FIGURE 1

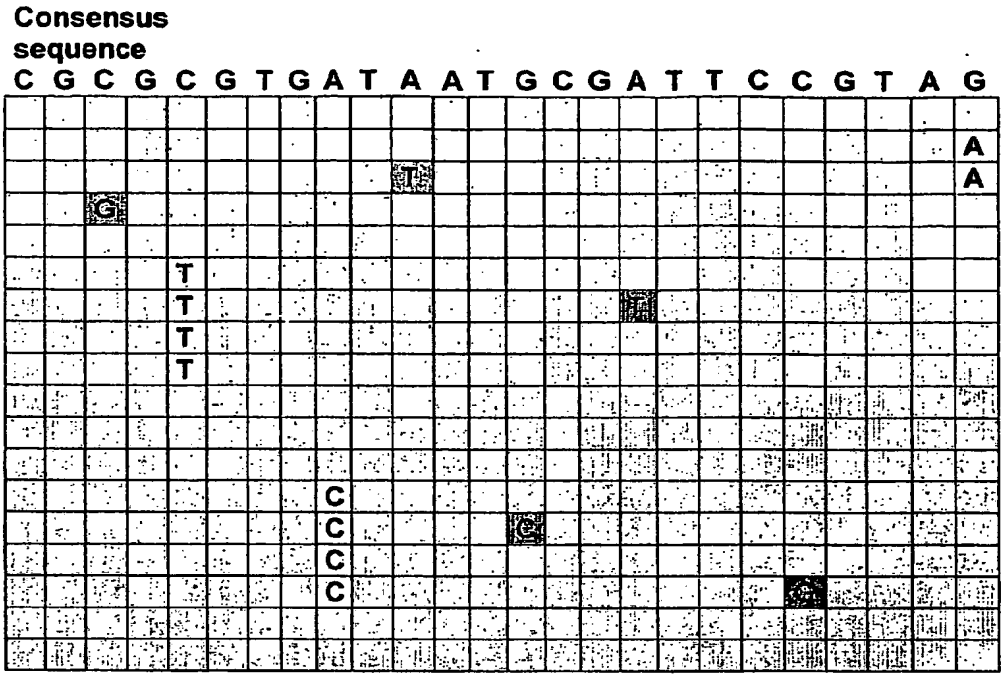
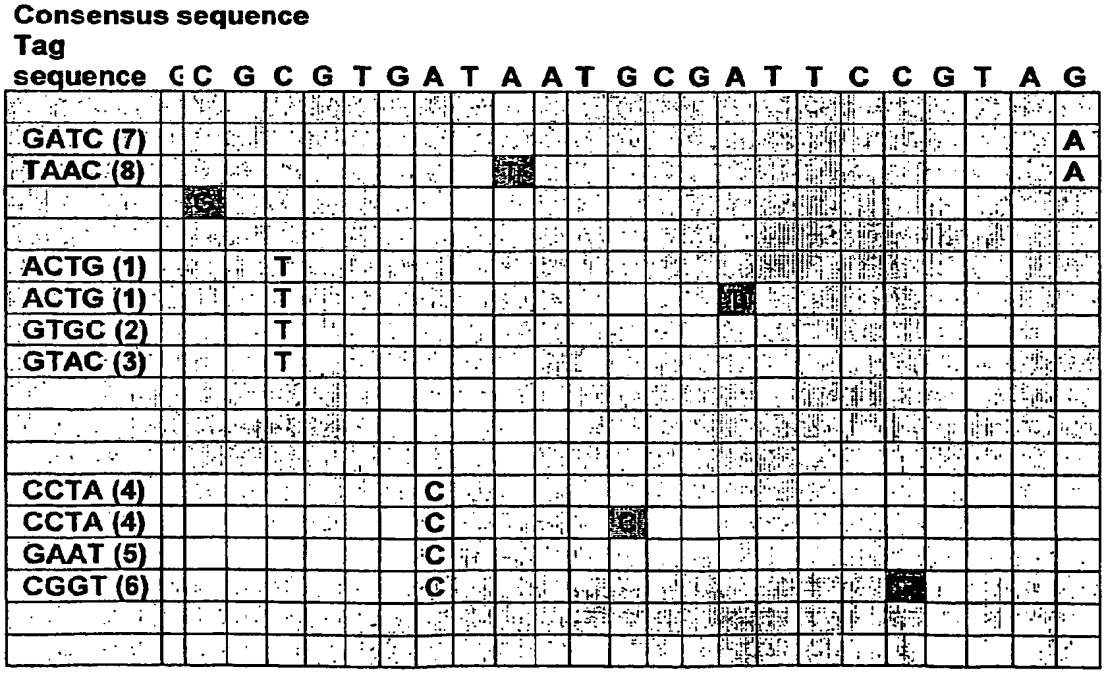


FIGURE 2



IDT Exhibit 1020.6455

Fig. 3

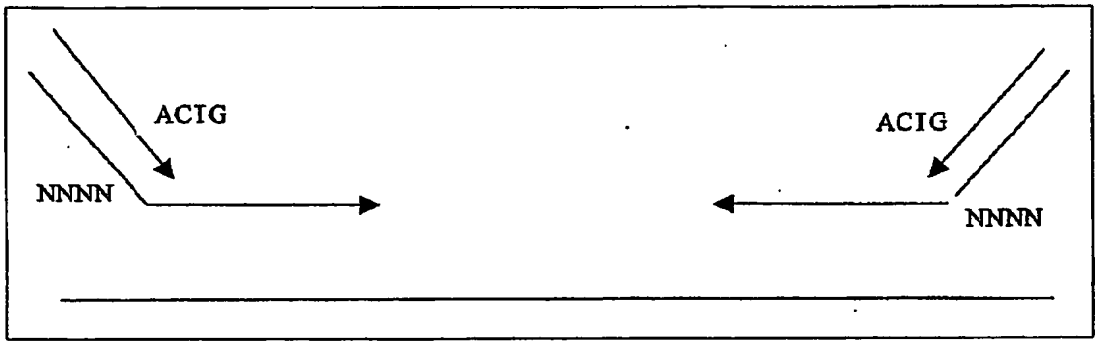
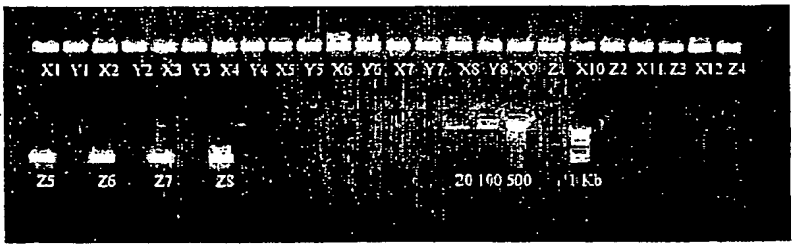


Fig 4



IDT Exhibit 1020.6456

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- WO 03004690 A **[0039] [0046]**
- WO 03054142 A **[0039] [0046]**
- WO 2004069849 A **[0039] [0046]**
- WO 2004070005 A **[0039] [0046]**
- WO 2004070007 A **[0039] [0046]**
- WO 2005003375 A **[0039] [0046]**
- US 60721528 B **[0114] [0115]**

