucleic acid sequencing methods according to this aspect of the invention comprise

contacting the nucleic acid molecule to be sequenced with one or more of the compositions of the invention. According to these methods, amplification and sequencing of the nucleic acid molecule may be accomplished by any of the above-described amplification and sequencing techniques, most preferably by PCR. The present amplification and sequencing methods are particularly useful for amplification and sequencing of large nucleic acid molecules (*e.g.*, by "long PCR"), preferably nucleic acid molecules that are larger than about 4-8 kilobases in size, more preferably larger than about 5-7 kilobases in size, and most preferably nucleic acid molecules that are larger than about 7 kilobases in size.

It will be readily apparent to those of ordinary skill in the relevant arts that other suitable modifications and adaptations to the methods and applications described herein are obvious and may be made without departing from the scope of the invention or any embodiment thereof. Having now described the present invention in detail, the same will be more clearly understood by reference to the following examples, which are included herewith for purposes of illustration only and are not intended to be limiting of the invention.

Example 1: Formulation of Standard Compositions

As an initial step in formulating stable, ready-to-use reagent compositions for nucleic acid amplification and sequencing, components were mixed in varying amounts as shown in Table 1 to provide 24 different formulations. The pH on all formulations was adjusted to about 8.3. After filtration through a low protein-binding filter, all compositions were formulated with 20 units/ml of *Taq* polymerase and were suitable for use as standard compositions for amplification or sequencing of nucleic acid fragments up to about 5-6 kilobases in size.

TABLE 1. FORMULATIONS OF STANDARD COMPOSITIONS

Formulation Number	1	2	3	4	5	6	7	8
Tris-HCl, pH 8.4	20 mM	20 mM	20 mM	20 mM	20 mM	20 mM	20 mM	20 mM
KCl	50 mM	50 mM	50 mM	50 mM	50 mM	50 mM	50 mM	50 mM
$(NH_4)_2SO_4$	0	0	0	0	0	0	0	0
dNTP	200 µM	200 µM	200 µM	200 µM	200 µM Na	200 µM Na	200 µM Li	200 µM Li
$MgCl_2$	1.5 mM	1.5 mM	1.5 mM	1.5 mM	1.5 mM	1.5 mM	1.5 mM	1.5 mM
Detergents	0.004% Tween20/NP40	0.1% Tween20/NP40	0.1% Tween20/NP40	0.1% Tween20/NP40	0.1% Tween20/NP40	0.004% Tween20/NP40	0.1% Tween20/NP40	0.004% Tween20/NP40
Taq, U/ml	20	20	20	20	20	20	20	20

9	10	11	12	13	14	15	16	17
20 mM	20 mM	20 mM	20 mM	20 mM	20 mM	20 mM	20 mM	20 mM
50 mM	50 mM	50 mM	50 mM	50 mM	50 mM	50 mM	50 mM	50 mM
0	0	0	0	0	0	0	0	0
200 µM	200 µM	200 µM	200 µM	200 µM	200 µM	200 µM	200 µM	200 µM
1.5 mM	3 mM	3 mM	3 mM	3 mM	3 mM	3 mM	3 mM	3 mM
0.1% Tween20/NP40	0.1% Tween20/NP40	0.1% Brij35	0.1% TritonX-100	0.01% Tween20/NP40	0.01% Brij35	0.01% TritonX-100	1% Tween20/NP40	1% Brij35
20	20	20	20	20	20	20	20	20

18	19	20	21	22	23	24
20 mM	20 mM	20 mM	20 mM	20 mM	50 mM (pH 9.0)	20 mM (pH 8.8)
50 mM	50 mM	50 mM	50 mM	50 mM	0	10 mM
0	0	0	0	0	20 mM	10 mM
200 µM	200 µM	200 µM	200 µM	200 µM	200 µM	200 µM
3 mM	3 mM	3 mM	3 mM	3 mM	1.5 mM	2 mM
1% TritonX-100	none	0.1% Tween20/NP40	0.1% Tween20/NP40	0.1% Tween20/NP40	0.1% Tween20/NP40	0.1% TritonX100
20	20	20	20	20	20	20

Example 2: Stability of Standard Compositions

To examine the stability of the standard compositions formulated in Example 1, samples of each formulation were aliquoted and stored, under diminshed light, at ambient temperature (about 20°-25°C), 4°C, -20°C and -70°C. Samples of each formulation from each temperature were taken daily for the first week, and weekly thereafter, and used in stability assays. These stability assays were performed by amplifying, via standard PCR, suboptimal amounts of human genomic DNA as a template using a template titration (if few samples were to be compared in a time point) or a fixed amount of template (if a larger number of samples were to be compared). To the desired amount of template DNA in a given formulation, 10 picomoles of primer was added, and the reaction mixtures were subjected to 35 cycles of PCR of 30 seconds at 94°C, 30 seconds at 55°C and 1 minute/kilobase at 72°C. A portion of each reaction was then subjected to agarose gel electrophoresis and visualized by ethidium bromide fluorescence under ultraviolet illumination.

The results of these assays indicated that certain of the formulations demonstrated enhanced stability upon storage. As shown in Figure 1, after about three months of storage at 4°C, formulations 15 and 19 had completely lost enzymatic activity, as evidenced by an absence of bands in the lanes corresponding to these samples. At this same time point, however, formulations 10-14, 16-18 and 20-24 demonstrated about the same levels of enzymatic activity as a freshly made ("control") formulation.

Storage temperature was also found to have a significant effect upon the stability of the formulations, even within a given formulation. As shown in Figure 2, when samples of formulation 4 were examined after about 15 weeks of storage at either 4°C or about 20-25°C, the samples stored at 4°C retained full enzymatic activity when compared to a control sample. Those stored at about 20-25°C, however, had lost some activity, as indicated by the lower yields of the target fragment obtained at all template concentrations.

The results for all of the formulations at the various storage temperatures are summarized in Table 2.

TABLE 2. STABILITY OF STANDARD COMPOSITIONS

Formulation Number[1]	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Storage Temperature, °C								
20-25	< 12 weeks	< 12 weeks	< 15 weeks	< 12 weeks	nd[2]	nd	nd	nd
4	> 26 weeks	> 26 weeks	> 56 weeks	> 27 weeks	> 8 weeks	> 8 weeks	> 8 weeks	> 8 weeks
-20	> 13 weeks	> 13 weeks	nd	> 15 weeks	nd	nd	nd	nd
-70	nd	nd	nd	> 15 weeks	nd	nd	nd	nd

9	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**
nd	< 10 weeks	< 6 weeks	< 6 weeks	< 12 weeks	< 10 weeks	< 7 days	< 12 weeks	< 6 weeks
> 40 weeks	> 26 weeks	> 26 weeks	> 26 weeks	> 26 weeks	> 26 weeks	< 1 day	> 26 weeks	> 26 weeks
> 10 weeks	> 26 weeks	> 26 weeks	> 26 weeks	> 26 weeks	> 26 weeks	< 7 days	> 26 weeks	> 26 weeks
nd	nd	nd	nd	nd	nd	nd	nd	nd

18	**19**	**20**	**21**	**22**	**23**	**24**
< 6 weeks	0 days	< 10 weeks	< 10 weeks	< 12 weeks	< 6 weeks	< 12 weeks
> 26 weeks	0 days	> 26 weeks	> 26 weeks	> 13 weeks	> 26 weeks	> 26 weeks
> 26 weeks	0 days	> 26 weeks	> 26 weeks	> 26 weeks	> 26 weeks	> 26 weeks
nd	nd	nd	nd	nd	nd	nd

[1] Formulation numbers correspond to those in Table 1; [2] nd = not done.

IDT Exhibit 1020.0683

These results indicate that several of the compositions unexpectedly maintained enzymatic activity for 6-12 weeks upon storage at 20-25°C, and for over one year at 4°C. Formulation 19, however, had completely lost activity within 24 hours of formulation. Formulation 15 also exhibited a rapid loss of activity.

For further analysis of the stability of these ready-to-use compositions, several formulations were stored at 20-25°C or at 4°C for up to six months, with samples taken monthly for stability assays performed by a determination of polymerase unit activity. The results of these assays are summarized in Tables 3 and 4.

TABLE 3. STABILITY OF STANDARD COMPOSITIONS (PERCENTAGE OF ENZYME ACTIVITY REMAINING) UPON STORAGE AT 20-25°C

Formulation No.[1]	1 Month	2 Months	3 Months	4 Months
10	89	65	nd[2]	nd
11	106	3	nd	nd
12	106	91	nd	nd
13	93	72	nd	nd
14	91	76	nd	nd
15	78	63	nd	nd
16	94	85	nd	nd
17	89	90	nd	nd
18	90	84	nd	nd
19	0	0	nd	nd
20	88	81	72	77
21	nd	103	nd	nd
22	97	84	nd	nd
23	83	77	nd	nd
24	81	106	nd	nd

[1] Formulation numbers correspond to those in Table 1.

[2] nd = not done

TABLE 4. STABILITY OF STANDARD COMPOSITIONS (PERCENTAGE OF ENZYME ACTIVITY REMAINING) UPON STORAGE AT 4°C

Formulation No.[1]	1 Month	2 Months	4 Months	5 Months	6 Months
10	84	86	98	nd[2]	105
11	94	97	98	nd	nd
12	94	97	106	nd	nd
13	86	93	93	nd	nd
14	85	91	105	nd	nd
15	89	89	98	nd	nd
16	88	103	104	nd	nd
17	83	94	91	nd	nd
18	90	97	99	nd	nd
19	95	59	nd	nd	38
20	97	100	94	94	103
21	93	97	100	nd	nd
22	100	109	35	nd	nd
23	94	97	89	nd	107
24	93	94	97	nd	nd

[1] Formulation numbers correspond to those in Table 1.

[2] nd - not done

Several of the formulations were stable upon storage at 20-25°C, most notable formulation 20 which retained >70% of its enzymatic activity even after storage for four months at this temperature. As described earlier, formulation 19 had completely lost activity within the first month, as determined by the PCR assay. Interestingly, however, as determined by the polymerase unit activity assay, formulation 19 was stable for one month when stored at 4°C but had lost substantial activity by the second month of storage at this temperature (Table 4). Formulation 20, shown previously to be stable upon extended storage at 20-25°C, was stable upon storage at 4°C for at least six months.

Taken together, these results indicate that the compositions of the present invention are readily suitable for use in nucleic acid amplification reactions, and demonstrate extended stability upon storage at 20-25°C or 4°C.

Example 3: Formulation and Stability of Large Sequence Compositions

For use in amplification and sequencing of nucleic acid fragments larger than 5-6 kilobases, it has been suggested as described above that a mixture of *Taq* and VENT™ or DEEPVENT™ polymerases (U.S. Patent No. 5,436,149; Barnes, *Id.*), or of *Tth* and either *Tli*, *Pyrococcus* or *Tma* (U.S. Patent No. 5,512,462), be used. Accordingly, *Taq* and DEEPVENT™ DNA polymerases were formulated, at activity ratios of 100:1 (for samples 1-3) or 50:1 (for sample 4) into solutions containing the buffer salts, cofactors and detergents shown in Table 5. Each of these formulations was adjusted to about pH 9.1, which is optimal for the activity of DEEPVENT™ DNA polymerase (Bej and Mahbubani, *Id.*). Samples were then stored at about 20-25°C or at 4°C and assayed weekly for 12 weeks, and monthly thereafter, in stability assays in which a 20 kilobase target in 100 nanograms of human genomic template was amplified by PCR. Reaction mixtures included 10 picomoles of primer and were subjected to 35 PCR cycles of 30 seconds at 94°C, 30 seconds at 62°C and 1 minute/kilobase at 68°C. Portions of the reaction were subjected to agarose gel electrophoresis and were visualized by ethidium bromide fluorescence under ultraviolet illumination, as shown above for Figures 1 and 2. Results are summarized in Table 5.

TABLE 5. STABILITY OF LARGE SEQUENCE COMPOSITIONS

Formulation No.	Formulation	Stability at:	
		20-25°C	4°C
1	66 mM Tris-SO_4 (pH 9.1) 19.8 mM $(NH_4)_2$ SO_4 2.2 mM $MgSO_4$ 22 units/ml *Taq* DNA Polymerase 0.22 units/ml DEEPVENT DNA Polymerase 0.11% Tween-20 0.011% NP-40	< 12 weeks	11 months
2	66 mM Tris-SO_4 (pH 9.1) 19.8 mM $(NH_4)_2$ SO_4 2.2 mM $MgSO_4$ 24.42 units/ml *Taq* DNA Polymerase 0.242 units/ml DEEPVENT DNA Polymerase 0.066% Tween-20 0.066% NP-40	< 12 weeks	>11 months
3	66 mM Tris-SO_4 (pH 9.1) 19.8 mM $(NH_4)_2$ SO_4 2.2 mM $MgSO_4$ 22 units/ml *Taq* DNA Polymerase 0.22 units/ml DEEPVENT DNA Polymerase 0.01% Tween-20 0.01% NP-40	nd[1]	11 months
4	66 mM Tris-SO_4 (pH 9.1) 19.8 mM $(NH_4)_2$ SO_4 2.2 mM $MgSO_4$ 22 units/ml *Taq* DNA Polymerase 0.44 units/ml DEEPVENT DNA Polymerase 0.01% Tween-20 0.01% NP-40	nd[1]	11 months

[1] nd = not done

Upon storage at ambient temperature (20°-25°C), all of the formulations were stable for 6-12 weeks. Storage of these formulations at 4°C provided enhanced stability of over 11 months. Similar results may be obtained with formulations in which *Taq* and *Tne* DNA polymerases were used in an activity ratio of 1:1, 1:2, 1:4, 1:5, 1:8, 1:10, 1:25, 1:50, 1:100, 1:250, 1:500, 1:1000 or 1:2000. These results indicate that the large sequence compositions of the present invention are readily suitable for use in amplification of nucleic acid sequences larger than 5-6 kilobases and demonstrate extended stability upon storage at 20° to 25°C, or at 4°C.

Example 4: Combinations of *Thermus flavis* (*Tfl*) DNA Polymerase and *Thermotoga neapolitana* (*Tne*) DNA Polymerase

To examine other DNA polymerase compositions for their utility in amplification of nucleic acid molecules, a mixture of *Tfl* and *Tne* DNA polymerases, at a 1:1 ratio, was used to amplify the 2.7 kilobase Puc19 plasmid. Amplification reactions were in a 50 µl final volume in buffer containing 1 mM magnesium acetate. 80 pg of Puc19 linearized by treatment with *Ad*II was used as the template, and was contacted with 1 µl of enzyme mixture. PCR conditions were 1 min at 94°C, followed by 35 cycles of 94°C for 30 seconds/60°C for 30 seconds/68°C for 5 minutes.

Upon analysis of the amplification products by gel electrophoresis, this composition comprising *Tfl* and *Tne* DNA polymerases was found to efficiently amplify the 2.7 kilobase Puc19 plasmid. The efficiency of amplification was comparable to amplification of Puc19 using 1 µl of *Taq* DNA polymerase.

Example 5: Amplification of Genomic DNA Using *Tfl/Tne* Compositions

Having demonstrated that compositions comprising *Tfl* and *Tne* DNA polymerases efficiently amplify plasmid-sized nucleic acid molecules, these compositions were examined for their ability to amplify nucleic acid molecules from genomic DNA templates. Six different primer sets were constructed (ranging in size from 0.25 to 4.1 kilobases) and used to amplify the human β-globin gene from a genomic DNA template from the K562 human leukemia cell line. Each reaction was performed in a volume of 50 µl comprising template at 40 ng/reaction, and *Tfl/Tne* mixture at either 0.5 unit/reaction or 1 unit/reaction (*Tfl* and *Tne* at a 1:1 ratio in both mixtures). PCR conditions

were 1 min at 94°C, followed by 35 cycles of 94°C for 30 seconds/55°C for 30 second/68°C for 5 minutes.

Upon analysis of the amplification products by gel electrophoresis, efficient amplification was observed for all primers. The sizes of the amplification products produced using the different primers were 0.25 kilobase, 0.7 kilobase, 1.1 kilobases, 2.0 kilobases and 4.1 kilobases. These results demonstrate that the *Tfl/Tne* compositions efficiently amplify nucleic acid molecules derived from genomic DNA templates.

Example 6: Use of Various Enzyme Ratios in DNA Polymerase Compositions

To determine the efficacy of different enzyme ratios in nucleic acid amplification, a composition comprising a higher amount of *Tne* (3' exo+) DNA polymerase than *Tfl* (3' exo-) DNA polymerase was made and tested for its ability to amplify Puc19 and the β-globin gene. A 1:3 mixture of *Tfl/Tne* was made and used to amplify Puc19 under the conditions described in Example 4, and β-globin under the conditions described in Example 5.

Upon analysis of the amplification products by gel electrophoresis, efficient amplification was observed for both Puc19 and for β-globin (all template sizes). These results demonstrate that compositions in which a 3' exo+ DNA polymerase (*Tne*) is present in higher quantity than a 3' exo- DNA polymerase (*Tfl*) efficiently amplify nucleic acid molecules derived from plasmid and genomic DNA templates.

Example 7: Amplification of Large Nucleic Acid Molecules Using DNA Polymerase Mixtures

Having demonstrated that compositions comprising mixtures of 3' exo+ and 3' exo- DNA polymerases efficiently amplify plasmid-sized and small genomic nucleic acid molecules, such compositions were examined for their ability to amplify larger nucleic acid molecules. For these studies, mixtures of *Taq* (3' exo+) DNA polymerase and *Tma* (3' exo-) DNA polymerase (ULTma™; LTI, Gaithersburg, Maryland) were prepared. Two sets of mixtures were prepared: one set contained 1 unit of *Taq* and varying amounts of ULTma (0.3 unit, 0.6 unit, 0.8 unit or 1 unit), and the other set contained no *Taq* and only varying amounts of ULTma (0.3 unit, 0.6 unit, 0.8 unit or 1 unit). These compositions were used to amplify 80 mg of human genomic DNA using primers

specific for a 7.5 kilobase region of the human β-globin gene. PCR conditions were 1 min at 94°C, followed by 35 cycles of 94°C for 30 seconds/60°C for 30 seconds/68°C for 5 minutes.

Upon analysis of the amplification products by gel electrophoresis, the 7.5 kilobase β-globin fragment was found to be efficiently amplified by all compositions comprising both *Taq* and *Tma* DNA polymerase. Compositions comprising only *Tma* DNA polymerase, however, were unable to amplify this large fragment. In control experiments, compositions comprising only *Taq* DNA polymerase were also unable to amplify this large fragment. These results demonstrate that compositions comprising a mixture of 3' exo+ and 3' exo- DNA polymerases are useful in efficiently amplifying large nucleic acid molecules, particularly in amplifying nucleic acid molecules larger than about 7 kilobases in size.

Example 8: Amplification of Large Nucleic Acid Molecules Using Various Ratios of *Taq* and *Tma* DNA Polymerases

Having demonstrated that compositions comprising mixtures of 3' exo+ (*Taq*) and 3' exo- (*Tma*) DNA polymerases efficiently amplify large nucleic acid molecules, compositions comprising these enzymes in various ratios were made and tested for their abilities to amplify the 7.5 kilobase β-globin gene fragment from Example 7. Two sets of mixtures were prepared: one set contained 1 unit of ULTma DNA polymerase and different amounts (0.25, 0.5, 0.6, 0.7, 0.8, 0.9 or 1.0 unit) of *Taq* DNA polymerase, and the other set contained 1 unit of *Taq* DNA polymerase and different amounts (1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 2.0, 3.0, 4.0 or 6.0 units) of ULTma DNA polymerase. Amplification templates, primers and cycling conditions were as described for Example 7.

Upon analysis of the amplification products by gel electrophoresis, efficient amplification of the 7.5 kilobase β-globin gene was observed for the first set of mixtures only in those compositions containing 0.5 to 1.0 unit of *Taq* DNA polymerase and 1.0 unit of ULTma. Compositions in this set containing less than 0.5 unit of *Taq* DNA polymerase did not amplify this fragment. However, all enzyme mixtures in the other set, *i.e.*, compositions comprising 1.0 unit of *Taq* DNA polymerase and 1.0 to 6.0 units of ULTma DNA polymerase, demonstrated efficient amplification of the 7.5 kilobase fragment. In separate experiments, a 13.5 kilobase fragment of the β-globin gene was efficiently amplified using mixtures containing a 1:1 or a 1:2 ratio of *Taq* to ULTma. Together, these results indicate that compositions in which *Tma* DNA polymerase is present in equal or higher quantity than

Taq DNA polymerase efficiently amplify large nucleic acid molecules, particularly those that are larger than about 7-13 kilobases in size.

Example 9: Use of *Taq* and *Tne* DNA Polymerase Mixtures for Long PCR

To determine if other DNA polymerases could be used in compositions also comprising *Taq* DNA polymerase in amplification of large nucleic acid molecules, various mixtures of *Taq* and *Tne* DNA polymerases were made. For these experiments, a *Tne* DNA polymerase deletion mutant (5' exo-; 3' exo+) was mixed in amounts ranging from 0.05 to 2.0 units with 1 unit of *Taq* DNA polymerase and used to amplify the 7.5 kilobase β-globin fragment under conditions described for Example 7.

Upon analysis of the amplification products by gel electrophoresis, all of the combinations of *Tne* and *Taq* DNA polymerases were found to efficiently amplify the 7.5 kilobase DNA fragment. These results indicate that compositions comprising combinations of *Taq* and *Tne* DNA polymerases are useful in amplifying large nucleic acid molecules, particularly those larger than about 7 kilobases in size.

Having now fully described the present invention in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious to one of ordinary skill in the art that the same can be performed by modifying or changing the invention within a wide and equivalent range of conditions, formulations and other parameters without affecting the scope of the invention or any specific embodiment thereof, and that such modifications or changes are intended to be encompassed within the scope of the appended claims.

All publications, patents and patent applications mentioned in this specification are indicative of the level of skill of those skilled in the art to which this invention pertains, and are herein incorporated by reference to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: Life Technologies, Inc.
(B) STREET: 9800 Medical Center Drive
(C) CITY: Rockville
(D) STATE: Maryland
(E) COUNTRY: USA
(F) POSTAL CODE (ZIP): 20850

(ii) TITLE OF INVENTION: Stable Compositions for Nucleic Acid Amplification and Sequencing

(iii) NUMBER OF SEQUENCES: 2

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/689,815
(B) FILING DATE: 14-AUG-1996

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/801,720
(B) FILING DATE: 14-FEB-1997

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 893 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
Met Ala Arg Leu Phe Leu Phe Asp Gly Thr Ala Leu Ala Tyr Arg Ala
1               5                  10                  15

Tyr Tyr Ala Leu Asp Arg Ser Leu Ser Thr Ser Thr Gly Ile Pro Thr
            20                  25                  30

Asn Ala Val Tyr Gly Val Ala Arg Met Leu Val Lys Phe Ile Lys Glu
        35                  40                  45
```

His Ile Ile Pro Glu Lys Asp Tyr Ala Ala Val Ala Phe Asp Lys Lys
50 55 60

Ala Ala Thr Phe Arg His Lys Leu Leu Val Ser Asp Lys Ala Gln Arg
65 70 75 80

Pro Lys Thr Pro Ala Leu Leu Val Gln Gln Leu Pro Tyr Ile Lys Arg
85 90 95

Leu Ile Glu Ala Leu Gly Phe Lys Val Leu Glu Leu Glu Gly Tyr Glu
100 105 110

Ala Asp Asp Ile Ile Ala Thr Leu Ala Val Arg Ala Ala Arg Phe Leu
115 120 125

Met Arg Phe Ser Leu Ile Thr Gly Asp Lys Asp Met Leu Gln Leu Val
130 135 140

Asn Glu Lys Ile Lys Val Trp Arg Ile Val Lys Gly Ile Ser Asp Leu
145 150 155 160

Glu Leu Tyr Asp Ser Lys Lys Val Lys Glu Arg Tyr Gly Val Glu Pro
165 170 175

His Gln Ile Pro Asp Leu Leu Ala Leu Thr Gly Asp Asp Ile Asp Asn
180 185 190

Ile Pro Gly Val Thr Gly Ile Gly Glu Lys Thr Ala Val Gln Leu Leu
195 200 205

Gly Lys Tyr Arg Asn Leu Glu Tyr Ile Leu Glu His Ala Arg Glu Leu
210 215 220

Pro Gln Arg Val Arg Lys Ala Leu Leu Arg Asp Arg Glu Val Ala Ile
225 230 235 240

Leu Ser Lys Lys Leu Ala Thr Leu Val Thr Asn Ala Pro Val Glu Val
245 250 255

Asp Trp Glu Glu Met Lys Tyr Arg Gly Tyr Asp Lys Arg Lys Leu Leu
260 265 270

Pro Ile Leu Lys Glu Leu Glu Phe Ala Ser Ile Met Lys Glu Leu Gln
275 280 285

Leu Tyr Glu Glu Ala Glu Pro Thr Gly Tyr Glu Ile Val Lys Asp His
290 295 300

Lys Thr Phe Glu Asp Leu Ile Glu Lys Leu Lys Glu Val Pro Ser Phe
305 310 315 320

Ala Leu Asp Leu Glu Thr Ser Ser Leu Asp Pro Phe Asn Cys Glu Ile
325 330 335

Val Gly Ile Ser Val Ser Phe Lys Pro Lys Thr Ala Tyr Tyr Ile Pro
340 345 350

Leu His His Arg Asn Ala His Asn Leu Asp Glu Thr Leu Val Leu Ser
355 360 365

Lys Leu Lys Glu Ile Leu Glu Asp Pro Ser Ser Lys Ile Val Gly Gln
370 375 380

Asn Leu Lys Tyr Asp Tyr Lys Val Leu Met Val Lys Gly Ile Ser Pro
385 390 395 400

Val Tyr Pro His Phe Asp Thr Met Ile Ala Ala Tyr Leu Leu Glu Pro
405 410 415

Asn Glu Lys Lys Phe Asn Leu Glu Asp Leu Ser Leu Lys Phe Leu Gly
420 425 430

Tyr Lys Met Thr Ser Tyr Gln Glu Leu Met Ser Phe Ser Ser Pro Leu
435 440 445

Phe Gly Phe Ser Phe Ala Asp Val Pro Val Asp Lys Ala Ala Glu Tyr
450 455 460

Ser Cys Glu Asp Ala Asp Ile Thr Tyr Arg Leu Tyr Lys Ile Leu Ser
465 470 475 480

Met Lys Leu His Glu Ala Glu Leu Glu Asn Val Phe Tyr Arg Ile Glu
485 490 495

Met Pro Leu Val Asn Val Leu Ala Arg Met Glu Phe Asn Trp Val Tyr
500 505 510

Val Asp Thr Glu Phe Leu Lys Lys Leu Ser Glu Glu Tyr Gly Lys Lys
515 520 525

Leu Glu Glu Leu Ala Glu Lys Ile Tyr Gln Ile Ala Gly Glu Pro Phe
530 535 540

Asn Ile Asn Ser Pro Lys Gln Val Ser Asn Ile Leu Phe Glu Lys Leu
545 550 555 560

Gly Ile Lys Pro Arg Gly Lys Thr Thr Lys Thr Gly Asp Tyr Ser Thr
565 570 575

Arg Ile Glu Val Leu Glu Glu Ile Ala Asn Glu His Glu Ile Val Pro
580 585 590

Leu Ile Leu Glu Phe Arg Lys Ile Leu Lys Leu Lys Ser Thr Tyr Ile
595 600 605

Asp Thr Leu Pro Lys Leu Val Asn Pro Lys Thr Gly Arg Phe His Ala
610 615 620

Ser Phe His Gln Thr Gly Thr Ala Thr Gly Arg Leu Ser Ser Ser Asp
625 630 635 640

Pro Asn Leu Gln Asn Leu Pro Thr Lys Ser Glu Glu Gly Lys Glu Ile
645 650 655

Arg Lys Ala Ile Val Pro Gln Asp Pro Asp Trp Trp Ile Val Ser Ala
660 665 670

Asp Tyr Ser Gln Ile Glu Leu Arg Ile Leu Ala His Leu Ser Gly Asp
675 680 685

Glu Asn Leu Val Lys Ala Phe Glu Glu Gly Ile Asp Val His Thr Leu
690 695 700

Thr Ala Ser Arg Ile Tyr Asn Val Lys Pro Glu Glu Val Asn Glu Glu
705 710 715 720

Met Arg Arg Val Gly Lys Met Val Asn Phe Ser Ile Ile Tyr Gly Val
725 730 735

Thr Pro Tyr Gly Leu Ser Val Arg Leu Gly Ile Pro Val Lys Glu Ala
740 745 750

Glu Lys Met Ile Ile Ser Tyr Phe Thr Leu Tyr Pro Lys Val Arg Ser
755 760 765

Tyr Ile Gln Gln Val Val Ala Glu Ala Lys Glu Lys Gly Tyr Val Arg
770 775 780

Thr Leu Phe Gly Arg Lys Arg Asp Ile Pro Gln Leu Met Ala Arg Asp
785 790 795 800

Lys Asn Thr Gln Ser Glu Gly Glu Arg Ile Ala Ile Asn Thr Pro Ile
805 810 815

Gln Gly Thr Ala Ala Asp Ile Ile Lys Leu Ala Met Ile Asp Ile Asp
820 825 830

Glu Glu Leu Arg Lys Arg Asn Met Lys Ser Arg Met Ile Ile Gln Val
835 840 845

His Asp Glu Leu Val Phe Glu Val Pro Asp Glu Glu Lys Glu Glu Leu
850 855 860

Val Asp Leu Val Lys Asn Lys Met Thr Asn Val Val Lys Leu Ser Val
865 870 875 880

Pro Leu Glu Val Asp Ile Ser Ile Gly Lys Ser Trp Ser
885 890

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 893 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Ala Arg Leu Phe Leu Phe Asp Gly Thr Ala Leu Ala Tyr Arg Ala
1 5 10 15

Tyr Tyr Ala Leu Asp Arg Ser Leu Ser Thr Ser Thr Gly Ile Pro Thr
20 25 30

Asn Ala Val Tyr Gly Val Ala Arg Met Leu Val Lys Phe Ile Lys Glu
35 40 45

His Ile Ile Pro Glu Lys Asp Tyr Ala Ala Val Ala Phe Asp Lys Lys
50 55 60

Ala Ala Thr Phe Arg His Lys Leu Leu Val Ser Asp Lys Ala Gln Arg
65 70 75 80

Pro Lys Thr Pro Ala Leu Leu Val Gln Gln Leu Pro Tyr Ile Lys Arg
85 90 95

Leu Ile Glu Ala Leu Gly Phe Lys Val Leu Glu Leu Glu Gly Tyr Glu
100 105 110

Ala Asp Asp Ile Ile Ala Thr Leu Ala Val Arg Ala Ala Arg Phe Leu
115 120 125

Met Arg Phe Ser Leu Ile Thr Gly Asp Lys Asp Met Leu Gln Leu Val
130 135 140

Asn Glu Lys Ile Lys Val Trp Arg Ile Val Lys Gly Ile Ser Asp Leu
145 150 155 160

Glu Leu Tyr Asp Ser Lys Lys Val Lys Glu Arg Tyr Gly Val Glu Pro
165 170 175

His Gln Ile Pro Asp Leu Leu Ala Leu Thr Gly Asp Asp Ile Asp Asn
180 185 190

Ile Pro Gly Val Thr Gly Ile Gly Glu Lys Thr Ala Val Gln Leu Leu
195 200 205

Gly Lys Tyr Arg Asn Leu Glu Tyr Ile Leu Glu His Ala Arg Glu Leu
210 215 220

Pro Gln Arg Val Arg Lys Ala Leu Leu Arg Asp Arg Glu Val Ala Ile
225 230 235 240

Leu Ser Lys Lys Leu Ala Thr Leu Val Thr Asn Ala Pro Val Glu Val
245 250 255

Asp Trp Glu Glu Met Lys Tyr Arg Gly Tyr Asp Lys Arg Lys Leu Leu
260 265 270

Pro Ile Leu Lys Glu Leu Glu Phe Ala Ser Ile Met Lys Glu Leu Gln
275 280 285

Leu Tyr Glu Glu Ala Glu Pro Thr Gly Tyr Glu Ile Val Lys Asp His
290 295 300

Lys Thr Phe Glu Asp Leu Ile Glu Lys Leu Lys Glu Val Pro Ser Phe
305 310 315 320

Ala Leu Asp Leu Glu Thr Ser Ser Leu Asp Pro Phe Asn Cys Glu Ile
325 330 335

Val Gly Ile Ser Val Ser Phe Lys Pro Lys Thr Ala Tyr Tyr Ile Pro
340 345 350

Leu His His Arg Asn Ala His Asn Leu Asp Glu Thr Leu Val Leu Ser
355 360 365

Lys Leu Lys Glu Ile Leu Glu Asp Pro Ser Ser Lys Ile Val Gly Gln
370 375 380

Asn Leu Lys Tyr Asp Tyr Lys Val Leu Met Val Lys Gly Ile Ser Pro
385 390 395 400

Val Tyr Pro His Phe Asp Thr Met Ile Ala Ala Tyr Leu Leu Glu Pro
405 410 415

Asn Glu Lys Lys Phe Asn Leu Glu Asp Leu Ser Leu Lys Phe Leu Gly
420 425 430

Tyr Lys Met Thr Ser Tyr Gln Glu Leu Met Ser Phe Ser Ser Pro Leu
435 440 445

Phe Gly Phe Ser Phe Ala Asp Val Pro Val Asp Lys Ala Ala Glu Tyr
450 455 460

Ser Cys Glu Asp Ala Asp Ile Thr Tyr Arg Leu Tyr Lys Ile Leu Ser
465 470 475 480

Met Lys Leu His Glu Ala Glu Leu Glu Asn Val Phe Tyr Arg Ile Glu
485 490 495

Met Pro Leu Val Asn Val Leu Ala Arg Met Glu Phe Asn Trp Val Tyr
500 505 510

Val Asp Thr Glu Phe Leu Lys Lys Leu Ser Glu Glu Tyr Gly Lys Lys
515 520 525

Leu Glu Glu Leu Ala Glu Lys Ile Tyr Gln Ile Ala Gly Glu Pro Phe
530 535 540

Asn Ile Asn Ser Pro Lys Gln Val Ser Asn Ile Leu Phe Glu Lys Leu
545 550 555 560

Gly Ile Lys Pro Arg Gly Lys Thr Thr Lys Thr Gly Asp Tyr Ser Thr
565 570 575

Arg Ile Glu Val Leu Glu Glu Ile Ala Asn Glu His Glu Ile Val Pro
580 585 590

Leu Ile Leu Glu Phe Arg Lys Ile Leu Lys Leu Lys Ser Thr Tyr Ile
595 600 605

Asp Thr Leu Pro Lys Leu Val Asn Pro Lys Thr Gly Arg Phe His Ala
610 615 620

Ser Phe His Gln Thr Gly Thr Ala Thr Gly Arg Leu Ser Ser Ser Asp
625 630 635 640

Pro Asn Leu Gln Asn Leu Pro Thr Lys Ser Glu Glu Gly Lys Glu Ile
645 650 655

Arg Lys Ala Ile Val Pro Gln Asp Pro Asp Trp Trp Ile Val Ser Ala
660 665 670

Asp Tyr Ser Gln Ile Glu Leu Arg Ile Leu Ala His Leu Ser Gly Asp
675 680 685

Glu Asn Leu Val Lys Ala Phe Glu Glu Gly Ile Asp Val His Thr Leu
690 695 700

Thr Ala Ser Arg Ile Tyr Asn Val Lys Pro Glu Glu Val Asn Glu Glu
705 710 715 720

Met Arg Arg Val Gly Lys Met Val Asn Tyr Ser Ile Ile Tyr Gly Val
725 730 735

Thr Pro Tyr Gly Leu Ser Val Arg Leu Gly Ile Pro Val Lys Glu Ala
740 745 750

Glu Lys Met Ile Ile Ser Tyr Phe Thr Leu Tyr Pro Lys Val Arg Ser
755 760 765

Tyr Ile Gln Gln Val Val Ala Glu Ala Lys Glu Lys Gly Tyr Val Arg
770 775 780

Thr Leu Phe Gly Arg Lys Arg Asp Ile Pro Gln Leu Met Ala Arg Asp
785 790 795 800

Lys Asn Thr Gln Ser Glu Gly Glu Arg Ile Ala Ile Asn Thr Pro Ile
805 810 815

Gln Gly Thr Ala Ala Asp Ile Ile Lys Leu Ala Met Ile Asp Ile Asp
820 825 830

Glu Glu Leu Arg Lys Arg Asn Met Lys Ser Arg Met Ile Ile Gln Val
835 840 845

His Asp Glu Leu Val Phe Glu Val Pro Asp Glu Glu Lys Glu Glu Leu
850 855 860

Val Asp Leu Val Lys Asn Lys Met Thr Asn Val Val Lys Leu Ser Val
865 870 875 880

Pro Leu Glu Val Asp Ile Ser Ile Gly Lys Ser Trp Ser
885 890

What Is Claimed Is:

1. A stable composition comprising a mixture of reagents at working concentrations, wherein said reagents are at least one thermostable enzyme and at least one buffer salt.

2. A stable composition for nucleic acid amplification comprising a mixture of reagents, wherein said reagents are at least one thermostable DNA polymerase, at least one buffer salt and at least one deoxynucleoside triphosphate.

3. A stable composition for nucleic acid sequencing comprising a mixture of reagents, wherein said reagents are at least one thermostable DNA polymerase, at least one deoxynucleoside triphosphate, at least one dideoxynucleoside triphosphate and at least one buffer salt.

4. The stable composition of claim 2 or claim 3, wherein said reagents are present at working concentrations.

5. The composition of claim 2 or claim 3, wherein said thermostable DNA polymerase is selected from the group of thermostable DNA polymerases consisting of a *Taq* DNA polymerase, a *Tne* DNA polymerase, a *Tma* DNA polymerase, and mutants thereof.

6. The composition of claim 2 or claim 3, wherein said thermostable DNA polymerase is selected from the group of thermostable DNA polymerases consisting of a *Pfu* DNA polymerase, a *Pwo* DNA polymerase, VENT™ DNA polymerase, DEEPVENT™ DNA polymerase, and mutants thereof.

7. The composition of claim 5, wherein said mixture further comprises DEEPVENT™ DNA polymerase or VENT™ DNA polymerase.

8. The composition of claim 5, wherein the concentration of *Taq* DNA polymerase or mutant thereof is about 0.1 to 200 units per milliliter.

9. The composition of claim 8, wherein the concentration is about 20 units per milliliter.

10. The composition of claim 5, wherein the concentration of *Tne* DNA polymerase or mutant thereof is about 0.1 to 200 units per milliliter.

11. The composition of claim 10, wherein the concentration is about 20 units per milliliter.

12. The composition of claim 5, wherein the concentration of *Tma* DNA polymerase or mutant thereof is about 0.1 to 200 units per milliliter.

13. The composition of claim 12, wherein the concentration is about 20 units per milliliter.

14. The composition of claim 6, wherein the concentration of VENT™ DNA polymerase or mutant thereof is about 0.1 to 200 units per milliliter.

15. The composition of claim 14, wherein the concentration is about 20 units per milliliter.

16. The composition of claim 6, wherein the concentration of DEEPVENT™ DNA polymerase or mutant thereof is about 0.1 to 200 units per milliliter.

17. The composition of claim 16 wherein the concentration is about 20 units per milliliter.

18. The composition of claim 6, wherein the concentration of *Pfu* DNA polymerase or mutant thereof is about 0.1 to 200 units per milliliter.

19. The composition of claim 18 wherein the concentration is about 20 units per milliliter.

20. The composition of claim 6, wherein the concentration of *Pwo* DNA polymerase or mutant thereof is about 0.1 to 200 units per milliliter.

21. The composition of claim 20 wherein the concentration is about 20 units per milliliter.

22. The composition of claim 7, wherein the concentration of DEEPVENT™ DNA polymerase or VENT DNA polymerase is about 0.002 to 200 units per milliliter.

23. The composition of claim 22, wherein the concentration is about 0.40 units per milliliter.

24. The composition of claim 2 or claim 3, wherein said DNA polymerase retains at least 90% of the enzymatic activity for at least four weeks when stored at about 20°C to 25°C.

25. The composition of claim 5, wherein said DNA polymerase retains at least 90% of the enzymatic activity for at least one year when stored at about 4°C.

26. The composition of claim 2 or claim 3, further comprising a magnesium salt.

27. The composition of claim 2 or claim 3, further comprising at least one nonionic detergent.

28. The composition of claim 2 or claim 3, wherein the concentration of said deoxynucleoside triphosphate is about 200 to about 300 micromolar.

29. The composition of claim 3, wherein the concentration of said dideoxynucleoside triphosphate is about 0.08 to about 5 micromolar.

30. A nucleic acid amplification kit comprising one or more containers, wherein a first container contains a stable composition comprising a mixture of reagents, wherein said reagents are at least one thermostable DNA polymerase, at least one buffer salt, and at least one deoxynucleoside triphosphate.

31. A nucleic acid sequencing kit comprising one or more containers, wherein a first container contains a stable composition comprising a mixture of reagents, wherein said reagents are at least one thermostable DNA polymerase, at least one buffer salt, at least one deoxynucleoside triphosphate and at least one dideoxynucleoside triphosphate.

32. The kit of claim 30 or 31, wherein said reagents are present at working concentrations.

33. A method of amplifying a nucleic acid molecule comprising contacting said nucleic acid molecule with the composition of claim 2.

34. A method of amplifying a nucleic acid molecule comprising contacting said nucleic acid molecule with a composition selected from the group consisting of :

a composition comprising a thermostable 3' exo+ DNA polymerase and a thermostable 3' exo- DNA polymerase wherein the concentrations of said 3' exo+ DNA polymerase and of said 3' exo- DNA polymerase are equal, and

a composition comprising a thermostable 3' exo+ DNA polymerase and a thermostable 3' exo- DNA polymerase wherein the concentration of said 3' exo+ DNA polymerase is higher than the concentration of said 3' exo- DNA polymerase.

35. A method of sequencing a nucleic acid molecule comprising contacting said nucleic acid molecule with the composition of claim 3.

36. A method of sequencing a nucleic acid molecule comprising contacting said nucleic acid molecule with a composition selected from the group consisting of :

a composition comprising a thermostable 3' exo+ DNA polymerase and a thermostable 3' exo- DNA polymerase wherein the concentrations of said 3' exo+ DNA polymerase and of said 3' exo- DNA polymerase are equal, and

a composition comprising a thermostable 3' exo+ DNA polymerase and a thermostable 3' exo- DNA polymerase wherein the concentration of said 3' exo+ DNA polymerase is higher than the concentration of said 3' exo- DNA polymerase.

37. The method of any one of claims 33-36, wherein said nucleic acid molecule is larger than about 4 kilobases in size.

38. The method of claim 37, wherein said nucleic acid molecule is larger than about 7 kilobases in size.

39. The method of claim 38, wherein said nucleic acid molecule is larger than about 8 kilobases in size.

40. A nucleic acid molecule amplified by the method of claim 33 or claim 34.

41. The nucleic acid molecule of claim 40, wherein said nucleic acid molecule is larger than about 4 kilobases in size.

42. The nucleic acid molecule of claim 41, wherein said nucleic acid molecule is larger than about 7 kilobases in size.

43. The nucleic acid molecule of claim 41, wherein said nucleic acid molecule is larger than about 8 kilobases in size.

Figure 1

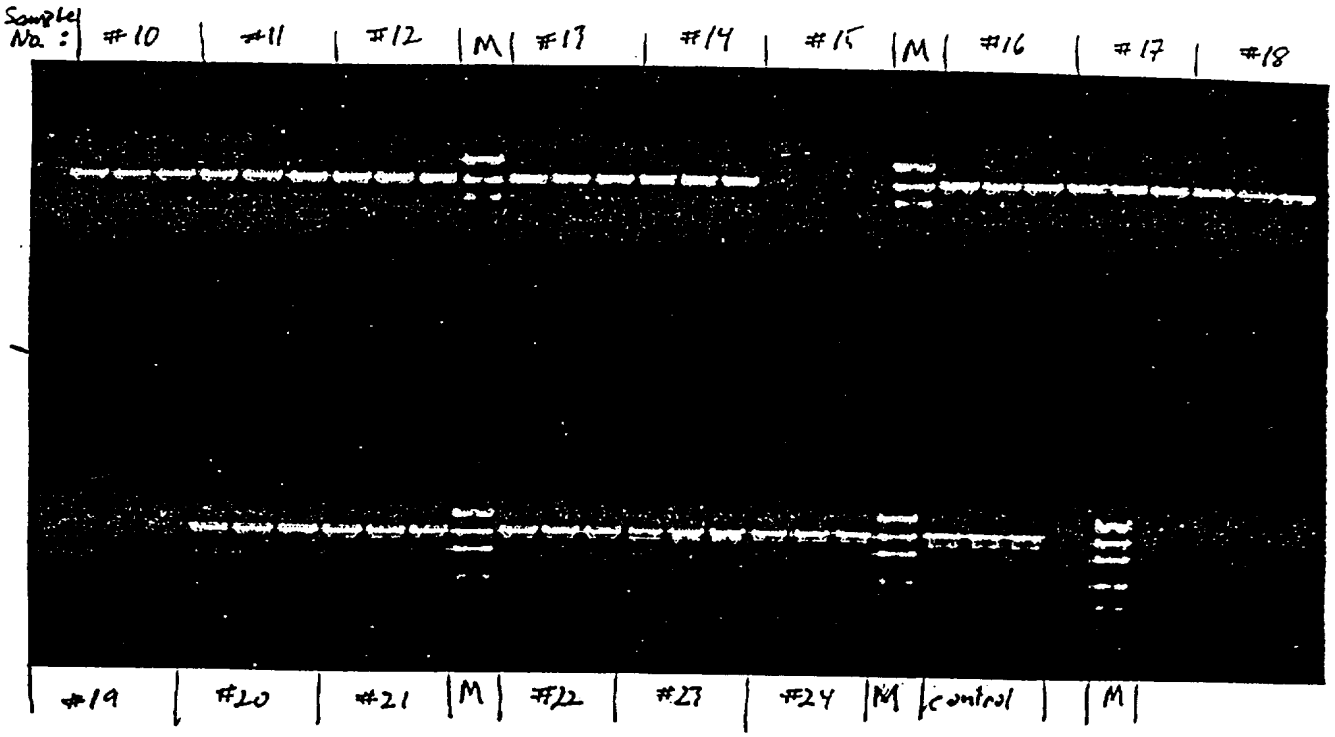
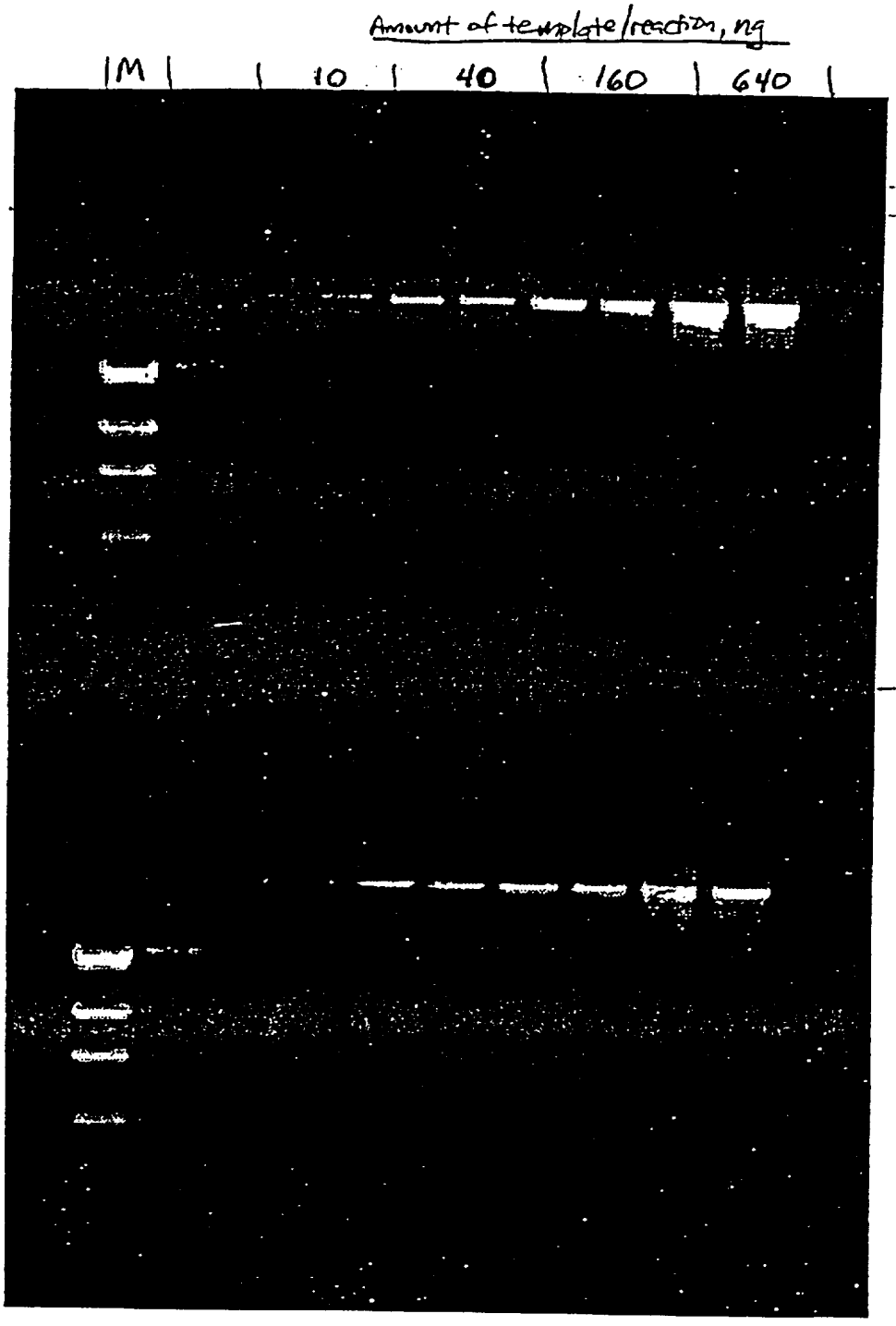

Sample No. :
#10
#11
#12
M
#13
#14
#15
M
#16
#17
#18
#19
#20
#21
M
#22
#23
#24
M
control
M

Figure 2



INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/14266

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C07H 21/04; C12P 19/34
US CL :435/91.1, 91.2; 536/22.1
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/91.1, 91.2; 536/22.1

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

NONE

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

BIOSIS, NTIS, AGRICOLA, CAB ABSTRACTS, EMBASE, LIFE SCIENCES, JICST, MEDLINE, PASCAL, DERWENT, CA, APS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P ----- Y	HUANG et al. Fidelity and Predominant Mutations Produced by Deep Vent Wild-Type and Exonuclease-Deficient DNA Polymerases during In Vitro DNA Amplification. DNA and Cell Biology. 1996, Vol. 17, No. 7, pages 589-594, especially page 590.	1, 2, 4 ---------- 3, 5-34, 37-43
X ----- Y	FLAMAN et al. A Rapid PCR Fidelity Assay. Nucleic Acids Research. 1994, Vol. 22, No. 15, pages 3259-3260, especially page 3259.	1, 2, 4 ---------- 3, 5-34, 37-43
X ----- Y	BARNES, W.M. PCR Amplification of up to 35-kb DNA with High Fidelity and High Yield from Lambda Bacteriophage Templates. Proc. Natl. Acad. Sci. March 1994, Vol. 91, pages 2216-2220, especially pages 2216 and 2217.	1, 2, 4 ---------- 3, 5-34, 37-43

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

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search
25 SEPTEMBER 1997

Date of mailing of the international search report
14 NOV 1997

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
PAUL B. TRAN, PH.D.
Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

IDT Exhibit 1020.0706

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/14266

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X ----- Y	US 5,489,523 A (MATHUR) 06 February 1996, columns 14-19.	1, 2, 4 --------- 3, 5-43
X ------ Y	US 5,466,591 A (ABRAMSON et al.) 14 November 1995, columns 19-23 and 29-33.	1, 2, 4 ---------- 3, 5-43
X ----- Y	US 5,420,029 A (GELFAND et al.) 30 May 1995, columns 21-23 and 38.	1, 2, 4 --------- 3, 5-34, 37-43
X ---- Y	WO 9516028 A1 (STRATAGENE) 15 June 1995, pages 2-7.	1, 2, 4 --------- 3, 5-34, 37-43

Form PCT/ISA/210 (continuation of second sheet)(July 1992)★

Electronic Acknowledgement Receipt

EFS ID:	46340839
Application Number:	17381888
International Application Number:	
Confirmation Number:	6188
Title of Invention:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ
First Named Inventor/Applicant Name:	Douglas A. Amorese
Customer Number:	21710
Filer:	Thomas C. Meyers/Kelley Warren
Filer Authorized By:	Thomas C. Meyers
Attorney Docket Number:	NUGN-001/04US 34079/550
Receipt Date:	04-AUG-2022
Filing Date:	21-JUL-2021
Time Stamp:	21:55:11
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Non Patent Literature	Xi-2011-PNAS.pdf	4125618 bbd1ecf756fab875d86d26b70451c96a33d54f08	no	9

Warnings:

Information:					
2	Non Patent Literature	Xu-2012-PLoSOne.PDF	345857 a9a50162448eb53c7650282cd0d4256b9bca4e1f	no	6
Warnings:					
Information:					
		Total Files Size (in bytes):	4471475		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Espacenet

Bibliographic data: WO2008033442 (A2) — 2008-03-20

METHODS AND COMPOSITIONS FOR PERFORMING LOW BACKGROUND MULTIPLEX NUCLEIC ACID AMPLIFICATION REACTIONS

Inventor(s): FREDRIKSSON JOHAN ERIK SIMON [SE]; DAHL CARL OSCAR FREDRIK [US] ± (FREDRIKSSON, JOHAN, ERIK, SIMON, ; DAHL, CARL, OSCAR, FREDRIK)

Applicant(s): UNIV LELAND STANFORD JUNIOR [US]; FREDRIKSSON JOHAN ERIK SIMON [SE]; DAHL CARL OSCAR FREDRIK [US] ± (THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY, ; FREDRIKSSON, JOHAN, ERIK, SIMON, ; DAHL, CARL, OSCAR, FREDRIK)

Classification: **- international:** ***C12Q1/68***
- cooperative: **C12Q1/686** more

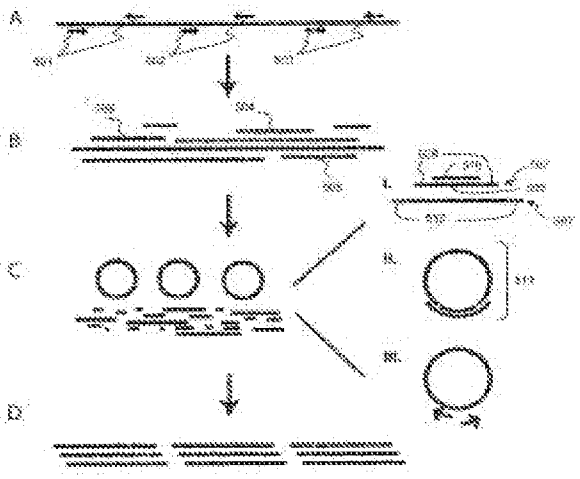
Application number: WO2007US19893 20070911 Global Dossier

Priority number(s): US20060844259P 20060912

Also published as: WO2008033442 (A3) US2011212490 (A1) US8293501 (B2)

Abstract of WO2008033442 (A2)

Methods and compositions for performing low background multiplex nucleic acid amplification reactions are provided. Aspects of the invention include contacting a nucleic acid sample with two or more primer pairs for two or more target nucleic acids under template dependent primer extension reaction conditions, e.g., polymerase chain reaction (PCR) conditions. The resultant amplified composition is then contacted with target nucleic acid circularizing reagents, and product circularized target



IDT Exhibit 1020.0710

nucleic acids are then selected, e.g., for further amplification. Also provided are systems and kits that find use in practicing embodiments of the inventions.

IDT Exhibit 1020.0711

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 March 2008 (20.03.2008)

PCT

(10) International Publication Number
WO 2008/033442 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2007/019893

(22) International Filing Date:
11 September 2007 (11.09.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/844,259 12 September 2006 (12.09.2006) US

(71) Applicant *(for all designated States except US)*: **THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY** [US/US]; 1705 El Camino Real, Palo Alto, California 94306-1106 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **FREDRIKSSON, Johan, Erik, Simon** [SE/SE]; Hallebergsvagen 12, S-167-37 Bromma (SE). **DAHL, Carl, Oscar, Fredrik** [SE/US]; 880 Roble Avenue #3, Menlo Park, CA 94025 (US).

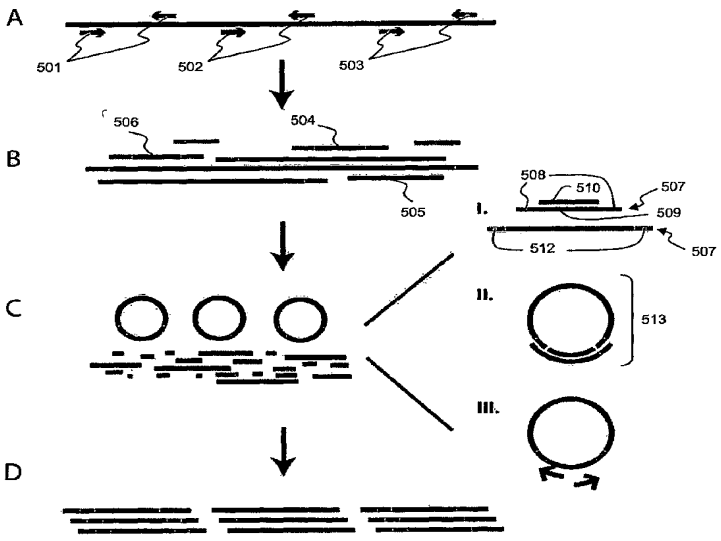
(74) Agent: **SCHERER, David, C.**; Bozicevic, Field & Francis LLP, 1900 University Avenue, Suite 200, East Palo Alto, California 94303 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL,

[Continued on next page]

(54) Title: METHODS AND COMPOSITIONS FOR PERFORMING LOW BACKGROUND MULTIPLEX NUCLEIC ACID AMPLIFICATION REACTIONS



(57) Abstract: Methods and compositions for performing low background multiplex nucleic acid amplification reactions are provided. Aspects of the invention include contacting a nucleic acid sample with two or more primer pairs for two or more target nucleic acids under template dependent primer extension reaction conditions, e.g., polymerase chain reaction (PCR) conditions. The resultant amplified composition is then contacted with target nucleic acid circularizing reagents, and product circularized target nucleic acids are then selected, e.g., for further amplification. Also provided are systems and kits that find use in practicing embodiments of the inventions.

WO 2008/033442 A2

PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

METHODS AND COMPOSITIONS FOR PERFORMING LOW BACKGROUND MULTIPLEX NUCLEIC ACID AMPLIFICATION REACTIONS

GOVERNMENT RIGHTS

This invention was made with Government support under contract 2P01HG000205 awarded by the National Institutes of Health. The Government has certain rights to this invention.

CROSS-REFERENCE TO RELATED APPLICATIONS

This applications claims benefit of priority to US provisional application 60/844,259, filed September 12, 2006, which is herein incorporated by reference.

BACKGROUND

To increase assay throughput and allow more efficient use of DNA samples, simultaneous amplification of many target nucleic acids in a sample of interest can be carried out by combining many specific primer pairs with a sample of interest and then subjecting the sample to polymerase chain reaction (PCR) conditions in a process known in the art as multiplex PCR. However, when multiple specific primer pairs are added to the same PCR reaction, non-target amplification products may be generated, with the risk of generating such products increasing with increasing numbers of specific primer pairs employed. These non-target "amplicons" significantly limit the utility of the amplification product for further analysis and/or assays. Even with careful attention paid to the design of the primers, multiplex PCR is typically limited to 10-20 specific primer pairs in a single multiplex reaction before amplification yield is compromised by the accumulation of non-target amplicons (see, e.g., Syvanen, A.C., *Toward genome-wide SNP genotyping*. Nature Genetics (2005) 37 Suppl: p. S5-10; and Broude, N.E., et al., *Multiplex allele-specific target amplification based on PCR suppression*. PNAS (2001) 98(1): p. 206-11). As such, there is a continued need for improved methods to reduce the impact of non-target amplicon generation in multiplex PCR.

IDT Exhibit 1020.0714

SUMMARY

Methods and compositions for performing low background multiplex nucleic acid amplification reactions are provided. Aspects of the invention include contacting a nucleic acid sample with two or more primer pairs for two or more target nucleic acid sequences under template dependent primer extension reaction conditions, e.g., polymerase chain reaction (PCR) conditions. The resultant amplified composition is then contacted with target nucleic acid circularizing reagents, and product circularized target nucleic acids are then selected, e.g., specifically amplified. Also provided are systems and kits that find use in practicing embodiments of the inventions.

BRIEF DESCRIPTION OF THE FIGURES

The accompanying drawings illustrate embodiments of the invention and, together with the description, serve to explain the invention. These drawings are offered by way of illustration and not by way of limitation.

Figure 1 illustrates target and non-target amplicons that are generated in a multiplex PCR reaction.

Figure 2 illustrates an exemplary target amplicon and circularization template oligonucleotide.

Figure 3A -3G each illustrate a distinct circularization complex that is formed using a specific circularization template oligonucleotide.

Figure 4A and 4B illustrate different DNA circle enrichment strategies.

Figure 5A to 5D together illustrate an exemplary target amplicon circularization and selection method according to one embodiment of the present invention.

Figure 6A shows an agarose gel of amplicons generated after 40 cycles of multiplex PCR. Numerous non-target amplicons can be seen (e.g., those greater than 500bp. Figure 6B shows an agarose gel of the amplicons generated in a multiplex PCR reaction using the same primers as in 6A in which a circularization and selection of the present invention is employed. Target amplicons have been selected while non-target amplicons have not.

Figure 7 shows the result of quantitative PCR assay for the EGFR exons that were amplified and selected as in Figure 6B. Numbers indicate copy number in the sample (estimated from an average abundance of 10 amplicons).

IDT Exhibit 1020.0715

Figure 8 shows an agarose gel of amplicons resulting from a multiplex PCR employing 170 primer pairs after circularization and selection using a method of the present invention. The resultant amplicons fall within the size range expected (between 200 and 800 bp).

Figure 9 shows the result of quantitative PCR reaction for 25 of the 170 target amplicons amplified in Figure 8. Each of the 25 targets from the whole pool of 170 was individually analyzed in a real-time PCR. As a control of the quantitative PCR variation, the same set of primer-pairs was used for amplifying and analyzing unprocessed genomic DNA. Relative fold deviation of amplified DNA (left) and non-amplified DNA (right) are shown.

Figure 10 is a diagrammatic representation of the multiplex amplification protocol of Example 4. (A) A multiplex PCR is carried out using target specific primer pairs, generating both correct and incorrect products. For clarity, only three of the 170 primer pairs are shown. (B) Guided by the collector probe, targets that contain matched primer pairs are circularized, leaving non-cognate products linear and thus susceptible to exonuclease degradation. In detail, (I) a collector probe contains complementary sequences to a cognate primer pair (shaded regions). (II) The collector probe and the DNA ligase enable circularization of correctly amplified targets. (C) A universal amplification is then carried out using a randomly primed rolling circle amplification, generating a final product of concatemers of correct target sequences.

Figure 11 demonstrates the evenness measurements of at various stages of the multiplex amplification process assessed by quantitative PCR. A subset of 48 targets, all successfully amplified according to the resequencing array, was chosen to represent the overall variation in amplification efficiency. The starting material of human genomic DNA, assumed to be perfectly uniform, is compared to the evenness after the multiplex PCR, the ligation and exonuclease treatment and finally the rolling circle amplified material. The Y-axis represents a log-scale with deviations from 1 being relative differences from the average abundance. No compensation for differences in real-time PCR efficiency between reactions was used. However, the genomic DNA starting material represents a measure of this variation and the general imprecision of the real-time PCRs. Here, 96% of the final amplicons analyzed was no less than one-fourth of the average abundance.

IDT Exhibit 1020.0716

Figure 12 shows the sequence of the forward and reveres amplification primers specific for exons of the human genes of interest as well as the circularization template oligonucleotides (listed as Gene-Collector Probes). Human genes of interest are as follows: v-akt murine thymoma viral oncogene homolog 1 (AKT-1); v-akt murine thymoma viral oncogene homolog 2 (AKT-2); adenomatosis polyposis coli (APC); epidermal growth factor receptor (EGFR); FK506 binding protein 12-rapamycin associated protein 1 (FRAP); v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (KRas); MAP/microtubule affinity-regulating kinase 3 (MARK3); SMAD, mothers against DPP homolog 4 (SMAD4); transforming growth factor, beta receptor II (TGF-beta R2); and tumor protein p53 (TP53).

DEFINITIONS

The term "nucleic acid" as used herein means a polymer composed of nucleotides, e.g., deoxyribonucleotides or ribonucleotides, or compounds produced synthetically (e.g., PNA as described in U.S. Patent No. 5,948,902 and the references cited therein) which can hybridize with naturally occurring nucleic acids in a sequence specific manner analogous to that of two naturally occurring nucleic acids, e.g., can participate in Watson-Crick base pairing interactions.

The terms "ribonucleic acid" and "RNA" as used herein mean a polymer composed of ribonucleotides.

The terms "deoxyribonucleic acid" and "DNA" as used herein mean a polymer composed of deoxyribonucleotides.

The term "oligonucleotide" or "oligo" as used herein denotes single-stranded nucleotide multimers of from about 10 up to about 400 nucleotides in length, e.g., from about 25 to about 200 nt, including from about 50 to about 175 nt, e.g. 150 nt in length.

The term "polynucleotide" as used herein refers to single- or double-stranded polymers composed of nucleotide monomers of generally greater than about 100 nucleotides in length.

The terms "hybridizing specifically to" and "specific hybridization" and "selectively hybridize to," as used herein refer to the binding, duplexing, or hybridizing of a nucleic acid molecule preferentially to a particular nucleotide sequence under stringent conditions.

IDT Exhibit 1020.0717

"Hybridizing" and "binding", with respect to polynucleotides, are used interchangeably.

The term "stringent assay conditions" as used herein refers to conditions that are compatible to produce binding pairs of nucleic acids of sufficient complementarity to provide for the desired level of specificity in the assay while being less compatible to the formation of binding pairs between binding members of insufficient complementarity to provide for the desired specificity. Stringent assay conditions are the summation or combination (totality) of both hybridization and wash conditions (e.g., when hybridization is between solution phase and surface bound nucleic acids).

"Stringent hybridization conditions" and "stringent hybridization wash conditions" in the context of nucleic acid hybridization (e.g., as in array, Southern or Northern hybridizations) are sequence dependent, and are different under different experimental parameters. Stringent hybridization conditions that can be used to identify nucleic acids within the scope of the invention can include, e.g., hybridization in a buffer comprising 50% formamide, 5×SSC, and 1% SDS at 42°C, or hybridization in a buffer comprising 5×SSC and 1% SDS at 65°C, both with a wash of 0.2×SSC and 0.1% SDS at 65°C. Exemplary stringent hybridization conditions can also include hybridization in a buffer of 40% formamide, 1 M NaCl, and 1% SDS at 37°C, and a wash in 1×SSC at 45°C. Alternatively, hybridization to filter-bound DNA in 0.5 M $NaHPO_4$, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1×SSC/0.1% SDS at 68°C can be employed. Yet additional stringent hybridization conditions include hybridization at 60°C or higher and 3×SSC (450 mM sodium chloride/45 mM sodium citrate) or incubation at 42°C in a solution containing 30% formamide, 1M NaCl, 0.5% sodium sarcosine, 50 mM MES, pH 6.5. Those of ordinary skill will readily recognize that alternative but comparable hybridization and wash conditions can be utilized to provide conditions of similar stringency.

In certain embodiments, the stringency of the wash conditions sets forth the conditions which determine whether a nucleic acid is specifically hybridized to a surface bound nucleic acid. Wash conditions used to identify nucleic acids may include, e.g.: a salt concentration of about 0.02 molar at pH 7 and a temperature of at least about 50°C or about 55°C to about 60°C; or, a salt concentration of about 0.15 M NaCl at 72°C for about 15 minutes; or, a salt concentration of about

IDT Exhibit 1020.0718

0.2xSSC at a temperature of at least about 50°C or about 55°C to about 60°C for about 15 to about 20 minutes; or, the hybridization complex is washed twice with a solution with a salt concentration of about 2×SSC containing 0.1% SDS at room temperature for 15 minutes and then washed twice by 0.1×SSC containing 0.1% SDS at 68°C for 15 minutes; or, equivalent conditions. Stringent conditions for washing can also be, e.g., 0.2×SSC/0.1% SDS at 42°C.

A specific example of stringent assay conditions is rotating hybridization at 65°C in a salt based hybridization buffer with a total monovalent cation concentration of 1.5 M (e.g., as described in U.S. Patent Application No. 09/655,482 filed on September 5, 2000, the disclosure of which is herein incorporated by reference) followed by washes of 0.5×SSC and 0.1×SSC at room temperature.

Stringent assay conditions are hybridization conditions that are at least as stringent as the above representative conditions, where a given set of conditions are considered to be at least as stringent if substantially no additional binding complexes that lack sufficient complementarity to provide for the desired specificity are produced in the given set of conditions as compared to the above specific conditions, where by "substantially no more" is meant less than about 5-fold more, typically less than about 3-fold more. Other stringent hybridization conditions are known in the art and may also be employed, as appropriate.

DETAILED DESCRIPTION

Methods and compositions for performing low background multiplex nucleic acid amplification reactions are provided. Aspects of the invention include contacting a nucleic acid sample with two or more primer pairs specific for two or more target nucleic acid sequences under template-dependent primer extension reaction conditions, e.g., polymerase chain reaction (PCR) conditions. The resultant amplified composition is then contacted with target nucleic acid circularizing reagents, and product circularized target nucleic acids are then selected, e.g., specifically amplified or enriched. Also provided are systems and kits that find use in practicing embodiments of the inventions.