Non-patent literature cited in the description

- **ALTSCHULER et al.** *Nature,* 2000, vol. 407, 513-516 **[0002]**
- **MCCALLUM et al.** *Nat. Biotechnol,* 2000, vol. 18, 455-457 **[0004]**
- **MCCALLUM et al.** *Plant Physiology,* 2000, vol. 123, 439-442 **[0004]**
- **TILL et al.** *Genome Research,* 2003, vol. 13, 524-530 **[0004]**
- **WIENHOLDS et al.** *Methods in Cell Biology,* 2004, vol. 70, 69-90 **[0004]**
- **LI et al.** *The Plant Journal,* 2001, vol. 27, 235-42 **[0004]**
- **SCHIMENTI et al.** *Genome Research,* 1998, vol. 8, 698-710 **[0018]**
- **COMAI et al.** *The Plant Journal,* 2004, vol. 37, 778-786 **[0018]**
- **LEHNINGER.** Principles of Biochemistry. Worth Pub, 1982, 793-800 **[0021]**
- **VANDENBUSSCHE et al.** *The Plant Cell,* 2003, vol. 15, 2680-93 **[0034]**
- **SEO et al.** *Proc. Natl. Acad. Sci. USA,* 2004, vol. 101, 5488-93 **[0039]**
- **SMITH ; WATERMAN.** *Adv. Appl. Math.,* 1981, vol. 2, 482 **[0049]**
- **NEEDLEMAN ; WUNSCH.** *J. Mol. Biol.,* 1970, vol. 48, 443 **[0049]**
- **PEARSON ; LIPMAN.** *Proc. Natl. Acad. Sci. USA,* 1988, vol. 85, 2444 **[0049]**
- **HIGGINS ; SHARP.** *Gene,* 1988, vol. 73, 237-244 **[0049]**
- **HIGGINS ; SHARP.** *CABIOS,* 1989, vol. 5, 151-153 **[0049]**
- **CORPET et al.** *Nucl. Acids Res.,* 1988, vol. 16, 10881-90 **[0049]**
- **HUANG et al.** *Computer Appl. in the Biosci,* 1992, vol. 8, 155-65 **[0049]**
- **PEARSON et al.** *Meth. Mol. Biol.,* 1994, vol. 24, 307-31 **[0049]**
- **ALTSCHUL et al.** *Nature Genet.,* 1994, vol. 6, 119-29 **[0049]**
- **SHENDURE et al.** *Science,* vol. 309, 1728-32 **[0065]**
- **MENDA et al.** *Plant J.,* 2004, vol. 38, 861-872 **[0092]**
- **STUART ; VIA.** *Biotechniques,* 1993, vol. 14, 748-750 **[0093]**
- **DUPRAT et al.** *Plant J.,* 2002, vol. 32, 927-934 **[0095]**
- **NICAISE et al.** *Plant Physiol.,* 2003, vol. 132, 1272-1282 **[0095]**
- **RUFFEL et al.** *Plant J,* 2002, vol. 32, 1067-1075 **[0095]**
- *Mol. Gen. Genomics,* 2005, vol. 274, 346-353 **[0095]**
- **MARGULIES et al.** *Nature,* 2005, vol. 437, 376-380 **[0102]**
- **COLBERT et al.** High-throughput screening for induced point mutations. *Plant Physiology,* 2001, vol. 126, 480-484 **[0113]**
- **DUPRAT et al.** The Arabidopsis eukaryotic initiation factor (iso)4E is dispensable for plant growth but required for susceptibility to potyviruses. *Plant J.,* 2002, vol. 32, 927-934 **[0113]**
- **EPINAT et al.** A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Research,* 2003, vol. 31 (11), 2952-2962 **[0113]**
- **HAVRE et al.** Targeted mutagenesis of DNA using triple helix-forming oligonucleotides linked to psoralen. *Proc.Natl.Acad.Sci. USA,* 1993, vol. 90, 7879-7883 **[0113]**
- **MCCALLUM et al.** Targeted screening for induced mutations. *Nature Biotechnology,* 2000, vol. 18, 455-457 **[0113]**
- **GREENE et al.** Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in Arabidopsis. *Genetics,* 2003, vol. 164, 731-740 **[0113]**
- **LLOYD et al.** Targeted mutagenesis using zinc-finger nucleases in Arabidopsis. *Proc.Natl.Acad.Sci. USA,* 2005, vol. 102, 2232-2237 **[0113]**
- **MARGULIES et al.** Genome sequencing in microfabricated high-density picolitre reactions. *Mature,* 2005, vol. 437, 376-380 **[0113]**

IDT Exhibit 1020.6457

- **MENDA et al.** In silico screening of a saturated mutation library of tomato. *Plant J.,* 2004, vol. 38, 861-872 **[0113]**
- **NICAISE et al.** The eukaryotic translation initiation factor 4E controls lettuce susceptibility to the p otyvirus lettuce mosaic virus1. *Plant Physiol.,* 2003, vol. 132, 1272-1282 **[0113]**
- **RUFFEL et al.** A natural recessive resistance gene against potato virus Y in pepper corresponds to the eukaryotic initiation factor 4E (eIF4E. *Plant J.,* 2002, vol. 32, 1067-1075 **[0113]**
- **RUFFEL et al.** The recessive potyvirus resistance gene pot-1 is the tomato orthologue of the pepper pvr2-eIF4E gene. *Mol. Gen. Genomics,* 2005, vol. 274, 346-353 **[0113]**
- **SHENDURE et al.** Accurate multiplex polony sequencing of an evolved bacterial genome. *Science-express Report,* 04 August 2005 **[0113]**
- **STUART ; VIA.** A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. *Biotechniques,* 1993, vol. 14, 748-750 **[0113]**
- **VANDENBUSSCHE et al.** Toward the analysis of the petunia MADS box gene family by reverse and forward transposon insertion mutagenesis approaches: B, C, and D floral organ identity functions require SEPALLATA-like MADS box genes in petunia. *The Plant Cell,* 2003, vol. 15, 2680-2693 **[0113]**

IDT Exhibit 1020.6458

Espacenet

Bibliographic data: WO03002736 (A2) — 2003-01-09

A WALK-THROUGH TECHNIQUE FOR IN VITRO RECOMTBINATION OF POLYNUCLEOTIDE SEQUENCES

Inventor(s): SHAO ZHIXIN [DE]; KRATZSCH PETER [DE]; SCHMUCK RAINER [DE]; VON DER ELTZ HERBERT [DE]; KENKLIES JANET [DE] ± (SHAO, ZHIXIN, ; KRATZSCH, PETER, ; SCHMUCK, RAINER, ; VON DER ELTZ, HERBERT, ; KENKLIES, JANET)

Applicant(s): ROCHE DIAGNOSTICS GMBH [DE]; HOFFMANN LA ROCHE [CH]; SHAO ZHIXIN [DE]; KRATZSCH PETER [DE]; SCHMUCK RAINER [DE]; ELTZ HERBERT VON DER [DE]; KENKLIES JANET [DE] ± (ROCHE DIAGNOSTICS GMBH, ; F. HOFFMANN-LA ROCHE AG, ; SHAO, ZHIXIN, ; KRATZSCH, PETER, ; SCHMUCK, RAINER, ; VON DER ELTZ, HERBERT, ; KENKLIES, JANET)

Classification: **- international:** ***C07K14/00; C12N15/09; C12N15/10; C12N9/06; C12N9/16; C12P21/02;*** (IPC1-7): C12N15/00
- cooperative: **C12N15/102; C12N9/0034; C12N9/16**

Application number: WO2002EP07060 20020626 Global Dossier

Priority number(s): EP20010115424 20010627

Also published as: WO03002736 (A3) CA2453864 (A1) EP1404825 (A2) JP2004533259 (A) JP3967319 (B2)

Abstract of WO03002736 (A2)

A simple and efficient method for random in vitro mutagenesis and recombination of polynucleo tide sequences based on walk-through with chain-elongating molecules and chain terminating molecules e.g. dNTPs/ddNTPs followed by reassembly and amplification is described. The utility of this method for improving protein structure and/or function is demonstrated by creating novel sarcosine oxidase variants and by recombining human placental alkaline phosphates and calf intestinal alkaline phosphatase for better mutants with improved properties.;A simple and efficient method for random in vitro mutagenesis and recombination of polynucleotide sequences based on walk-through with chain-elongating molecules and chain terminating molecules e.g. dNTPs/ddNTPs followed by reassembly and amplification is described. The utility of

IDT Exhibit 1020.6459

this method for improving protein structure and/or function is demonstrated by creating novel sarcosine oxidase variants and by recombining human placental alkaline phosphates and calf intestinal alkaline phosphatase for better mutants with improved properties.

IDT Exhibit 1020.6460

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
9 January 2003 (09.01.2003)

PCT

(10) International Publication Number
WO 03/002736 A2

(51) International Patent Classification[7]: C12N 15/00

(21) International Application Number: PCT/EP02/07060

(22) International Filing Date: 26 June 2002 (26.06.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
01115424.2 27 June 2001 (27.06.2001) EP

(71) Applicant *(for DE only)*: **ROCHE DIAGNOSTICS GMBH** [DE/DE]; Sandhofer Strasse 116, 68305 Mannheim (DE).

(71) Applicant *(for all designated States except DE, US)*: **F. HOFFMANN-LA ROCHE AG** [CH/CH]; Grenzacherstrasse 124, CH-4070 Basel (CH).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **SHAO, Zhixin** [CN/DE]; Westend 43, 82377 Penzberg (DE). **KRATZSCH, Peter** [DE/DE]; Dorfstrasse 5, 82387 Antdorf (DE). **SCHMUCK, Rainer** [DE/DE]; Am Klosterweiher 7, 83671 Benediktbeuern (DE). **VON DER ELTZ, Herbert** [DE/DE]; In der Au 21, 82362 Weilheim (DE). **KENKLIES, Janet** [DE/DE]; Am alten Bahnhof 1, 82377 Penzberg (DE).

(81) Designated States *(national)*: CA, JP, US.

(84) Designated States *(regional)*: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).

Declarations under Rule 4.17:
— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations CA, JP, European patent (AT, BE, CH, CY, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR)*
— *of inventorship (Rule 4.17(iv)) for US only*

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/002736 A2

(54) Title: A WALK-THROUGH TECHNIQUE FOR *IN VITRO* RECOMBINATION OF POLYNUCLEOTIDE SEQUENCES

(57) Abstract: A simple and efficient method for random *in vitro* mutagenesis and recombination of polynucleotide sequences based on walk-through with chain-elongating molecules and chain terminating molecules e.g. dNTPs/ddNTPs followed by reassembly and amplification is described. The utility of this method for improving protein structure and/or function is demonstrated by creating novel sarcosine oxidase variants and by recombining human placental alkaline phosphates and calf intestinal alkaline phosphatase for better mutants with improved properties.

IDT Exhibit 1020.6461

A walk-through technique for *in vitro* recombination of polynucleotide sequences

A simple and efficient method for random *in vitro* mutagenesis and recombination of polynucleotide sequences based on walk-through with chain-elongating molecules and chain terminating molecules e.g. dNTPs/ddNTPs followed by reassembly and amplification is described. The utility of this method for improving protein structure and/or function is demonstrated by creating novel sarcosine oxidase variants and by recombining human placental alkaline phosphatase and calf intestinal alkaline phosphatase.

Various optimization procedures such as genetic algorithms (Holland, J. H. 1975. Adaptation in natural and artificial systems. The University Press, Ann Arbor, Goldberg, D. E. 1989. Genetic algorithms in search, optimization and machine learning. Addison-Wesley. Reading) and evolutionary strategies (Eigen, M. 1971. *Naturwissenschaften 58*: 465-523, Rechenberg, I. 1973. Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution. Frommann-Holzboog, Stuttgart) have been inspired by natural evolution. These procedures employ mutation, which makes small random changes in each single member of the population, as well as crossover, which combines properties of different individuals, to achieve a specific optimization goal. There also exist strong interplay's between mutation and crossover, as shown by computer simulations of different optimization problems (Brady, R. M. 1985. *Nature* 317: 804-806, Muhlenbein, H. 1991. *Parallel Computing 17*:619-632, Pál, K. F. 1993. *Bio. Cybern. 69*: 539-546, Pál, K. F. 1995. *Bio.Cybern. 73*, 335-341). Developing efficient and practical experimental techniques to mimic these key processes is a scientific challenge. The application of such techniques should allow us, for example, to explore and optimize the functions of biological molecules such as proteins and nucleic acids, *in vivo* or even completely free from the constraints of a living system (Joyce, G. F. 1992. *Scientific American*, 267:90-97, Shao, Z. and Arnold F. H. 1996. *Curr. Opin. Struct. Biol.* 6: 513-518).

Proteins are engineered with the goal of better understanding the molecular basis for their functions as well as much improving their performance for practical applications. A variety of approaches, involving both 'rational' and 'irrational' design, have been successfully used to optimize protein func-

tions (Shao, Z. and Arnold F. H. 1996. *Curr. Opin. Struct. Biol. 6*: 513-518). The choice of approach for a given optimization problem will depend upon the degree of understanding of the relationships between sequence, structure and function. The rational redesign of an enzyme's catalytic site, for example, often requires extensive knowledge of the enzyme structure, the structures of its complexes with various ligands and analogs of reaction intermediate and details of the catalytic mechanism. Such information is available only for a very few well-studied systems; little is known about the vast majority of potentially interesting enzymes.

For optimizing protein functions, several 'irrational' optimization approaches are being used (recently reviewed by Shao, Z. and Arnold F. H. 1996. *Curr. Opin. Struct. Biol. 6*: 513-518). These optimization procedures do not require much specific knowledge about the enzyme itself, only values of its function to be optimized. These procedures are even fairly tolerant to inaccuracies and noise in the function evaluation (Moore, J. C. and Arnold, F. H. 1996. *Chem. Eng. Sci. 51*:5091-5102). A particularly promising approach is to use both mutation and crossover synchronously to probe populations of solutions simultaneously for a specific optimization problem. Through multiple generations of screening or selection, bad local optima are evaded and better and better solutions are "bred" (Stemmer, W. P. C. 1994a. *Nature*, 370: 389-391, 1994; Stemmer, W. P. C. 1994b. *Proc. Natl. Acad. Sci.*, USA, 91:10747-10751). During the past few years, more and more people have recognized the merits of evolutionary search strategies. Only a few practical techniques, however, such as

- sequential mutagenic PCR (Moore, J. C. and Arnold, F. H. 1996. *Chem. Eng. Sci. 51*:5091-5102),
- combinatorial cassette mutagenesis (Reidhaar-Olson, J. F. and Sauer, R. T. 1988. *Science 241*: 53-57),
- 'DNA shuffling' (Stemmer, W. P. C. 1994a. *Nature, 370*: 389-391; Stemmer, W. P. C. 1994b. *Proc. Natl. Acad. Sci., USA, 91*:10747-10751),
- random-priming recombination (Shao, Zhixin, H. Zhao, Lori Giver and Frances H. Arnold. 1998. *Nucleic Acids Res. 26(2):* 681-683),
- staggered-extension (StEP). H. Zhao et al. 1998, *Nature Biotechnol. 16(3):*258-2661),
- artificial recombination (Shao, Zhixin. 1999. *Protein Science 8(Suppl.2)*: 55; Volkov, Alexander, Zhixin Shao, and Frances H. Arnold. *1999 Nucleic Acids Res. 27(18)*e18:i-vi)

have been successfully applied to these problems.

One of the most commonly used technique of evolutionary optimization in procedure is "DNA-shuffling" (US 5,834,252; US 5,605,793; US 5,830,721; US 5,837,458; US 5,811,238) DNA-shuffling comprises the following steps:

1) creation of random double-stranded fragments from different double-stranded template polynucleotides wherein said different template polynucleotides contain preferably areas of identity and areas of heterology
2) reassembling of the random fragments
3) cloning
4) expression and
5) screening.

The different double-stranded template polynucleotides may belong to the same family of nucleic acids or proteins (i.e. are related) but which differ in their sequence (i.e. are not identical) and hence in their biological activity. The use of DNAse I is described for random fragmentation in the prior art.

The DNA-shuffling using DNAse fragmentation described by Stemmer et al. has a number of disadvantages. First of all, DNAse I cleaves the DNA non-randomly. During an early phase of reaction cleavage occurs preferably in the middle of a DNA sequence and in the later phase of fragmentation cleavage occurs preferably between purine – and pyrimidine analogues. This leads to a fragmentation procedure which can be hardly controlled. Furthermore the minimal length of the desired gene is limited because of the non-controllable digestion of the template polynucleotide. Moreover, DNAse has to be removed completely after the digestion step because it introduces disturbance into the further reassembly reaction.

The aim of the present invention was, therefore, developing a novel technique for the recombination of DNA sequences.

The subject of the present invention is a novel technique for *in vitro* walk-through recombination of DNA sequences. The technique involves walking through a template gene with a mixture of chain-elongating molecules and chain-terminating molecules, e.g. dNTPs/ddNTPs, to generate a pool of 3'

end-randomly distributed DNA fragments with a low level of point mutations. Thus, a fragment ladder is generated similar to fragment ladders during sequencing reactions. After removing the terminating molecules, e.g. ddNMP ends, these short DNA fragments can prime one another under appropriate reaction conditions based on homology and thus can be reassembled to form full-length genes by repeated thermocycling in the presence of thermostable DNA polymerase. These genes can be further amplified by a conventional PCR and cloned into a proper vector for expression of the encoded proteins. Screening or selection of the expressed mutants leads to new variants with improved or even novel functions. These variants can be immediately used as partial solutions to a practical problem, or they can serve as new starting points for further cycles of walk-through mutagenesis and recombination. This technique has been proven by creating *Bacillus subtilis* BMTU3420 sarcosine oxidase variants and by recombining human placental alkaline phosphatase and calf intestinal alkaline phosphatase. It was found that the technique is both – simple and efficient.