As noted in the background section, conventional multiplex PCR reactions often generate significant amounts of non-target amplicons which most often are generated by mis-priming events (i.e., primers hybridizing to non-target nucleic acid sequences in the sample). Non-target amplicons are often referred to in the art as

IDT Exhibit 1020.0719

"background" products. The present invention provides methods and compositions for multiplex nucleic acid amplification that produce low to no non-target amplicons, and as such may be referred to as low background multiplex nucleic acid amplification reactions. In certain embodiments, a low background multiplex nucleic acid amplification reaction of the invention produces non-target amplicons in an amount of about 50% or less by weight of the target amplicons produced in the reaction, such as about 25% or less, about 15% or less, about 10% or less, and including about 5% or less by weight of the target amplicons produced in the reaction.

Before the present invention is described in greater detail, it is to be understood that this invention is not limited to particular embodiments described, as such may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting, since the scope of the present invention will be limited only by the appended claims.

Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the upper and lower limit of that range and any other stated or intervening value in that stated range, is encompassed within the invention. The upper and lower limits of these smaller ranges may independently be included in the smaller ranges and are also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the invention.

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can also be used in the practice or testing of the present invention, representative illustrative methods and materials are now described.

All publications and patents cited in this specification are herein incorporated by reference as if each individual publication or patent were specifically and individually indicated to be incorporated by reference and are incorporated herein by

IDT Exhibit 1020.0720

reference to disclose and describe the methods and/or materials in connection with which the publications are cited. The citation of any publication is for its disclosure prior to the filing date and should not be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention. Further, the dates of publication provided may be different from the actual publication dates which may need to be independently confirmed.

It is noted that, as used herein and in the appended claims, the singular forms "a", "an", and "the" include plural referents unless the context clearly dictates otherwise. It is further noted that the claims may be drafted to exclude any optional element. As such, this statement is intended to serve as antecedent basis for use of such exclusive terminology as "solely," "only" and the like in connection with the recitation of claim elements, or use of a "negative" limitation.

As will be apparent to those of skill in the art upon reading this disclosure, each of the individual embodiments described and illustrated herein has discrete components and features which may be readily separated from or combined with the features of any of the other several embodiments without departing from the scope or spirit of the present invention. Any recited method can be carried out in the order of events recited or in any other order which is logically possible.

In further describing various aspects of the invention, embodiments of methods that may be performed in accordance with the invention are reviewed first in greater detail, followed by a review of various applications in which the methods may find use, as well as a review of systems and kits that may find use in practicing methods of the invention.

METHODS

As summarized above, aspects of the invention include methods of performing low background multiplex nucleic acid amplification reactions. By low background is meant a multiplex nucleic acid amplification reaction that produces low to no non-target amplicons. In certain embodiments, a low background multiplex nucleic acid amplification reaction of the invention produces non-target amplicons at about 50% or less by weight of the target amplicons produced in the reaction, such as about 25% or less, about 15% or less, about 10% or less, and including about 5% or less by weight of the target amplicons produced in the

IDT Exhibit 1020.0721

reaction, e.g., as determined using gel electrophoresis, such as described in the experimental section below. In certain embodiments, no non-target amplicons are produced in the low background multiplex amplification reactions of the present invention.

By multiplex nucleic acid amplification reaction is meant that more than one primer pair specific for a distinct target nucleic acid sequence is included in the reaction. In certain embodiments, the number of target-specific primer pairs in a multiplex amplification reaction is 2 or more, e.g., about 5 or more, about 10 or more, about 25 or more, about 50 or more, about 100 or more, about 500 or more, about 1000 or more and including up to about 2000 or more target-specific primer pairs. In such multiplex amplification reactions, each target specific primer pair is designed to produce a specific target amplicon that contains its cognate target nucleic acid sequence. As such, a multiplex amplification reaction produces 2 or more species of target amplicon, e.g., about 5 or more, about 10 or more, about 25 or more, about 50 or more, about 100 or more; about 500 or more, about 1000 or more, and up to about 2000 or more species of target amplicon.

In certain embodiments, one or more primer pair in a multiplex amplification reaction is a control primer pair, i.e., a negative or positive control primer pair. By negative control primer pair is meant that the primer pair should not produce a target amplicon in the multiplex amplification reaction (i.e., the nucleic acid sample used in the multiplex amplification reaction does not contain a target nucleic acid sequence for which the negative control primer pair is specific). By positive control primer pair is meant that the primer pair should produce a target amplicon in the multiplex amplification reaction (i.e., the nucleic acid sample used in the multiplex amplification reaction is known to contain a target nucleic acid sequence for which the positive control primer pair is specific).

Embodiments practiced in accordance with the invention may include the following steps:

multiplex amplification of two or more target nucleic acids of a nucleic acid sample;

specific circularization of the target amplicons in the produced amplified nucleic acid sample; and

selection of the circularized target nucleic acids.

IDT Exhibit 1020.0722

The resultant selected circularized target nucleic acids can be analyzed directly or further manipulated, e.g., subjected to further amplification, as desired.

Multiplex Amplification

As indicated above, the low background nucleic acid amplification reaction of the present invention includes a multiplex amplification step. Multiplex amplification reactions that find use in this step include, but are not limited to, strand displacement amplification (SDA), nucleic acid sequence based amplification (NASBA), or polymerase chain reaction (PCR; described in more detail below).

In certain embodiments, the multiplex nucleic acid amplification reaction employed is PCR. In standard PCR reactions known in the art, two primers, often referred to as a forward primer and a reverse primer, work in pairs to generate multiple copies of a specific target nucleic acid sequence present in a nucleic acid sample. In such standard PCR assays, the forward and reverse primers are designed to prime nucleic acid synthesis toward each other on opposite strands of the desired target sequence. By performing repeated cycles of melting, priming; and extending (i.e., nucleic acid synthesis), multiple specific target amplicons are formed. In such standard PCR reactions, the target amplicons formed are double stranded target nucleic acids bounded by the forward and reverse primer sequences.

There are a number of distinct types of PCR reactions that find use in the subject invention, and as such, no limitation in this regard is intended. Examples include reverse-transcription PCR (RT-PCR), inverse PCR, asymmetric PCR, etc., each of which are known in the art. Any convenient nucleotide polymerase or combination of polymerases may be used in the methods of the present invention, including but not limited to reverse transcriptases (e.g., Moloney Murine Leukemia Virus RT) and thermostable DNA polymerases (e.g., Taq, Pfu, Vent, etc.).

The nucleic acid sample employed in the nucleic acid amplification reaction of the subject invention contains a nucleic acid having a sequence for which at least one primer pair is specific. As such, the nucleic acid sample can be any of a wide variety types, including, but not limited to, genomic DNA, cDNA , total RNA, mRNA, siRNA, micro-RNA, plasmid DNA, viral DNA, viral RNA, etc. In certain embodiments, the nucleic acid sample is derived from whole organisms or from specific cells from an organism, e.g., cells from a specific tissue or cell line. In

IDT Exhibit 1020.0723

certain embodiments, the nucleic acid sample is derived from cells known to or suspected of having a disease condition and/or genetic abnormality. In certain embodiments, the nucleic acid sample is derived from an infectious agent or a tissue or organism known to or suspected of being infected by an infectious agent. In certain embodiments, the nucleic acid sample is enriched for certain nucleic acid species prior to the amplification reaction, e.g., by size fractionation, subtraction methods, ligand binding properties, nucleic acid type (e.g., DNA versus RNA) or any other convenient nucleic acid enrichment method.

While primer pairs for use in nucleic acid amplification reactions are designed to produce only target amplicons, non-target amplicons can also be produced. In general, a non-target amplicon is a population of amplified nucleic acids that vary in any way from the desired target amplicon. Non-target amplicons are generally considered amplified nucleic acid products that are the result of one or more of the primers in a nucleic acid amplification reaction annealing to non-target sites in the nucleic acid in the sample. Examples of non-target amplicons include, but are not limited to, non-target sequences bounded by the same primer (e.g., two forward or two reverse primers) and non-target sequences bounded by the forward and reverse primers. In multiplex amplification embodiments, the potential for non-target amplicons increases dramatically with increasing numbers of primer pairs, as the potential for mis-priming events generating amplifiable non-target amplicons is greatly increased. In these embodiments, non-target amplicons also include those having non-target sequences bound by any combination of primers in the sample.

Figure 1 provides a schematic of a multiplex amplification reaction employing three primer pairs and potential amplicons, both target and non-target, that can be produced. Nucleic acid 100 contains target sequences for which three distinct primer pairs are specific. The first primer pair contains forward primer 101 and reverse primer 102, the second primer pair contains forward primer 103 and reverse primer 104, and the third primer pair contains forward primer 105 and reverse primer 106. Upon performing the multiplex amplification reaction, the expected target amplicons are produced as well as a number of non-target amplicons.

Specifically, as shown in Figure 1, the target amplicons will have the unique characteristic of having the target nucleic acid sequence bounded by the two cognate primers (i.e., forward and reverse primers from a specific pair). In contrast, the non-target amplicons will not have this configuration. For example, non-target

IDT Exhibit 1020.0724

amplicon 107 contains a nucleic acid sequence bounded by the forward primer of the first primer pair (101) and the reverse primer of the second primer pair (104). Any configuration that does not conform to the description of the target amplicon is considered a non-target amplicon. Only a few representative non-target amplicons are shown in Figure 1.

Circularization of Target Amplicons

As noted above, the methods of the subject invention include selective circularization of only target amplicons produced in the multiplex amplification reaction. In certain embodiments, circularization involves 1) hybridizing target amplicons produced in the multiplex amplification reaction to cognate circularization template oligonucleotides (sometimes referred to as Collector probes; e.g., as in Example 4 below) to form circularization complexes; and 2) ligating the ends of the hybridized target amplicons together (i.e., intramolecularly) by contacting the circularization complex to a ligating agent (or agents). In these embodiments, each circularization template oligonucleotide is designed to form circularization complexes with a specific target amplicon generated in the multiplex amplification reaction but not with non-target amplicons (described in detail below). Thus, the circularization template oligonucleotides of the invention facilitate circularization of target amplicons but not non-target amplicons.

In certain embodiments, free primers are removed from the multiplex-amplified nucleic acid sample prior to the hybridization step. By "free primer" is meant any primer that has not served as a primer for a nucleic acid synthesis reaction in the multiplex nucleic acid amplification reaction. Any convenient primer removal method may be employed. For example, the multiplex-amplified nucleic acid sample may be passed over a size selection column that retains short primer oligonucleotides but allows larger nucleic acids (e.g., amplicons) to pass through. Alternatively, the amplified nucleic acid sample can be contacted with an exonuclease that specifically degrades single stranded DNA (e.g., *E. coli* exonuclease I). As another example, the amplified nucleic acid sample may be subjected to a precipitation reaction in which smaller nucleic acids remain in the supernatant whereas the larger nucleic acids are in the pellet.

Any convenient hybridization method can be used in hybridizing the target amplicons to the circularization template oligonucleotide. In certain embodiments,

IDT Exhibit 1020.0725

the multiplex-amplified nucleic acid sample and the specific circularization template oligonucleotides are combined and incubated under conditions that promote specific hybridization between target amplicons and their cognate circularization template oligonucleotides, e.g., stringent hybridization conditions. In certain embodiments, the target amplicons are denatured prior to placing the sample under stringent hybridization conditions to promote strand separation of the double stranded amplicons, which allows the circularization template oligonucleotide to anneal to complementary sequences present at the ends of the cognate target amplicons (similar to the annealing in standard PCR reactions). Denaturation (or strand separation) may be carried out by any convenient method.

In certain embodiments, one strand of the double-stranded target amplicons is enriched prior to contact with the cognate circularization template oligonucleotides. In these embodiments, the target amplicon strands enriched are the ones that are predicted to hybridize to their cognate circularization template oligonucleotide to form the circularization complex (described in more detail below). Any convenient method for enriching for single-stranded target amplicons may be employed. For example, single stranded target amplicons may be selected by biotinylating one of the primers of each primer pair used in the multiplex nucleic acid amplification reaction. The resultant double stranded target amplicons will thus have one biotinylated and one non-biotinylated strand. The non-biotinylated strands can then be enriched by denaturing the multiplex-amplified nucleic acid sample, binding the denatured sample to a streptavidin -conjugated solid support, and eluting the unbound fraction (e.g., using sodium hydroxide).

As indicated above, circularization template oligonucleotides of the invention are designed to promote intramolecular ligation of target amplicons that results in circular target amplicon products. The circularization template oligonucleotides accomplish this by serving as a template that brings the distal ends of one strand of a target amplicon into proximity such that a ligating agent (or agents) can act on the target amplicon and form a circular product (i.e, a circularized target amplicon). In certain embodiments, a circularization template oligonucleotide of the invention contains a first domain and a second domain, where the first domain contains a nucleic acid sequence that is complementary to a sequence at or near the first end of one strand of a target amplicon and the second domain contains a nucleic acid sequence that is complementary to a sequence at or near the second end of the

IDT Exhibit 1020.0726

same strand of the target amplicon. In certain embodiments, the nucleic acid sequence of the first domain and the second domain contains is about 5 or more nucleotides in length, including about 10 or more nucleotides, about 15 or more nucleotides, about 20 or more nucleotides, about 30 or more nucleotides, about 50 or more nucleotides, and including up to about 100 or more nucleotides. In certain embodiments, the first and second domains of the circularization template oligonucleotide are of similar size, whereas in other embodiments the first and second domains are different sizes. The sequences at or near the end of the target amplicon to which the first and second domain of the circularization template oligonucleotide are designed to hybridize can include primer sequences (e.g., forward primer sequence or its complement and reverse primer sequences or its complement), sequences adjacent to the primer sequences, or combinations thereof. The length of the circularization template oligonucleotide may vary widely and may depend on the specifics of the target amplicon and the desired circularized product thereof. In certain embodiments, circularization template oligonucleotides contain an intervening domain that is between the first and second domains. In certain embodiments, intervening domain contains a nucleic acid sequence that is not designed to hybridize to sequences in the target amplicon (e.g., universal amplification sequences, described below). In certain embodiments, the intervening domain can be a single nucleotide or more in length, including up to about 5 nucleotides or more in length, such as from about 10 nucleotides or more, about 15 nucleotides or more, about 20 nucleotides or more, about 30 nucleotides or more, about 50 nucleotides or more, about 100 nucleotides or more, and including up to about 200 nucleotides or more in length. As such, in certain embodiments, the circularization template oligonucleotide is up to about 10 nucleotides or more in length, such as about 20 nucleotides or more, about 30 nucleotides or more, about 50 nucleotides or more, about 100 nucleotides or more, about 200 nucleotides or more, about 400 nucleotides or more and including up to about 400 nucleotides or more in length.

Figure 2 provides one example of sequences for the primer regions of a target amplicon and a circularization template oligonucleotide specific for the target amplicon. In Figure 2, target amplicon 206 was amplified using forward primer 204 and reverse primer 203. Circularization template oligonucleotide 201 contains, in the 5' to 3' direction, the complementary sequence 202 of reverse primer 203

IDT Exhibit 1020.0727

followed by the sequence of the forward primer 204. In this orientation, the circularization template oligonucleotide 201 will hybridize to the lower strand 205 of target amplicon 206 and brings the ends into close proximity to form a circularization complex, which allows them to be ligated by a ligating agent 207 (described in further detail below).

Circularization template oligonucleotides that find use in the present invention can have a wide variety of configurations and as such, can form a variety of distinct types of circularization complexes. Certain embodiments of such circularization template oligonucleotides hybridized to single-stranded target amplicons (i.e., circularization targets) to form circularization complexes are shown in Figure 3. Figure 3A illustrates circularization template oligonucleotides similar to that shown in Figure 2. In these embodiments, the circularization template oligonucleotide 301 contains a first domain 312 and a second domain 313 containing sequences that hybridize to the primer regions 302 and 303, respectively, of a single stranded target amplicon. In Figure 3B, the circularization template oligonucleotide 304 contains a first domain 312 that is complementary to one of the primer regions 302 in the target amplicon and a second domain 314 that is complementary to an internal sequence of the target amplicon. As can be seen from Figure 3B, in certain embodiments, hybridization of a target amplicon to a circularization template oligonucleotide results in one matched end and one end having an overhang sequence. Removal of the overhang sequence [e.g., "flap"-based cleavage as described in Dahl et al Nucleic Acids Res. 2005 Apr 28;33(8):e71] results in juxtaposed 3′and 5′ends on the target amplicon that are suitable for ligation. As such, an internal site can be used for selective circularization. This internal ligation site can be within the primer binding region of the target amplicon or internal to the primer binding regions (as shown in Figure 3B). In Figure 3C, circularization template oligonucleotide 305 contains first and second domains that are complementary to internal sequences of the target amplicon. As with the example in Figure 3B, removal of the overhang sequences on both ends will result in juxtaposed 5' and 3' ends suitable for ligation.

In each of Figures 3D-G, the circularization template oligonucleotide (306, 307, 308 and 309, respectively) contains an intervening domain 310 that is between the regions that hybridize to the target amplicon (described above). The intervening domain of the circularization template oligonucleotide can have virtually any nucleic acid sequence provided that it does not interfere with circularization of the target

IDT Exhibit 1020.0728

nucleic acid for which it was designed. In certain embodiments, the intervening domain is included to serve a particular function in either the selection process or in downstream applications. For example, specially designed intervening domains can be used as primer binding sites for subsequent nucleic acid amplification of the target amplicon. In Figure 3D, the circularization template oligonucleotide includes an intervening domain 310 that has a complementary oligonucleotide 311 hybridized thereto. Hybridization of this circularization template oligonucleotide and hybridized oligonucleotide to its cognate target amplicon forms a circularization complex that produces a circularized target amplicon having additional known sequences (i.e., the complement of the intervening domain). Figures 3F and 3G show circularization complexes that are analogous to those shown in Figures 3B and 3C, respectively, except that the circularization template oligonucleotides include intervening domain 310 hybridized to a complementary oligonucleotide 311. The circularization complex shown in Figure 3E is similar to that shown in 3D except that the intervening domain 310 is not hybridized to a complementary oligonucleotide. In this example, formation of the circularized target amplicon is achieved by performing a nucleic acid synthesis reaction (a gap fill reaction) prior to closing the circle by ligation. Any convenient nucleic acid synthesis method for filling in the gap may be employed. In certain embodiments, the nucleic acid polymerases employed in such gap fill reactions lack exonuclease activity (e.g., 5' to 3' exonuclease activity). Non-limiting examples of such polymerases include the Klenow fragment of *E. coli* DNA polymerase and the large fragment of the thermostable *B. stearothermophilus* DNA polymerase.

As noted above, once circularization complexes are formed, the target amplicon is circularized by contacting the circularization complex to a ligating agent(s). Any convenient ligating agent may be used in the methods of the invention, where the choice of ligating agent is guided in part by the type of circularization complex formed. In certain embodiments, the ligating agent is a DNA ligase, e.g., T4 DNA ligase. Certain ligases require the presence of a 5' end phosphate group to achieve DNA ligation. Any convenient method for ensuring that the target amplicon is appropriately 5' phosphorylated may be employed. For example, one or both of the primers of a primer pair can be 5' phosphorylated prior to performing the nucleic acid amplification reaction. As another example, the target amplicon can be 5' phosphorylated by being contacted to a nucleotide kinase (e.g.,

IDT Exhibit 1020.0729

T4 polynucleotide kinase) at any time in the assay after the amplification step and prior to the ligation step.

There are many possible variations in the target amplicon circularization methods described above. Non-limiting representative variations are described in brief below.

In certain embodiments, each primer of a cognate primer pair contains a compatible restriction enzyme recognition sequences that promote amplified target circularization when cleaved (prior to formation of the circularization complex). By compatible means that the cleaved sites have compatible cohesive ends that facilitate ligation.

In certain embodiments, one or more primers of a primer pair in a nucleic acid amplification reaction include additional non-target binding sequences that serve as all or part of the sequences that are complementary to the circularization template oligonucleotide. In certain of these embodiments, such additional non-target complementary sequences are present at the 5' end of the primer. In general, primers having these additional non-target binding sequences will be longer than primers that do not have these sequences. These non-target sequences in the primers may be so called molecular bar codes, frequently used in multiplex DNA detection on DNA microarrays (see, e.g., Hardenbol et al., Nature Biotechnology 2003, 21(6):673-8).

In certain embodiments, the circularization template oligonucleotide is modified to allow for its targeted degradation after the ligation step. Any convenient modification can be used. For example, the circularization template oligonucleotide can incorporate deoxyuridine residues (instead of deoxythymidine) which serve as targets for Uracil-N-Glycosylase digestion. Upon completion of the target amplicon ligation reaction, the sample can be treated with Uracil-N-Glycosylase which will selectively degrade the circularization template oligonucleotide.

In certain embodiments, one or more steps of the method are carried out on a solid support, e.g., a bead or array surface. In certain of these embodiments, the circularization template oligonucleotide is immobilized on a solid support. In such embodiments, the circularization complex is formed on the solid support (e.g., incubated with a sample having cognate target amplicons under hybridization conditions), contacted to a ligating agent, and then released form the solid support, e.g., under denaturing conditions. In certain embodiments, wash steps are be

IDT Exhibit 1020.0730

included. In certain other embodiments, the primers used in the nucleic acid amplification reaction are immobilized on a solid support. In certain embodiments, primer pairs are synthesized in multiplex on a DNA microarray, employed in a solid-support-based amplification reaction, and then the products are cleaved off for further use.

In certain embodiments, more than one circularization target oligonucleotide is provided for target amplicons generated by a primer pair. For example, Taq polymerase is known to sometimes add an additional "A" nucleotide to the 3'end of an amplified product (this is the reason TA cloning of Taq-amplified targets works). As such, two distinct target amplicons are generated when using Taq (i.e., with regard to the strand of interest which will bind to a circularization target): those that have an additional "A" and those that don't (the blunt end amplicon). To circularize both of these target amplicon species, two circularization oligonucleotides can be employed: one that will form a circularization complex with the additional "A" target amplicon and one that will form a circularization complex with the blunt end target amplicon.

When the nucleic acid amplification reaction is a multiplex amplification reaction, the target amplicons generated can be at varying concentrations in the amplified nucleic acid sample. This may be due to many factors, including, but not limited to, varying amounts of the targets in the original nucleic acid sample, differences in the efficiency of target amplification during the multiplex amplification reaction, or both. In certain embodiments, it is desirable to normalize such a pool of target amplicons to contain substantially the same or similar amounts of each before analysis [e.g., high throughput sequencing analysis [Margulies M. et al Nature 2005 Sep 15;437(7057):376-80)]. In certain of these embodiments, normalization can be accomplished by providing the same limiting amount of the circularization template oligonucleotides for each target amplicon. By limiting amount is meant that the molar ratio of the lowest frequency target amplicon in the amplified sample to the circularization template oligonucleotide is less than 1. Having a limiting amount of the circularization template oligonucleotides effectively normalizes the target amplicons because the surplus of each target amplicon will not be circularized and thus not selected for (as described below).

IDT Exhibit 1020.0731

Selection of Circularized Target Amplicons

The circularization step described above results in a circularized sample which contains circularized target amplicons and non-circularized (i.e., linear) non-target amplicons. The linear non-target amplicons are generally referred to herein as "background constituents". In addition to the circularized target amplicons and background constituents, the circularization sample may also contain one or more of: non-circularized target amplicons, non-ligated circularization complexes, circularization template oligonucleotides, and oligonucleotides complementary to the intervening domain.

To reduce the amount of background (and other) constituents in the circularized sample, the methods of the invention may include selection of the circularized target amplicons to produce a low-background amplification sample. Selection of circularized target amplicons can be carried out using any convenient method, and can include positive and/or negative selection methods. In certain embodiments, background constituents in the circularized sample are removed by contacting the circularization sample with an exonuclease that degrades linear, but not circularized, nucleic acids. One non-limiting example of such an exonuclease is *E. coli* exonuclease I. This embodiment is illustrated in Figure 4A and is described, for example, in Hardenbol et al. (Nat Biotechnology, 2003, 21(6):673-8). In certain embodiments, after degradation of the linear nucleic acid constituents, the exonuclease is inactivated so that it does not interfere with subsequent analysis or assays performed on the sample (e.g., further amplification reactions).

In certain embodiments, circularization of target amplicon(s) is enhanced by modulating the relative concentration of the circularization template oligonucleotide and the amplicons in the amplified sample. For example, circularization of target amplicons can be enhanced is by using a low relative concentration of the circularization template oligonucleotide. In these embodiments, the intramolecular effects of two matched hybridization events make matched circularization favored over hybridization/ligation to non-target amplicons.

In certain embodiments, the circularized sample is subjected to another round of nucleic acid amplification that selectively amplifies the circularized target amplicons contained therein. For example, circularized target amplicons can be amplified by rolling circle replication as shown in Figure 4B and described, for example, in Baner J. et al (Nucleic Acid Res. 1998 Nov 15;26(22):5073-8). The

IDT Exhibit 1020.0732

primer binding site for rolling circle amplification can be virtually anywhere in the circularized target, including in amplification primer regions, intervening domain regions, and target regions between the primer regions. In certain embodiments, hyperbranched rolling-circle replication is employed (e.g., as described in Zhang et al., Mol Diagn. 2001, 6(2):141-50). Commercial kits for performing rolling circle amplification are available (e.g., TempliPhi, GE Healthcare; method described in Dean et al., Genome Res. 2001, 11: 1095-1099). As another example, circularized targets can be amplified by PCR. In embodiments in which an intervening domain is present in the circularization template oligonucleotide, cognate primer pairs that bind in this region can be employed to re-amplify circularized target amplicons (see Figure 5, described in more detail below). In certain embodiments, each cognate circularization template oligonucleotide used to circularize target amplicons in a multiplex amplification reaction contains an intervening domain containing an identical nucleic acid region such that a single primer pair can be employed to amplify all circularized target amplicons. In certain embodiments, the nucleic acid amplification reaction is performed after exonuclease digestion of the linear nucleic acids in the circularized sample (as described above). One non-limiting example is described in Dahl et al (Nucleic Acids Res. 2005 Apr 28;33(8):e71).

In certain embodiments, the circularized target amplicons are linearized prior to amplification. Linearization can be achieved using any convenient method. In certain embodiments, the circularized target amplicon is contacted to a restriction endonuclease that cleaves the amplicon at a unique site that was engineered into one of the primers or into the intervening domain of the circularization template oligonucleotide. In other embodiments, a deoxy-Uridine residue included in one of the primers or the circularization template oligonucleotide can be used for linearizing the circularized target amplicon with Uracil-N-Glycosylase. In embodiments in which the sample is contacted to exonuclease prior to amplification, the exonuclease is removed and/or inactivated prior to any linearization step (e.g., by heat inactivation). The linearized target amplicon may be used for further downstream applications or analyzed as desired. For example, the linearized amplicon can have linkers ligated to the ends for further amplification, sequencing, purification, etc.

Figure 5 provides an exemplary diagram of one embodiment of the present invention. In this figure, a multiplex amplification reaction is carried out using three