Therefore, subject of the present invention is a method for forming a polynucleotide sequence comprising the steps of

a) generating a nucleic acid fragment ladder by nucleic acid synthesis which is carried out in presence of a reaction mixture comprising at least one primer, an enzyme having nucleic acid-synthesizing activity and a mixture of nucleic acid chain-terminating and nucleic acid chain-elongating molecules and a template polynucleotide,

b) removing the chain-terminating molecules or changing it into non-terminating molecules,

c) reassembly of the polynucleotide by hybridizing these fragments to one another or to template polynucleotides in presence of a thermostable enzyme having nucleic acid-synthesizing activity and conducting nucleic acid synthesis either
   i) in presence of a mixture of nucleic acid chain-terminating molecules and nucleic acid chain elongating molecules or
   ii) in presence of a reaction mixture of nucleic acid chain elongating molecules which does not contain chain-terminating molecules
   whereas in case of i) steps b) and c) will be repeated subsequently.

The step of removing chain-terminators and the step of fragment reassembly step can also be done at the same time. The subdivision of the process in steps a, b, c does not imply that three separate steps have to be performed. The three steps may also take place at the same time. It is also possible that the whole WalkThrough recombination may be performed in more preferred forms, such as in a single container, vessel or tube.

Even though one primer is enough for the fragment synthesis (or chain elongation), two or more primers are necessary for different fragment synthesis reactions in order to get correctly orientated fragments for the afterwards reassembly. That is, two or more primer are needed for the whole walk through process.

Only one template is enough for the WalkThrough process, as diversity can be introduced during the chain synthesis and fragment reassembly (resulting from, for example, synthesis errors of DNA polymerase). That means in case the template polynucleotide is not a mixture of different polynucleotides the DNA fragments may contain a low level of local mutations due to some synthesis errors and thus new sequences will be created.

Using more than one template may significantly increase the diversity of the variant library, and, consequently, finding potentially positive variants may be easier. Preferably, the template polynucleotide in step a) and/or c) is a mixture consisting of different template polynucleotides which may belong to the same family of nucleic acids or proteins (i.e. related) but which differ in their sequence (i.e. not identical) and hence differ in their biological activity. In this case gene shuffling may occur.

Conditions for the WalkThrough technique may vary from case to case and can be optimized at three main different levels:

(a) fragment synthesis in the presence of chain-terminators,
(b) Removing the chain terminators incorporated, and
(c) Reassembling the fragments from the step (b).

a) Fragment synthesis in the presence of chain-terminators
This step is somehow similar to normal polynucleotide sequencing reaction, and, therefore, various conditions suitable for sequencing different targets may be applied to this step.

To carry out the WalkThrough procedure, the template can be single- or double-stranded polynucleotides in linear or closed circular form. The template may also be in the form of genomic DNA, that is, in native, intact, unpurified, or uncloned forms. Since, in most cases, the template genes are cloned in vectors into which no additional mutations should be introduced, they are usually first cleaved with restriction endonuclease(s) and purified from the vectors through agarose gel electrophoresis. For example linear DNA molecules are denatured, annealed to oligodeoxynucleotides for walkthrough reactions in the presence of an appropriate amount of chain-elongating/chain-terminating molecules as e.g. dNTPs/ddNTPs. Thus the oligonucleotides prime the DNA of interest at different positions along the entire target region extend to generate DNA fragments complementary to each strand of the template DNA. Due to some synthesis errors, these DNA fragments also contain a low level of local mutations. After removing the chain-terminating ends of ddNMP's, these DNA fragments can prime one another under appropriate reaction conditions based on homology and be reassembled into full-length genes by repeated thermocycling in the presence of thermostable DNA polymerase. The resulting full-length genes will have diverse sequences, most of which, however, still resemble that of the original template DNA.

The fragment synthesis can be carried out with either chain elongation under mild conditions using mesophilic DNA polymerases or thermocycle sequencing using thermostable polymerase. To adapt PCR to the WalkThrough synthesis provides a more convenient way for more nascent DNA fragments and makes our technique more robust.

If thermocycling is used for the fragment synthesis, the number of thermocycling can be from about 15 to about 55 cycles, depending on the amount of template and its purity. The concrete cycle number can be easily determined according to the quality and quantity of the newly synthesized fragments.

The amount of initial DNA template(s) may vary according to the chain elongation conditions. Normally, 0.1-10 pmol of ds-DNA is sufficient for this kind of reaction.

The length of the primers used may vary from 14-mer to 28-mer. The lengths of the selected primers should be long enough to prevent annealing to unspecific DNA binding positions and guarantee a good signal-to-noise ratio. We usually use primers of 18- to 21-mer.

The distances between the WalkThrough primers depend on the chain reaction modes and may vary from 100 to 800 nucleotides. The average distance between two primers in our cases is normally around 500 nucleotides.

The concentration of each primer can vary from 0.05 pmol to 2 pmol, depending on what kind of chain reaction is used. Normally, 0.1 pmol primer is used for the chain elongation of *B. subtilis* sarcosine oxidase gene, while 1.0 pmol each primer is used for the chain elongation of human placental alkaline phosphatase and calf intestinal alkaline phosphatase.

There are dozen of polymerases currently available, the synthesis of the nascent DNA fragments with randomly distributed 3' end can be achieved in more different fashions. For example, bacteriophage T4 DNA polymerase (Nossal 1974) or T7 sequenase version 2.0 DNA polymerase (Tabor and Richardson 1987, 1989), and /or thermostable DNA polymerases can be used for the WalkThrough synthesis. More preferably, other modified or mixture of DNA polymerases can be used which incorporate ddNTPs more efficiently than the wild type polymerases. For single-stranded RNA template, reverse transcriptase is preferred for the WalkThrough synthesis. Each of these enzymes is usually used under its optimized conditions to promote the chain elongation. Enzymes having nucleic acid-synthesizing activity can be DNA or RNA dependent polymerases.

It has been known for long time that different DNA polymerases can use 2', 3'-dideoxynucleotide triphosphates (ddNTPs) as substrates (Sanger, F., Nicklen, S., and Coulson, A.R. 1977. *Proc. Natl. Acad. Sci. U.S.A. 74*:5463-5467, Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.M., and Roe, B.A. 1980. *J. Mol. Biol. 143*:161-178) for DNA synthesis on single-stranded templates (Klenow, H. and I. Henningsen. 1970. *Proc. Natl. Acad. Sci. 65*:168).

When a ddNMP is incorporated at the 3' end of the growing chain, chain elongation is terminated at A, C, G, or T because the chain now lacks a 3' hydroxyl group. To generate the four chain ladders, only one of the four possible ddNTPs is included in each of the four reactions. The ddNTP/dNTP

ratio in each reaction is adjusted such that a portion of the elongating chains terminate at each occurrence of the base in the template corresponding to the included complementary ddNTP. In this way, each of the four elongation reactions contains a population of extended chains, all of which have a fixed 5' end determined by the annealed primers and a variable 3' end terminated at a specific dideoxynucleotide. The chain elongation and termination can also be done in one reaction in the presence of proper amount of ddNTPs/dNTPs.

Chain terminating molecules means that these molecules terminate nucleic acid synthesis and, therefore, a fragment ladder is generated. After generation of the fragment ladder suitable chain-terminating molecules can be removed. Suitable chain-terminating nucleotides are e.g. dideoxynucleotides as ddGTP, ddATP, ddCTP, ddTTP or derivatives thereof. Derivatives of dideoxynucleotides are defined as those dideoxynucleotides that are able to be incorporated by a thermostable DNA polymerase into growing DNA molecules that are synthesized in a thermocycling reaction. These dideoxynucleotides and derivatives are preferably used at a concentration of about 20 μM to 1.0 mM. Such derivatives can include thionucleotides, 7-deaza-2'-dGTP, 7-deaza-2'-dATP as well as deoxyinosine triphosphate that can also be used as a substitute deoxynucleotide for dATP, dGTP, dTTP or dCTP, but are not limited to these. These deoxynucleotides and derivatives are preferably used at a concentration of about 4 μM to 400 μM.