IDT Exhibit 1020.0733

target specific primer pairs (panel A; 501, 502, and 503), generating both target amplicons (panel B; 504, 505, and 506) and non-target amplicons (panel B; not numbered). Targets amplicons that contain matched primer pairs are circularized (panel C), leaving non-target amplicons linear and therefore substrates for exonuclease. The circularization template oligonucleotide, circularization complex and resultant circularized target amplicon are shown on the right of panel C. (I) Circularization template oligonucleotide 507 has sequences 508 complementary to the ends 512 of target amplicon 511 as well as intervening domain 509 which is hybridized to its complement 510. In this example, each of the three circularization template oligonucleotides contains the same intervening domain and hybridized complement. (II) When hybridized, the target amplicon and circularization template oligonucleotide form a circularization complex 513. Circularization of the target amplicon results in a circularized target product that includes the complement of the intervening domain; (III) Once ligated, a universal amplification reaction is carried out using a primer pair that primes in the intervening domain region. In certain embodiments, the amplification is performed after removal of non-circularized species (e.g., by exonuclease digestion). This second amplification generates a final amplification sample that has low background (i.e., contains primarily target amplicons as defined above). In this example, all of the resulting target amplicons are flanked by intervening domain sequences (panel D).

SYSTEMS

The present invention provides systems for performing low background multiplex nucleic acid amplification reactions as described above. In certain embodiments, a system of the invention includes the following components: two or more target-specific primer pairs; a circularization template oligonucleotide specific for the predicted target amplicon for each primer pair; and ligating reagents. In certain embodiments, the system includes reagents for performing a multiplex amplification reaction including, but not limited to, one or more of: nucleotide triphosphates, reaction buffers, and one or more nucleic acid polymerase (e.g., thermostable DNA polymerase, reverse transcriptase, etc.). In certain embodiments, the system includes reagents for removing free primers form a multiplex amplification reaction (e.g., spin columns, nucleic acid precipitation reagents). In certain embodiments, one or more of the primers of the target-

IDT Exhibit 1020.0734

specific primer pairs in the system is modified (e.g., 5' phosphorylated, biotinylated, etc.). Such modifications will depend on the particular method of the invention for which the system is designed.

In certain embodiments, the system includes up to 5 or more target-specific primer pairs, including, e.g., about 10 or more, about 25 or more, about 50 or more, about 100 or more, about 500 or more, about 1000 or more and including up to about 2000 or more target-specific primer pairs. In certain embodiments, ligating reagents of the system include one or more of: DNA ligase, reagents for gap fill reaction (e.g., DNA polymerase, nucleotide triphosphates, etc.), reagents for 5' phosphorylation reaction (e.g., T4 polynucleotide kinase), and reagents for "flap"-based cleavage reaction. In certain embodiments, the system includes reagents for removing linear nucleic acids from a circularized sample, e.g., *E. coli* exonuclease I and buffers for use. In certain embodiments, the system includes reagents for performing an amplification reaction on the circularized sample. Such reagents include, but are not limited to, one or more of: universal primer pair, rolling circle primer, nucleotide triphosphates, reaction buffers, and one or more nucleic acid polymerase (e.g., thermostable DNA polymerase, Klenow, etc.).

KITS

The present invention provides kits that find use in performing low background multiplex nucleic acid amplification reactions as described above. In certain embodiments, a kit of the invention contains the following components: two or more primer pairs specific for distinct target nucleic acid sequences; and a circularization template oligonucleotide specific for the predicted target amplicon for each target-specific primer pair. In certain embodiments, the kit contains circularizing agent(s) (e.g., DNA ligase). In certain embodiments, the kit contains reagents for performing a multiplex amplification reaction including, but not limited to, one or more of: nucleotide triphosphates, reaction buffers, and one or more nucleic acid polymerase (e.g., thermostable DNA polymerase, reverse transcriptase, etc.). In certain embodiments, the kit contains positive and negative control nucleic acid samples for use as templates in the multiplex amplification reaction. In certain embodiments, the kit contains control primers (positive and/or negative). In certain embodiments, the kit contains reagents for removing free primers form a multiplex amplification reaction (e.g., spin columns, nucleic acid precipitation reagents, etc.).

IDT Exhibit 1020.0735

In certain embodiments, one or more of the primers of the target-specific primer pairs in the kit is modified (e.g., 5' phosphorylated, biotinylated, etc.). Such modifications will depend on the particular method of the invention for which the kit is designed to perform.

In certain embodiments, the kit contains up to 5 or more target-specific primer pairs, including, e.g., about 10 or more, about 25 or more, about 50 or more, about 100 or more, about 500 or more, about 1000 or more and including up to about 2000 or more target-specific primer pairs. In certain embodiments, circularizing agent(s) of the kit contain one or more of: DNA ligase, reagents for gap fill reaction (e.g., DNA polymerase, nucleotide triphosphates, etc.), reagents for 5' phosphorylation reaction (e.g., T4 polynucleotide kinase), and reagents for "flap"-based cleavage reaction. In certain embodiments, the kit contains reagents for removing linear nucleic acids from a circularized sample, e.g., *E. coli* exonuclease I and buffers for use. In certain embodiments, the kit contains reagents for performing an amplification reaction on the circularized sample. Such reagents include, but are not limited to, one or more of: universal primer pair, rolling circle primer, nucleotide triphosphates, reaction buffers, and one or more nucleic acid polymerase (e.g., thermostable DNA polymerase, Klenow, etc.).

In certain embodiments, the kits contain instructions for practicing the subject methods. These instructions may be present in the subject kits in a variety of forms, one or more of which may be present in the kit. One form in which these instructions may be present is as printed information on a suitable medium or substrate, e.g., a piece or pieces of paper on which the information is printed, in the packaging of the kit, in a package insert, etc. Yet another means would be a computer readable medium, e.g., diskette, CD, etc., on which the information has been recorded. Yet another means that may be present is a website address which may be used via the internet to access the information at a removed site. Any convenient means may be present in the kits.

UTILITY

As is evident form the description of the subject invention above, the low background multiplex amplification compositions and methods described herein find use in a wide range of protocols, e.g., assays that employ amplification of nucleic

IDT Exhibit 1020.0736

acids. While certain specific examples are discussed below, no limitation in this regard is intended.

In certain embodiments, target amplicons in a low background amplification reaction as described herein are analyzed directly using any convenient method. Exemplary assays include, but are not limited to, gel electrophoresis, mass spectrometry, nucleic acid hybridization assays (e.g., southern blot analysis, microarray analysis, etc.), restriction enzyme digestion analysis, nucleic acid quantitation assays, single strand nucleic acid polymorphism analysis, etc.

In certain embodiments, target amplicons are employed in assays that include additional manipulation. In certain of these embodiments, one or more of the selected target amplicons is sub-cloned into a vector prior to use in downstream assays (e.g., plasmid vector, viral vector, sequencing vector, expression vector, library vector, combinations thereof, etc.). Sub-cloning can be achieved using any convenient method, and may include contacting the target amplicons to agents that facilitate the sub-cloning reaction (e.g., restriction enzymes, polynucleotide kinases, phosphatases, ligases etc.). In certain embodiments, the target amplicons in a multiplex reaction are fractionated (e.g., separated from each other) prior to use in a downstream assay. Non-limiting fractionation methods include fractionation by size (e.g., gel electrophoresis, size selection column, etc.), by affinity (e.g., to a nucleic acid binding protein, complementary nucleic acid sequence, etc.), by molecular conformation, etc. In certain embodiments, target amplicons generated in the low background amplification reaction of the invention are used in a PCR reaction. In certain of these embodiments, the PCR reaction is a real-time PCR reaction (e.g., as shown in the examples section below). Such PCR reactions may be employed to determine which (if any) of the target primer pairs used in a multiplex amplification reaction produced a product (i.e., the nucleic acid sample contained a target nucleic acid for which the primer pair is specific).

In certain embodiments, one or more of the target amplicons in a low background amplification reaction of the present invention are subjected to nucleic acid sequence analysis. As indicated above, certain sequencing embodiments include a target amplicon fractionation and/or sub-cloning step prior to performing the sequencing reaction. For example, Sanger type sequencing can be performed using one or both of the initial multiplex amplification primers as a sequencing primer(s). In addition, pyro-sequencing or other sequencing by synthesis is also

IDT Exhibit 1020.0737

applicable. Regardless of the sequencing system used, having a low background amplified sample produced according to the present invention as a starting material (as opposed to higher background amplified samples) enhances their efficiency and accuracy. In certain embodiments, the sequencing assay employed is a highly parallel DNA analysis platform. For example, this process can be used in conjunction with large-scale sequencing and re-sequencing platforms [e.g., as reviewed in Shendure, J., et al., Nat Rev Genet, 2004. 5(5): p. 335-44]. Representative examples include sequencing by hybridization [e.g., Chee, M., et al., Science, 1996. 274(5287): p. 610-4; and Patil, N., et al., Science, 2001. 294(5547): p. 1719-23.], sequencing by ligation [Shendure, J., et al., Science, 2005: p. 1117389.], and sequencing by synthesis [e.g., Margulies, M., et al., Nature, 2005; Brenner, S., et al., Nat Biotechnol, 2000. 18(6): p. 630-4; and Mitra, R.D., et al., Anal Biochem, 2003. 320(1): p. 55-65] systems.

The low background amplification compositions and methods of the present invention can also be used in any of a variety of genotyping and diagnostic methods [e.g., see Syvanen, A.C., Nat Genet, 2005. 37 Suppl: p. S5-10; and Syvanen, A.C., Nat Rev Genet, 2001. 2(12): p. 930-42]. The functionality of genotyping and diagnostic methods that analyze amplified target nucleic acids, especially in multiplex format, are significantly hampered by background non-target amplicons present in the sample. As such, employing low-background samples amplified in accordance with the present invention can improve the accuracy and efficiency of these types of assays. In addition, the present invention can reduce the amount of the original nucleic acid sample required for analyzing multiple distinct target nucleic acids by allowing the use of increased numbers of distinct primer pairs in a single multiplex reaction without leading to the generation of significant levels of non-target (background) amplicons. Genotyping and diagnostic assays in which the low background amplification methods and compositions of the invention find use include, but are not limited to, mini-sequencing [e.g., Syvanen, A.C., et al., Genomics, 1990. 8(4): p. 684-92; and Pastinen, T., et al., Genome Res, 2000. 10(7): p. 1031-42], primer extension-based methods in concert with mass spectrometry analysis [e.g., Tost, J. and I.G. Gut, Mass Spectrom Rev, 2002. 21(6): p. 388-418], and pathogen detection, including high throughput pathogen detection (e.g., detection of the presence of multiple pathogen-derived nucleic acids in multiple samples). Furthermore, the multiplex amplification reactions of the present

IDT Exhibit 1020.0738

invention can increase throughput of mutation detection techniques that typically use many single amplification reactions by enabling multiplexing of the various primer pairs without leading to the production of interfering levels of background amplicons [see, e.g., the assays described in Faham, M., et al., Hum Mol Genet, 2001. 10(16): p. 1657-64; and Fakhrai-Rad, H., et al., Genome Res, 2004. 14(7): p. 1404-12].

While the methods and compositions of the invention described above are drawn primarily to multiplex amplification, the subject invention also finds use in low background amplification reactions in which a single target nucleic acid sequence of interest is amplified (e.g., a single primer pair is employed in a PCR rection).

EXPERIMENTAL

EXAMPLE 1

Non-target amplicons formed with multiplex PCR

To illustrate the amplification artifacts produced during a traditional multiplex PCR amplification of many selected fragments, 12 PCR primer pairs specific for exons of the EGFR gene were combined in one reaction (see Table 1 for primer pair sequences). PCR was performed in 50 μL, 50 nM of each primer, 5 Units of *Pfu* polymerase, 1X *Pfu* buffer (Stratagene), 25 mM dNTP's and 0.5 μg of human genomic DNA. Temperature cycling: 1) 95°C 5min; 2) 40 cycles of: 95°C, 30sec; 55°C, 30sec; 72°C, 60sec; 3) 72°C 10min.

IDT Exhibit 1020.0739

Table 1. EGFR primer pair sequences (5' to 3') for first amplification reaction.

Primer Pair #	Forward	SEQ ID NO:	Reverse [5' Phosphorylated]	SEQ ID NO:
1	ATC CTG CAT GGG ATG GTG	1	AAA CAG GAA AGG ACG GGC	2
2	ACC TGG ACC TTG AGG GAT TG	3	CTT CAA GTG GAA TTC TGC CC	4
3	CCC TGG ACC CAT TTT AGA CC	5	CCA TCG GAA CTG CTG TCT G	6
4	AGC ACA TGC ATC CTT CAT GG	7	AGT GCT GTA GAG CTG TCC CC	8
5	GAA AGG GCG TCA TCA GTT TC	9	CAA GTG AAG GAA GAG AGG GG	10
6	CCC TGG GAA ATG ATC CTA CC	11	GTC TTC TGT CCT GGT GTG GG	12
7	CGC TTC CTC CGT GTG TG	13	AGG AGA CAG AGC GGG ACA AG	14
8	CTC AAG AGG ACC TGG ACC G	15	GAG CCC AGC CTC AGC AG	16
9-10	GGA TCC CTA GCT ATT CTT AAT CCA AC	17	GGA AAT ATG TCG AAA AGT TCT CTC TC	18
11	CAG AGT CCC TGA GAG TCT AGA GTA ATG	19	GAG CTC TGT GCC CTA TCT TAG C	20
12	CTC CCA CAG CAT GAC CTA CC	21	GGA ATT CAC ATG GTA ATT TCA CAG	22
13	AAG GTG CCG TCT CCT CC	23	GCT ATA ACA ACA ACC TGG AGC C	24

The products of this multiplex PCR reaction were analyzed by agarose gel electrophoresis (1.5% agarose gel stained with SYBR® green I) and run with a 50bp latter (Figure 6A). The expected sizes of the target amplicons are listed on the left. As can be seen from the gel, numerous amplicons were generated that do not correspond to the expected sizes for target amplicons (e.g., no target amplicons should be larger than 500 base pairs, yet a number of amplicons are much larger). Therefore, this multiplex PCR reaction generated many non-target amplicons.

Selection of target amplicons by circularization

Multiplex PCR was performed using the same regents and primers as above except that the reaction was subjected to 8 cycles in step 2 rather than 40. After the reaction, excess primers were removed by three rounds of Centricon YM-100 filtration (500 μL TE buffer was used for each round). Circularization was performed as shown in Figure 5 employing 12 specific circularization template oligonucleotides.

IDT Exhibit 1020.0740

Sequences of the circularization template oligonucleotide, including the intervening domain and the hybridized complement, are shown in Table 2.

Table 2. Circularization templates for EGFR primer pairs listed in Table 1 having a intervening domain (in bold and underlined).**

Primer Pair #	5' to 3' Sequence [U = deoxyuridine]	SEQ ID NO:
1	GCC CGU CCU UUC CUG UUU **GCA AUG GAA CUA AGG GCA GGU GGG UAC UUG UGU UCA ACG C**AU CCU GCA UGG GAU GGU G	25
2	GGG CAG AAU UCC ACU UGA AG**G CAA UGG AAC UAA GGG CAG GUG GGU ACU UGU GUU CAA CGC** ACC UGG ACC UUG AGG GAU UG	26
3	CAG ACA GCA GUU CCG AUG G**GC AAU GGA ACU AAG GGC AGG UGG GUA CUU GUG UUC AAC GC**C CCU GGA CCC AUU UUA GAC C	27
4	GGG GAC AGC UCU ACA GCA CU**G CAA UGG AAC UAA GGG CAG GUG GGU ACU UGU GUU CAA CGC** AGC ACA UGC AUC CUU CAU GG	28
5	CCC CUC UCU UCC UUC ACU UG**G CAA UGG AAC UAA GGG CAG GUG GGU ACU UGU GUU CAA CGC** GAA AGG GCG UCA UCA GUU UC	29
6	CCC ACA CCA GGA CAG AAG AC**G CAA UGG AAC UAA GGG CAG GUG GGU ACU UGU GUU CAA CGC** CCC UGG GAA AUG AUC CUA CC	30
7	CUU GUC CCG CUC UGU CUC CU**G CAA UGG AAC UAA GGG CAG GUG GGU ACU UGU GUU CAA CGC** CGC UUC CUC CGU GUG UG	31
8	CUG CUG AGG CUG GGC UC**G CAA UGG AAC UAA GGG CAG GUG GGU ACU UGU GUU CAA CGC** CUC AAG AGG ACC UGG ACC G	32
9-10	GAG AGA GAA CUU UUC GAC AUA UUU CC**G CAA UGG AAC UAA GGG CAG GUG GGU ACU UGU GUU CAA CGC** GGA UCC CUA GCU AUU CUU AAU CCA AC	33
11	GCU AAG AUA GGG CAC AGA GCU C**GC AAU GGA ACU AAG GGC AGG UGG GUA CUU GUG UUC AAC GC**C AGA GUC CCU GAG AGU CUA GAG UAA UG	34
12	CUG UGA AAU UAC CAU GUG AAU UCC **GCA AUG GAA CUA AGG GCA GGU GGG UAC UUG UGU UCA ACG C**CU CCC ACA GCA UGA CCU ACC	35
13	GGC UCC AGG UUG UUG UUA UAG C**GC AAU GGA ACU AAG GGC AGG UGG GUA CUU GUG UUC AAC GC**A AGG UGC CGU CUC CUC C	36

** A 5' phosphorylated oligo complementary to the intervening domain above and having central UU nucleotides was included in circularization reaction. The sequence is as follows: [Phos] GCG TTG AAC ACA AGT ACC CAU UCC TGC CCT TAG TTC CAT TGC (SEQ ID NO: 37).

Formation of the circularization complex and ligation of the target amplicon were accomplished by combining 10nM of each of the twelve circularization templates and 400 nM of the oligo complementary to the intervening domain in a sample of 1X ampligase buffer. To this sample, 5 Units ampligase (Epicentre) and 5μL of the multiplex PCR product were added. This mixture was incubated at 95°C for 5 min, 75°C for 10min, 70°C for 10min, 65°C for 30min, 60°C for 10min, and 55°C for 10min. To enrich for the circular DNA, these reactions were then treated with a combination of exonuclease I and III in exo I buffer (New England Biolabs) in

IDT Exhibit 1020.0741

a volume of 50μL (with 15uL of the ligation reaction) for 1 hour at 37°C then heat inactivated at 95°C for 10min.

In order to enable the selective degradation of circularization template oligonucleotides, each contains deoxyuridine residues instead of the standard deoxythymidine, which makes them sensitive to Uracil-N-glycosylase. Also, the oligo complementary to the intervening domain contains two deoxyuridine residues centrally located, which allows Uracil-N-glycosylase to open the circular structure to ease the following PCR amplification (see Table 2). Two μL of the exonuclease reaction above was treated with 2 μL Uracil-N-glycosylase enzyme mix in 1X USER buffer (10 μL total volume) and incubated 60 min at 37°C.

Amplification of circularized pool

After Uracil-N-glycosylase treatment, the entire pool of target amplicons was amplified with a common primer pair which is specific for the universal amplification sequence (see Table 3).

Table 3. Common primer pairs for amplifying EGFR templates circularized using the circularization templates in Table 2 (5' to 3').

Forward	CCT GCC CTT AGT TCC ATT GC (SEQ ID NO: 38)
Reverse	TGG GTA CTT GTG TTC AAC GC (SEQ ID NO: 39)

The protocol was the same as the PCR reaction above except that 200nM of each primer was used. The products of the reaction were run on a 2% agarose gel stained with SYBR®green I (Figure 6B). As can be seen on the gel, all the expected amplification product sizes have appeared with minimal or no non-target amplicon contamination. (The strongest staining band resolves into two bands when the intensity of the UV light source is lower).

Control of uniformity

For many analytical applications the evenness of concentration of all the included DNA targets after amplification is important. The selected PCR products above were analyzed by quantitative real-time PCR (qPCR) using specific internal primers for each exon (see Table 4).

IDT Exhibit 1020.0742

Table 4. Primers for real-time quantitative PCR for EGFR Exons amplified using common primer pairs shown in Table 3 (5' to 3').

Exon	Forward	SEQ ID NO:	Reverse	SEQ ID NO:
2	GGA GGC TGA GAA AAT GAT CTT CA	40	CCA AGG CAC GAG TAA CAA GCT	41
3	CCC AGC CTC TCA CCC TGT AA	42	GAT GCA AAT AAA ACC GGA CTG AA	43
4	TGT AGA GCT GTC CCC CAT AGG A	44	CAC CTG GGC AGC TGT AAG TGT	45
5	TTT TAC ATT TCA GGC CAA AAG	46	GGC TGT TCA CTG ACT TAC GTT TC	47
6	AGC AGT CAC TGG GGG ACT T	48	CTC ACA GGG AAC CTT TGC TC	49
7	AGG TGG CAC CAA AGC TGT ATT T	50	GAG ACG AAG CCA CGT GCA A	51
8	GGG CCG ACA GCT ATG AGA TG	52	GGC CCT TCG CAC TTC TTA CA	53
9-10	TGT GGA GAT CGC CAC TGA TG	54	CTC ACT CTC CAT AAA TGC TAC GAA TAT T	55
11	CCA TGC CTT TGA GAA CCT AGA AA	56	GCT GTG GTC AAC TTA CTG TTG CTT	57
12	CCA AAA TTA TAA GCA ACA GAG GTG AA	58	GAC CCA TTA GAA CCA ACT CCA TAA A	59

Each qPCR was performed with a 2X SYBR® green master mix from Bio-Rad, 200 nM of each primer, and run on an ABI 7900. Figure 7 displays normalized data where the optimal value for each exon is 10 copies.

EXAMPLE 2

Oligonucleotides

All probes and primers synthesized at the Stanford Genome Technology Center. PCR primers were pooled and phosphorylated by optikinase (USB) according to manufacturer's protocol.

Target amplicon selection

First multiplex PCR was run in 50 μL with all primers at 100 nM concentration using 10 units *Pfu* polymerase in 1x Pfu buffer (Stratagene), 200 μM each dNTP and human genomic DNA. Temperature cycling program was: 1) 95°C, 5min; 2) 8 cycles of 95°C, 30sec; 55°C 60sec; 72°C 8 min; and 3) 72°C 10 min. Excess primers after PCR were removed by filtration on centricon YM-100 with three washes of 500 μL TE buffer prior to ligation. Amplicon circularization by ligation on

IDT Exhibit 1020.0743

100 pM of each circularization template oligonucleotide and 52 nM of the complementary universal amplification sequence oligo was performed in 1x ampligase buffer (Epicentre), 5 Units ampligase, at 95°C for 5 minutes, then 75°C 10 min, 70°C 10 min, 65°C 30 min, 60°C 10 min, 55°C 10 min. To enrich for the circular DNA and remove excess circularization template oligos (which contained deoxyuridine residues), 15 μL of the ligation was treated in exonuclease buffer (Epicenter) with 24 units exonuclease I and 2 μL uracil DNA excision mix (Epicenter) in 50 μL for 1 hour at 37°C then heated to 95°C for 10 minutes. Finally all amplicons were amplified by PCR in a 25 μL reaction using a universal primer pair at 200 nM concentration with 5 Units platinum Taq DNA polymerase (Invitrogen), 3 mM MgCl, and 200 μM of each dNTP.

Quantitative PCR of amplicons

PCR primers were designed for individual exons of the EGFR gene and used to quantitative the relative abundance of each exon in 10 μL reactions with BioRad SYBR®Green master mix on an ABI 7900 instrument.

Results

Coding sequence specific PCR primer pairs were designed using the ExonPrimer online software for 10 cancer genes. The resulting 170 primer pairs were synthesized, pooled into one tube, and enzymatically phosphorylated. A multiplexed PCR was then run for 8 cycles using *Pfu* polymerase which generates blunt end PCR products suitable for ligation. Excess primers were then removed using filtration. A pool of circularization template oligos, each specific for one target amplicon, along with an oligo complementary to the universal amplification sequence then guided a circularization reaction of DNA targets with matched primer pair ends. By adding DNA ligase to the circularization reaction, closed circular molecules were formed that all contained the universal amplification sequence complement. The circularization reaction was then followed by an exonuclease treatment to degrade the remaining linear amplicons. Simultaneously, the uracil containing circularization template oligos were degraded by a combination of uracil-N-glycosylase and an apurinic endonuclease. The circularized sequences were

IDT Exhibit 1020.0744

finally PCR amplified using a universal primer pair specific for the universal amplification sequence.

The specificity of this target amplicon enrichment process was assessed using a 1.5% agarose gel stained with SYBR®green (Figure 8). The gel analysis shows a smear of product with the expected size range for the 170 amplicons (i.e., 200-800 bp), indicating high specificity of the corrector-PCR method.

To determine whether the target sequences of interest are efficiently amplified, we analyzed 25 specific target amplicons using quantitative PCR. As shown in Figure 9, all of the 25 amplicons were present in the selected sample (shown on the left side, black diamonds) and are present within a 30-fold concentration range around the overall average abundance. Moreover, 76% of the 25 targets are within 4-fold of the average. Quantitative PCR results using non-amplified genomic DNA and the same 25 primer pairs are shown on the right (open circles).

The target amplicon selection process described above could be simplified or modified in many ways. For example, removal of the excess multiplex PCR primers could be done by using a single stranded exonuclease or vacuum manifold as apposed to the centrifugation method used above. Further, the final universal primer pair could be made exonuclease resistant to enable a simultaneous removal of non-target amplicons (with exonuclease) and second PCR reaction, eliminating one step and reducing hands on time.

EXAMPLE 3

Single Reaction Tube

This example provides a method for performing the entire reaction in one tube with sequential additions of reagents.

A multiplex PCR is performed as described in Example 2 except that it is performed in a 10 μL volume. After completion, the primers are degraded by adding exonuclease I and the dNTPs are inactivated by adding apyrase (a phosphatase may be used instead). After incubation, the exonuclease and apyrase are inactivated by heating (85°C for 15-30 minutes).

Circularization template oligonucleotide(s) are added and the volume is adjusted to approximately 15μL. Hybridization and ligation is performed as described in Example 2 (using ampligase and specific incubation conditions).

IDT Exhibit 1020.0745

Removal of the linear nucleic acids by exonuclease I digestion is performed as described above. A final PCR reaction is then performed as described above in a final volume of 50μl using universal primer pairs.

This one-tube example enables automated high throughput analyses by removing the need for sample manipulation between steps of the low background amplification reaction of the invention.

EXAMPLE 4

This Example provides a method for single tube multiplex amplification sequences from multiple genes. In this example, the genes of interest are the following human cancer-related genes: v-akt murine thymoma viral oncogene homolog 1 (AKT-1); v-akt murine thymoma viral oncogene homolog 2 (AKT-2); adenomatosis polyposis coli (APC); epidermal growth factor receptor (EGFR); FK506 binding protein 12-rapamycin associated protein 1 (FRAP); v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (KRas); MAP/microtubule affinity-regulating kinase 3 (MARK3); SMAD, mothers against DPP homolog 4 (SMAD4); transforming growth factor, beta receptor II (TGF-beta R2); and tumor protein p53 (TP53).

MATERIALS AND METHODS

Oligonucleotide probes and target amplicons

All oligonucleotides were synthesized at the Stanford Genome Technology Center (see Figure 12 for primer sequences and circularization template oligos used). The thymidines were substituted with uracil bases in the circularization template oligonucleotides (also called Collector probes) for degradation purposes by uracil-DNA glycosylase. However, this enzymatic procedure was later found not to be necessary and removed from the protocol.

IDT Exhibit 1020.0746

Table 5. Analysis of failed amplifications.

	Failures	Total	% Success	Fraction
Targeted amplicons		170		
Human design error	1	169		
No collector probe *negative control*	2	167		
Failed at Mux-PCR	5	162	97%	162/167
Failed at ligation	3	159	95%	159/167
Failed at final amplification	5	154	92%	154/167
Unknown failures (no qPCR) *2 with 75% GC*	4	150	90%	150/167

One primer pair was incorrectly designed through human error and two target sequences lacked Collector probes as a negative control leaving a total of 167 amplicons with a chance of successful amplification. Quantitative PCR revealed at what stage the Gene-Collector protocol failed. The failure reason for the final four amplicons still remains unknown as no successful quantitative PCR primers could be designed.

Multiplex Amplification Protocol (Gene-Collector protocol)

First, multiplex PCR was run in 50 µl with all 340 primers (170 pairs) at 100 nM concentration each using 10 units pfu polymerase in 1 x pfu buffer (Stratagene), 200 µM each dNTP and 200 ng human genomic DNA, at 95°C for 5 min—[(95°C for 30 s; 55°C for 2 min; 72°C for 8 min) x 8] followed by 72°C for 10 min. Excess primers were removed by the addition of exonuclease I and incubated for 30 min at 30°C, followed by removal of enzymes by a Qiagen PCR purification column. Amplicon circularization by ligation was performed on 20 nM of each circularization template oligonucleotide (or collector probe) in 1x Ampligase buffer (Epicentre), 5 units Ampligase, 5 units OptiKinase (USB), 1 mM ATP, 1 mM DTT at 37°C for 30 min—[(95°C for 30 s; 65°C for 2 min; 55°C for 1 min, 60°C for 5 min) x 10] in 50 µl. A combination of exonuclease I, exonuclease T7 gene 6 and λ exonuclease reduced the amount of linear DNA during 45 min at 37°C and then stopped when heated for 20 min at 80°C. The circular DNA was concentrated by a second Qiagen PCR purification column eluted in the supplied elution buffer and set to evaporate for 45 min at 65°C. One microliter of the 10-fold concentrated circles were added to a 10 µl TempliPhi reaction (GE) supplemented with 10% DMSO and run at 30°C for 16 h, then inactivated at 65°C for 10 min.

IDT Exhibit 1020.0747

Resequencing by hybridization

A 50-kb high-density DNA array was designed by Affymetrix to match the 10-gene reference sequences. The collector amplified product was purified in a PCR purification column (Qiagen). One hundred and fifty nanograms of purified product was fragmented, labeled and finally hybridized according to the protocol provided by Affymetrix (GeneChip CustomSeq Resequencing Array Protocol). The array was washed and stained using the Affymetrix GeneChip Fluidics Station 450 and scanned using GeneChip Scanner 3000 according to the protocol. The scanned probe array image was analyzed using Affymetrix GeneChip Sequence Analysis Software.

Quantitative PCR of amplicons

Ten microliter reactions containing 400 nM of qPCR primers specific for the individual amplicons with 2 µl of the TempliPhi reaction diluted 1000-fold in TE buffer were performed to assay their relative abundance. Bio-Rad Sybr Green master mix (1x) was used on an ABI 7900 instrument (see Table 7 for primers).

RESULTS

Coding-sequence-specific PCR primer pairs were designed using ExonPrimer [see http(colon)//ihg(dot)gsf(dot)de/ihg/ExonPrimer(dot)html] for 10 cancer-related genes (see Figure 12). The resulting 170 primer pairs were synthesized and pooled into one tube. A multiplexed PCR was then run for eight cycles using pfu polymerase which generates blunt-end PCR products suitable for circularization by ligation. Excess primers were then removed using a single strand-specific exonuclease followed by a Qiagen PCR product purification column. A pool of Collector probes, each specific to one correct amplicon (shown in Figure 12) then guided a circularization reaction of matched PCR primer pair ends and closed circles were formed by a DNA ligase enzyme. The ligation reaction also involved a pre-step at 37°C for phosphorylation of 5'-ends by a kinase enzyme prior to ligation. Circularization was then followed by the addition of an exonuclease cocktail to degrade linear DNA such as amplification artifacts, genomic DNA and excess Collector probes. The circularized sequences were finally amplified using hyper-branched rolling circle amplification with random hexamers and phi-29 polymerase

IDT Exhibit 1020.0748

(TempliPhi). An outline of the above described multiplex amplification procedure (the Gene-Collector procedure) is displayed in shown in Figure 10.

The success rate of the amplification was assessed by hybridizing the final product on an Affymetrix custom-designed resequencing array containing probes scanning the coding sequence of these 10 genes with four variant probes for each nucleotide position, A, T, G and C. The array revealed that 90% of the target sequences had been successfully amplified as assessed by providing accurately read sequence for at least 30% of the nucleotides in each individual amplicon located in continuous stretches of sequence.

The initial multiplex PCR is conducted under very non-stringent conditions in order to give all target sequences the best chance of efficient amplification. This would normally generate many amplification artifacts but these are efficiently removed by circularization and exonuclease degradation. To ensure uniformity of the multiplex-PCR, extension times were long (i.e., 8 min.), with primer hybridizations conducted at 55°C for 2 min. Each stage of the reaction was analyzed for evenness by quantitative PCR (see Figure 11). As shown in Figure 11, the multiplex Gene-Collector protocol amplifies target sequences in a substantially uniform manner (i.e., the Gene-Collector protocol does not lead to significant over- or under-sampling of the target amplicons).

It is noted here that a number of primer pairs that did not work in individual PCRs under standard conditions (e.g., as analyzed by agarose gel) did produce the correct product with the Gene-Collector procedure (data not shown).

The final amplification by TempliPhi was supplemented with a 10% final DMSO concentration to reduce potential skewing effects of varied amplicon GC content. The average abundance of each final product was estimated to be at 10 nM in a 10 µl reaction volume with 96% of all amplicons having no less than one-fourth of the average abundance.

In order to measure the levels of false amplification products generated by the Gene-Collector protocol, the final product was cloned and sequenced. The TempliPhi reaction produces concatemeric products of 10 kb each, which were fragmented by sonication, gel purified and cloned into a sequencing vector. When 96 colonies were picked and Sanger sequenced, 93 reads showed that 58% of the reads were of expected products (see Table 6). As cloning selects the sequence representation randomly, it provides an additional measure of frequency distribution.

IDT Exhibit 1020.0749