The concentrations and the ratios of the ddNTP/dNTP may vary from case to case in order to get shorter or longer fragments. Though a ratio of dNTP/ddNTP between 1/50 and 5/1 works well for the chain elongation, it has to be optimized for different individual template in order to get fragment pools with randomly distributed 3'-ends. This is very important for allowing every nucleotide of the template should be copied at a similar frequency into products, and therefore, providing possibility to recombine or dissect two or more mutations although they may be very close to each other.

Preferred WalkThrough buffer systems normally include Tris-HCl at a concentration of about 50 to 500 mM, preferably of about 100 to 250 mM. The pH values of these buffer systems range from pH6.5 to pH 10.0, depending on polymerases used. $MaCl_2$ is generally included in these buffer systems at a concentration ranging from 1.0 to 5.0 mM. KCl may also be included at a concentration of 2-80 mM. Certain amount of mercaptoethanol (0.5-1.5%), Tween 20 (0.2-0.4%) and DMSO (1-5%) may also exist in the buffers. Other agents, such as glycerol, betaine, etc., which can lower the

melting point of the templates may also be included in the buffers to facilitate the WalkThrough reaction.

The Walk-Through DNA synthesis is based on the chain elongation guided by template, and the nascent strand is synthesized from the 3'-OH termini at the primers using polymerase and the four deoxynucleoside triphosphates and stop after the ddNTP incorporation. Thus the reaction is independent of the length of the DNA template.

(b) Removing the chain terminators incorporated
This step is the second key step for successful WalkThrough recombination. The end nucleotide or incorporated terminator at each newly generated fragments must be removed so that each such fragment gains a free 3'-OH group necessary for further fragment reassembly.

Methods of removing chain-terminating molecules are known to a person skilled in the art. Chain-terminating molecules can be removed, for example, by DNA polymerases, many DNA polymerases have a 3'-5' exonuclease activity. This activity removes a single nucleotide at a time, releasing a nucleotide 5' monophosphate. In the absence of dNTPs, this activity will catalyze stepwise degradation from 3' end of both single- and double-stranded DNA. The polymerase which may serve this purpose include Klenow fragment of E. coli DNA polymerase I (Jacobsen, H., Klenow, H., and Overgaard-Hansen, K. 1974. The N-terminal amino-acid sequences of DNA polymerase I from Escherichia coli and of the large and the small fragments obtained by a limited proteolysis. Eur. J. Biochem. 45:623-627. Joyce, C.M. and Gridley, N.D.F. 1983. Construction of a plasmid that overproduces the large proteolytic fragment (Klenow fragment) of DNA polymerase I of Escherichia coli. Proc. Natl. Acad. Sci. USA 80:1830-1834), bacteriophage T4 DNA polymerase (Nossal, N.G. 1984, Prokaryotic DNA replication systems. Annu. Rev. Biochem. 53:581-615. Lin, T.C., Rush, J., Spicer, E. K., and Konigsberg, W.H. 1987. Cloning and expression of T4 DNA polymerase. Proc. Natl. Acad. Sci. USA 84:7000-7004), bacteriophage T7 DNA polymerase (Tabor, S. and Richardson, C.C. 1987. DNA sequencing analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA. 84:4767-4771). The 3'-end terminators may be removed from the fragments with DNA polymerase with 3'→5' proof reading activity in the presence of very limited or no dNTPs under mild conditions. Alternatively, chain-terminating molecules can also be removed by exonucleases, such as Exonuclease III. This enzyme is a multifunctional enzyme that catalyzes hydrolysis of several types of

phosphodiester bonds in double stranded DNA. The main application of Exo III is as a 3'-5' double-stranded specific exonuclease that catalyzes release of 3' nucleotides from the 3'- end of double stranded DNA (Roger, S.G. and Weiss, B. 1980. Exonuclease III of Escherichia coli K-12, an AP exonuclease. Meth. Enzymol. 65:201-211).

Preferred reaction buffers used for terminator-removing reactions normally include Tris-HCl at a concentration of about 40 to 200 mM, preferably of about 50 to 100 mM. The pH values of these buffer systems range from pH 6.5 to pH9.0, depending on enzymes with 3'-5' exonuclease activity used. $MgCl_2$ is generally included in these buffer systems at a concentration ranging from 1.0 to 3.0 mM. KCl may also be included at a concentration of 5-50 mM. Certain amount of mercaptoethanol (0.5-1.5%) may also exist in the buffers.

The reaction normally is carried out at 37°C for 0.5-2.0 hours.

Using thermostable DNA polymerases and thermostable enzymes with 3'->5' exonuclease activity in the proper buffer and under the optimized conditions, the steps of terminator removal and re-assembly can also be combined, so that no separate (B) and (C) steps are necessary. This method of the present invention is particularly preferred, as demonstrated in the example of recombining human placental alkaline phosphatase and calf intestinal alkaline phosphatase.

(c) Reassembling the fragments

The reassembling of the fragments is a PCR like reaction, comprising cycles of DNA-denaturation, annealing and DNA synthesis. During the annealing ssDNA-fragments hybridize in homologues areas. The overlapping ends of the ssDNA are extended by a polymerase. In case ssDNA random fragments from different polynucleotides hybridize, gene-shuffling occurs. Gene-shuffling means recombination between homologues but non-identical sequences. The term "identical" means that two nucleic acid sequences have the same sequence or a complementary sequence. Thus, "areas of identity" means that regions or areas of a nucleic acid fragment or polynucleotide are identical or complementary to another polynucleotide or nucleic acid fragment. The term "homologues" means that one single-stranded (ss) nucleic acid sequence may hybridize to a complementary ss nucleic acid sequence. The degree of hybridization may depend on a number of factors including the amount of

identity between sequences and the hybridization conditions such as temperature and salt concentration as discussed later.

The reassembly step is the most critical one in the whole WalkThrough process, as all synthesized fragments during Step (A) prime one another for elongation without any additional primers added.

Since the concentration of the DNA is the most important variable, it is useful to set up three separate reactions with different concentrations (high, middle, low). According to our experience, the amount of DNA fragments between 0.1 μg to 2.0 μg usually gives satisfied reassembly results in a reaction volume of 20-50 μl.

The deoxynucleotides during reassembly are preferably used at a concentration of about 100 μM to 400 μM, and the chosen concentration depends on the cycle number of the reaction.

Preferred reassembly buffers normally include Tris-HCl at a concentration of about 5 to 50 mM, preferably of about 10 mM. The pH values of these buffer systems range from pH 7.5 to pH 10.0, depending on polymerases used. $MgCl_2$ is generally included in these buffer systems at a concentration ranging from 1.0 to 5.0 mM. KCl may also be included at a concentration of 20-80 mM. Certain amount of mercaptoethanol (0.5-1.5%), Tween 20 (0.2-0.4%) or Triton-X 100 (0.1-0.5%) may also exist in the buffers.

A typical reassembly cycle consists of three steps: the first step is heat denaturation of the double-stranded target nucleic acid. The exact conditions required for denaturation of the sample nucleic acid depend on the length and composition of the sample nucleic acid. Typically, an incubation at 90°C-100°C for about 10 seconds up to 5 minutes is efficient to denature the sample nucleic acid. The annealing temperature used in reassembly reaction is about 40°C to 70°C, usually ranging from about 55°C to 65°C and lasting for a period of 15 second to 60 seconds. The elongation is usually done under conditions sufficient to provide for polymerization of nucleotides to the fragment ends. To achieve polymerization conditions, the temperature of the reaction mixture will typically be maintained at a temperature ranging from about 65°C to 75°C, more preferably at 68°C to 72°C for

about 15 second to 2 minutes, preferably for 30 seconds to 1 minute depending on the length of the finally reassembled DNA.

The number of thermocycling can be from about 15 to about 55 cycles, depending on the amount of template and the length of the finally reassembled DNA. The concrete cycle number can be easily determined according to the quality and quantity of the newly synthesized fragments.