Most amplicons appeared only once showing even representation. Nine amplicons appeared twice and two of the targets three times. No non-specific products appeared more than once. The fraction of paired matched primers found among the non-specific products was much lower than for the specific ones. As can be seen in Table 6, few non-specific products were formed by two matched primer pairs amplifying a non-target sequence. This type of false product would still become circularized by the Collector probe but are not the main source of errors. As expected from cloned rolling-circle-amplified material, many sequencing reactions produced concatemeric reads of repeated elements. Interestingly, this provided redundant sequencing within one and the same read with up to 3-fold coverage.

Table 6. Analysis of amplification product by cloning and sequencing.

	Reads	% of total	Fraction
Total sequence reads	93		
Correct products	54	58%	54/93
two matched primers	52		
one primer	52		
Non-specific products	39	42%	39/93
two matched primers	4		
one primer	8		
two non matched primers	2		
not found in human genome	1		

From the 93 total reads produced, 58% of these were of the expected products. Primer sequences were only rarely found within the non-specific products either as single primers, non-matched pairs or as matched pairs suggesting that the TempliPhi reaction produced the majority of the artifacts or that they were simply caused by remaining genomic DNA

We have amplified all the coding sequences located in 10 cancer genes using a multiplexed amplification procedure (termed Gene-Collector). Resequencing of large numbers of cancer-related genes has recently shown to provide important biological insights into the disease (see, e.g., Sjoblom et al. (2006) The consensus coding sequences of human breast and colorectal cancers Science, 314, 268–274). Even with extensive optimization, standard multiplex PCR is not a feasible approach to large-scale genetic studies as the failure rate is too high due to the many false amplicons out competing the correct ones for the amplification reagents. However, even though these false amplicons do result, the correct products are also present

IDT Exhibit 1020.0750

and at uniform abundance early in the amplification. The present invention reduces the presence of false products enabling further amplification of the correct ones.

The presented initial multiplex PCR had very relaxed conditions in order to give all primer pairs the ability to hybridize through the use of low hybridization temperature and long duration. Polymerization of all templates was assured by a long extension time and an ample amount of DNA polymerase. This condition was suitable for all amplicons as the Collector procedure removes artifacts by exonuclease degradation. Primer-dimer artifacts, which are a major problem in traditional multiplexed PCR, are of little concern for Gene-Collector as the circularization process is impossible of such short DNA strands due to the lower limit size constraints of partially double stranded circular DNA (Baneret al. (1998) Signal amplification of padlock probes by rolling circle replication Nucleic Acids Res, 26, 5073–5078). As shown above, the relative abundance of products from the rolling circle reaction was very even.

Alternatively, one may use PCR in the final amplification of the circularized amplicons, which then gives distinct bands on standard agarose gel (see, e.g., Figure 6B). Certain embodiments of the Gene-Collector protocol employ a general primer pair motif within the Collector probe and as such can generate a purer product than the randomly primed rolling circle amplification. This could, for example, be suitable for rapid multiplex pathogen detection using electrophoretic separation.

It is noted that target sequences could be arrayed if the circularization is performed on immobilized Collector probes.

The present invention finds use in a wide range of amplification-based applications, particularly in combination with highly parallel DNA analysis platforms. One class of parallel DNA analysis is large-scale sequencing and resequencing platforms [see, e.g., Shendure et al. (2004) Advanced sequencing technologies: methods and goals Nat. Rev, 5, 335–344], such as sequencing by hybridization, sequencing by ligation or sequencing by synthesis systems [see, e.g., Chee et al. (1996) Accessing genetic information with high-density DNA arrays Science, 274, 610–614; Patil et al. (2001) Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21 Science, 294, 1719–1723; Margulies et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors

IDT Exhibit 1020.0751

Nature, 437, 376–380; and Shendure et al. (2005) Accurate multiplex polony sequencing of an evolved bacterial genome Science, 309:1728-32]. The multiplex amplification technology disclosed herein also finds use in combination with PCR-intense genotyping methods, like mini-sequencing [see, e.g., Syvanen et al. (1990) A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E Genomics, 8, 684–692; and Pastinen et al. (2000) A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays Genome Res, 10, 1031–1042] and primer extension-based methods in concert with mass spectrometry analysis [see, e.g., Tost, J and Gut, IG. (2002) Genotyping single nucleotide polymorphisms by mass spectrometry Mass Spectrom. Rev, 21, 388–418], as well as high throughput pathogen detection. Gene-Collector can be combined with genetic variation detection techniques that require many single PCRs [see, e.g., Faham et al. (2001) Mismatch repair detection (MRD): high-throughput scanning for DNA variations Hum. Mol. Genet, 10, 1657–1664; and Fakhrai-Rad et al. (2004) SNP discovery in pooled samples with mismatch repair detection Genome Res, 14, 1404–1412] to increase assay throughput.

The present invention also enables analysis of small and precious sample materials, reduces enzyme consumption and offers higher throughput of DNA amplification.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it is readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

Accordingly, the preceding merely illustrates the principles of the invention. It will be appreciated that those skilled in the art will be able to devise various arrangements which, although not explicitly described or shown herein, embody the principles of the invention and are included within its spirit and scope. Furthermore, all examples and conditional language recited herein are principally intended to aid the reader in understanding the principles of the invention and the concepts contributed by the inventors to furthering the art, and are to be construed as being

IDT Exhibit 1020.0752

without limitation to such specifically recited examples and conditions. Moreover, all statements herein reciting principles, aspects, and embodiments of the invention as well as specific examples thereof, are intended to encompass both structural and functional equivalents thereof. Additionally, it is intended that such equivalents include both currently known equivalents and equivalents developed in the future, i.e., any elements developed that perform the same function, regardless of structure. The scope of the present invention, therefore, is not intended to be limited to the exemplary embodiments shown and described herein. Rather, the scope and spirit of the present invention is embodied by the appended claims.

IDT Exhibit 1020.0753

What is claimed is:

1. A method of selectively amplifying two or more target nucleic acids of a nucleic acid sample, said method comprising:

a) contacting said nucleic acid sample with primer pairs specific for two or more target nucleic acid sequences of interest to produce a first reaction mixture;

(b) subjecting said first reaction mixture to template-dependent primer extension reaction conditions to produce an amplified composition comprising target amplicons;

(c) circularlizing said target amplicons in said amplified composition; and

(d) selecting for said circularized target amplicons in said amplified composition.

2. The method according to Claim 1, wherein said template dependent primer extension reaction conditions are polymerase chain reaction conditions (PCR).

3. The method according to Claim 1, wherein said amplified composition further comprises non-target amplicons.

4. The method according to Claim 1, wherein said circularizing comprises:

(i) contacting said amplified composition under hybridization conditions with circularlizing template nucleic acids specific for each of said target amplicons to produce circularization complexes; and

(ii) contacting said circularization complexes with a circularizing agent to circularize said target amplicon.

5. The method according to Claim 4, wherein each of said circularization template nucleic acids comprises a first domain and a second domain, wherein said first domain and said second domain comprise sequences complementary to a 5' end sequence and a 3' end sequence of the same strand of its corresponding target amplicon, respectively.

IDT Exhibit 1020.0754

6. The method according to Claim 5, wherein each of said circularization template nucleic acids comprises an intervening domain between said first domain and said second domain.

7. The method according to Claim 4, wherein said circularizing agent is a DNA ligase.

8. The method according to Claim 4, wherein said circularization template nucleic acid is immobilized on a solid support.

9. The method according to Claim 1, wherein said method employs primer pairs specific for 5 or more different target nucleic acid sequences.

10. The method according to Claim 9, wherein said method employs primer pairs specific for 50 or more different target nucleic acid sequences.

11. The method according to Claim 1, wherein said selecting step comprises contacting said amplified composition comprising circularized target amplicons to an exonuclease.

12. The method according to Claim 1, wherein said selecting step further comprises amplifying said circularized target amplicons.

13. The method according to claim 12, wherein said amplifying is by rolling circle amplification.

14. The method according to claim 12, wherein said amplifying is by PCR.

15. The method according to claim 14, wherein said method further comprises linearizing said circularized target nucleic acids prior to said amplifying.

16. The method according to Claim 1, wherein said method produces an amplified composition comprising less than about 50% by weight non target-amplicons.

IDT Exhibit 1020.0755

17. A composition comprising two or more distinct circularization complexes, wherein each of said distinct circularization complexes comprises a target amplicon hybridized to its cognate circularization template nucleic acid.

18. The composition according to Claim 17, wherein said composition further comprises non-target amplicons.

19. A composition comprising two or more distinct circularized target amplicons, wherein each of said distinct circularized target amplicons is hybridized to its cognate circularization template nucleic acid.

20. The composition according to Claim 19, further comprising a circularizing agent.

21. The composition according to Claim 20, further comprising non-target amplicons.

22. A kit comprising:

a set of two or more primer pairs specific for distinct target nucleic acid sequences; and

a set of circularization template nucleic acids specific for the predicted target amplicons produced by each of said two or more primer pairs.

23. The kit according to Claim 22, wherein said kit further comprises a circularization agent.

24. The kit according to Claim 23, wherein said circularization agent comprises a DNA ligase.

25. The kit according to Claim 22, wherein said kit further comprises a reagent for removing linear nucleic acids from a circularized amplification composition.

26. The kit according to Claim 25, wherein said reagent for removing linear nucleic acids from a circularized amplification composition is an exonuclease.

IDT Exhibit 1020.0756

27. The kit according to Claim 22, wherein said kit further comprises reagents for performing an amplification reaction on a circularized amplification composition.

28. The kit according to Claim 22, wherein said set of primer pairs comprises 20 or more primer pairs.

29. The kit according to Claim 28, wherein said set of primer pairs comprises 100 or more primer pairs.

30. A system comprising:

a set of two or more primer pairs specific for distinct target nucleic acid sequences;

a set of circularization template nucleic acids specific for the predicted target amplicons produced by each of said two or more primer pairs; and

a circularizing agent.

IDT Exhibit 1020.0757

Figure 1

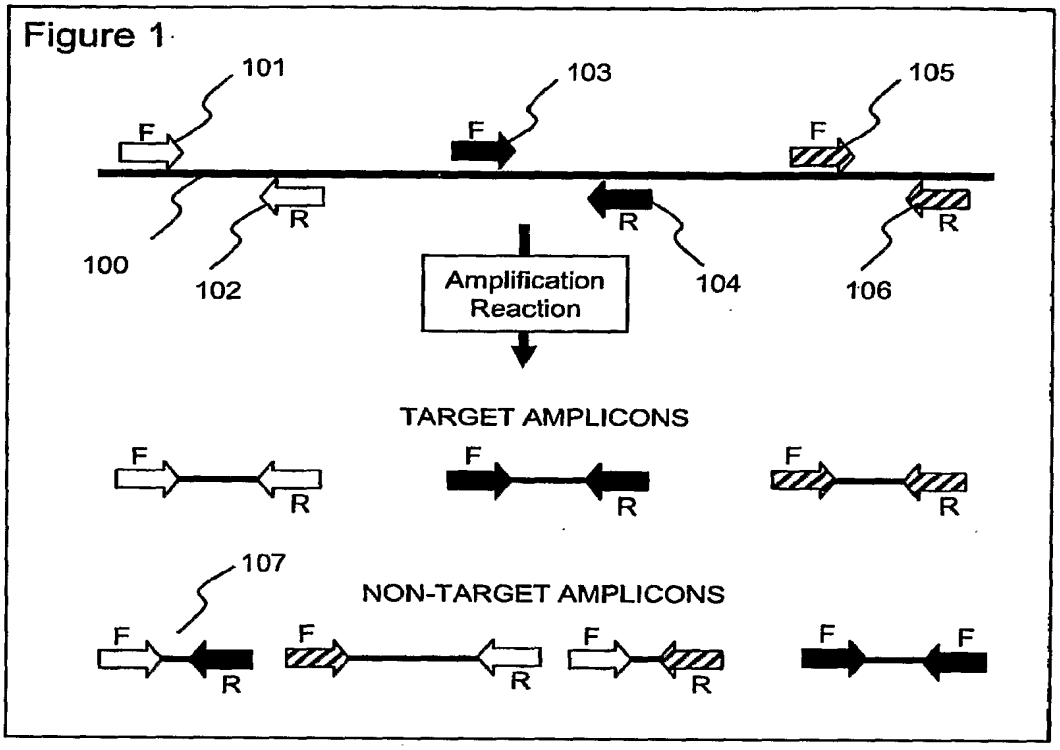

101
103
105
F
F
F
R
R
R
100
102
104
106
Amplification Reaction
TARGET AMPLICONS
F
F
F
R
R
R
107
NON-TARGET AMPLICONS
F
F
F
F
R
R
R
R

Figure 2

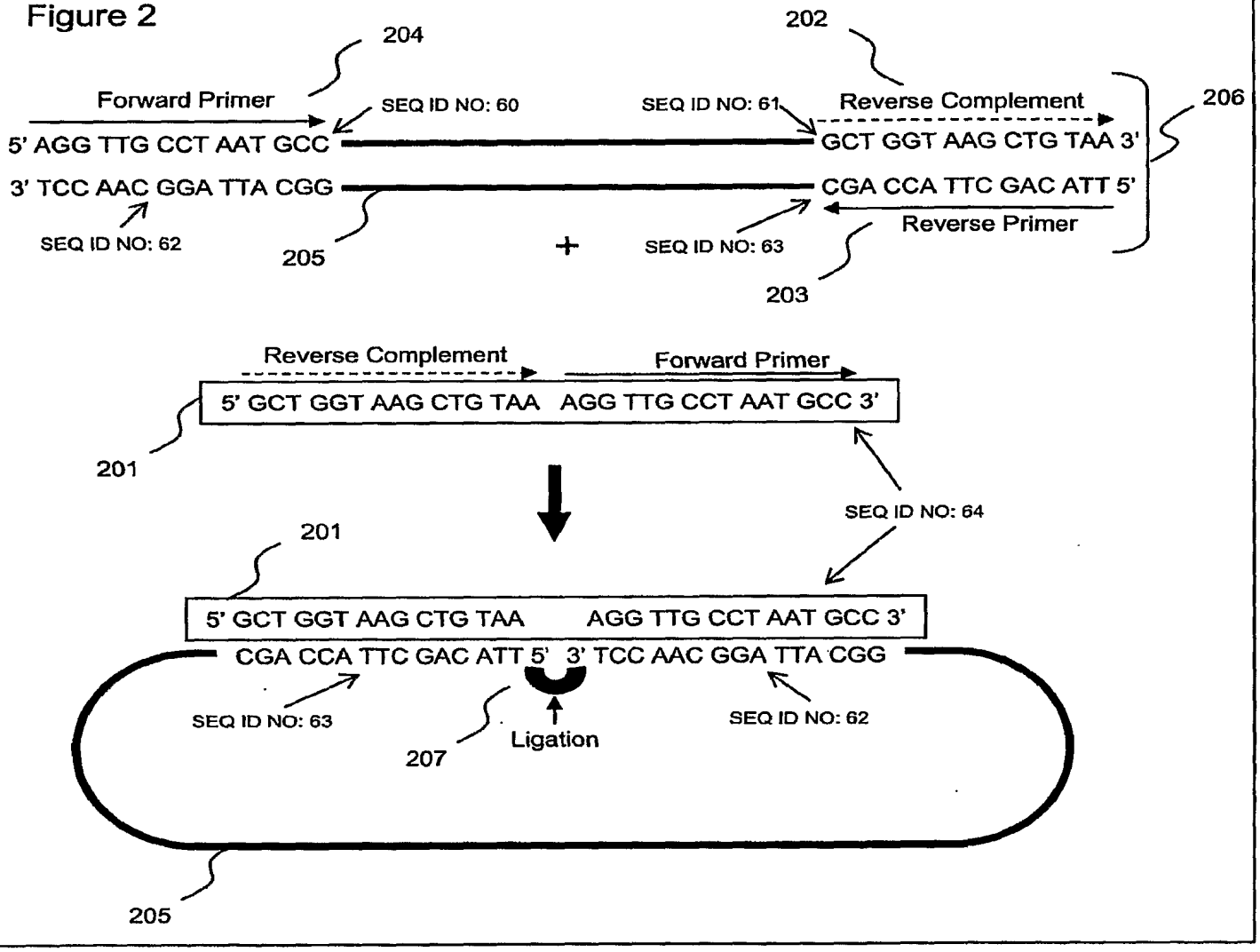

204
202
Forward Primer
SEQ ID NO: 60
SEQ ID NO: 61
Reverse Complement
206
5' AGG TTG CCT AAT GCC
GCT GGT AAG CTG TAA 3'
3' TCC AAC GGA TTA CGG
CGA CCA TTC GAC ATT 5'
SEQ ID NO: 62
205
+
SEQ ID NO: 63
Reverse Primer
203
Reverse Complement
Forward Primer
5' GCT GGT AAG CTG TAA AGG TTG CCT AAT GCC 3'
201
SEQ ID NO: 64
201
5' GCT GGT AAG CTG TAA AGG TTG CCT AAT GCC 3'
CGA CCA TTC GAC ATT 5' 3' TCC AAC GGA TTA CGG
SEQ ID NO: 63
207
Ligation
SEQ ID NO: 62
205

Figure 3

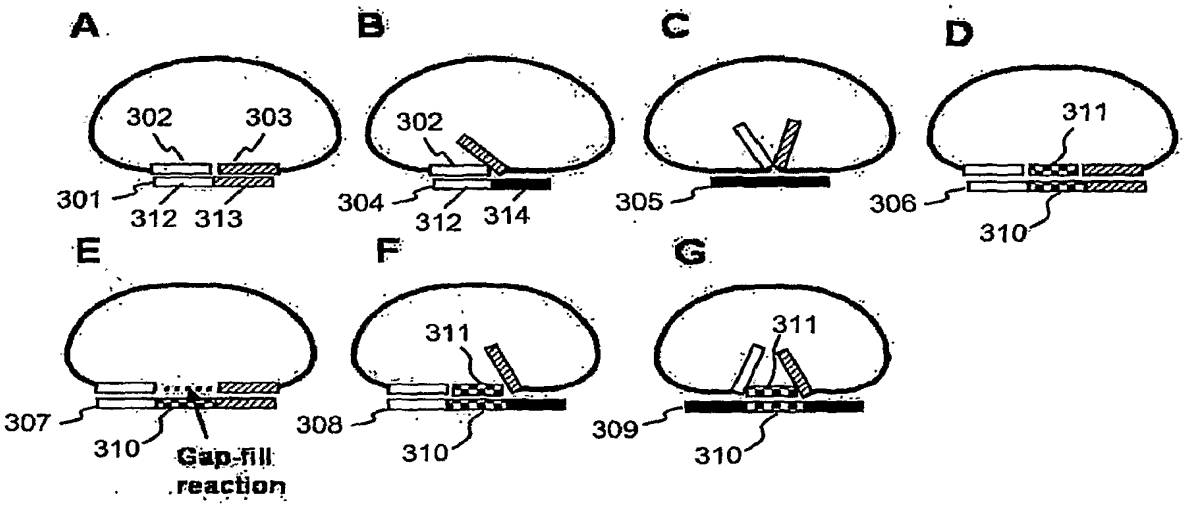


Figure 4

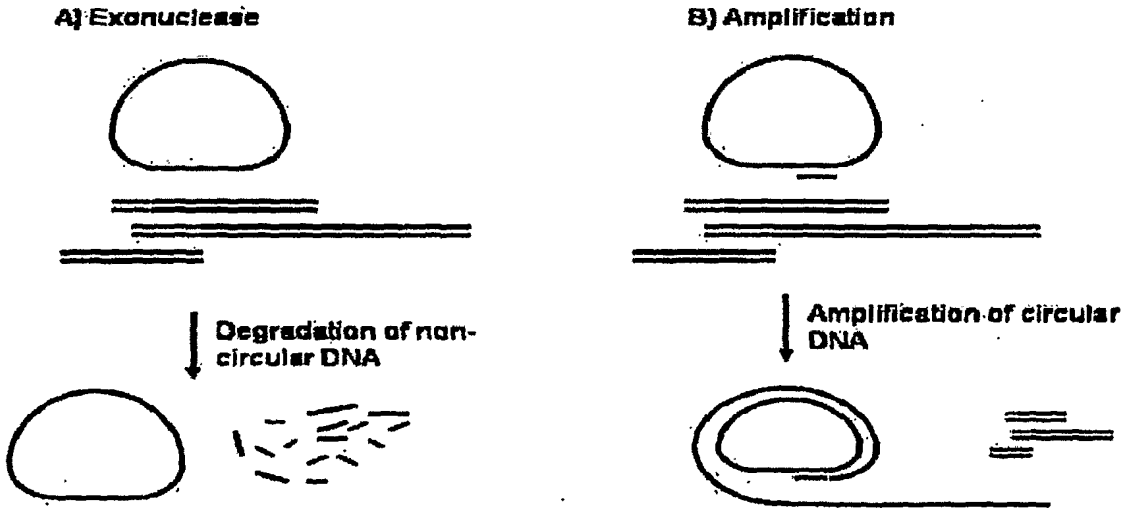


Figure 5

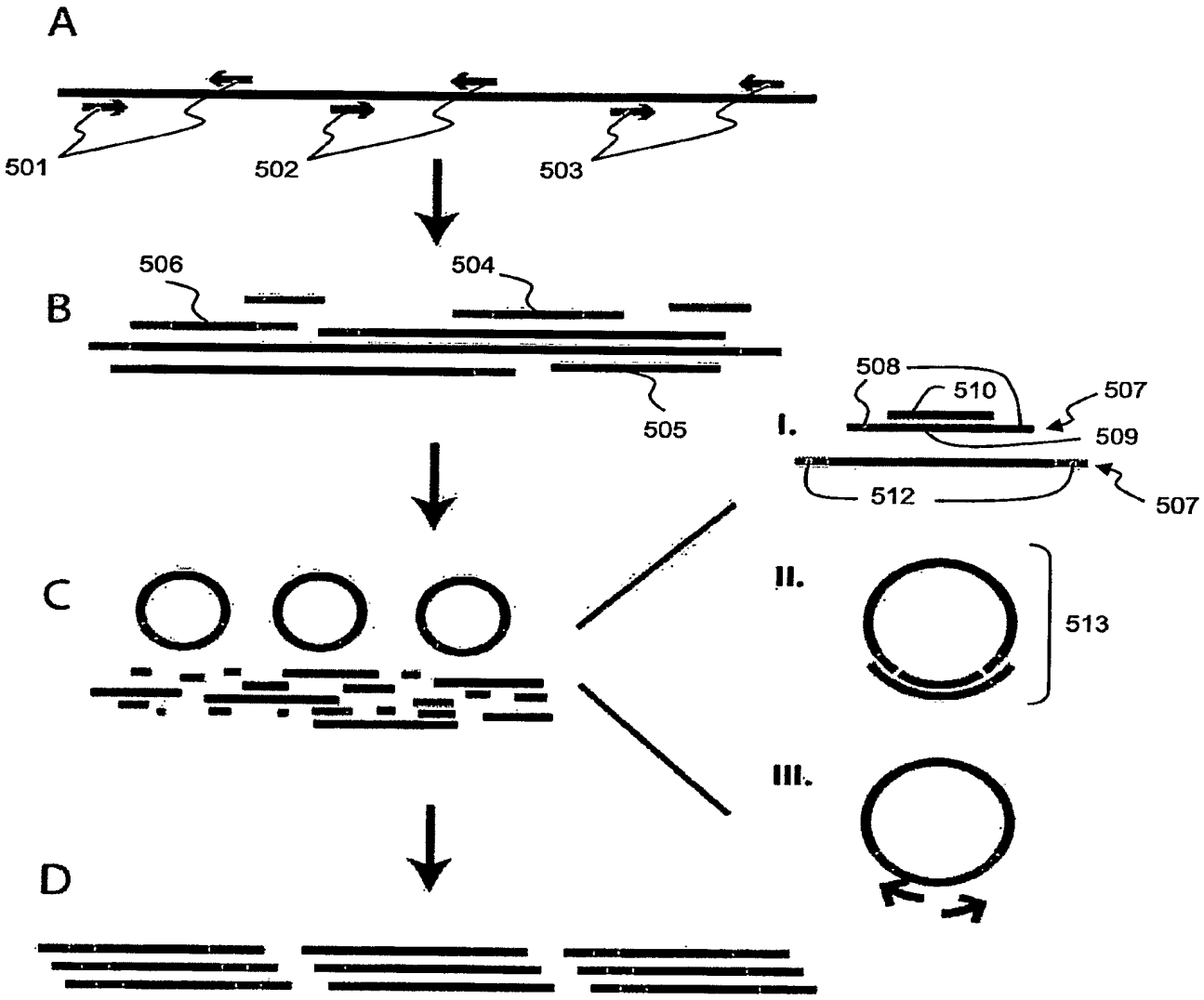


Figure 6

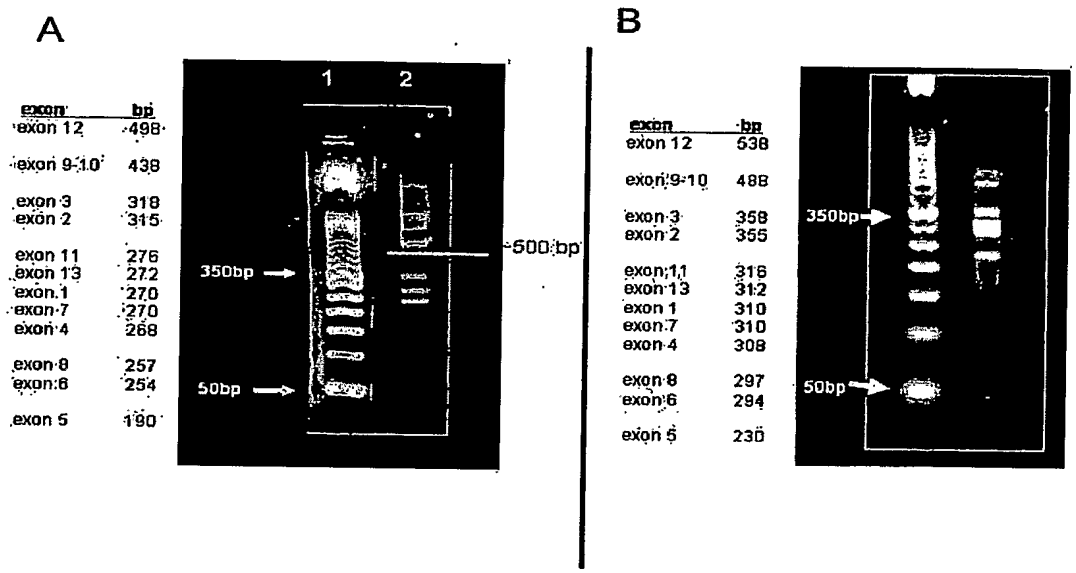


Figure 7

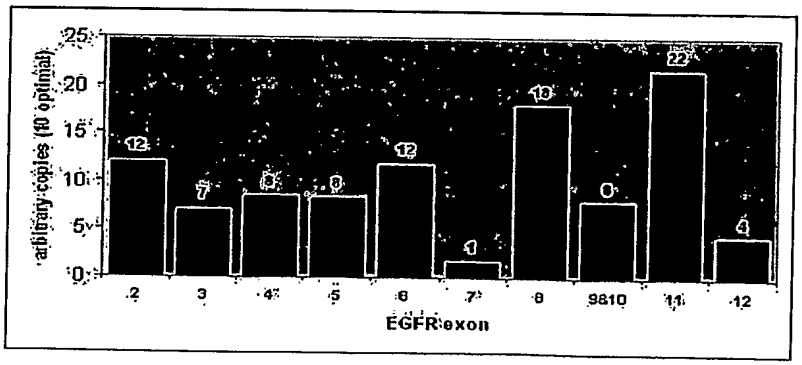


Figure 8

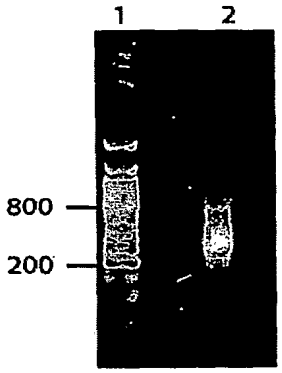


Figure 9

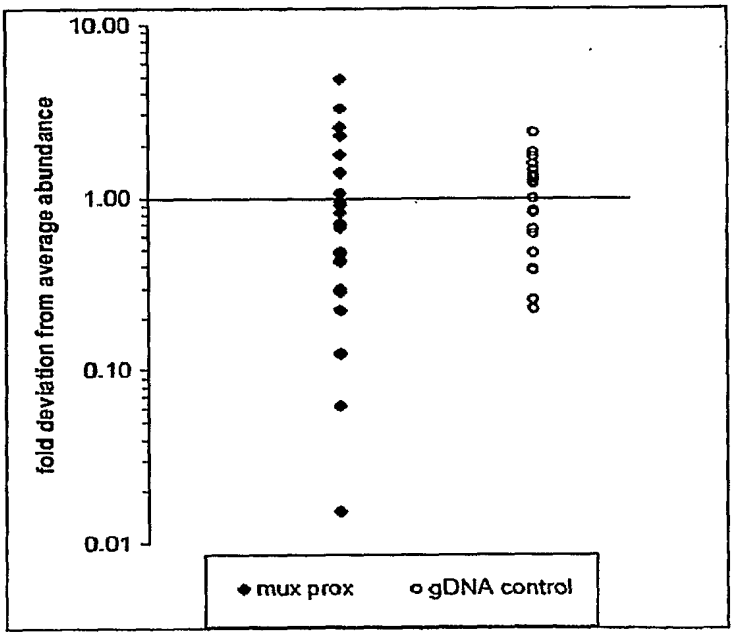


Figure 10

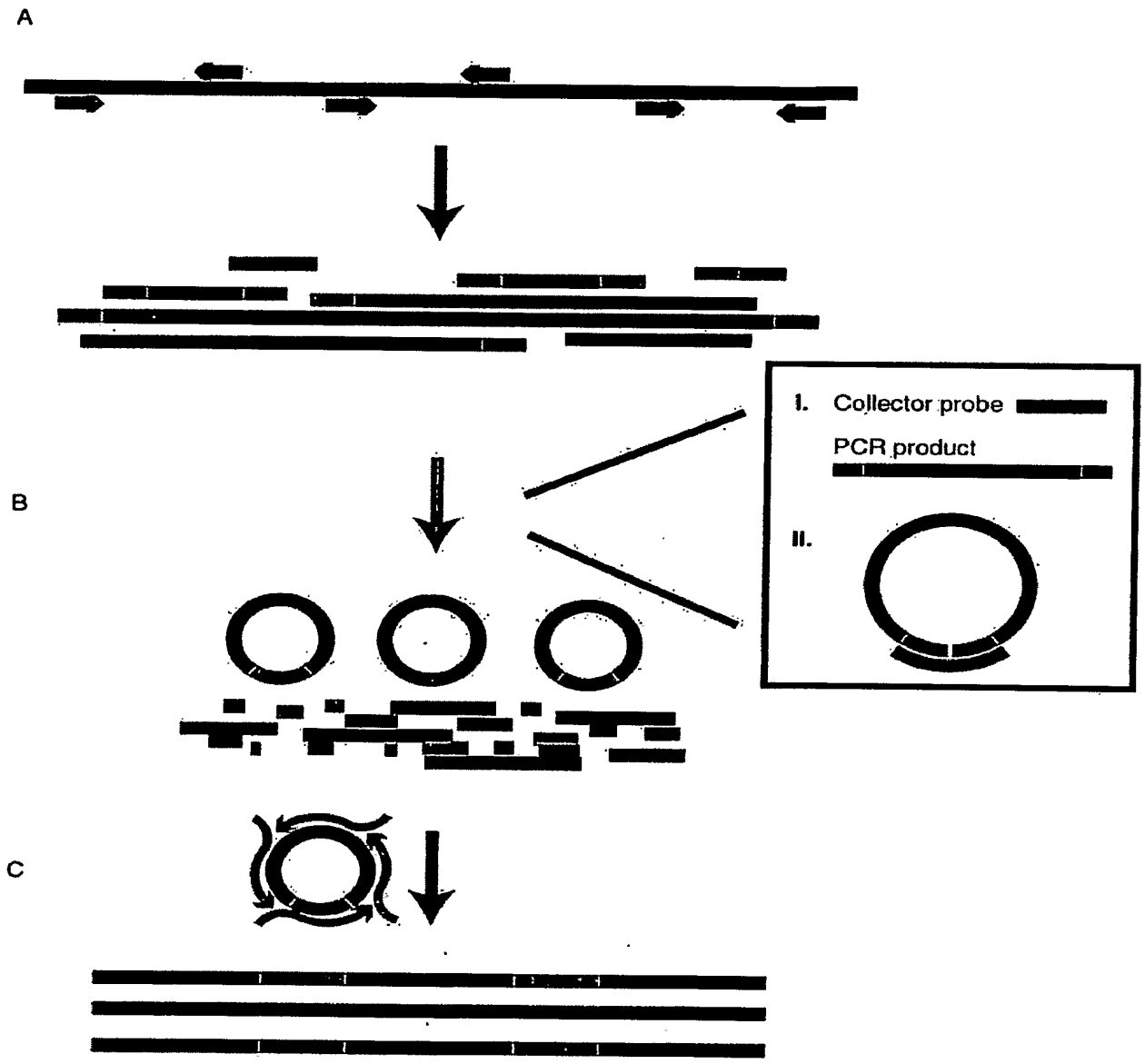


Figure 11

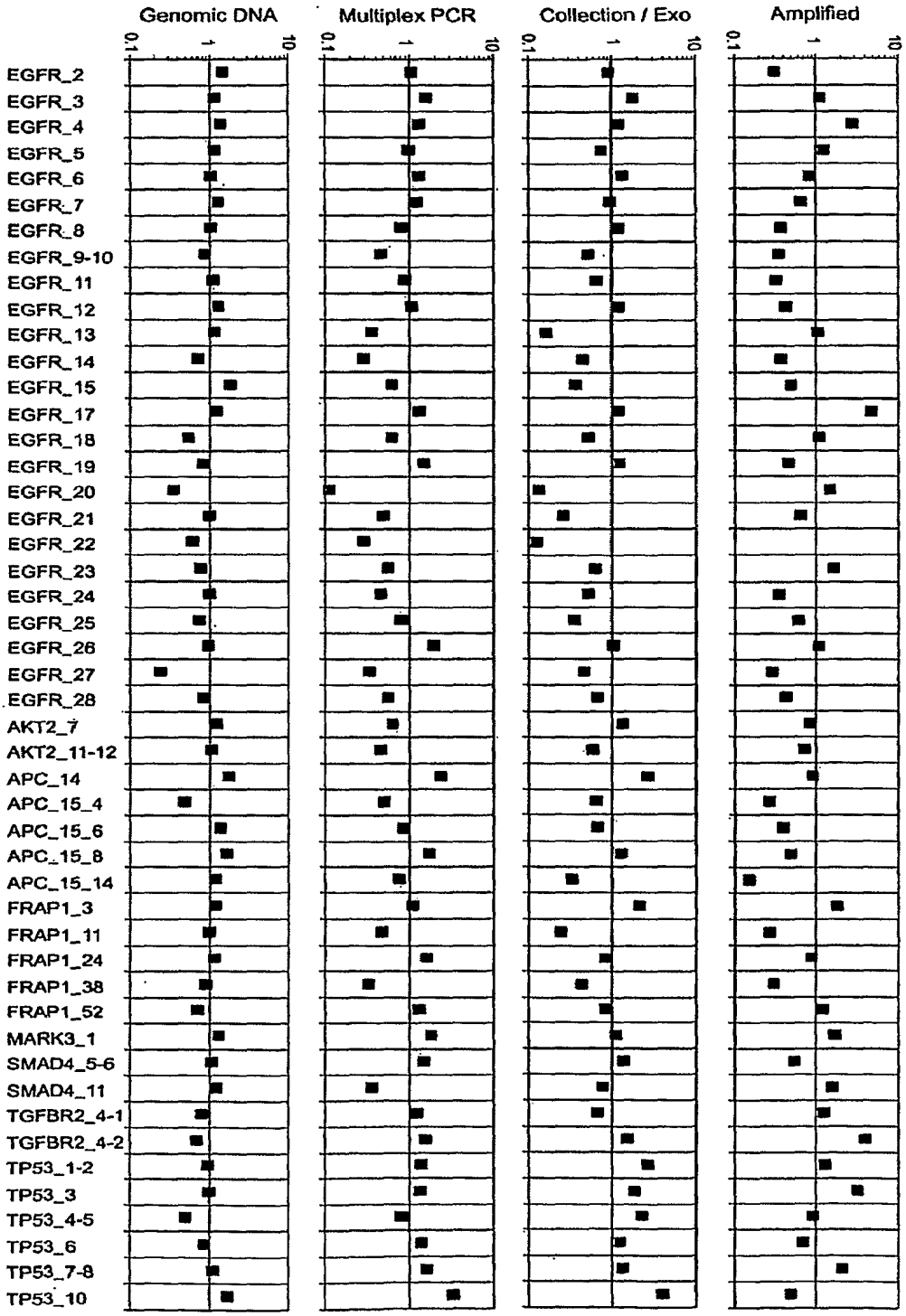

Genomic DNA
Multiplex PCR
Collection / Exo
Amplified
0.1
1
10
EGFR_2
EGFR_3
EGFR_4
EGFR_5
EGFR_6
EGFR_7
EGFR_8
EGFR_9-10
EGFR_11
EGFR_12
EGFR_13
EGFR_14
EGFR_15
EGFR_17
EGFR_18
EGFR_19
EGFR_20
EGFR_21
EGFR_22
EGFR_23
EGFR_24
EGFR_25
EGFR_26
EGFR_27
EGFR_28
AKT2_7
AKT2_11-12
APC_14
APC_15_4
APC_15_6
APC_15_8
APC_15_14
FRAP1_3
FRAP1_11
FRAP1_24
FRAP1_38
FRAP1_52
MARK3_1
SMAD4_5-6
SMAD4_11
TGFBR2_4-1
TGFBR2_4-2
TP53_1-2
TP53_3
TP53_4-5
TP53_6
TP53_7-8
TP53_10

Figure 12

Forward Primer	SEQ ID NO:	Reverse Primer	SEQ ID NO:	Gene-Collector probe	SEQ ID NO:
AKT-1					
TCAGCTTCCTTTGCTTCTCC	65	CAGACCCTGGGGCTACTACC	237	negative control (no probe)	
CGAGGGTGTGACGGGTAG	66	CCTGGTGGGCAAAGAGG	238	negative control (no probe)	
GTGGGTGGTATGCAAGGG	67	GAGGATGGCTACAGGCAGAG	239	CUCUGCCUGUAGCCAUCCUCUGUGGGUGGUAUGC AAGGG	409
GTGAAAGACGTGGGGTGG	68	CAGCCCTCCACAGTCCAAG	240	GUCGGGAGGUGUCAGGUUCGUGAAAGACGUGGG GUGG	410
CAGCTGGGCACTGTTGG	69	CATCGTCCCCTAGAGACAGC	241	GUAGCAGGGGAUCUCUGUCGCAGCUGGGCACUG UUGG	411
GTATCAAGCAGCGTGGTGTC	70	TCTGGTGCCATGGAGAGTAG	242	AGACCACGGUACCUCUCAUCGUAUCAAGCAGCGU GGUGUC	412
ACCCGGTCTGAGAAACCC	71	GCTCAGGACGTGGGGAC	243	CGAGUCCUGCACCCCUGACCCGGUCUGAGAAACC C	413
CAAGCACGTCACACCTCC	72	CTGGGGATGAGGGGATG	244	GACCCCUACUCCCCUACCAAGCACGUCACACCUC C	414
TACATCACAGGAGGAAGGGG	73	AGTGTGGATATGTGGGGAGC	245	UCACACCUAUACACCCCUCGUACAUCACAGGAGG AAGGGG	415
ACACTTGAGGGTGTGCTGG	74	TGAACACTGCAGGCCTCTC	246	ACUUGUGACGUCCGCAGAGACACUUGAGGGUGU GCUGG	416
GTGGGTGGGTGGAGGTG	75	AATACAGATCATGGCACGAGG	247	UUAUGUCUAGUACCGUGCUCCGUGGGUGGGUGG AGGUG	417
AKT-2					
ATGTCCTGCTGCCCTGAG	76	GCACAGCTTTCCAGGAGGAG	248	CUCCUCCUGGAAAGCUGUGCAUGUCCUGCUGCC CUGAG	418
ACCTGGGACATCACTCAACC	77	AAGTCCCACAAGCCCCTAAG	249	CUUAGGGGCUUGUGGGACUUACCUUGGCACAUCA CUCAACC	419
CTGCAGGCTAGCAGGGC	78	CCTTTTCTCCTCACACCAGG	250	CCUGGUGUGAGGAGAAAAGGCUGCAGGCUAGCA GGGC	420
ATTTGCTTGCCTGTTGCTG	79	AGGGCAGCCTTGTCTCTCAG	251	CUGAGAGACAAGGCUGCCCUAUUUGCUUGCCUG UUGCUG	421
GTTAAAATTTCCTCCCACGG	80	CTCAGGGTCAGGCTCCAG	252	CUGGAGCCUGACCCUGAGGUUAAAAUUUCCUCCC ACGG	422
GAGTGGCAGGTGTGGTG	81	TCTGAGCTCTGTCCAAAGGC	253	GCCUUUGGACAGAGCUCAGAGAGUGGGCAGGUG UGGUG	423
CAGCAGAGCCCTCCTCC	82	CATCTCACCCACAGCTCCTC	254	GAGGAGCUGUGGGUGAGAUGCAGCAGAGCCCUC CUCC	424
AAGGTGAGGGCAGGTGG	83	AACTTCCCCAGTGTGAGTCC	255	GGACUCACACUGGGGAAGUUAAGGUGAGGGCAG GUGG	425
TTCCCTGGAAGGAAAGGC	84	CAGGGACAGTGGCAGCAG	256	CUGCUGCCACUGUCCCCUGUUCCCCUGGAAGGAAA GGC	426

Figure 12 (cont.)

CTCCGAAAGCCCGTCTG	85	AAGAGCAGATCCCATCCCTC	257	GAGGGAUGGGAUCUGCUCUUCUCCGAAAGCCCG UCUG	427
GTTTCCCCAGGGAGTCTGG	86	ACCACCCAGCGGTGATG	258	CAUCACCGCUGGGUGGUGUUUCCCCAGGGAGUC UGG	428
APC					
TTTTGTTTCCTTTACCCCTTTC	87	GCCATTGGAGTTTTACACTTATTTTC	259	GAAAAUAAGUGUAAAACUCCAAUGGCUUUUGUUU CCUUUACCCCUUUC	429
TAAGGTGCGTGCTTTGAGAG	88	ACCAACACCCAAATCGAGAG	260	CUCUCGAUUUGGGUGUUGGUUAAGGUGCGUGCU UUGAGAG	430
TGTTTTCAGTCATGTATATTTGTGG	89	GCAATCAATAACATGCTATCTTTGAG	261	CUCAAAGAUAGCAUGUUAUUGAUUGCUGUUUUCA GUCAUGUAUAUUUGUGG	431
TGCTCTTCTGCAGTCTTTATTAGC	90	CAGGCCTAAAGTTGGGTAAAAC	262	GUUUUACCCAACUUUAGGCCUGUGCUCUUCUGCA GUCUUUAUUAGC	432
TCATGCACCATGACTGACG	91	AGAGCCAAAATAAACACAGCC	263	GGCUGUGUUUAUUUUGGCUCUUCAUGCACCAUG ACUGACG	433
TGCGGTGAGCTGAGATTATG	92	GGGGTTTCTGGAGTAAACACAG	264	CUGUGUUUACUCCAGAAACCCCUGCGGUGAGCU GAGAUUAUG	434
CCTTGGGCTAAGAAAGCCTAC	93	TTCTTAGAACCATCTTGCTTCATAC	265	GUAUGAAGCAAGAUGGUUCUAAGAACCUUGGGCU AAGAAAGCCUAC	435
TGCTCATATGCAAGAAACTCTC	94	AGAGATGGGGTTTTGCCAC	266	GUGGCAAAACCCCAUCUCUUGCUCAUAUGCAAGA AACUCUC	436
CACTGATTACTTCATCCTGGAAAG	95	TGGCTGATATGAATTTTCTCCTC	267	GAGGAGAAAAUUCAUAUCAGCCACACUGAUUACU UCAUCCUGGAAAG	437
AGGGTTATATTAGTGATCCCTGC	96	TTGGTGGCCTTATATCCTAATTC	268	GAAUUAGGAUAUAAGGCCACCAAAGGGUUAUAUU AGUGAUCCCUGC	438
TGGCATAAAATGGAATAATTGTC	97	AGCGAATGTGAAGCACAGG	269	CCUGUGCUUCACAUUCGCUUGGCAUAAAAUGGAA UAAUUGUC	439
AAAGCTTGGCTTCAAGTTGTC	98	AGTGAGACCCTGCCTCAAAG	270	CUUUGAGGCAGGGUCUCACUAAAGCUUGGCUUCA AGUUGUC	440
TGTTACCCAGAAGGTCTTGAAC	99	GAAATCTCATGGCTAAAAGAAGG	271	CCUUCUUUUAGCCAUGAGAUUUCUGUUACCCAGA AGGUCUUGAAC	441
AGTGAGGGACGGGCAATAG	100	CCTCCACCTATGGGCTACAC	272	GUGUAGCCCAUAGGUGGAGGAGUGAGGGACGGG CAAUAG	442
TTTGTTGTTACTGCATACACATTG	101	GGGTAACACTGTAGTATTCAAATATGG	273	CCAUAUUUGAAUACUACAGUGUUACCCUUUGUUG UUACUGCAUACACAUUG	443
AAGCAGAGACACAAGCAAAGTC	102	CAATCGAGGGTTTCATTTGAC	274	GUCAAAUGAAACCCUCGAUUGAAGCAGAGACACA AGCAAAGUC	444
AAGTCGGAAAATTCAAATAGGAC	103	TTCAGAGTAACGTTCACTATAATTGG	275	CCAAUUAUAGUGAACGUUACUCUGAAAAGUCGGA AAAUUCAAAUAGGAC	445
TTCTCCATACAGGTCACGGG	104	CCTGTGTCGTCTGATTACATCC	276	GGAUGUAAUCAGACGACACAGGUUCUCCAUACAG GUCACGGG	446
TGCAAAGTTTCTTCTATTAACCAAG	105	CGCTTGGTTTGAGCTGTTTG	277	CAAACAGCUCAAACCAAGCGUGCAAAGUUUCUUC UAUUAACCAAG	447
AGCCCCAGTGATCTTCCAG	106	AGCTGACTTGGTTTCCTTGC	278	GCAAGGAAACCAAGUCAGCUAGCCCCAGUGAUCU UCCAG	448

IDT Exhibit 1020.0767

Figure 12 (cont.)

AAAAGGACCTATTAGATGATTCAGATG	107	TTTGTTGGGTGCAGAAGAAG	279	CUUCUUCUGCACCCAACAAAAAAAGGACCUAUUA GAUGAUUCAGAUG	449
ATTCTGCTATGCCCAAAGGG	108	GCTGACCTCGATTTATTGGC	280	GCCAAUAAAUCGAGGUCAGCAUUCUGCUAUGCCC AAAGGG	450
AATAAGGAATCAGAGGCTAAAGTTAC	109	AAGTGTCAGATCTTCACCTAATATGC	281	GCAUAUUAGGUGAAGAUCUGACACUUAAUAAGGA AUCAGAGGCUAAAGUUAC	451
GATGACCTGTTGCAGGAATG	110	GAATCATTGTCCTGCCTCG	282	CGAGGCAGGACAAUGAUUCGAUGACCUGUUGCAG GAAUG	452
AAAGTTTGATTACTGGAAAAGTTCG	111	TCCTTTGGAGGCAGACTCAC	283	GUGAGUCUGCCUCCAAAGGAAAAGUUUGAUUACU GGAAAAGUUCG	453
TCAACTAAGTCCTCAGGTTCTGG	112	ACGTTTCCAGGTTCCTGACC	284	GGUCAGGAACCUGGAAACGUUCAACUAAGUCCUC AGGUUCUGG	454
CTCCCACCTAATCTCAGTCCC	113	ATGGGAACACTGCCATTACC	285	GGUAAUGGCAGUGUUCCCAUCUCCCACCUAAUCU CAGUCCC	455
TTATCAAATGGCACCTGCTG	114	AAACCCTCTAACAAGAATCAAACC	286	GGUUUGAUUCUUGUUAGAGGGUUUUUAUCAAAUG GCACCUGCUG	456
EGFR					
CCCCTGACTCCGTCCAG	115	AAACAGGAAAGGACGGGC	2	GCCCGUCCUUUCCUGUUUCCCCUGACUCCGUCC AG	457
ACCTGGACCTTGAGGGATTG	3	CTTCAAGTGGAATTCTGCCC	4	GGGCAGAAUUCCACUUGAAGACCUGGACCUUGAG GGAUUG	458
CCCTGGACCCATTTTAGACC	5	CCATCGGAACTGCTGTCTG	6	CAGACAGCAGUUCCGAUGGCCCUGGACCCAUUUU AGACC	459
AGCACATGCATCCTTCATGG	7	AGTGCTGTAGAGCTGTCCCC	8	GGGGACAGCUCUACAGCACUAGCACAUGCAUCCU UCAUGG	460
GAAAGGGCGTCATCAGTTTC	9	CAAGTGAAGGAAGAGAGGGG	10	CCCCUCUCUUCCUUCACUUGGAAAGGGCGUCAUC AGUUUC	461
CCCTGGGAAATGATCCTACC	11	GTCTTCTGTCCTGGTGTGGG	12	CCCACACCAGGACAGAAGACCCCUGGGAAAUGAU CCUACC	462
CGCTTCCTCCGTGTGTG	13	AGGAGACAGAGCGGGACAAG	14	CUUGUCCCGCUCUGUCUCCUCGCUUCCUCCGUG UGUG	463
CTCAAGAGGACCTGGACCG	15	GAGCCCAGCCTCAGCAG	16	CUGCUGAGGCUGGGCUCCUCAAGAGGACCUGGA CCG	464
GGATCCCTAGCTATTCTTAATCCAAC	17	GGAAATATGTCGAAAAGTTCTCTCTC	18	GAGAGAGAACUUUUCGACAUAUUUCCGGAUCCCU AGCUAUUCUUAAUCCAAC	465
CAGAGTCCCTGAGAGTCTAGAGTAATG	19	GAGCTCTGTGCCCTATCTTAGC	20	GCUAAGAUAGGGCACAGAGCUCCAGAGUCCCUGA GAGUCUAGAGUAAUG	466
CTCCCACAGCATGACCTACC	21	GGAATTCACATGGTAATTTCACAG	22	CUGUGAAAUUACCAUGUGAAUUCCCUCCCACAGC AUGACCUACC	467
AAGGTGCCGTCTCCTCC	23	GCTATAACAACAACCTGGAGCC	24	GGCUCCAGGUUGUUGUUAUAGCAAGGUGCCGUC UCCUCC	468
AGAGGTGATTTGTGTTCCTGC	127	TCATCACTGTTCGGCTTCTG	299	CAGAAGCCGAACAGUGAUGAAGAGGUGAUUUGUG UUCCUGC	469
TATCATTTGGCTTTCCCCAC	128	ATTTGTTGCCGGAAAACTTG	300	CAAGUUUUCCGGCAACAAAUUAUCAUUUGGCUUU CCCCAC	470

IDT Exhibit 1020.0768

Figure 12 (cont.)

TCTCCAAAATATATGCCAAAGAAG	129	CCACAGCAGTGTGGTCATTC	301	GAAUGACCACACUGCUGUGGUCUCCAAAAUAUAU GCCAAAGAAG	471
GGCCATGGAATCTGTCAGC	130	AACTGCTAATGGCCCGTTC	302	GAACGGGCCAUUAGCAGUUGGCCAUGGAAUCUG UCAGC	472
TTTCCAGCATGGTGAGGG	131	ACAGCTTGCAAGGACTCTGG	303	CCAGAGUCCUUGCAAGCUGUUUUCCAGCAUGGU GAGGG	473
AGCATGTGGCACCATCTCAC	132	AGACATGAGAAAAGGTGGGC	304	GCCCACCUUUUCUCAUGUCUAGCAUGUGGCACCA UCUCAC	474
CATGTGCCCCTCCTTCTG	133	CTATCCCAGGAGCGCAGAC	305	GUCUGCGCUCCUGGGAUAGCAUGUGCCCCUCCU UCUG	475
AATTCGGATGCAGAGCTTC	134	TACAGCTAGTGGGAAGGCAG	306	CUGCCUUCCCACUAGCUGUAAAUUCGGAUGCAGA GCUUC	476
TCGTCTGTGTGTGTCACTCG	135	AAAGAAAATACTTGCATGTCAGAGG	307	CCUCUGACAUGCAAGUAUUUUCUUUUCGUCUGUG UGUGUCACUCG	477
GCAAGGGATTGTGATTGTTC	136	CTGTTTGGCTAAGAGCAGCC	308	GGCUGCUCUUAGCCAAACAGGCAAGGGAUUGUGA UUGUUC	478
GCAATGCCATCTTTATCATTTC	137	CAGAACACAGTGACATGAGATGC	309	GCAUCUCAUGUCACUGUGUUCUGGCAAUGCCAUC UUUAUCAUUUC	479
AGACCCCTGCTCCTATAGCC	138	GAGTTTCACTAGATGGTTATTTTCCC	310	GGGAAAAUAACCAUCUAGUGAAACUCAGACCCCU GCUCCUAUAGCC	480
CCTGCATTCAGGAAAAGTGG	139	TCCATGTAAAATAGAGCCATAGTG	311	CACUAUGGCUCUAUUUUACAUGGACCUGCAUUCA GGAAAAGUGG	481
TTGCAAACACTGAAGTTGGG	140	TTGGACACTGGAGACTGGC	312	GCCAGUCUCCAGUGUCCAAUUGCAAACACUGAAG UUGGG	482
ATCCTGCATGGGATGGTG	141	GTGGCTTGGTCCTGGGTATC	313	GAUACCCAGGACCAAGCCACAUCCUGCAUGGGAU GGUG	483
FRAP					
AGATCTGTGTTGGGTGACCAG	142	TGATGGTACATTTAGCCCACAC	314	GUGUGGGCUAAAUGUACCAUCAAGAUCUGUGUUG GGUGACCAG	484
CACTCCCACCACCACAGTTAG	143	TAAGTGGCAGACACAGGGTG	315	CACCCUGUGUCUGCCACUUACACUCCCACCACCA CAGUUAG	485
AGTCTGCGTACTTTCCTGGC	144	AGAAGGAAGCAAAAGACCCC	316	GGGGUCUUUUGCUUCCUUCUAGUCUGCGUACUU UCCUGGC	486
TCCTGAGAGAACAAAACTCTGG	145	AGGCCAGGTGATTCTCTACG	317	CGUAGAGAAUCACCUGGCCUUCCUGAGAGAACAA AACUCUGG	487
ATGGGATGGGCCTGTATTC	146	CGCTCACAGAATGGTACACG	318	CGUGUACCAUUCUGUGAGCGAUGGGAUGGGCCU GUAUUC	488
GTAGAATCCACAGTGCCCAG	147	CCTGTTCCCTGTTTACCCTG	319	CAGGGUAAACAGGGAACAGGGUAGAAUCCACAGU GCCCAG	489
GCTTTATAGGACCTTTGACTGTTG	148	AGCGTCCTTCCTCTCCAACC	320	GGUUGGAGAGGAAGGACGCUGCUUUAUAGGACC UUUGACUGUUG	490
CAAGGTGATTTTGAGGTGGG	149	TAACAGTCCCAACACTGGGG	321	CCCCAGUGUUGGGACUGUUACAAGGUGAUUUUG AGGUGGG	491
GAGAAAGCTGTTTCTTCCCAAG	150	TTCCCAAAGTTTCCAGCATC	322	GAUGCUGGAAACUUUGGGAAGAGAAAGCUGUUUC UUCCCAAG	492

IDT Exhibit 1020.0769

Figure 12 (cont.)

CAAGCTGCAGACCTACCTGTC	151	CTGGCTACTCCCAATTGTCC	323	GGACAAUUGGGAGUAGCCAGCAAGCUGCAGACCU ACCUGUC	493
TGGAGGTAGAAGCTGGGAAG	152	TCAAGCAGTTCTCACAGCG	324	CGCUGUGAGAACUGCUUGAUGGAGGUAGAAGCU GGGAAG	494
TTTTCTGTTGGACACAGTCTCTC	153	TCCCCATATGAGCTGATGAC	325	GUCAUCAGCUCAUAUGGGGAUUUUCUGUUGGACA CAGUCUCUC	495
TTATGACTGAGATGGTCTCTTGG	154	GAGGTTAGTTTCTTTTCCACCC	326	GGGUGGAAAAGAAACUAACCUCUUAUGACUGAGA UGGUCUCUUGG	496
CCCGGCCTCCAAAATAC	155	AAAGAGAAGCAAAGTCCCTGG	327	CCAGGGACUUUGCUUCUCUUUCCCGGCCUCCAAA AUAC	497
GAAACCTCTTTTCTTCTTACAGCC	156	GCAACCATCTCTCTCTTGCC	328	GGCAAGAGAGAGAUGGUUGCGAAACCUCUUUUCU UCUUACAGCC	498
GCAAGGCTCTGTGAGTGAG	157	GCAAGGGCTCTGTGAGTGAG	329	CUCACUCACAGAGCCCUUGCGCAAGGGCUCUGU GAGUGAG	499
TAACCATGCTCTGCCTGAAG	158	GATCTGTGCATGTGTGGTGC	330	GCACCACACAUGCACAGAUCUAACCAUGCUCUGC CUGAAG	500
AGCTTAAGGTAAGCCTGGGG	159	TTCAAACTTGCTTCTGAGCC	331	GGCUCAGAAGCAAGUUUGAAAGCUUAAGGUAAGC CUGGGG	501
CTTTTCTGGCCATCTTGATTC	160	CCAGGGACTCAGAGGAAATC	332	GAUUUCCUCUGAGUCCCUGGCUUUUCUGGCCAU CUUGAUUC	502
TGTGCTTAGATACTTGGGACCTG	160	GCCTGGGAACTTAAGAAATGAAC	333	GUUCAUUUCUUUAAGUUCCCAGGCUGUCCUUAGAU ACUUGGGACCUG	503
CCCTCAGTGACGTCTCTTCC	162	TGGGTCACAGACAAAGTCTTC	334	GAAGACUUUGUCUGUGAGCCACCCUCAGUCACGU CUCUUCC	504
CCCAGAAGAAATGGTCATAATGTAG	163	CACAAACTCCCATAGCCAAG	335	CUUGGCUAUGGGAGUUUGUGCCCAGAAGAAAUG GUCAUAAUGUAG	505
TCTACCAGGTCTGCACCTCC	164	CTGGTTTTCGGTTGCCC	336	GGGCAACCGAAAACCAGUCUACCAGGUCUGCACC UCC	506
AGGGGAATTTTATCGTGAAAG	165	CCAGCCCCTTGATTATTACTTC	337	GAAGUAAUAAUCAAGGGGCUGGAGGGGAAUUUUA UCGUGAAAG	507
AGGTGTGACTTTGAGGCAGC	166	GCAGTGGGAGAAGAGAGGTC	338	GACCUCUCUUCUCCCACUGCAGGUGUGACUUUGA GGCAGC	508
CCCACATCTCCTACCCTCAG	167	TGAAGGTCAGGGCCAATAAC	339	GUUAUUGGGCCCUGACCUUCACCCACAUCUCCUAC CCUCAG	509
GCGTACAGCAGCACATTAGG	168	AAGTGAGAACTCCGTGTGGG	340	CCCACACGGAGUUCUCACUUGCGUACAGCAGCAC AUUAGG	510
GGGCAGGAGAGGAAGATTG	169	GAGCCCCTAGCCTCACTCAC	341	GUGAGUGAGGCUAGGGGCUCGGGCAGGAGAGGA AGAUUG	511
CTGCCTTTAGCCCAACCAG	170	TCTTTACCAAAGCACCGTGG	342	CCACGGUGCUUUGGUAAAGACUGCCUUUAGCCCA ACCAG	512
CTGTAGCTAGTTGGGGTGCC	171	GTCCTGGAAGGGGTAGGG	343	CCCUACCCCUUCCAGGACCUGUAGCUAGUUGGG GUGCC	513
ACAGAGCAAGACTCTGCCAC	172	GCAGAGCGAAGCAGATTAGG	344	CCUAAUCUGCUUCGCUCUGCACAGAGCAAGACUC UGCCAC	514
GAACGGTAGCTCCCTTCCTC	173	GCAGAAGCTGCTGGGATG	345	CAUCCCAGCAGCUUCUGCGAACGGUAGCUCCCUU CCUC	515

Figure 12 (cont.)

TTAGGAGGGCTGTTTTGAGG	174	TCAGCTGTAACCACGAGCAC	346	GUGCUCGUGGUUACAGCUGAUUAGGAGGGCUGU UUUGAGG	516
GTTCCTCAGCATCGACCTTG	175	CTCAGAGAGCCTGGCACC	347	GGUGCCAGGCUCUCUGAGGUUCCUCAGCAUCGA CCUUG	517
CCACCCTTTGAAGTAGGTACAG	176	AGGGTCAGGAAGGGAAAGAG	348	CUCUUUCCCUUCCUGACCCUCCACCCUUUGAAGU AGGUACAG	518
CTGCCAGCTCTCTTCTCAGG	177	AGCTTTTGGAAAGGCTGACC	349	GGUCAGCCUUUCCAAAAGCUCUGCCAGCUCUCUU CUCAGG	519
ATACCAGCTCTTCCCCAACC	178	CCCACCAGCTAAGGGACC	350	GGUCCCUUAGCUGGUGGGAUACCAGCUCUUCCC CAACC	520
TTCCTCTGACTGCTGGAAATAG	179	TGGGTCACGTCCTTTCATTC	351	GAAUGAAAGGACGUGACCCAUUCCUCUGACUGCU GGAAAUAG	521
ATATCTGCAGTCAGCCTGGG	180	CCAGATGCTTTGGAATGAGTG	352	CACUCAUUCCAAAGCAUCUGGAUAUCUGCAGUCA GCCUGGG	522
GCGTTTAAATTCTTCCCTGG	181	GAAGAGGGAAGGGGTCTCAG	353	CUGAGACCCCUUCCCUCUUCGCGUUUAAAUUCUU CCCUGG	523
CAGCTGAAGAGCTGAGGACC	182	ATGACTACACGAGACAAATGTAGG	354	CCUACAUUUGUCUCGUGUAGUCAUCAGCUGAAGA GCUGAGGACC	524
TCTGCCTGTGTTCTGAGCTG	183	AGCAGCGCTACGGAGATTC	355	GAAUCUCCGUAGCGCUGCUUCUGCCUGUGUUCU GAGCUG	525
ACCCCAAGCCTTGTTTCTTC	184	CATCCTATTGCGAGTGGGG	356	CCCCACUCGCAAUAGGAUGACCCCAAGCCUUGUU UCUUC	526
CTCGGTCTCAAAAGTACAAACC	185	AAACCAAATGAAACCATTCAGG	357	CCUGAAUGGUUUCAUUUGGUUUCUCGGUCUCAAA AGUACAAACC	527
CGTTTGCCAACTCCTAGCTTAC	186	TGGGCTTAAGTCTGCCTACAG	358	CUGUAGGCAGACUUAAGCCCACGUUUUGCCAACUC CUAGCUUAC	528
AGAGGAAAGCCACCTGCTC	187	AAGGCTGTAACTGTCCTGATCC	359	GGAUCAGGACAGUUACAGCCUUAGAGGAAAGCCA CCUGCUC	529
ACATGGCCTGTGTCTGCTTC	188	CAAGCAACTCCTCTGCCTTG	360	CAAGGCAGAGGAGUUGCUUGACAUGGCCUGUGU CUGCUUC	530
GACTGGAAGAAAATAACCAAGTTTC	189	ACAAACACTCTGCACAAGGG	361	CCCUUGUGCAGAGUGUUUGUGACUGGAAGAAAAU AACCAAGUUUC	531
ATGCTAACACCAACAGTGGC	190	TCAGAGAGGAAAGTGTGCTCAG	362	CUGAGCACACUUUCCUCUCUGAAUGCUAACACCA ACAGUGGC	532
AGGGAACAAGAAGTGCATCG	191	TTTGCTACTCTGGCTTTGGG	363	CCCAAAGCCAGAGUAGCAAAAGGGAACAAGAAGU GCAUCG	533
KRas					
TTAAAAGGTACTGGTGGAGTATTTG	192	CCTTTATCTGTATCAAAGAATGGTC	364	GACCAUUCUUUGAUACAGAUAAAGGUUAAAAGGU ACUGGUGGAGUAUUUG	534
TCTTTGGAGCAGGAACAATG	193	TGCATGGCATTAGCAAAGAC	385	GUCUUUGCUAAUGCCAUGCAUCUUUGGAGCACGA ACAAUG	535
AGAAGGAAGGAAAATTTGGTG	194	AGAAGCAATGCCCTCTCAAG	366	CUUGAGAGGGCAUUGCUUCUAGAAGGAAGGAAAA UUUGGUG	536
AACTTCTTGCACATGGCTTTC	195	GTGGTTGCCACCTTGTTACC	367	GGUAACAAGGUGGCAACCACAACUUCUUGCACAU GGCUUUC	537

Figure 12 (cont.)

MARK-3					
CCTAGGGCTGTGCTGTTTTG	196	GAGGAGCACGAATGCCAC	368	GUGGCAUUCGUGCUCCUCCCUAGGGCUGUGCUG UUUUG	538
TGGTTAAATTCCTTTGAAGTGC	197	ATGGTGTGAACCCGGGAG	369	CUCCCGGGUUCACACCAUUGGUUAAAUUCCUUUG AAGUGC	539
TGCCATATATCTTGGCATTTATC	198	TTAGATGGCAATTAGCACCG	370	CGGUGCUAAUUGCCAUCUAAUGCCAUAUAUCUUG GCAUUUAUC	540
AAATGGAAGCATTTGGGAATAC	199	AAGCTTTTAACCTGTCTTCAGC	371	GCUGAAGACAGGUUAAAAGCUUAAAUGGAAGCAU UUGGGAAUAC	541
TGCATGGTTTGTGCATACAG	200	CCATATGGGCAGTATGGTTG	372	CAACCAUACUGCCCAUAUGGUGCAUGGUUUGUGC AUACAG	542
CAAATTCCACATATTTCTGGCTAAC	201	AAACACTTGTTTGCTTCAGGG	373	CCCUGAAGCAAACAAGUGUUUCAAAUUCCACAUA UUUCUGGCUAAC	543
ATTTTCATCTTAATTACGAATCTGC	202	TGACCACTGAGAGAACTGCAC	374	GUGCAGUUCUCUCAGUGGUCAAUUUUCAUCUUAA UUACGAAUCUGC	544
AAAGCTTTTCTAAAATGCCTAATCC	203	CAGAGGAAAGAAAACGAGTAAGTC	375	GACUUACUCGUUUUCUUUCCUCUGAAAGCUUUUC UAAAAUGCCUAAUCC	545
AACATTATATCAGTGCGGGG	204	CCATTATTCACATTTTAGGCACAG	376	CUGUGCCUAAAAUGUGAAUAAUGGAACAUUAUAU CAGUGCGGGG	546
TTGCTTGAATCTGGGAGGTG	205	GCCAATATCCATCATTAAGGG	377	CCCUUAAUGAUGGAUAUUGGCUUGCUUGAAUCUG GGAGGUG	547
ATAAGCCATTTGGGTTCGTG	206	AGGAGACCTCTTGCCTGGAC	378	GUCCAGGCAAGAGGUCUCCUAUAAGCCAUUUGGG UUCGUG	548
TGGTTACTGTCACAAAATAAAACTTG	207	CCTGGTGTCTAGCTGGTTCC	379	GGAACCAGCUAGACACCAGGUGGUUACUGUCACA AAAUAAAACUUG	549
GGTGACAGAGTAAGACCTTGCC	208	GCAGTAAATAACATGCTTCCATTTC	380	GAAAUGGAAGCAUGUUAUUUACUGCGGUGACAGA GUAAGACCUUGCC	550
TTTATACCCGATTTTCTCCACTG	209	CCAGGTAGCAGTGGCTTCAC	381	GUGAAGCCACUGCUACCUGGUUUAUACCCGAUUU UCUCCACUG	551
CATTTTATGGTGTTGGTGTTGG	210	AAAATAGTCGCCTCTGCCAC	382	GUGGCAGAGGCGACUAUUUUCAUUUUAUGGUGU UGGUGUUGG	552
TGAAGTGTAAGAGGTTGATTTTCC	211	ACATTTCCATCAGTGTTCAGG	383	CCUGAACACUGAUGGAAAUGUUGAAGUGUAAGAG GUUGAUUUUCC	553
SMAD-4					
CAAAGGATCAAAATTGCTTCAG	212	GTAGCTTGAAAGGAAACGTAGC	384	GCUACGUUUCCUUUCAAGCUACCAAAGGAUCAAA AUUGCUUCAG	554
CTGAGTTGGTAGGATTGTGAGG	213	GTCGCGGGCTATCTTCC	385	GGAAGAUAGCCCGCGACCUGAGUUGGUAGGAUU GUGAGG	555
GCGTTTATGCTACTTCTGAATTG	214	TTAATGTTACTGCCTGCCGC	386	GCGGCAGGCAGUAACAUUAAGCGUUUAUGCUACU UCUGAAUUG	116
CCGCTGAATAAATGACTTTTGC	215	TTCCAAGTGATTGTGCATACC	387	GGUAUGCACAAUCACUUGGAACCGCUGAAUAAAU GACUUUUGC	117
CCCATCTTTATAGTTGTGCATTATC	216	AAAACAGAAAACAAAGCCCTACC	388	GGUAGGGCUUUGUUUUCUGUUUUCCCAUCUUUA UAGUUGUGCAUUAUC	118

IDT Exhibit 1020.0772

Figure 12 (cont.)

TTGGCAGATAGCACTGAAATG	217	TTAAAGCCTGTGTTTGTGCG	389	CGCACAAACACAGGCUUUAAUUGGCAGAUAGCAC UGAAAUG	119
TGTGGAGTGCAAGTGAAAGC	218	TGTACATGGGAAAACATAACCTTG	390	CAAGGUUAUGUUUUCCCAUGUACAUGUGGAGUGC AAGUGAAAGC	120
GAATTCATACTACATGCTCCTGACAC	219	TTTTCCATTCCTTCCACCC	391	GGGUGGAAGGAAUGGAAAAGAAUUCAUACUACAU GCUCCUGACAC	121
TCCAAGCCACCTTTCCTAAC	220	ACAGAAGAACAGATTTTACCAATTC	392	GAAUUGGUAAAAUCUGUUCUUCUGUUCCAAGCCA CCUUUCCUAAC	122
AGGATGGGAAGAGATCACCC	221	CAGGATTGTATTTTGTAGTCCACC	393	GGUGGACUACAAAAUACAAUCCUGAGGAUGGGAA GAGAUCACCC	123
TGF-beta R2					
CATCTGGCCCGCACATC	222	GAAACTTTCCTCGTTTCCGC	394	GCGGAAACGAGGAAAGUUUCCAUCUGGCCCGCAC AUC	124
TTCAGGAATTCATTGGCAGG	223	GGAAAGGGAAATGGAACAGG	395	CCUGUUCCAUUUCCCUUUCCUUCAGGAAUUCAUU GGCAGG	125
TTCCAGATTGCCTTTCTGTC	224	ATGCAATCCACCACAGGAG	396	CUCCUGUGGUGGAUUGCAUUUCCAGAUUGCCUU UCUGUC	126
GCATGAACCCACTTCCTGAC	225	CATGCTTCAGATTGATGTCTGAG	397	CUCAGACAUCAAUCUGAAGCAUGGCAUGAACCCA CUUCCUGAC	287
AGAGCAGTTTGAGACAGTGGC	226	AAGAGGTAGGGTGAGGCCAG	398	CUGGCCUCACCCUACCUCUUAGAGCAGUUUGAGA CAGUGGC	288
TCTGCACGTGTCAGGGG	227	CCCTGGAATAATGCTCGAAG	399	CUUCGAGCAUUAUUCCAGGGUCUGCACGUGUCA GGGG	289
CATGCTCATTTCCTTTGGC	228	TTCCAGAATTCTCTGCCACC	400	GGUGGCAGAGAAUUCUGGAACAUGCUCAUUUCCU UUGGC	290
TAGCAACAAGGTCAGCAGGC	229	CTGTTCTTTGGTGAGAGGGG	401	CCCCUCUCACCAAAGAACAGUAGCAACAAGGUCA GCAGGC	291
TP53					
ATGCTGGATCCCCACTTTTC	230	GACCAGGTCCTCAGCCC	402	GGGCUGAGGACCUGGUCAUGCUGGAUCCCCACU UUUC	292
GACAAGGGTTGGGCTGG	231	CCAAAGGGTGAAGAGGAATC	403	GAUUCCUCUUCACCCUUUGGGACAAGGGUUGGG CUGG	293
TCTTTGCTGCCGTCTTCC	232	AGGGCCACTGACAACCAC	404	GUGGUUGUCAGUGGCCCUUCUUUGCUGCCGUCU UCC	294
TGCTTGCCACAGGTCTCC	233	GTCAGAGGCAAGCAGAGGC	405	GCCUCUGCUUGCCUCUGACUGCUUGCCACAGGU CUCC	295
GGACAGGTAGGACCTGATTTCC	234	AAACAGTCAAGAAGAAAACGGC	406	GCCGUUUUCUUCUUGACUGUUUGGACAGGUAGG ACCUGAUUUCC	296
AACTTGAACCATCTTTTAACTCAGG	235	GGAATCCTATGGCTTTCCAAC	407	GUUGGAAAGCCAUAGGAUUCCAACUUGAACCAUC UUUUAACUCAGG	297
AGGGGCACAGACCCTCTC	236	AGACCCAAAACCCAAAATGG	408	CCAUUUUGGGUUUUGGGUCUAGGGGCACAGACC CUCUC	298

IDT Exhibit 1020.0773

Espacenet

Bibliographic data: WO2009102878 (A2) — 2009-08-20

METHOD FOR ARCHIVING AND CLONAL EXPANSION

Inventor(s): KURN NURITH [US] ± (KURN, NURITH)

Applicant(s): NUGEN TECHNOLOGIES INC [US]; KURN NURITH [US] ± (NUGEN TECHNOLOGIES, INC, ; KURN, NURITH)

Classification: **- international:** ***C12N15/10***
- cooperative: **C12Q1/6844** more

Application number: WO2009US33936 20090212 Global Dossier

Priority number (s): US20080028146P 20080212 ; US20080038696P 20080321 ; US20080074991P 20080623 ; US20080085811P 20080801

Also published as: WO2009102878 (A3) EP2247727 (A2) EP2247727 (A4) EP2250283 (A2) WO2009102896 (A2) more

Abstract of WO2009102878 (A2)

The present method provides methods, libraries, and kits related to the archiving and clonal expansion of sequences related to target polynucleotide sequences. The method allow for the attachment of polynucleotides with defined 3' and or 5' sequences to solid surfaces. The polynucleotides attached to the solid substrates can be stored or archived as libraries and can subsequently be retrieved for analysis, for example by clonal expansion. In some embodiments, nucleotides attached to solid surfaces can be used for sequencing of nucleotide sequences related to target RNA or target RNA. The methods are applicable to total RNA and/or total DNA analysis.

IDT Exhibit 1020.0774

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 August 2009 (20.08.2009)

PCT

(10) International Publication Number
WO 2009/102878 A2

(51) International Patent Classification:
C12N 15/10 (2006.01)

(21) International Application Number:
PCT/US2009/033936

(22) International Filing Date:
12 February 2009 (12.02.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

61/028,146	12 February 2008 (12.02.2008)	US
61/038,696	21 March 2008 (21.03.2008)	US
61/074,991	23 June 2008 (23.06.2008)	US
61/085,811	1 August 2008 (01.08.2008)	US

(71) Applicant *(for all designated States except US)*: **NUGEN TECHNOLOGIES, INC.** [US/US]; 821 Industrial Road, Unit A, San Carlos, CA 94070 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **KURN, Nurith** [IL/US]; 2876 Ramona Street, Palo Alto, CA 94306 (US).

(74) Agents: **SKUBATCH, Maya** et al.; Wilson Sonsini Goodrich & Rosati, 650 Page Mill Road, Palo Alto, CA 94304-1050 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

(54) **Title**: METHOD FOR ARCHIVING AND CLONAL EXPANSION

(57) **Abstract**: The present method provides methods, libraries, and kits related to the archiving and clonal expansion of sequences related to target polynucleotide sequences. The method allow for the attachment of polynucleotides with defined 3' and or 5' sequences to solid surfaces. The polynucleotides attached to the solid substrates can be stored or archived as libraries and can subsequently be retrieved for analysis, for example by clonal expansion. In some embodiments, nucleotides attached to solid surfaces can be used for sequencing of nucleotide sequences related to target RNA or target RNA. The methods are applicable to total RNA and/or total DNA analysis.

WO 2009/102878 A2

METHOD FOR ARCHIVING AND CLONAL EXPANSION

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Application Nos. 61/028,146, filed February 12, 2008, 61/074,991, filed June 23, 2008, and 61/085,811, filed August 1, 2008, which applications are incorporated herein by reference in their entirety. This application is also related to the co-pending patent application [Attorney Docket No 25115-731.201] filed February 12, 2009, which is incorporated herein by reference in its entirety.

BACKGROUND OF THE INVENTION

[0002] The quality and quantity of nucleic acid sample is important for many studies. High-throughput genomic analysis requires large amounts of template for testing, yet typically the yield of nucleic acids from individual patient samples is limited. Forensic and paleoarcheology work also can be severely limited by nucleic acid sample size. The limitation of starting material impacts the ability to carry out large scale analysis of multiple parameters, as is required for, for example, the genotyping of multiple loci in the study of complex diseases. Moreover, it is well accepted that molecular analysis determination of genomic instability in various pathological condition such as cancer, is most precisely carried out in well defined cell populations, such as that obtained by laser capture micro-dissection or cell sorting. Nucleic acid amplification technologies that provide global amplification of very small polynucleotide samples, for example, from one or a very few cells, may provide a solution to the limited starting materials generally available for analysis.

[0003] Likewise, the ability to amplify ribonucleic acid (RNA) is an important aspect of efforts to elucidate biological processes. Total cellular mRNA represents gene expression activity at a defined time. Gene expression is affected by cell cycle progression, developmental regulation, response to internal and external stimuli and the like. The profile of expressed genes for any cell type in an organism reflects normal or disease states, response to various stimuli, developmental stages, cell differentiation, and the like. Non-coding RNAs have been shown to be of great importance in regulation of various cellular functions and in certain disease pathologies. Such RNAs are often present in very low levels. Thus, amplification methods capable of amplifying low abundance RNAs, are of great importance.

[0004] In addition to the need for amplifying RNA and DNA, there is a need for being able to archive samples, and later retrieve the samples for analysis.

[0005] Various methods for global amplification of DNA target molecules (e.g., whole genome amplification) have been described, including methods based on the polymerase chain reaction (PCR). See, e.g., U.S. Pat. Nos. 5,731,171; 6,365,375; Daigo et al., (2001) Am. J. Pathol. 158 (5):1623-1631; Wang et al, (2001); Cancer Res. 61:4169-4174; Zheng et al, (2001) Cancer Epidemiol. 10:697-700; Dietmaier et al (1999) Am. J. Pathol. 154 (1) 83-95; Stoecklein et al (2002) Am. J. Pathol. 161 (1):43-51; U.S. Pat. Nos. 6,124,120; 6,280,949; Dean et al (2002) PNAS 99 (8):5261-5266. However, PCR-based global amplification methods, such as whole genome amplification (WGA), may generate non-specific amplification artifacts, give incomplete coverage of loci, or generate DNA of insufficient length that cannot be used in many applications. PCR-based methods also suffer from the propensity of the PCR reaction to generate products that are

IDT Exhibit 1020.0776

preferentially amplified, and thus resulting in biased representation of genomic sequences in the products of the amplification reaction. Methods of global amplification of DNA using composite primers have been described. See e.g. U. S. Patent application 10/824,829.

[0006] Additionally, a number of methods for the analysis of gene expression have been developed in recent years. See, for example, U.S. Pat. Nos. 6,251,639, 6,692,918, 6,686,156, 5,744,308; 6,143,495; 5,824,517; 5,829,547; 5,888,779; 5,545,522; 5,716,785; 5,409,818; EP 0971039A2; EP0878553A2; and U.S. published patent applications nos. 2002/0115088, 2003/0186234, 2003/0087251, and 2004/0023271. These include quantification of specific mRNAs, and the simultaneous quantification of a large number of mRNAs, as well as the detection and quantification of patterns of expression of known and unknown genes. RNA amplification is most commonly performed using the reverse transcriptase-polymerase chain reaction (RT-PCR) method and variations thereof. These methods are based on replication of RNA by reverse transcriptase to form single stranded DNA complementary to the RNA (cDNA), which is followed by polymerase chain reaction (PCR) amplification to produce multiple copies of double stranded DNA. However, the total amount of sample RNA that is available is frequently limited by the amount of biological sample from which it is derived. Biological samples are often limited in amount and precious. Moreover, the amount of the various RNA species is not equal; some species are more abundant than others are, and these are more likely and easier, to analyze. The ability to amplify RNA sequences enables the analysis of less abundant, rare RNA species. The ability to analyze small samples, by means of nucleic acid amplification, is also advantageous for design parameters of large scale screening of effector molecule libraries, for which reduction in sample volume is a major concern both for the ability to perform very large scale screening or ultra high throughput screening, and in view of the limiting amounts of library components. Methods of amplification from RNA templates have been described, for example in U. S. Patent 6,946,251.

[0007] Sequencing of nucleic acids continues to be one of the most important and useful ways to analyze DNA and RNA samples. Recent developments have made possible highly parallel high throughput sequencing. Many of these approaches use an in vitro cloning step to generate many copies of each individual molecule. Emulsion PCR is one method, isolating individual DNA molecules along with primer-coated beads in aqueous bubbles within an oil phase. A polymerase chain reaction (PCR) then coats each bead with clonal copies of the isolated library molecule and these beads are subsequently immobilized for later sequencing. See, e.g. WO04069849A2, WO05010145A2. In other cases, surface methods of clonal amplification have been developed, for example, by the use of polonies (PCR colonies), or by bridge PCR where fragments are amplified upon primers attached to a solid surface. These methods produce many physically isolated locations which each contain many copies of a single fragment. While these methods have provided improvements in sequencing throughput, there is a continuing need to improve the methods of obtaining samples appropriate for sequencing, and of handling, storing, and amplifying such samples.

[0008] Therefore, there is a need for improved methods of obtaining, storing, amplifying, and analyzing DNA and RNA samples, including methods which can globally amplify DNA or RNA polynucleotide targets. The invention described herein fulfills this need.

IDT Exhibit 1020.0777

SUMMARY OF THE INVENTION

[0009] One aspect of the invention comprises amplifying a target nucleic acid sequence (DNA or RNA) or the complement of a target nucleic acid sequence on a solid support such as a bead, a magnetic bead, a substantially planar array, an isolated surface, or a well in a plate to form a plurality (e.g. 2; 3; 4; 5; 10; 25; 50; 100; 150; 500; 1,000; 5,000; 25,000 or more) of amplified products. In some cases, the plurality of amplified products comprise clonally amplified products such that a given bead, isolated surface, or well contains a plurality of amplified products of substantially identical sequence. The amplification may be performed by linear amplification such as for example single primer isothermal amplification (SPIA), amplification using a single primer such as for example single primer PCR, rolling circle amplification, or SPIA, amplification from a double stranded nucleic acid having a single stranded 3' overhang, amplification using a DNA-RNA chimeric amplification primer, or any combination thereof. In some embodiments of the method, the amplification step results in at least 1,000; 5,000; 10,000; 100,000; 1,000,000; or more copies of the target nucleic acid or a portion thereof. In some embodiments of the method, the amplification step is performed on a linear template. In some embodiments of the method, the target nucleic acid sequence and/or the amplified product is greater than 100; 200; 400; 500; 1,000; 2,000; 5,000 or more base pairs in length.

[0010] One aspect of the invention comprises a method for clonally amplifying a target nucleic acid sequence or its complement by delivering a target nucleic acid sequence or a set of target nucleic acid sequences into the aqueous phase of a microdroplet of an emulsion and performing linear amplification of the target nucleic acid or a portion thereof in the aqueous phase of the emulsion such that on average each microdroplet amplifies one or less than one of the target nucleic acid sequence or sequences, or such that on average the step of amplification provides microdroplets that comprise a plurality of amplified products that are substantially identical in sequence. In some cases, the amplification within an emulsion is performed in the presence of a solid substrate such as a bead or isolated surface. In some cases, the solid substrate may comprise capture moieties useful for capturing target nucleic acid and/or amplified product. In some cases, the amplification is performed such that on average, a given bead or isolated surface captures one or less than one target nucleic acid, or such that a given bead or isolated surface captures a plurality of amplified products that are substantially identical in sequence.

[0011] One aspect of the invention comprises a kit for performing the methods of the present invention. The kit may be useful for amplifying DNA or RNA for subsequent analysis such as expression analysis including quantitative PCR or microarray analysis; sequencing including thermocycle sequencing, dye terminator sequencing, sequencing using the methods of Illumina/Solexa, SOLiD (ABI), Roche/454 LifeSciences, Helicos, or Sequenom; alternatively the kit may be useful for archiving DNA or RNA sequencing. In another aspect, the kit may be useful for generating clonally expanded target sequences. In yet another aspect, the kit may be useful for generating one or a plurality of amplified products having a defined 3' and 5' end. The kit comprises an RNA-DNA chimeric first primer for creating a first primer extension product, and an RNA-DNA chimeric oligonucleotide for annealing to an end of a double stranded nucleic acid product having a single stranded 3' overhang at one end. In some embodiments, the kit may further

IDT Exhibit 1020.0778

comprise a polymerase having substantial strand-displacement activity. In some embodiments, the kit may further comprise a second primer. In some embodiments, the kit may further comprise RNase H.

[0012] One aspect of the invention comprises a method for creating a double stranded nucleic acid comprising an RNA-DNA heteroduplex, wherein the double stranded nucleic acid is suitable for creating amplified products by single primer isothermal linear amplification that have a defined 5' and 3' end. The method comprises (a) annealing to a template nucleic acid and extending a first primer or a set of first primers with a DNA polymerase comprising a DNA segment and a 5' RNA segment, wherein a 3' portion of the primer is complementary to a target nucleic acid sequence of interest, and a portion of the 5' RNA segment, sequence (A), is not complementary to the target nucleic acid. The annealing and the extension produces a first primer extension product or set of products that is hybridized to the target nucleic acid. The extension step may be performed with an RNA-dependant DNA polymerase from an RNA template or a DNA-dependent DNA polymerase from a DNA template. In some cases, step (a) may be performed in the presence of RNA and DNA and an inhibitor such as for example actinomycin to selectively inhibit the formation of an extension product complementary to the DNA.

[0013] The method further comprises (b) separating or removing the first primer extension product from the target nucleic acid. The separating step may be performed by heat, enzyme, chemical treatment, or a combination thereof.

[0014] The method further comprises step (c) annealing and extending a second primer with a DNA-dependent DNA polymerase. The second primer may comprise a 3' DNA segment and a 5' DNA segment, wherein a portion of the 3' DNA segment is complementary to the first primer extension product or set of first primer extension products and a 5' portion, sequence (B), is not complementary to the first primer extension product or the set of first primer extension products, to produce a double-stranded DNA product or a set of double-stranded DNA products, each product comprising the first primer extension product hybridized to a second primer extension product. The second primer extension product may comprise a DNA sequence (A') that is complementary to and hybridized to the RNA sequence (A) of the first primer extension product at its 3' end, thereby forming a double stranded nucleic acid product or products with an RNA-DNA heteroduplex at one end.

[0015] The method further comprises step (d) cleaving the RNA from the RNA-DNA heteroduplex. The cleaving step may be performed using an enzyme that is specific for the RNA portion of an RNA-DNA heteroduplex such as for example RNase H. The cleaving step may provide a double stranded nucleic acid product or products with a single stranded 3' overhang at one end.

[0016] The method further comprises step (e) annealing a chimeric oligonucleotide comprising a 5' end, wherein the 5' end comprises sequence (C), and the 5' end further comprises RNA, and a 3' end wherein the 3' end comprises the DNA sequence (A), to the single stranded 3' overhang of the double stranded nucleic acid product or products formed in step (d).

[0017] The method further comprises step (f), extending the double stranded nucleic acid product or products with a DNA polymerase to produce a double stranded nucleic acid product with a (C)-

IDT Exhibit 1020.0779

(C') RNA-DNA heteroduplex at one end and a (B)-(B') double stranded DNA sequence at the other end.

[0018] In some embodiments of the method, the second primer further comprises a ligand at a 5' end of the 5' DNA segment.

[0019] In some embodiments of the method, the method further comprises binding the ligand to a solid surface, whereby the first and second primer extension product or products are bound to the solid surface.

[0020] In some embodiments of the method, the method further comprises step (g), cleaving the RNA portion of the chimeric oligonucleotide in the DNA-RNA heteroduplex to generate a double stranded nucleic acid or a set of double stranded nucleic acids with a single stranded 3' overhang at one end of sequence (C'). The method may further comprise step (h), annealing a DNA-RNA chimeric amplification primer to the single stranded portion of the second primer extension product. The chimeric amplification primer may comprise a sequence (C) that is complementary to the 3' overhang formed in step (g). The chimeric amplification primer may comprise a DNA portion and a 5' RNA portion. The method may further comprises step (i), extending the amplification primer with a DNA polymerase having strand displacement activity to produce an amplified product hybridized to the second primer extension product. The method may further comprise step (j), cleaving the RNA from the amplified product hybridized to the second primer extension product in the RNA-DNA heteroduplex; and step (k), repeating steps (h) to (j) to produce multiple copies of amplified product comprising a 5' end and a 3' end, wherein the 5' end comprises sequence (A) and the end comprises sequence (B').

[0021] In some embodiments of the method, the method comprises use of a first primer that comprises a random annealing sequence such as for example random hexamers or random decamers. Alternatively, the method comprises use of a set of first primers, wherein each member of the set of first primers comprises a distinct 3' DNA annealing sequence, each specific for a target or a region of template nucleic acid. Alternatively, the method comprises use of a first primer that comprises a degenerate annealing sequence for binding to multiple related target sequences. Alternatively, the method comprises use of a first primer that comprises a poly-T sequence, or any sequence that substantially hybridizes to the poly-sequence of messenger RNA. Alternatively, the first primer or set of first primers may comprise a combination thereof.

[0022] In some embodiments of the method, the method comprises use of a second primer that comprises a random annealing sequence such as for example random hexamers or random decamers. Alternatively, the method comprises use of a set of second primers, wherein each member of the set of first primers comprises a distinct 3' DNA annealing sequence, each specific for a target or a region of template nucleic acid. Alternatively, the method comprises use of a second primer that comprises a degenerate annealing sequence for binding to multiple related target sequences. Alternatively, the second primer or set of second primers may comprise a combination thereof.

[0023] In some embodiments of the method, the amplified products are attached, bound, or covalently linked to a solid surface such as a bead. In some cases, the attached, bound, or covalently linked amplified products comprise a clonally amplified sequence on the surface such as a bead or an isolated area on a surface. In some cases, the bead or isolated area on a surface is the only bead

IDT Exhibit 1020.0780

or isolated area on a surface within an isolated liquid volume such as an aqueous droplet in a water/oil emulsion or a liquid volume in the well of a plate, such that the amplified product is contained with such a liquid volume. In some cases, the bead or isolated area comprises covalently attached oligonucleotides comprising sequence (B), at their 3' ends, whereby upon the amplification of step (k) multiple copies of amplified product comprising sequence (B') at their 5' end are hybridized to the bead or isolated area.

[0024] In some embodiments of the method, the sequence (B) covalently attached to the bead or isolated area is extended to produce multiple polynucleotides covalently attached to the bead or isolated area that are substantially complementary to the amplified product, wherein the multiple polynucleotides comprise sequence (A') near their 5' ends and sequence (B) near their 3' ends.

[0025] In some embodiments of the method, the method further comprises removing the amplified product from the bead or isolated area to render the covalently attached polynucleotides single stranded, and extending a sequencing primer, a portion of which is complementary to a portion of sequence (C') to produce a detectable signal or detectable oligonucleotide fragments characteristic of the sequence of the polynucleotide bound to the bead or isolated area, and thereby perform sequencing. In some cases, the sequencing method comprises the use of cleavable labeled terminators, dye terminators, pyrophosphate detection, an isothermal sequencing method, cycle sequencing, or sequencing by ligation.

[0026] One aspect of the invention comprises a method for attaching a target nucleic acid (DNA or RNA) to a solid surface such as a bead or an isolated area. The target nucleic acid attached thereto may be useful for example for clonally amplifying the sequence or a portion thereof. The method comprises step (a), providing a first nucleic acid or a set of first nucleic acids comprising DNA corresponding to a region of a target nucleic acid and further comprising a 5' end and a 3' end, wherein said 5' end of the first nucleic acid or the set of first nucleic acids comprises a sequence (A) and a sequence (C), wherein said sequence (A) is 3' of sequence (C), and wherein sequence (C) comprises RNA, and wherein said 3' end of the first nucleic acid comprises sequence (B'). The first nucleic acid may be from a variety of sources including but not limited to the methods provided herein, single primer polymerase chain reaction products, products, and endonuclease treated rolling circle amplification products. The method further comprises step (b), hybridizing the first nucleic acid or set of first nucleic acids to a solid support comprising an oligonucleotide comprising sequence (B) complementary to sequence (B') of the first nucleic acid. The solid support may be a bead, a substantially planar array, or a magnetic bead. In some cases, the bead or an isolated area of the solid support may comprise only one copy of the first nucleic acid. The method further comprises step (c), extending the oligonucleotide to produce a double stranded product comprising a second nucleic acid hybridized to the first nucleic acid, wherein the second nucleic acid comprises a 3' segment complementary to a portion of the first nucleic acid and a 5' segment complementary to a portion of the first nucleic acid, whereby a portion of the 3' segment comprises a sequence (A') and a portion of the 3' segment comprises a sequence (C'), and wherein the sequence (A') is 5' of the sequence (C'), and whereby a portion of the 5' segment comprises sequence (B). In some cases, the bead or an isolated area of the solid support may thus comprise only one copy of the first and second nucleic acid.

IDT Exhibit 1020.0781

[0027] In some embodiments, a plurality of first and second nucleic acids are provided corresponding to different sequences in the target nucleic acid. In some cases, the plurality of first and second nucleic acids are bound to one or a plurality of beads or isolated areas on a surface. In some cases, the plurality of first and second nucleic acids are bound to a plurality of beads or isolated areas on a surface such that generally, or on average, one or fewer than one first and second nucleic acids are bound to one bead or isolated area.

[0028] In some embodiments, the method further comprises amplifying the target nucleic acid by treating the first and second nucleic acids bound to the beads or isolated areas with reagents to produce multiple copies of amplification product in an clonal fashion (i.e. multiple copies of substantially one amplification product on each bead or isolated area) that are complementary to all or a portion of the second nucleic acid product. In some cases, the reagents comprise SPIA amplification reagents including but not limited to a DNA-RNA chimeric amplification primer, a DNA-polymerase with substantial strand-displacement activity, and RNase H. In some cases, the bead or isolated area are stored prior to the step of amplification.

[0029] One aspect of the invention comprises a method for attaching a polynucleotide sequence that is representative of a sequence within a nucleic acid target molecule to a solid surface comprising: (a) extending a first primer comprising a DNA segment and a 5' RNA segment, wherein a 3' portion of the primer, sequence (P), is complementary to a target nucleic acid and a 5' portion of the of the primer, sequence (A), is not complementary to the target nucleic acid, to form a first primer extension product hybridized to the target nucleic acid; (b) separating or removing the first primer extension product from the target nucleic acid; (c) extending a second primer to produce a double stranded product comprising a second primer extension product hybridized to the first primer extension product, wherein the second primer comprises a 3' segment complementary to a portion of the first primer extension product and a ligand, whereby a portion of the 3' end of the second primer extension product comprises a sequence (A') that is complementary to the sequence (A) of the of the first primer; and (d) binding the ligand to a receptor bound to a solid surface whereby the second primer extension product is attached to the solid surface.

[0030] One aspect of the invention comprises a method comprising: (a) extending a first primer comprising a DNA segment and a 5' RNA segment, wherein a 3' portion of the primer is complementary to a target RNA and a 5' portion, sequence (A), of the of the primer is not complementary to the target RNA; to form a first primer extension product hybridized to the target RNA, forming an RNA/DNA hybrid; (b) cleaving the target RNA from the RNA/DNA hybrid; and (c) extending a second primer, comprising a ligand and a 3' segment complementary to a portion of the first primer extension product, to produce a double stranded product with a DNA/RNA heteroduplex at one end; wherein the double stranded product comprises a second primer extension product hybridized to the first primer extension product, and whereby a portion of the 3' end of the second primer extension product comprises a sequence (A') that is complementary to the sequence (A) of the of the first primer.

[0031] In some embodiments of the method, the 3' portion of the primer that is complementary to the target RNA comprises a random nucleotide sequence. In some embodiments of the method the 3' portion of the primer that is complementary to the target RNA comprises a sequence that is

IDT Exhibit 1020.0782

complementary to polyadenosine (poly-A). In some embodiments of the method the 3' portion of the primer that is complementary to the target RNA comprises a specific sequence that is complementary to a multiplicity of targets. In some embodiments of the method, in step (b), the target RNA is cleaved by heat, enzyme treatment, or chemical treatment.

[0032] In some embodiments of the method the RNA target is in a sample that also comprises DNA, and wherein actinomycin is added prior to step (a) to selectively inhibit the production of extension product complementary to the DNA during step (a).

[0033] In some embodiments of the method the 3' segment of the second primer complementary to a portion of the first primer extension product comprises a random nucleotide sequence, a specific sequence complementary to a specific sequence of the first primer extension product, or a sequence common to multiple first primer extension products. In some embodiments of the method the second primer further comprises a nucleotide sequence (B) that is not complementary to the first primer extension product sequence. In some embodiments of the method the method further comprises step (d) of binding the ligand to a solid surface, whereby the second primer extension product becomes bound to the solid surface. In some embodiments of the method the binding of the ligand results in the double stranded product being bound to the solid surface. In some embodiments of the method the binding of the ligand results in the single stranded second primer extension product being bound to the solid surface.

[0034] In some embodiments of the method the method further comprises treating the solid surface with reagents to produce multiple copies of an amplification product that are substantially complementary the second primer extension product. In some embodiments of the method the solid surface comprises a substantially planar array. In some embodiments of the method the solid surface comprises a bead. In some embodiments of the method the bead comprises a magnetic bead. In some embodiments of the method the bead comprises only one copy of second primer extension product. In some embodiments of the method the second primer extension product is single-stranded. In some embodiments of the method the second primer extension product is double-stranded.

[0035] In some embodiments of the method a plurality of second primer extension products are produced corresponding to different sequences in the target RNA. In some embodiments of the method the plurality of second primer extension products is bound to one or a plurality of beads or a plurality of isolated areas on a surface. In some embodiments of the method the plurality of second primer extension products is bound to a plurality of beads under conditions such that generally, one or fewer second primer extension products is bound to one bead or one isolated area on a surface.

[0036] In some embodiments of the method the method further comprises treating the beads or isolated areas with reagents to produce multiple copies of an amplification product complementary to all or a portion of the second primer extension products. In some embodiments of the method the plurality of beads or isolated areas are contained within a plurality of isolated volumes such that generally one or fewer beads or isolated area is associated with each isolated volume, and whereby the production of multiple copies of amplification product results in multiple copies of substantially one amplification product in each volume. In some embodiments of the method the amplification is

IDT Exhibit 1020.0783

carried out with a reaction mixture comprising RNase H, an amplification primer comprising a DNA portion and a 5' RNA portion, and a DNA polymerase with strand displacement activity.

[0037] In some embodiments of the method the method further comprises the step of storing the beads or isolated areas comprising generally one or fewer primer extension products per bead or isolated area.

[0038] In some embodiments of the method the method further comprises clonally amplifying the primer extension product bound to the beads or isolated areas after storing them.

[0039] One aspect of the invention comprises a method for amplifying a nucleic acid representative of a target RNA comprising carrying out steps (a) through (d) above and further comprises the steps of: (e) cleaving the RNA in the heteroduplex from the first primer extension product such that a portion of the second primer extension product that is complementary to sequence (A) is single stranded; (f) annealing an amplification primer to the single stranded portion of the second primer extension product complementary to sequence (A), wherein the amplification primer has a DNA portion and a 5' RNA portion; (g) extending the amplification primer with a DNA polymerase having strand displacement activity to produce an amplified product hybridized to the second primer extension product; (h) cleaving the RNA from the amplified product hybridized to the second primer extension product; and repeating steps (f) to (h) to produce multiple copies of amplified product.

[0040] In some embodiments of the method the second primer further comprises a sequence (B that is not complementary to the first primer extension product sequence, whereby the amplification product comprises a sequence at its 3' end which is complementary to (B). In some embodiments of the method the target RNA comprises messenger RNA.

[0041] One aspect of the invention comprises a method comprising:(a) denaturing a double-stranded target DNA; (b) annealing to the target DNA and extending with a DNA polymerase comprising strand displacement activity, a first primer comprising a DNA segment and a 5' RNA segment, wherein a 3' portion of the primer comprises a random sequence, and a 5' portion of the of the primer comprises sequence (A), which is not complementary to the target DNA; to form a plurality of first primer extension products, each with sequence (A) at its 5' end; and (c) extending a second primer comprising a ligand and a 3' DNA region that comprises a random sequence, wherein the primer is optionally a tailed primer comprising a nucleic acid sequence (B) that is 5' of the random sequence, to form a plurality of double-stranded products each comprising a first primer extension product and a second primer extension product whereby the second primer extension product comprises a ligand;

[0042] In some embodiments of the method step (b) comprises a first incubation at a temperature below about 30 °C, and a second incubation at a temperature above about 40 °C. In some embodiments of the method a DNA polymerase which is active at temperatures above about 45 °C is used to extend the first primer.

[0043] In some embodiments of the method the method further comprises step (d) of binding the ligand to a solid surface, whereby the plurality of second primer extension products become bound to the solid surface.

IDT Exhibit 1020.0784

[0044] In some embodiments of the method the binding of the ligand results in the double stranded product being bound to the solid surface. In some embodiments of the method the binding of the ligand results in the single stranded second primer extension product being bound to the solid surface.

[0045] In some embodiments of the method the method further comprises treating the solid surface with reagents to produce multiple copies of amplification products that are substantially complementary to the plurality of second primer extension products. In some embodiments of the method the solid surface comprises a substantially planar array. In some embodiments of the method the solid surface comprises a plurality of beads or a plurality of isolated areas on a surface. In some embodiments of the method the plurality of second primer extension products is bound to a plurality of beads or isolated areas under conditions such that generally, one or fewer copies of a single second primer extension product is bound to one bead or one isolated area.

[0046] In some embodiments of the method the method further comprises treating the beads or isolated areas with reagents to produce multiple copies of an amplification product substantially complementary to the second primer extension products. In some embodiments of the method the plurality of beads or isolated areas are contained within a plurality of isolated volumes such that generally one or fewer beads or isolated area is associated with each isolated volume, and whereby the production of multiple copies of amplification product results in multiple copies of substantially one amplification product in each volume. In some embodiments of the method the amplification is carried out with a reaction mixture comprising RNase H, an amplification primer with a DNA portion and a 5' RNA portion, and a DNA polymerase with strand displacement activity.

[0047] In some embodiments of the method the method further comprises the step of storing the beads or isolated areas comprising generally one or fewer primer extension products.

[0048] In some embodiments of the method the method further comprises clonally amplifying the primer extension product bound to the beads or isolated areas after storing them.

[0049] One aspect of the invention comprises a method for amplifying a nucleic acid representative of a target DNA comprising carrying out steps (a) through (d) and further comprising the steps of: (e) cleaving the RNA from the first primer extension products such that a portion of the second primer extension product that is complementary to sequence (A) is single stranded; (f) annealing an amplification primer to the single stranded portion of the second primer extension products complementary to sequence (A), wherein the amplification primer has a DNA portion and a 5' RNA portion; (g) extending the amplification primer with a DNA polymerase having strand displacement activity to produce an amplified product hybridized to the second primer extension product; (h) cleaving the RNA from the amplified product hybridized to the second primer extension product; and repeating steps (f) to (h) to produce multiple copies of amplified product

[0050] In some embodiments of the method the second primer comprises a tailed primer comprising a nucleic acid sequence (B) that is 5' of the random sequence, whereby the amplified product comprises a portion complementary to sequence (B) at or near its 3' end. In some embodiments of the method the target DNA is genomic DNA. In some embodiments of the method the target DNA comprises multiple genomes.

IDT Exhibit 1020.0785

[0051] One aspect of the invention comprises a method for archiving and/or clonal expansion of a nucleotide sequence comprising the steps of: (a) obtaining a plurality of partially double-stranded DNA products comprising a first polynucleotide and a second polynucleotide, wherein the second polynucleotide comprises a sequence (A') at its 3' end and a ligand, wherein the sequence (A') portion of the second polynucleotide is single-stranded, wherein optionally the second polynucleotide comprises a sequence (B) at or near its 5' end; (b) attaching the partially double stranded DNA products to a plurality of beads or a plurality of isolated areas on a surface by binding the ligands to the bead or isolated area; (c) annealing an amplification primer to the single stranded portion of the second polynucleotide complementary to sequence (A'), wherein the amplification primer has a DNA portion and a 5' RNA portion; (d) extending the amplification primer with an enzyme having strand displacement activity to produce a plurality of amplified products hybridized to the second polynucleotide products; (e) cleaving the RNA from the amplified product hybridized to the second polynucleotide products using RNase H; and repeating steps (c) to (e) to produce multiple copies of amplified products.