Several thermostable polymerases with different synthesis fidelity can be used for the reassembly, depending on what kind of the error rate the final reassembly product should have. Each of these enzymes is usually used under its optimized conditions to promote the reassembly. The goal is that the resulting full-length genes will have diverse sequences, most of which, however, still resemble that of the original template DNA.

These sequences obtained after the reassembling step can be further amplified by a conventional PCR and cloned into a vector for expression. Suitable vectors are known to a person skilled in the art and vectors cited in the following references are herein incorporated by reference Kingsman SM, Kingsman AJ. Philos Trans. R. Soc. Lond B. Biol. Sci. 1989, 324(1224):477-485; Bailey JE. Adv. Biochem. Engl. Biotechnol. 1993, 48:29-52. Suitable expression systems are known in the art as well and expression systems cited in the following references are herein incorporated by reference Shatzman AR, Rosenbrg M. Methods Enzymol. 1987; 152:661-73. Screening or selection of the expressed mutants should lead to variants with improved or even new specific functions. Suitable screening and selection systems are known in the art and screening and selection systems cited in the following references are herein incorporated by Kuchner O, Arnold FH. Trends Biotechno. 1997 Dec.; 15(12):523-30; Patel PH, Loeb LA. Procc. Natl. Acad. Sci. USA, 2000 May 9; 97(10):5095-100. These variants can be immediately used as partial solutions to a practical problem, or they can serve as new starting points for further cycles of directed evolution.

Compared to other techniques used for protein optimization, such as combinatorial cassette and oligonucleotide-directed mutagenesis (Reidhaar-Olson, J. F. and Sauer, R. T. 1988. Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. Science 241: 53-57; Oliphant, A. R., Nussbaum, A. L., and Struhl, K. 1986. Cloning of random-sequence oligodeoxynucleotides. Gene 44: 177-183, Hermes, J. D., Blacklow, S. C., and Knowles, J. R. 1990. Searching

sequence space by definably random mutagenesis - improving the catalytic potency of an enzyme. Proc. Natl. Acad. Sci. USA 87: 696-700), error-prone PCR (Leung, D. W., Chen, E., and Goeddel, D. V. 1989. A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. BioTechnique 1: 11-15; Chen K. and Arnold, F. H. 1993 . Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. Proc. Natl. Acad. Sci. USA 90:5681-5622), or DNA 'shuffling' (Stemmer, W. P. C. 1994a. Rapid evolution of a protein *in vitro* by DNA shuffling. Nature, 370: 389-391; Stemmer, W. P. C. 1994b. DNA shuffling by random fragmentation and reassembly -*in vitro* recombination for molecular evolution. Proc. Natl. Acad. Sci., USA, 91:10747-10751), this walk-through based technique shows several advantages for *in vitro* protein optimization:

(1) Since the template used for walkthrough synthesis may be either single- or double-stranded polynucleotides, one limitation of DNA 'shuffling' (Stemmer, W. P. C. 1994a. Rapid evolution of a protein *in vitro* by DNA shuffling. Nature, 370: 389-391;), which has to employ double-stranded polynucleotide as a template, does not exist any more. With the technique described here, any potential mutations and/or crossovers can be introduced at DNA level by using different DNA-dependent DNA polymerases, or even directly from mRNA level by using different RNA-dependent DNA polymerases. This provides more opportunities and realities to achieve the goal of optimizing protein specific functions.

(2) Different from the DNA 'shuffling' procedure which requires fragmenting the double-stranded DNA template with DNAse I to get random fragments (Stemmer, W. P. C. 1994a. *Nature*, 370: 389-391; Stemmer, W. P. C. 1994b. *Proc. Natl. Acad. Sci., USA, 91*:10747-10751), the technique described here employs WalkThrough synthesis to obtain size-controllable DNA fragments with randomly distributed 3' end as "breeding blocks" for further reassembly (Fig. 1).

(3) Since the WalkThrough chains are a population of fragments that each stops in every position, they are uniform in their positional preference and lack a sequence bias. The sequence heterogeneity allows both, mutations and crossover may happen more randomly than, for example, with error-prone PCR or DNA 'shuffling'.

(4) Normal error-prone PCR and DNA shuffling can not efficiently recombine or dissect two or more mutations if they are very close to each other. (Stemmer, W. P. C. 1994a. Rapid evolution of a protein in vitro by DNA shuffling. Nature, 370: 389-391). In contrast, Walk Through approach allows recombination occuring at every position of templates and therefore, provides possibility to recombine or dissect two or more mutations although they may be very close to each other.

(5) The Walk-Through DNA synthesis is based on the chain elongation guided by template, and the nascent strands are synthesized from the 3'-OH termini at the primers using polymerase and the four deoxynucleoside triphosphates and stop after the ddNTP incorporation. Thus the reaction is independent of the length of the DNA template. This is particularly useful for engineering small peptides or large enzymes or even enzyme pathways.

(6) Since DNase I is an endonuclease that hydrolyzes double-stranded DNA preferentially at sites adjacent to pyrimidine nucleotides, its use in DNA shuffling (Stemmer, W. P. C. 1994a.. *Nature*, 370: 389-391; Stemmer, W. P. C. 1994b. *Proc. Natl. Acad. Sci.*, USA, 91:10747-10751) may result in bias (particularly for genes with high G+C or high A+T content) at the step of template gene digestion. Effects of this potential bias on the overall mutation rate and recombination frequency have not yet been investigated, but they may be avoided by using the Walk-Through approach.

(7) Since there are dozen of polymerases currently available, the synthesis of the nascent DNA fragments with randomly distributed 3' end can be achieved in more different fashions. For example, bacteriophage T4 DNA polymerase (Nossal, N.G. 1974. J. Biol. Chem. 249: 5668-5676) or T7 sequenase version 2.0 DNA polymerase (Tabor, S. and Richardson, C. C. 1987. Proc. Natl. Acad. Sci., USA, 84:4767-4771, Tabor, S. and Richardson, C. C. 1989. J. Biol. Chem. 264:6447-6458), and /or thermostable DNA polymerases can be used for the WalkThrough synthesis.

For single-stranded polynucleotide template (particularly for RNA template), reverse transcriptase is preferred for the WalkThrough synthesis. Since this enzyme lacks 3'→5' exonuclease activity, it is therefore prone to error. In the presence of high concentrations of dNTPs and

Mn^{2+}, about 1 base in every 500 is misincorporated (Sambrook, J., Fritsch, E. F. and Maniatis, T. 1989. Molecular cloning: A Laboratory Manual. 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

(8) One of the key steps in our technique is to control the 3' end of the nascent, single-strand DNA synthesized during the WalkThrough process. Under certain conditions, this step may be used for efficient terminal and/or internal insertion/deletion, resulting in molecules with different sizes. Efficient changing target molecule sizes is normally unachievable through error-prone PCR or DNA shuffling.

(9) By modifying reaction conditions, PCR can be adjusted for the WalkThrough synthesis using thermostable polymerase for the short, nascent DNA fragments. To adapt PCR to the Walk Through synthesis provides a more convenient way for more nascent DNA fragments and makes our technique more robust.

The inventive method of the present invention may, however, be combined with other known methods of the art if this seems to be advantageously.

WalkThrough DNA synthesis was used to generate DNA fragments with randomly distributed 3' ends from denatured, linear, double-stranded DNA (e.g., restriction fragments purified by gel electrophoresis) to Bacillus subtilis BMTU3420 sarcosine oxidase gene. The purified DNA, mixed with a molar excess of primers, was denatured, and synthesis was then carried out using the Stoffel fragment. This enzyme lacks 5'→3' exonuclease activity, so that the WalkThrough product was synthesized exclusively by extension and was not degraded by exonuclease (see Example 1).

The WalkThrough recombination technique is further demonstrated with another example of recombining human placental alkaline phosphatase and calf intestinal alkaline phosphatase. (see Example 2). In example 2, the step of removing chain-terminators and the step of fragment reassembly step can be done at the same time.

The technique described here may be used to explore the vast space of potentially useful catalysts for their optimal performance in a wide range of applications as well as to develop or evolve new enzymes for basic structure-function studies.

Most of the experimental conditions described here may potentially be used with little modification for optimizing native enzymes, part (or all) of the enzyme pathways, as well as other biomolecules. For example, in terms of reducing proteolytic activity, probing activator genes, enhancing secretion mechanisms, or solving other fundamental enzyme engineering problems, the technique described here may find immediate applications.