[0052] In some embodiments of the method the ligands of the DNA products are attached to beads or isolated areas, and on average, one DNA product is attached to one or fewer beads or isolated areas.

[0053] In some embodiments of the method the method further comprises storing the plurality of beads or isolated areas for later analysis.

[0054] In some embodiments of the method the beads or isolated areas are stored after step (b), then later amplified with steps (c) through (f). In some embodiments of the method the amplification is a clonal amplification carried out in multiple isolated volumes wherein on average, one isolated volume has one or fewer beads or isolated areas. In some embodiments of the method the multiple isolated volumes are droplets in a non-aqueous phase.

[0055] One aspect of the invention comprises a method comprising: (a) extending a first primer comprising a DNA segment and a 5' RNA segment, wherein a 3' portion of the primer is complementary to a target RNA and a 5' portion, sequence (A), of the of the primer is not complementary to the target RNA; to form a first primer extension product hybridized to the target RNA, forming an RNA/DNA hybrid; (b) cleaving the target RNA from the RNA/DNA hybrid; (c) extending a second primer, comprising a ligand and a 3' segment complementary to a portion of the first primer extension product, to produce a double stranded product with a DNA/RNA heteroduplex at one end; wherein the double stranded product comprises a second primer extension product hybridized to the first primer extension product, and wherein a portion of the 3' end of the second primer extension product comprises a sequence (A') that is complementary to the sequence (A) of the of the first primer; (d) cleaving the RNA in the heteroduplex from the first primer extension product such that a portion of the second primer extension product that is complementary to sequence (A) is single stranded; (e) annealing to the second primer extension product an oligonucleotide comprising a 3'- DNA segment that is complementary to sequence (A') and a 5' RNA segment comprising sequence (C); (f) optionally extending the oligonucleotide to form an oligonucleotide extension product hybridized to the second primer extension product; (g) extending the second primer extension product to create a heteroduplex such that the second primer comprises

IDT Exhibit 1020.0786

a DNA sequence (C') that is complementary to sequence (C); and (h) cleaving the RNA from the heteroduplex created in step (g) to produce a single-stranded portion of the second primer extension product corresponding to sequence (C').

[0056] In some embodiments of the method the 3' portion of the first primer that is complementary to the target RNA comprises a random nucleotide sequence. In some embodiments of the method the 3' portion of the first primer that is complementary to the target RNA comprises a sequence that is complementary to polyadenosine (poly-A). In some embodiments of the method the 3' portion of the primer that is complementary to the target RNA comprises a specific sequence that is complementary to a multiplicity of targets.

[0057] In some embodiments of the method the RNA target is contained within a sample that also comprises DNA, and actinomycin is added prior to step (a) to selectively inhibit the production of extension product complementary to the DNA during step (a).

[0058] In some embodiments of the method, in step (b), the target RNA is cleaved by chemical heat, or enzyme treatment.

[0059] In some embodiments of the method the 3' segment of the second primer complementary to a portion of the first primer extension product comprises a random nucleotide sequence, a specific sequence complementary to a specific sequence of the first primer extension product, or a sequence common to multiple first primer extension products.

[0060] In some embodiments of the method the method further comprises: (i) binding the ligand on the second primer extension product to a solid surface. In some embodiments of the method, step (i) of binding the ligand to the solid surface is performed before step (h). In some embodiments of the method, step (i) of binding the ligand to the solid surface is performed after step (h).

[0061] One aspect of the invention comprises a method of amplifying a sequence representative of an sequence within an RNA target molecule comprising carrying out steps (a) through (i), and further comprising the steps of: (j) annealing an amplification primer, wherein the amplification primer has a DNA portion and a 5' RNA portion, to the single stranded portion of the second primer extension product complementary to sequence (C'); (k) extending the amplification primer with an enzyme having strand displacement activity to produce an amplified product; (l) cleaving the RNA from the amplified product; and (m) repeating steps (j) to (l) to produce multiple copies of amplified product wherein the 5' portion of the amplified product has a sequence complementary to sequence (A').

[0062] In some embodiments of the method, the second primer further comprises a segment (B) that is not complementary to the first primer extension product sequence, whereby the amplified product comprises a sequence (B') at or near its 3' end that is substantially complementary to sequence (B), and a sequence (A) near its 5' end that is complementary to sequence (A'). In some embodiments of the method, the amplification is a clonal amplification. In some embodiments of the method, the solid surface is a bead or an isolated area on a surface. In some embodiments of the method, the bead or isolated area is the only bead or isolated area associated with a isolated liquid volume such that the amplified product is contained within such liquid volume. In some embodiments of the method, the liquid volume is an aqueous droplet within a non-aqueous fluid. In some embodiments of the method, the solid surface is a bead and the droplet is part of a

IDT Exhibit 1020.0787

microemulsion. In some embodiments of the method, the liquid volume is a well in a plate. In some embodiments of the method, the solid surface is a substantially planar substrate.

[0063] In some embodiments of the method, the bead or isolated area comprises covalently attached thereto multiple oligonucleotides comprising the sequence (B) at their 3' ends, whereby upon the amplification of step (m) multiple copies of amplified product comprising sequence (B') at their 5' end are hybridized to the bead or isolated area.

[0064] One aspect of the invention comprises a method of producing a bead or isolated area with multiple copies of a nucleotide sequence covalently attached thereto by attaching the amplified product as described above, and further comprising extending the (B) sequences to produce a multiple polynucleotides covalently attached to the bead or isolated area that are substantially complementary to the amplified product and that comprise sequence (A') near their 5' ends.

[0065] One aspect of the invention comprises a sequencing method comprising extending the (B) sequences as described above, further comprising the steps of removing the amplified product to render the covalently attached polynucleotides single-stranded, and extending a primer to sequence (A') to produce detectable oligonucleotide fragments characteristic of the sequence of the polynucleotide bound to the bead or isolated area. In some embodiments of the method, the sequencing method comprises cleavable labeled terminators. In some embodiments of the method, the sequencing method comprises pyrophosphate detection. In some embodiments of the method, the sequencing method is an isothermal sequencing method. In some embodiments of the method the sequencing method comprises cycle sequencing.

[0066] One aspect of the invention comprises a method of performing bridge PCR comprising creating amplified product with defined 3' and 5' ends as described herein, and further comprising the steps of exposing the amplified product to a solid substrate comprising oligonucleotide sequences attached thereto complementary to the A and B' sequences on the amplified product in the presence of components necessary for polymerase chain reaction, and thermal cycling the system to perform bridge PCR amplification.

[0067] One aspect of the invention comprises a method of performing rolling circle amplification comprising creating amplified product with defined 3' and 5' ends as described herein, and further comprising the steps of: (n) hybridizing the amplified product to a nucleic acid sequence comprising regions complementary to A and B' sequences in close proximity; (o) optionally extending the gap with a DNA polymerase enzyme; (p) ligating to form a circular nucleic acid comprising the amplified product, and performing rolling circle amplification by extending a primer that is complementary to a sequence in the circular nucleic acid. In some embodiments of the method, the primer is complementary to sequence (A), sequence (B'), or a sequence that was between sequences (A) and (B') in the amplified product. In some embodiments of the method, the primer is an oligonucleotide attached to a solid surface.

[0068] One aspect of the invention comprises a method of PCR amplification comprising creating amplified product with defined 3' and 5' ends as described herein further comprising the steps of amplifying the amplified product using primers complementary to sequences (A) and (B), or using primers complementary to sequences (A') and (B').

IDT Exhibit 1020.0788

[0069] One aspect of the invention comprises a method of strand displacement amplification (SDA) creating amplified product with defined 3' and 5' ends as described herein wherein sequences (A) and (B') in the amplified product are designed to be cleaved by a restriction enzyme, and performing strand displacement amplification on the amplified product.

[0070] One aspect of the invention comprises a method comprising: (a) denaturing a double-stranded target DNA; (b) annealing to the target DNA and extending with and enzyme comprising strand displacement activity, a first primer comprising a DNA segment and a 5' RNA segment, wherein a 3' portion of the primer comprises a random sequence, and a 5' portion of the of the primer comprises sequence (A), which is not complementary to the target DNA; to form a plurality of first primer extension products, each with sequence (A) at its 5' end; (c) extending a second primer comprising a ligand and a 3' DNA region that comprises a random sequence, wherein the primer is optionally a tailed primer comprising a nucleic acid sequence (B) that is 5' of the random sequence, to form a plurality of double-stranded products each comprising a first primer extension product and a second primer extension product whereby the second primer extension product comprises a ligand; (d) cleaving the RNA from the first primer extension products such that a portion of the second primer extension products that are complementary to sequence (A) are single stranded; (e) annealing to the second primer extension product an oligonucleotide comprising a 3'-DNA segment that is complementary to sequence (A') and a 5'-RNA segment comprising sequence (C); (f) optionally extending the oligonucleotide to form a plurality of oligonucleotide extension products hybridized to the second primer extension products; (g) extending the second primer extensions product to create a heteroduplex such that the second primer extension products comprise a DNA sequence (C') that is complementary to sequence (C); and (h) cleaving the RNA from the heteroduplex created in step (g).

[0071] In some embodiments of the method the method further comprises: (i) binding the ligand on the second primer extension products to a solid surface. In some embodiments of the method, step (i) of binding the ligand to the solid surface is performed before step (h). In some embodiments of the method, step (i) of binding the ligand to the solid surface is performed after step (h).

[0072] One aspect of the invention comprises a method of amplifying a sequence representative of an sequence within n DNA target molecule carrying out steps (a) through (h) above, and further comprising the steps of: (j) annealing an amplification primer, wherein the amplification primer has a DNA portion and a 5' RNA portion, to the single stranded portion of the second primer extension products complementary to sequence (C'); (k) extending the amplification primer with an enzyme having strand displacement activity to produce a amplified products; (l) cleaving the RNA from the amplified products; and (m) repeating steps (j) to (l) to produce multiple copies of amplified products wherein the 5' portion of the amplified product has a sequence complementary to sequence (A').

[0073] In some embodiments of the method, the second primer comprises the segment (B) that is not complementary to the first primer extension product sequence, whereby the amplified products comprise a sequence (B') at or near their 3' ends that is substantially complementary to sequence (B), and a sequence (A) near their 5' ends that is complementary to sequence (A'). In some embodiments of the method, the amplification is a clonal amplification. In some embodiments of

IDT Exhibit 1020.0789

the method, the solid surface is a bead or isolated area on a surface. In some embodiments of the method, the bead or isolated area is the only bead or isolated area within isolated liquid volume such that the amplified product is contained within such liquid volume. In some embodiments of the method, the liquid volume is an aqueous droplet within a non-aqueous fluid. In some embodiments of the method, the solid surface is a bead and the droplet is part of a microemulsion. In some embodiments of the method, the liquid volume is a well in a plate. In some embodiments of the method, the solid surface is a substantially planar substrate.

[0074] In some embodiments of the method, the bead or isolated area comprises covalently attached multiple oligonucleotides comprising the sequence (B) at their 3' ends, whereby upon the amplification of step (m) multiple copies of amplified products comprising sequence (B') at their 5' end are hybridized to the bead or isolated area.

[0075] One aspect of the invention comprises a method of producing a bead or isolated area with multiple copies of a nucleotide sequence covalently attached thereto by comprising hybridizing amplified product as described herein further comprising extending the (B) sequences to produce a multiple polynucleotides covalently attached to the bead or isolated area that are substantially complementary to the amplified product and that comprise sequence (A') near their 5' ends.

[0076] One aspect of the invention comprises a sequencing method comprising hybridizing amplified product to a solid surface as described herein, further comprising the steps of removing the amplified product to render the covalently attached polynucleotides single-stranded, and extending a primer to sequence (A') to produce detectable oligonucleotide fragments characteristic of the sequence of the polynucleotide bound to the bead or isolated area. In some embodiments of the method, the sequencing method comprises cleavable labeled terminators. In some embodiments of the method the sequencing method comprises pyrophosphate detection. In some embodiments of the method, the sequencing method is an isothermal sequencing method. In some embodiments of the method, the sequencing method comprises cycle sequencing.

[0077] One aspect of the invention comprises a method of performing bridge PCR comprising creating amplified product with defined 3' and 5' ends as described herein further comprising the steps of exposing the amplified products to a solid substrate comprising oligonucleotide sequences attached thereto complementary to the A and B' sequences on the amplified products in the presence of components necessary for polymerase chain reaction, and thermal cycling the system to perform bridge PCR amplification.

[0078] One aspect of the invention comprises a method of performing rolling circle amplification comprising creating amplified product with defined 3' and 5' ends as described above further comprising the steps of: (n) hybridizing the amplified products to a target nucleic acid comprising regions complementary to A and B' sequences in close proximity; (o) optionally extending the gap with a polymerase enzyme; (p) ligating to form a circular nucleic acid comprising the amplified product, and performing rolling circle amplification by extending a primer that is complementary to a sequence in the circular nucleic acid. In some embodiments of the method, the primer is complementary to sequence (A), sequence (B'), or a sequence that was between sequences (A) and (B') in the amplified product. In some embodiments of the method, the primer is an oligonucleotide attached to a solid surface.

IDT Exhibit 1020.0790

[0079] One aspect of the invention comprises a method of PCR amplification comprising creating amplified product with defined 3' and 5' ends as described above further comprising the steps of amplifying the amplified product using primers complementary to sequences (A) and (B), or using primers complementary to sequences (A') and (B').

[0080] One aspect of the invention comprises a method of strand displacement amplification (SDA) comprising creating amplified product with defined 3' and 5' ends as described above wherein sequences (A) and (B') in the amplified product are designed to be cleaved by a restriction enzyme, and performing strand displacement amplification on the amplified product.

[0081] One aspect of the invention provides an alternative method to produce DNA with defined 3' and 5' sequences from RNA.

[0082] One aspect of the invention comprises a method comprising: (a) extending a first primer comprising a 3' portion complementary to a target RNA and a 5' portion, sequence (D), not complementary to the target RNA, to form a first primer extension product hybridized to the target RNA, forming an RNA/DNA hybrid; (b) cleaving the target RNA from the RNA/DNA hybrid; (c) extending a second primer comprising a DNA segment and a 5' RNA segment, wherein a 3' portion of the primer is complementary to the first primer extension product and a 5' portion, sequence (E), of the of the second primer is not complementary to the first primer extension product, to produce a double-stranded DNA product comprising the first primer extension product hybridized to a second primer extension product, whereby the second primer extension product has a sequence (D') that is complementary to sequence (D) at its 3' end; (d) denaturing the double-stranded DNA product; (e) annealing to the second primer extension product and extending a third primer comprising, from its 5' end, a ligand, optionally a sequence (F), and a sequence (D), wherein sequence (D) is complementary to sequence (D') on the second primer extension product to produce a double-stranded DNA product comprising the second primer extension product hybridized to a third primer extension product, whereby the third primer extension product comprises a sequence (E') at its 3' end complementary to sequence (E).

[0083] In some embodiments of the method the method further comprises binding the ligand to a solid surface, whereby the third primer extension product is bound to the solid surface.

[0084] One aspect of the invention comprises an amplification method comprising carrying out steps (a) through (e) further comprising the steps of: (f) cleaving the RNA portion of the second primer extension product in the DNA-RNA heteroduplex, whereby sequence (E') of the third primer extension product is single stranded. (g) annealing an oligonucleotide comprising a 3' DNA segment (E) that is complementary to sequence (E') and a 5' RNA segment comprising sequence (G); (h) extending the third primer extension product to produce a sequence (G') at its 3' end complementary to sequence (G); cleaving the RNA from the heteroduplex created in step (h) to produce a single-stranded portion of the third primer extension product corresponding to sequence (G'). In some embodiments of the method the method further comprises binding the ligand to a solid surface, whereby the third primer extension product comprising sequence (G') is bound to the solid surface. In some embodiments of the method, the 3' portion of the first primer that is complementary to the target RNA comprises a random nucleotide sequence. In some embodiments of the method, the 3' portion of the first primer that is complementary to the target RNA comprises a sequence that is

IDT Exhibit 1020.0791

complementary to polyadenosine (poly-A). In some embodiments of the method, the RNA target is contained within a sample that also comprises DNA, and actinomycin is added prior to step (a) to selectively inhibit the production of extension product complementary to the DNA during step (a).

[0085] In some embodiments of the method, in step (b), the target RNA is cleaved by heat, enzyme treatment, or chemical treatment or enzymes.

[0086] In some embodiments of the method, the 3' segment of the second primer complementary to a portion of the first primer extension product comprises a random nucleotide sequence, a specific sequence complementary to a specific sequence of the first primer extension product, or a sequence common to multiple first primer extension products.

[0087] One aspect of the invention comprises a method of amplifying a sequence representative of an sequence within an RNA target molecule carrying out steps (a) through (i) and further comprising the steps of: (j) annealing an amplification primer, wherein the amplification primer has a DNA portion and a 5' RNA portion, to the single stranded portion of the third primer extension product complementary to sequence (G'); (k) extending the amplification primer with an enzyme having strand displacement activity to produce an amplified product; (l) cleaving the RNA from the amplified product; and (m) repeating steps (j) to (l) to produce multiple copies of amplified product wherein the 5' portion of the amplified product has a sequence (E) complementary to sequence (E') and the 3' end of the amplified product has sequence (D') complementary to sequence (D) and optionally sequence (F') complementary to sequence (F).

[0088] In some embodiments of the method, the amplification is a clonal amplification. In some embodiments of the method, the solid surface is a bead or an isolated area on a surface. In some embodiments of the method, the bead or isolated area is the only bead or isolated area within isolated liquid volume such that the amplified product is contained within such liquid volume. In some embodiments of the method, the liquid volume is an aqueous droplet within a non-aqueous fluid. In some embodiments of the method, the droplet is part of a microemulsion. In some embodiments of the method, the liquid volume is a well in a plate. In some embodiments of the method, the solid surface is a substantially planar substrate.

[0089] In some embodiments of the method, the bead or isolated area comprises covalently attached thereto multiple oligonucleotides comprising the sequence (D), and/or sequence (F) at their 3' ends, whereby upon the amplification of step (m) multiple copies of amplified product comprising sequence (D') (and/or F') at their 5' end are hybridized to the bead or isolated area.

[0090] One aspect of the invention comprises a method of producing a bead or isolated area with multiple copies of a nucleotide sequence covalently attached thereto by hybridizing amplified product as described above, further comprising extending the oligonucleotide at the (D), and/or (F) sequences to produce a multiple polynucleotides covalently attached to the bead or isolated area that are substantially complementary to the amplified product comprising sequence (E') near their 5' ends.

[0091] One aspect of the invention comprises a sequencing method comprising producing a bead with multiple copies of a nucleotide sequence covalently attached thereto, further comprising the steps of removing the amplified product to render the covalently attached polynucleotides single-stranded, and extending a primer to sequence (E') to produce detectable oligonucleotide fragments

IDT Exhibit 1020.0792

characteristic of the sequence of the polynucleotide bound to the bead or isolated area. In some embodiments of the method, the sequencing method comprises cleavable labeled terminators. In some embodiments of the method, the sequencing method comprises pyrophosphate detection. In some embodiments of the method, the sequencing method is an isothermal sequencing method. In some embodiments of the method, the sequencing method comprises cycle sequencing.

[0092] One aspect of the invention comprises alternative methods to produce DNA with defined 3' and 5' sequences from a DNA target

[0093] One aspect of the invention comprises a method comprising: (a) denaturing a double-stranded target DNA; (b) annealing to the target DNA and extending a first primer comprising a 3' portion comprising a random sequence and a 5' portion, sequence (D), which is not complementary to the target DNA, to form a plurality of first primer extension products, each comprising sequence (D) at its 3' end; (c) extending a second primer comprising a DNA segment and a 5' RNA segment, wherein a 3' portion comprises a random sequence, and a 5' portion, sequence (E), of the of the second primer is not complementary to the first primer extension products, to produce a plurality of double-stranded DNA products comprising a first primer extension product hybridized to a second primer extension product, whereby the second primer extension products have a sequence (D') that is complementary to sequence (D) at their 3' ends; (d) denaturing the double-stranded DNA products; e) annealing to the second primer extension products and extending a third primer comprising, from its 5' end, a ligand, optionally a sequence (F), and a sequence (D), wherein sequence (D) is complementary to sequence (D') on the second primer extension products to produce double-stranded DNA products comprising second primer extension products hybridized to third primer extension products, whereby the third primer extension products comprise a sequence (E') at its 3' end complementary to sequence (E) in a DNA-RNA heteroduplex.

[0094] In some embodiments of the method the method further comprises binding the ligand to a solid surface, whereby the third primer extension products are bound to the solid surface.

[0095] In some embodiments of the method the method further comprises the steps of: (f) cleaving the RNA portion of the second primer extension products in the DNA-RNA heteroduplex, whereby sequence (E') of the third primer extension products is single stranded. (g) annealing an oligonucleotide comprising a 3' DNA segment (E) that is complementary to sequence (E') and a 5' RNA segment comprising sequence (G); (h) extending the third primer extension products to produce a sequence (G') at their 3' ends complementary to sequence (G); (i) cleaving the RNA from the heteroduplex created in step (h) to produce single-stranded portions of the third primer extension products corresponding to sequence (G').

[0096] In some embodiments of the method the method further comprises binding the ligand to a solid surface, whereby the third primer extension products comprising sequence (G') are bound to the solid surface.

[0097] One aspect of the invention comprises a method of amplifying a sequence representative of a sequence within a DNA target molecule comprising following the steps (a) through (i), further comprising the steps of: (j) annealing an amplification primer to the single stranded portion of the third primer extension products complementary to sequence (G'); wherein the amplification primer has a DNA portion and a 5' RNA portion, (k) extending the amplification primer with an enzyme

IDT Exhibit 1020.0793

having strand displacement activity to produce a amplified products; (l) cleaving the RNA from the amplified products; and (m) repeating steps (j) to (l) to produce multiple copies of amplified products wherein the 5' portion of the amplified products have a sequence (E) complementary to sequence (E') and the 3' end of the amplified products have sequence (D') complementary to sequence (D) and optionally sequence (F') complementary to sequence (F).

[0098] In some embodiments of the method, the amplification is a clonal amplification. In some embodiments of the method, the solid surface is a plurality of beads or a plurality of isolated areas on a surface.

[0099] In some embodiments of the method, each bead or isolated area in the plurality of beads or isolated areas is the only bead or isolated area within isolated liquid volume such that the amplified product corresponding to the sequence on that bead or isolated area is associated with such liquid volume. In some embodiments of the method, the liquid volume is an aqueous droplet within a non-aqueous fluid. In some embodiments of the method, the droplet is part of a microemulsion. In some embodiments of the method, the liquid volume is a well in a plate. In some embodiments of the method, the solid surface is a substantially planar substrate.

[00100] In some embodiments of the method, the bead or isolated area comprises covalently attached thereto multiple oligonucleotides comprising the sequence (D), and/or sequence (F) at their 3' ends, whereby upon the amplification of step (m) multiple copies of amplified product comprising sequence (D') (and/or sequence (F') at their 5' end are hybridized to the bead or isolated area.

[00101] One aspect of the invention comprises a method of producing a bead or isolated area with multiple copies of a nucleotide sequence covalently attached thereto further comprising extending the (D), and/or (F) sequences to produce a multiple polynucleotides covalently attached to the bead or isolated area that are substantially complementary to the amplified product comprising sequence (E') near their 5' ends.

[00102] One aspect of the invention comprises a sequencing method comprising creating amplified product with defined 3' and 5' ends as described above, further comprising the steps of removing the amplified product to render the covalently attached polynucleotides single-stranded, and extending a primer to sequence (E') to produce detectable oligonucleotide fragments characteristic of the sequence of the polynucleotide bound to the bead or isolated area. In some embodiments of the method, the sequencing method comprises cleavable labeled terminators. In some embodiments of the method, the sequencing method comprises pyrophosphate detection. In some embodiments of the method, the sequencing method is an isothermal sequencing method. In some embodiments of the method, the sequencing method comprises cycle sequencing.

[00103] In one aspect the invention comprises a library of nucleic acid sequences.

[00104] One aspect of the invention comprises a library comprising a plurality of double-stranded oligonucleotides each of the oligonucleotides comprising: (a) a first strand comprising DNA which has, proceeding from its 5' end (i) a ligand, (ii) a specific sequence (B, D, or DF), (iii) a sequence corresponding to or complementary to a sequence within a nucleic acid target, (iv) a specific sequence (A' or E'); and a specific sequence (C' or G'); and (b) a second strand having from its 5' end, (i) a specific RNA sequence (C or G) complementary to specific sequence C' or G', (ii) a specific sequence (A or E) complementary to sequence (A' or E'), (iii) a sequence complementary to

IDT Exhibit 1020.0794

or corresponding to a sequence within a target nucleic acid. In some embodiments, the ligands are bound to a solid surface. In some embodiments, the solid surface comprises a plurality of beads. In some embodiments, each of the plurality of beads comprises a single molecule of double-stranded oligonucleotide. In some embodiments, the solid surface comprises a plurality of isolated areas on a surface.

[00105] One aspect of the invention relates to kits comprising reagents that can be used, for example, for carrying out the methods of the invention.

[00106] One aspect of the invention comprises kit comprising: (a) a first primer comprising a 3'-DNA portion and a 5'-RNA portion, wherein the 5' RNA portion further comprises sequence (A); (b) a second primer comprising a 5'-ligand; (c) an RNA dependent DNA polymerase; (d) a DNA dependent DNA polymerase with strand displacement activity; (e) RNase H; and (f) an amplification chimeric primer comprising a 3'-DNA portion and a 5'-RNA portion wherein the sequence of the amplification primer is the substantial the same sequence as the (A) sequence. In some embodiments, the 3'-DNA portion of the first primer comprises a random sequence. In some embodiments, the second primer is a DNA primer that comprises a random sequence at its 3' end. In some embodiments, the 3'-DNA portion of the first primer comprises a random sequence, and the second primer is a DNA primer that comprises a random sequence at its 3' end.

[00107] One aspect of the invention comprises a kit comprising: (a) a first primer comprising a 3'-DNA portion and a 5'-RNA portion, wherein the 5' RNA portion further comprises sequence (A); (b) a second primer comprising a 5'-ligand; (c) an RNA dependent DNA polymerase; (d) a DNA dependent DNA polymerase with strand displacement activity; (e) RNase H; (f) a chimeric oligonucleotide comprising a 3'-DNA portion substantially comprising sequence (A) and a 5'-RNA sequence (C); and (g) a chimeric amplification primer comprising a 3'-DNA portion and a 5'-RNA portion, wherein the chimeric amplification primer comprises a sequence which is substantially the same as sequence (C). In some embodiments, the 3'-DNA portion of the first primer comprises a random sequence. In some embodiments, the second primer is a DNA primer that comprises a random sequence at its 3' end. In some embodiments, the 3'-DNA portion of the first primer comprises a random sequence, and the second primer is a DNA primer that comprises a random sequence at its 3' end. In some embodiments, the second primer further comprises a sequence (B) at or near the 5'-end. In some embodiments, the kit further comprises solid support with immobilized receptor to the ligand on it surface. In some embodiments, the kit further comprises solid surface with an oligonucleotide attached to the surface by the 5'-end and comprising a sequence substantially the same as sequence (B).

[00108] In some embodiments, the kit further comprises solid surface with an oligonucleotide attached to the surface by the 5'-end hybridizable to sequence (A).

[00109] In some embodiments, the kit further comprises an inhibitor of the DNA dependent DNA polymerase. In some embodiments, the inhibitor of the DNA dependent DNA polymerase is Actinomycin.