While this protocol describes using DNA-dependent DNA polymerase and single-stranded DNA as template, alternative protocols are also feasible for using single-stranded RNA as a template. By using specific enzyme mRNA as the template and RNA-dependent DNA polymerase (reverse transcriptase) as the catalyst, our approach may be modified to introduce mutations and crossovers into cDNA clones and to create molecular diversity directly from mRNA level to achieve the goal of optimizing enzyme functions. Reverse transcriptases are derived from retrovirus, such as avian myeloblastosis virus (AMV) or Moloney murine leukemia virus (MMLV), which use them to make DNA copies of their RNA genomes. AMV and MMLV reverse transcriptases (Verma IM. The reverse transcriptase. Biochim. Biophys. Acta. 1977 Mar. 21; 473(1):1-38. Roth MJ. Tanese N. Goff SP. Gene product of Moloney murine leukemia virus required for proviral integration is a DNA-binding protein. J. Mol. Biol. 1988; 203(1):131-9) are mainly used as RNA-directed DNA polymerase. Specifically, deoxyoligonucleotides are used as primers for extension on RNA (usually messenger RNA) templates to generate complementary DNA.

Figure 1:
Walking through the single-stranded polynucleotide template with dNTPs/ddNTPs to generate fragments with randomly-distributed 3' end, removing ddNMP from the 3' end, reassembly of the full length DNA by thermocycling in the presence of DANN polymerase and nucleotides and amplification of reassembled products by conventional PCR for further cloning and screening.

Figure 2:
Primer arrangement along human placental alkaline phosphatase (hpap) and calf intestinal alkaline phosphatase (ciap) genes

Figure 3:
The N-terminal sequence alignment of parental ciap and hpap genes and their recombination variants (AP01, AP03, AP05, AP06, AP11 and AP15)

Figure 4:
The C-terminal sequence alignment of parental parental ciap and hpap genes and their recombination variants (AP2, AP9 and AP13)

Figure 5:
Sequences of hpap, ciap, AP01, AP03, AP05, AP06, AP11, AP15, AP2, AP9 and AP13.

Example 1

1. The DNA of interest with appropriate restriction endonuclease(s) and purify the DNA fragment of interest was purified by gel electrophoresis using Roche High Pure PCR Prep Kit (Roche Diagnostics GmbH, Germany)). As an example, the *Bacillus subtilis* BMTU3420 sarcosine oxidase gene was cleaved as a 1.2 kb-long *Pst*I-*Nsi*I fragment from the recombinant plasmid pBMTU5823.

2. About 0.5 pmol of the double-stranded DNA dissolved in H_2O was mixed with 0.1 pmol of each SODF1, SODF2, SODR1, and SODR2 primers. After immersion in boiling water for 3 minutes, the mixture was placed immediately in an ice/ethanol bath.

3. Ten µl of 10 x reaction buffer [10X buffer: 900 mM HEPES, pH 6.6; 0.1 M magnesium chloride, 10 mM dithiothreitol, 5 mM each dATP, dCTP, dGTP and dTTP, and 2mM each ddATP, ddCTP, ddGTP and ddTTP) was added to the denatured sample, and the total volume of the re-

action mixture was brought up to 98 µl with H_2O.

4. Ten units (about 2 µl) of the Stoffel fragment was added. All the components were mixed by gently tapping the outside of the tube and were centrifuged at 12,000 g for 1-2 seconds in a microfuge to move all the liquid to the bottom. The reaction was carried out under standard PCR cycling sequencing conditions using Taq DNA Sequencing Kit for Standard and Cycle Sequencing Kit (Roche Diagnostics GmbH, Germany).

5. The reaction products were subjected to Wizard DNA Clean up System (Promega, WI, USA) to remove the enzymes, primers, reaction buffer components and dNTPs/ddNTPs. The purified products were incubated in 1 x Klenow buffer and 10 U of Klenow at 37°C for 40 minutes. Two µl of DpnI (10 U/µl) was added to the mixture, and the incubation was continued at 37°C for further 40 minutes.

6. The digested WalkThrough products were purified with High Pure PCR Purification Kit (Roche Diagnostics GmbH, Germany) and was used for whole gene reassembly.

Reassembly of the whole gene

1. For SarcOD gene reassembly by PCR, 5µl of the purified WalkThrough DNA fragments, 20 µl of 2X PCR pre-mix (5-fold diluted cloned *Pfu* buffer, 0.5 mM each dNTP, 0.1U/µl cloned *Pfu* polymerase (Stratagene, La Jolla, CA)) and 15 µl of H_2O were mixed on ice.

2. After incubation at 96°C for 5 min, 40 thermocycles were performed, each with 1.0 min at 95°C, 1.0 min at 55°C and 0 + 5 sec/cycle at 72°C, with the extension step of the last cycle proceeding at 72°C for 10 min, in an Eppendorf Mastercycler Gradient (Eppendorf, Germany).

3. 3 µl aliquots at cycles 20, 30 and 40 were removed from the reaction mixture and analyzed by agarose gel electrophoresis. The reassembled PCR product at 40 cycles contained the correct size product in a smear of larger and smaller sizes.

Amplification

The correctly reassembled product of this first PCR was further amplified in a second PCR reaction which contained the PCR primers complementary to the ends of the template DNA.

1. 2.0 µl of the PCR reassembly aliquots were used as template in 100-µl standard PCR reactions, which contained 0.2 µM each primers of *pstF* (5' GGTAGAGCGAG-TCTCGAGGGGGAGATG C 3', SEQ ID NO: 1) and *NsiR* (5' AGCCGGCGTGACGTGGGTCAGC 3', SEQ ID NO: 2), 1.5 mM $MgCl_2$, 10 mM Tris-HCl [pH 9.0], 50 mM KCl, 200 µM each of the four dNTPs, 2.5 U of *Taq* polymerase (Roche, Germany) and 2.5 U of *Pfu* polymerase (Stratagene, La Jolla, CA).

2. After incubation at 96°C for 5 min, 30 thermocycles were performed, each with 1.0 min at 95°C and 1.0 min at 72°C, with the extension step of the last cycle proceeding at 72°C for 10 min, in an Eppendorf Mastercycler Gradient (Eppendorf GmbH, Germany).

3. The amplification resulted in a large amount of PCR product with the correct size of the sarcosine oxidase whole gene.

Cloning

1. The PCR product of Bs SarcOD gene was digested with *Pst* I and *Nsi* I restriction enzymes, and cloned into pBMTU5823 (*Pst/Nsi*).

2. *E. coli* XL1 F' cells were transformed with the above ligation mixture to form a mutant library.

Results:

After only one round of applying this technique on the wild-type SarcOD gene, two Hind III and Pvu II resistant mutants have been found to be enzymatic active.

Example 2

Recombination of human placental alkaline phosphatase and calf intestinal alkaline phosphatase genes

1. Description of the method

Walk through recombination' is a method used to recombine two or more DNA-sequences based on their homology. Furthermore point mutations can also be introduced during the recombination steps. This method consists mainly of five steps:

Fragment synthesis
Template removal
Terminator removal
Reassembly
Amplification

1.1 Fragment synthesis

Fragment synthesis results from a DNA synthesis reaction where extension is terminated by the incorporation of dideoxy nucleotides. Target DNA sequences selected for recombination serve as templates in this reaction. The number of primers should be chosen depending on the length of the target DNA sequences and the length of the resulting fragments. The fragment length may also be controlled by the reaction conditions. Furthermore, a ,Cycle Sequencing reaction' is favorable because of its higher product yield and easy use.

1.2 Template removal

To avoid interference with the reassembly step, template DNA has to be completely removed after the fragment synthesis. Several methods can be used for this purpose. we have chosen to use U-DNA and Uracil-DNA Glycosylase. Uracil-DNA Glycosylase can be used to remove uracil base at any site where a deoxyuridylate has been incorporated (U-DNA). The resulting abasic site can subsequently be hydrolyzed by alkali-treatment, high temperatures or specific endonucleases. In our approach, a simple temperature treatment is sufficient. U-DNA can be prepared by a PCR reaction using dUTP's instead of dTTP's. After the U-DNA has served as the template for the fragment synthesis, the whole

reaction is subjected to the uracil-DNA glycosylase treatment and temperature treatment. Another control PCR can be used to ensure the complete U-DNA removal.