[00110] One aspect of the invention comprises a kit comprising: (a) a first primer that is a tailed DNA primer comprising a 5'-tail sequence (D); (b) a second primer that is a chimeric primer comprising a 3'-DNA portion and a 5'-RNA portion wherein the 5'-end comprises a tail sequence

IDT Exhibit 1020.0795

(E); (c) a third primer which is a tailed primer comprising a 3'-sequence that comprises a sequence substantially the same as sequence (D), optionally a 5'-tail sequence (F), and 5'-ligand; (d) an RNA dependent DNA polymerase; e) a DNA dependent DNA polymerase with strand displacement activity; (f) RNase H; and (g) a chimeric amplification primer comprising a 3'-DNA portion and a 5'-RNA portion wherein the chimeric amplification primer comprises a sequence which is substantially the same a sequence (E).

[00111] In some embodiments, the 3'-DNA portion of the first primer comprises a random sequence. In some embodiments, the second primer is a DNA primer that comprises a random sequence at its 3' end. In some embodiments, the 3'-DNA portion of the first primer comprises a random sequence, and the second primer is a DNA primer that comprises a random sequence at its 3' end.

[00112] One aspect of the invention comprises a kit comprising: (a) a first primer that is a tailed DNA primer comprising a 5'-tail sequence (D); (b) a second primer that is a chimeric primer comprising a 3'-DNA portion and a 5'-RNA portion wherein the 5'-end comprises a tail sequence (E); (c) a third primer which is a tailed primer comprising a 3'-sequence that comprises a sequence substantially the same as sequence (D), optionally a 5'-tail sequence (F), and 5'-ligand; (d) an RNA dependent DNA polymerase; e) a DNA dependent DNA polymerase with strand displacement activity; (f) RNase H; (g) a chimeric oligonucleotide comprising a 3'-DNA sequence (E) and a 5'-RNA sequence G; and (h) a chimeric amplification primer comprising a 3'-DNA portion and a 5'-RNA portion wherein the chimeric amplification primer comprises a sequence which is substantially the same as sequence (G). In some embodiments, the 3'-DNA portion of the first primer comprises a random sequence. In some embodiments, the second primer is a DNA primer that comprises a random sequence at its 3' end. In some embodiments, the 3'-DNA portion of the first primer comprises a random sequence, and the second primer is a DNA primer that comprises a random sequence at its 3' end. In some embodiments, the kit further comprises an inhibitor of the DNA dependent DNA polymerase, In some embodiments, the inhibitor of the DNA dependent DNA polymerase is Actinomycin.

[00113] One aspect of the invention comprises a method for attaching a polynucleotide sequence that is representative of a sequence within a nucleic acid target molecule to a solid surface comprising: (a) extending a first primer comprising a DNA segment and a 5' RNA segment, wherein a 3' portion of the primer, sequence (P), is complementary to a target nucleic acid and a 5' portion of the of the primer, sequence (A), is not complementary to the target nucleic acid, to form a first primer extension product hybridized to the target nucleic acid; (b) separating or removing the first primer extension product from the target nucleic acid; (c) extending a second primer to produce a double-stranded product comprising a second primer extension product hybridized to the first primer extension product, wherein the second primer comprises a 3' segment complementary to a portion of the first primer extension product and 5' segment non-complementary sequence (B) to the first primer extension product, whereby a portion of the 3' end of the second primer extension product comprises a sequence (A') that is complementary to the sequence (A) of the of the first primer and a portion of the 5' end of the second primer extension product comprises non-complementary sequence (B); (d) adding an exonuclease to the double-stranded DNA/RNA hybrid, whereby single

IDT Exhibit 1020.0796

stranded 3' nucleotides are removed from the first primer extension product; (e) extending the first primer extension product to produce a sequence (B'), complementary to sequence (B) on the second primer extension product; (f) denaturing the first and second primer extension products; (g) binding the sequence (B') of the first primer extension product to a third primer comprising sequence (B) bound to a solid surface, whereby the first primer extension product is attached to the solid surface; and (h) extending the sequence (B) of the third primer to produce a double-stranded product comprising a third primer extension product hybridized to the first primer extension product, wherein the 5'end of the third primer comprises a sequence (B) complementary to the sequence (B') of the first primer extension product, whereby a portion of the 3' end of the third primer extension product comprises a sequence (A') that is complementary to the sequence (A) of the of the first primer.

[00114] One aspect of the invention comprises a method comprising: (a) extending a first primer comprising a DNA segment and a 5' RNA segment, wherein a 3' portion of the primer is complementary to a target RNA and a 5' portion, sequence (A), of the of the primer is not complementary to the target RNA; to form a first primer extension product hybridized to the target RNA, forming an RNA/DNA hybrid; (b) removing the target RNA from the RNA/DNA hybrid; (c) extending a second primer, comprising a 3' segment complementary to a portion of the first primer extension product and a 5' segment non-complementary to the first primer extension product comprising sequence (B), to produce a double-stranded DNA product with a DNA/RNA heteroduplex at one end, wherein the double-stranded product comprises a second primer extension product hybridized to the first primer extension product and wherein a portion of the 3' end of the second primer extension product comprises a sequence (A') that is complementary to the sequence (A) of the of the first primer; (d) adding an exonuclease to the double-stranded DNA product, whereby single stranded 3' nucleotides are removed from the 3' region of the first primer extension product that is not hybridized to the second primer extension product; (e) extending the first primer extension product to produce a sequence (B'), complementary to sequence (B) on the second primer extension product; (f) denaturing the double-stranded DNA product; (g) attaching the single-stranded first primer extension product to a solid support by annealing sequence (B') to the solid support comprising an oligonucleotide attached thereto, comprising a sequence (B); and (h) extending sequence (B) on the solid support to produce a third primer extension product, hybridized to the first extension product, wherein the third primer extension product comprises a 3' sequence (A'), whereby a DNA/RNA heteroduplex at one end is generated.

[00115] One aspect of the invention comprises a method for amplifying a nucleic acid representative of a target RNA comprising carrying out steps (a) through (h) above and further comprising the steps of: (i) cleaving the RNA region from the first polynucleotide product hybridized to the third primer extension product using RNase H; (j) annealing an amplification primer to sequence (A') on the single-stranded portion of the third primer extension product, wherein the amplification primer has a DNA portion and a 5' RNA portion; (k) extending the amplification primer with an enzyme having strand displacement activity to produce an amplified product hybridized to the third primer extension product on the solid support; (l) repeating steps (i) to (k) to produce multiple copies of an amplified product wherein the amplified product comprises

IDT Exhibit 1020.0797

sequence (B') at its 3' end; and (m) capturing the amplified product on the solid support wherein the solid support comprises sequence (B).

[00116] In some embodiments, the 3' portion of the primer that is complementary to the target RNA comprises a random nucleotide sequence. In some embodiments, the 3' portion of the primer that is complementary to the target RNA comprises a sequence that is complementary to polyadenosine (poly-A). In some embodiments, the 3' portion of the primer that is complementary to the target RNA comprises a specific sequence that is complementary to a multiplicity of targets. In some embodiments, the target RNA is cleaved by heat, enzyme treatment, or chemical treatment in step (b).

[00117] In some embodiments, the RNA target is in a sample that also comprises DNA, and wherein actinomycin is added prior to step (a) to selectively inhibit the production of extension product complementary to the DNA during step (a). . In some embodiments, the 3' portion of the second primer complementary to a portion of the first primer extension product comprises a random nucleotide sequence, a specific sequence complementary to a specific sequence of the first primer extension product, or a sequence common to multiple first primer extension products.

[00118] In some embodiments, the solid support comprises a bead. In other embodiments, the solid support comprises an isolated area. In some embodiments, the solid support comprises a plurality of beads or a plurality of isolated areas on a surface. In some embodiments, the solid support comprises a substantially planar array. In some embodiments, the bead comprises a magnetic bead. In some embodiments, the bead comprises only one copy of the first primer extension product.

[00119] In some embodiments, the first primer extension product is single-stranded. In some embodiments, a plurality of first primer extension products are produced corresponding to different sequences in the target RNA. In some embodiments, the plurality of first primer extension products is bound to a solid support comprising either one or a plurality of beads or a plurality of isolated areas on a surface. In some embodiments, the plurality of first primer extension products is bound to a plurality of beads under conditions such that generally, one or fewer first primer extension products is bound to one bead or one isolated area on a surface. In some embodiments, the plurality of beads or isolated areas are contained within a plurality of isolated volumes such that generally one or fewer beads or isolated area is associated with each isolated volume, and whereby the production of multiple copies of amplification product results in multiple copies of substantially one amplification product in each volume.

[00120] In some embodiments, the method further comprises the step of storing the beads or isolated areas comprising generally one or fewer primer extension products per bead or isolated area. In some embodiments, the method further comprises clonally amplifying the primer extension product bound to the beads or isolated areas after storing them. In other embodiments, the target RNA comprises messenger RNA.

[00121] One aspect of the invention comprises a method comprising: (a) denaturing a double-stranded target DNA; (b) annealing to the target DNA and extending with a DNA polymerase comprising strand displacement activity, a first primer comprising a DNA segment and a 5' RNA segment, wherein a 3' portion of the primer comprises a random sequence, and a 5' portion of the primer comprises sequence (A), which is not complementary to the target DNA; to form a plurality

IDT Exhibit 1020.0798

of first primer extension product hybridized to the target DNA and comprising sequence (A) at its 5' end; (c) separating the first primer extension product from the target DNA; (d) annealing to the first primer extension product and extending a second primer comprising a 3' complementary DNA region that comprises a random sequence, wherein the second primer is a tailed primer comprising a 5' sequence (B), to form a double-stranded product comprising a first primer extension product and a second primer extension product, whereby a double-stranded product with a DNA/RNA heteroduplex at one end is generated; (e) adding an exonuclease to the double-stranded DNA product, whereby single stranded 3' nucleotides are removed from the 3' region of the first primer extension product that is not hybridized to the second primer extension product; (f) extending the first primer extension product to produce a sequence (B'), complementary to sequence (B) on the second primer extension product; (g) denaturing the double-stranded DNA product; (h) attaching the single-stranded first primer extension product to a solid support by annealing sequence (B') to the solid support comprising an oligonucleotide attached thereto, comprising a sequence (B), whereby a plurality of first primer extension products become bound to the solid surface; and (i) extending sequence (B) on the solid support to produce a third primer extension product, hybridized to the first primer extension product, comprising a 3' sequence (A'), whereby a DNA/RNA heteroduplex at one end is generated.

[00122] One aspect of the invention comprises a method for amplifying a nucleic acid representative of a target RNA comprising carrying out steps (a) through (i) above and further comprising the steps of: (j) cleaving the RNA from the first polynucleotide product hybridized to the amplified product using RNase H; (k) annealing an amplification primer to the single-stranded portion of the amplified product complementary to sequence (A'), wherein the amplification primer has a DNA portion and a 5' RNA portion; (l) extending the amplification primer with an enzyme having strand displacement activity to produce an amplified product hybridized to the third primer extension product on the bead or isolated area; (m) repeating steps (j) to (l) to produce multiple copies of an amplified product wherein the amplified product comprises sequence (B') at its 3' end; and (n) capturing the amplified product on the solid support comprising sequence (B).

[00123] In some embodiments, step (b) above comprises a first incubation at a temperature below about 30 °C, and a second incubation at a temperature above about 40 °C. In some embodiments, a DNA polymerase which is active at temperatures above about 45 °C is used to extend the first primer.

[00124] In some embodiments, the solid support comprises a bead. In some embodiments, the solid support comprises an isolated area. In some embodiments, the solid support comprises a plurality of beads or a plurality of isolated areas on a surface. In some embodiments, the solid support comprises a substantially planar array. In some embodiments, the bead comprises a magnetic bead.

[00125] In some embodiments, the plurality of first primer extension products is bound to a plurality of beads or isolated areas under conditions such that generally, one or fewer copies of a single first primer extension product is bound to one bead or one isolated area. In some embodiments, the plurality of beads or isolated areas are contained within a plurality of isolated volumes such that generally one or fewer beads or isolated area is associated with each isolated volume, and whereby

IDT Exhibit 1020.0799

the production of multiple copies of amplification product results in multiple copies of substantially one amplification product in each volume.

[00126] In some embodiments, the method further comprises the step of storing the beads or isolated areas comprising generally one or fewer primer extension products. In some embodiments, the method further comprises clonally amplifying the primer extension product bound to the beads or isolated areas after storing them. In some embodiments, the target DNA is genomic DNA. In some embodiments, the target DNA comprises multiple genomes.

[00127] One aspect of the invention for preparing DNA with defined 3' and 5' sequences comprises a method comprising: (a) extending a first primer comprising a DNA segment and a 5' RNA segment, wherein a 3' portion of the primer is complementary to a target RNA and a 5' portion, sequence (A), of the of the primer is not complementary to the target RNA; to form a first primer extension product hybridized to the target RNA, forming an RNA/DNA hybrid; (b) removing the target RNA from the RNA/DNA hybrid; (c) extending a second primer, comprising a 3' segment complementary to a portion of the first primer extension product and a 5' segment non-complementary to the first primer extension product comprising sequence (B), to produce a double-stranded product with a DNA/RNA heteroduplex at one end; wherein the double-stranded product comprises a second primer extension product hybridized to the first primer extension product, and whereby a portion of the 3' end of the second primer extension product comprises a sequence (A') that is complementary to the sequence (A) of the of the first primer; (d) cleaving the RNA in the heteroduplex from the first primer extension product such that a portion of the second primer extension product that is complementary to sequence (A) is single-stranded; (e) annealing to the second primer extension product an oligonucleotide comprising a 3'- DNA sequence (A) that is complementary to sequence (A') and a 5'-RNA segment comprising sequence (C) that is non-complementary to the second primer extension product; (f) extending the oligonucleotide at the 3' segment to form an oligonucleotide extension product hybridized to the second primer extension product; (g) denaturing the double-stranded DNA product; (h) attaching the single-stranded first primer extension product to a solid support by annealing sequence (B') to the solid support comprising a sequence (B); and (i) extending sequence (B) on the solid support to produce a third primer extension product, comprising a 3' sequence (A') and (C'), whereby a DNA/RNA heteroduplex at one end is generated.

[00128] One aspect of the invention comprises a method for amplifying a nucleic acid representative of a target RNA comprising carrying out steps (a) through (i) above and further comprising the steps of: (j) cleaving the RNA from the heteroduplex polynucleotide product hybridized to the amplified product using RNase H to produce a single-stranded portion of the third primer extension product corresponding to sequence (C'); (k) annealing an amplification primer to the single-stranded portion of the third primer extension product complementary to sequence (C'), wherein the amplification primer has a DNA portion and a 5' RNA portion; (l) extending the amplification primer with an enzyme having strand displacement activity to produce an amplified product hybridized to the third primer extension product on the solid support; (m) repeating steps (j) to (l) to produce multiple copies of the amplified product comprising sequences (A) and (B'); and

IDT Exhibit 1020.0800

(n) capturing the amplified product on the solid support wherein the solid support comprises sequence (B).

[00129] In some embodiments, the method further comprises extending sequence (B), whereby multiple copies are bound to solid support through sequence (B) having sequence (A') at 3' end. In some embodiments, the method further comprises sequencing by synthesis using a sequence (A) complementary to (A') as the priming sequence. In some embodiments, the 3' portion of the first primer that is complementary to the target RNA comprises a random nucleotide sequence. In some embodiments, the 3' portion of the first primer that is complementary to the target RNA comprises a sequence that is complementary to polyadenosine (poly-A). In some embodiments, the 3' portion of the primer that is complementary to the target RNA comprises a specific sequence that is complementary to a multiplicity of targets. In some embodiments, the RNA target is contained within a sample that also comprises DNA, and actinomycin is added prior to step (a) to selectively inhibit the production of extension product complementary to the DNA during step (a). In some embodiments, the target RNA is cleaved by chemical heat, or enzyme treatment in step (b)

[00130] In some embodiments, the 3' segment of the second primer complementary to a portion of the first primer extension product comprises a random nucleotide sequence, a specific sequence complementary to a specific sequence of the first primer extension product, or a sequence common to multiple first primer extension products. In some embodiments, the amplification is a clonal amplification.

[00131] In some embodiments, the solid support is a bead. In some embodiments, the solid support is an isolated area on a surface. In some embodiments, the bead or isolated area is the only bead or isolated area associated with an isolated liquid volume such that the amplified product is contained within such liquid volume. In some embodiments, the liquid volume is an aqueous droplet within a non-aqueous fluid. In some embodiments, the solid surface is a bead and the droplet is part of a microemulsion. In some embodiments, the liquid volume is a well in a plate. In some embodiments, the solid support is a substantially planar substrate. In some embodiments, a method of producing a solid support with multiple copies of a nucleotide sequence covalently attached thereto by performing the method described above, and further comprising extending the (B) sequences to produce multiple polynucleotides covalently attached to the solid support that are substantially complementary to the amplified product and that comprise sequence (A') near their 3' ends.

[00132] In some embodiments, a sequencing method comprises performing the method described above and further comprises the steps of removing the third primer extension product to render the covalently attached polynucleotides single-stranded, and extending a primer to sequence (A) to produce detectable oligonucleotide fragments characteristic of the sequence of the polynucleotide bound to the bead or isolated area. In some embodiments the sequencing method comprises cleavable labeled terminators. In some embodiments the sequencing method comprises pyrophosphate detection. In some embodiments the sequencing method is an isothermal sequencing method. In some embodiments the sequencing method comprises cycle sequencing.

[00133] In some embodiments a method of performing bridge PCR comprising performing the method described above and further comprising the steps of exposing the amplified product to a solid substrate comprising oligonucleotide sequences attached thereto complementary to the A and

IDT Exhibit 1020.0801

B' sequences on the amplified product in the presence of components necessary for polymerase chain reaction, and thermal cycling the system to perform bridge PCR amplification.

[00134] In some embodiments a method of performing rolling circle amplification comprising performing the method described previously and further comprising the steps of: (o) hybridizing the amplified product to a nucleic acid sequence comprising regions complementary to A and B' sequences in close proximity; (p) optionally extending the gap with a DNA polymerase enzyme; (q) ligating to form a circular nucleic acid comprising the amplified product, and performing rolling circle amplification by extending a primer that is complementary to a sequence in the circular nucleic acid.

[00135] In some embodiments the primer is complementary to sequence (A), sequence (B'), or a sequence that was between sequences (A) and (B') in the amplified product. In some embodiments the primer is an oligonucleotide attached to a solid surface.

[00136] In some embodiments, a method of PCR amplification comprising performing the method previously described and further comprising the steps of amplifying the amplified product using primers complementary to sequences (A) and (B), or using primers complementary to sequences (A') and (B'). In some embodiments, a method of strand displacement amplification (SDA) comprising performing the method previously described wherein sequences (A) and (B') in the amplified product are designed to be cleaved by a restriction enzyme, and performing strand displacement amplification on the amplified product.

[00137] One aspect of the invention for preparing DNA with defined 3' and 5' sequences comprises a method comprising: (a) denaturing a double-stranded target DNA; (b) annealing to the target DNA and extending with a DNA polymerase comprising strand displacement activity, a first primer comprising a DNA segment and a 5' RNA segment, wherein a 3' portion of the primer comprises a random sequence, and a 5' portion of the of the primer comprises sequence (A), which is not complementary to the target DNA; to form a first primer extension product hybridized to the target DNA and comprising sequence (A) at its 5' end; (c) separating the first primer extension product from the target DNA; (d) annealing to the first primer extension product and extending a second primer comprising a 3' complementary DNA region that comprises a random sequence, wherein the second primer is a tailed primer comprising a 5' sequence (B), to form a double-stranded product comprising a first primer extension product and a second primer extension product, whereby a double-stranded product with a DNA/RNA heteroduplex at one end is generated; (e) cleaving the RNA in the heteroduplex from the first primer extension product such that a portion of the second primer extension product that is complementary to sequence (A) is single stranded; (f) annealing to the second primer extension product an oligonucleotide comprising a 3'- DNA segment that is complementary to sequence (A') and a 5' RNA segment comprising sequence (C); (g) extending the oligonucleotide along the second primer extension product to form an oligonucleotide extension product comprising a sequence (B'), complementary to sequence (B) on the second primer extension product; (h) denaturing the double-stranded DNA product; (i) attaching the single-stranded first primer extension product to solid support by annealing sequence (B') to the bead or isolated area comprising a sequence (B); and (j) extending sequence (B) on the solid support to produce a third

IDT Exhibit 1020.0802

primer extension product, hybridized to the oligonucleotide extension product, comprising a 3' sequence (A') and (C'), whereby a DNA/RNA heteroduplex at one end is generated.

[00138] One aspect of the invention comprises a method for amplifying a nucleic acid representative of a target DNA comprising carrying out steps (a) through (j) above and further comprising the steps of: (k) cleaving the RNA from the heteroduplex polynucleotide product hybridized to the amplified product using RNase H to produce a single-stranded portion of the second primer extension product corresponding to sequence (C'); (l) annealing an amplification primer to the single-stranded portion of the amplified product complementary to sequence (C'), wherein the amplification primer has a DNA portion and a 5' RNA portion; (m) extending the amplification primer with an enzyme having strand displacement activity to produce an amplified product hybridized to the amplified product on the bead or isolated area; and (n) repeating steps (k) to (m) to produce multiple copies of the second polynucleotide product comprising sequences (A) and (B').

[00139] In some embodiments, the amplification is a clonal amplification. In some embodiments, the solid support is a bead or isolated area on a surface. In some embodiments, the bead or isolated area is the only bead or isolated area within isolated liquid volume such that the amplified product is contained within such liquid volume. In some embodiments, the liquid volume is an aqueous droplet within a non-aqueous fluid. In some embodiments, the solid surface is a bead and the droplet is part of a microemulsion. In some embodiments, the liquid volume is a well in a plate. In some embodiments, the solid support is a substantially planar substrate. In some embodiments, the bead or isolated area comprises covalently attached multiple oligonucleotides comprising the sequence (B) at their 3' ends, whereby upon the amplification of step (m), multiple copies of amplified products comprising sequence (B') at their 5' end are hybridized to the bead or isolated area.

[00140] In some embodiments, a method of producing a bead or isolated area with multiple copies of a nucleotide sequence covalently attached thereto by performing the method previously described and further comprising extending the (B) sequences to produce a multiple polynucleotides covalently attached to the bead or isolated area that are substantially complementary to the amplified product and that comprise sequence (A') near their 5' ends.

[00141] In some embodiments, a sequencing method comprising performing the method previously described, further comprising the steps of removing the amplified product to render the covalently attached polynucleotides single-stranded, and extending a primer to sequence (A') to produce detectable oligonucleotide fragments characteristic of the sequence of the polynucleotide bound to the bead or isolated area. In some embodiments, the sequencing method comprises cleavable labeled terminators. In some embodiments, the sequencing method comprises pyrophosphate detection. In some embodiments, the sequencing method is an isothermal sequencing method. In some embodiments, the sequencing method comprises cycle sequencing.

[00142] In some embodiments, a method of performing bridge PCR comprising performing the method previously described and further comprising the steps of exposing the amplified products to a solid substrate comprising oligonucleotide sequences attached thereto complementary to the A and B' sequences on the amplified products in the presence of components necessary for polymerase chain reaction, and thermal cycling the system to perform bridge PCR amplification.

IDT Exhibit 1020.0803

[00143] In some embodiments, a method of performing rolling circle amplification comprising performing the method previously described and further comprising the steps of: (o) hybridizing the amplified products to a target nucleic acid comprising regions complementary to A and B' sequences in close proximity; (p) optionally extending the gap with a polymerase enzyme; (q) ligating to form a circular nucleic acid comprising the amplified product, and performing rolling circle amplification by extending a primer that is complementary to a sequence in the circular nucleic acid.

[00144] In some embodiments, the primer is complementary to sequence (A), sequence (B'), or a sequence that was between sequences (A) and (B') in the amplified product. In some embodiments, the primer is an oligonucleotide attached to a solid surface.

[00145] In some embodiments, a method of PCR amplification comprising performing the method previously described and further comprising the steps of amplifying the amplified product using primers complementary to sequences (A) and (B), or using primers complementary to sequences (A') and (B').

[00146] In some embodiments, a method of strand displacement amplification (SDA) comprising performing the method previously described wherein sequences (A) and (B') in the amplified product are designed to be cleaved by a restriction enzyme, and performing strand displacement amplification on the amplified product.

INCORPORATION BY REFERENCE

[00147] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[00148] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[00149] Figure 1 illustrates a method of producing a polynucleotide bound to a solid support wherein the polynucleotide comprises sequences related to a target nucleic acid and comprises a defined sequence (A') at its 3' end.

[00150] Figure 2 illustrates a method of producing a polynucleotide comprising a ligand, sequences related to a target RNA, and a defined sequence (A') at its 3' end. The Figure also illustrates binding such polynucleotide to a solid surface.

[00151] Figure 3 illustrates a method of producing a polynucleotide comprising a ligand, sequences related to a target DNA and a defined sequence (A') at its 3' end. The Figure also illustrates binding such polynucleotide to a solid surface.

[00152] Figure 4 shows a method of producing amplified product from a nucleotide, bound to the bead through a ligand, wherein the amplified product comprises sequences related to the target nucleic acid and a defined sequence (B') at its 3' end.

IDT Exhibit 1020.0804

[00153] Figure 5A shows a method of producing a second primer extension product comprising, from its 5' end, a ligand, a defined sequence (B), a sequence representative of a target polynucleotide, a sequence (A') and a sequence (C'). The Figure also illustrates binding the second primer extension product to a solid surface.

[00154] Figure 5B illustrates an isothermal amplification using a composite primer utilizing the second primer extension product produced as illustrated in Figure 5A bound to the bead.

[00155] Figure 6 illustrates the capture by the solid surface comprising oligonucleotides of sequence (B) of amplification product having defined 3' and 5' sequences by hybridization to (B'), the 3' defined sequence of the amplified product. The Figure also shows the extension of the sequence (B) bound to the solid surface to produce a polynucleotide bound to the solid surface. The Figure also illustrates the use of sequencing primers (A) to sequence the bound polynucleotide.

[00156] Figure 7 shows the capture by the solid surface comprising oligonucleotides of sequence (A') of amplification product having defined 3' and 5' sequences by hybridization to (A), the 5' defined sequence of the amplified product.

[00157] Figure 8 shows how the of amplification product having defined 3' and 5' sequences can be used for analysis of sequence D by hybridization, gap ligation, and rolling circle amplification.

[00158] Figure 9 illustrates an alternative method of producing a polynucleotide comprising a ligand that can be bound to a solid support that comprises sequences related to a target nucleic acid and comprises a defined sequence (E') at its 3' end.

[00159] Figure 10 shows how the polynucleotide produced as shown in Figure 9 can be used to produce amplified product.

[00160] Figure 11 shows how the amplification illustrated in Figure 10 can be carried out while the polynucleotide is bound to a solid surface.

[00161] Figure 12 shows how a chimeric RNA/DNA oligonucleotide (G/E) can be used with the product of the method illustrated in Figure 9 to produce amplified product with defined sequences at its 3' and 5' ends.

[00162] Figure 13 illustrates an alternative method of producing a polynucleotide bound to a solid support wherein the polynucleotide comprises sequences related to a target nucleic acid and comprises a defined sequence (A') at its 3' end.

[00163] Figure 14 illustrates an alternative method of producing a polynucleotide bound to a solid support wherein the polynucleotide comprises sequence B, sequences related to a target RNA, and a defined sequence (A') at its 3' end.

[00164] Figure 15 shows a method of producing amplified product from a polynucleotide, bound to the bead through sequence B, wherein the amplified product comprises sequences complementary to the target nucleic acid and a defined sequence (B') at its 3' end.

[00165] Figure 16 shows an alternative method of producing a polynucleotide bound to a solid support wherein the polynucleotide comprises sequence B, sequences related to a target DNA and a defined sequence (A') at its 3' end. The figure also illustrates production of amplified product from the bound polynulcleotide, wherein the amplified product comprises defined sequence B' at its 3' end.

IDT Exhibit 1020.0805