1.3 Terminator removal

The fragments need to have 3'-OH end for the reassembly reaction. Since our fragments are terminated with dideoxynucleotides, they do not carry 3'-OH ends. These terminators can be removed by 3'-5'-exonuclease activity of several nucleases or DNA polymerases resulting in 3'-OH end. We combine terminator removal with the reassembly by employing a thermostable exonuclease III and *Taq* polymerase for this step. Exonuclease III cuts the terminator by it's 3'-5'-exonuclease activity (enzyme activity only at ds DNA, and therefor only after DNA annealing) and *Taq* polymerase extends the fragment by it's 5'-3'-polymerase activity.

1.4. Reassembly

Reassembly is a PCR-like reaction without primers, there the fragments anneal to each other based on their homologies and extend. Through all cycles of denaturation, annealing and extension fragments growth or reassemble up to the length of the original DNA sequences.

1.5 Amplification

Amplifying the reassembled DNA takes place in a PCR with sequence flanking primers in order to provide enough material for subsequent cloning and analysing steps.

2. Experimental Protocol

2.1 Preparation of U-DNA

dUTP containing DNA template is generated by PCR using 2.5 mM $MgCl_2$ according to the manual (‚Uracyl-DNA Glycosylase' Roche # 1 444 646), 600 µM dUTP and three other nucleotides dATP, dCTP and dGTP at a concentration of 200 µM each. A typical reaction mixture and cycling conditions used for recombining the alkaline phosphatase are shown below:

Template DNA (plasmid)	40 ng
Primer APhpaF	20 pmol

Primer APxbaR	20 pmol	
10 x *Taq* Puffer	10 µl	
Taq DNA polymerase	0,5 U	
$MgCl_2$	2,5 mM	
dATP, dCTP, dGTP (Roche # 1 969 064)	0,2 mM	each
dUTP (Roche # 1 420 470)	0,6 mM	
H_2O	ad 100 µl	

Cycling conditions:

95 °C	5 min	
95 °C	1 min	
60 °C	1 min	30 x
72 °C	2 min	
72 °C	10 min	
4 °C	∞	

The synthesized U-DNA is separated from the template DNA by preparative agarose gel electrophoresis (1% agarose/TAE) and gel extraction (‚QIAquick Gel Extraction Kit' Qiagen # 28706).

2.2 Fragment synthesis

Fragments are synthesized using Cycle sequencing reactions (‚DIG Taq DNA Sequencing Kit for Standard and Cycle Sequencing' Roche # 1449443) with some modifications. Twelve reaction mixtures are set up for a total of six primers and two U-DNA templates. A typical reaction mixture with cycling conditions used for recombining the two alkaline phosphatase genes are shown as following:

U-DNA	175 ng
Primer (1µM)	1 pmol
Puffer	3 µl
Taq DNA polymerase (5U/µl)	0,6 µl
Termination mixture ddATP/dGTP	2,5 µl

Termination mixture ddCTP/dGTP	2,5 µl
Termination mixture ddGTP/dGTP	2,5 µl
Termination mixture ddTTP/dGTP	2,5 µl
H_2O	ad 30 µl

Cycling conditions:

95 °C	5 min	
95 °C	30 sec	
60 °C	30 sec	30 x
72 °C	1 min	
4 °C	∞	

The calf intestinal alkaline phosphatase gene (*ciap*) and the human placental alkaline phosphatase gene (*hpap*) show a sequence similarity of 81 % and have lengths of about 1530 bps. Six primers are used for fragment synthesis. Two external primers (APhpaF and APxbaR) share the same sequence. Figure 2 shows the primer arrangement along human placental alkaline phosphatase (hpap) and calf intestinal alkaline phosphatase (ciap) genes.

Primer sequences

Flanking primers

APhpaF	CTT CGG CGT TCA GTA ACA CGC (SEQ ID NO.: 3)
ApxbaR	GCT TTC GAG GTG AAT TTC GAC C (SEQ ID NO.: 4)

Primers for ciap gene

CIAPint1F	GGT CAC GTC TGT GAT CAA CCG (SEQ ID NO.: 5)
CIAPint1R	CGG TTG ATC ACA GAC GTG AC C (SEQ ID NO.: 6)
CIAPint2F	CGC AAA GCT TAT ATG GCA CTG AC (SEQ ID NO.: 7)

CIAPint2R	GTC AGT GCC ATA TAA GCT TTG CC
	(SEQ ID NO.: 8)
Primers for hpap gene	
HPAPint1F	CCA AGA AAG CAG GGA AGT CAG TG
	(SEQ ID NO.: 9)
HPAPint1R	CAC TGA CTT CCC TGC TTT CTT GG
	(SEQ ID NO.: 10)
HPAPint2F	CAT GTT CGA CGA CGC CAT TGA G
	(SEQ ID NO.: 11)
HPAPint2R	CTC AAT GGC GTC GTC GAA CAT G
	(SEQ ID NO.: 12)

2.3 Uracil-DNA Glycosylase treatment and temperature treatment

This step is accomplished by addition of 1 µl Uracil-DNA Glycosylase (1 U/µl; Roche , #1775375) to each fragment synthesis mixture and by incubation at 37 °C for 4 h to cleavage the U-bases followed by incubation at 95 °C for 2 min to hydrolyse the abasic sites and to inactivate the enzyme. Another PCR may be used to control the complete removal of the template U-DNA.

2.4 Reassembly

It is important to use proper amount of purified DNA in the critical reassembly step. DNA fragments are purified from the cleavage products by using Microcon (Microcon 50; Millipore # 42415) and are retained in H_2O. A typical mixture and cycling conditions used for recombining the two alkaline phosphatase genes are shown below:

Purified fragments	x µl	(proper amount)
dNTP's (Roche # 1 969 064)	0,4 mM	each
Expand High Fidelity 10x buffer (Roche # 1 759 175)	5 µl	
Taq/ Exo III mix	1 U	
H_2O	ad 50 µl	

After 30 cycles 1 µl of Taq/ExoIII is added to the reaction mixture for further reasembly.

Cycling conditions:

95 °C	5 min	
95 °C	1 min	
45 °C	1 min	50 cycles
72 °C	30 sec	
72 °C	10 min	
4 °C	∞	

2.5 Amplification PCR (also used as control PCR)

The amplification PCR is a standard PCR with the two sequence flanking primers (AphpaF and ApxbaR).The following is a typical mixture with cycling conditions used for recombining the two alkaline phosphatase genes:

Reasembly mixture	0,5 µl	
Primer APhpaF	10 pmol	
Primer APxbaR	10 pmol	
10 x *Taq* Puffer (+ Mg; Roche # 1 271 318)	5 µl	
Taq DNA polymerase (Roche # 1 418 432)	2,5 U	
dATP, dCTP, dGTP, dTTP (Roche # 1 969 064)	0,2 mM	each
H_2O	ad 50 µl	

Cycling conditions:

95 °C	5 min	
95 °C	1 min	
60 °C	1 min	30 x
72 °C	2 min	
72 °C	10 min	
4 °C	∞	

2.6 Cloning and expression

The amplification product is purified by preparative gel electrophoresis (1% agarose/TAE) and gel extraction (‚QIAquick Gel Extraction Kit' Quiagen # 28706), cloned into vector pCR®-XL-TOPO® according to the suppliers manual (Invitrogen # K 475020) and expressed in E. coli TOP 10 cells delivered with the vector.

2.7 Sequencing results

Nine recombination clones have been analyzed by sequencing using one sequencing primer. The results are summarized in Table 1. Eight of these clones show one ore more recombination events within the sequenced region (445 – 645 bp). One variant contains 4 crossovers within the mentioned region. The overall mutation rate is about 0,25 %. Since the reassembly products can be ligated into the vector in two different orientations, sequencing these clones with only one primer could give both sequences of the N-terminal and coding for the C-terminal regions.

Clone	Direction of sequencing	Length of sequenced region	Crossovers	Mutations
AP 1	N-terminus	509 bp	1	1
AP 3	N-terminus	445 bp	1	0
AP 5	N-terminus	624 bp	4	3
AP 6	N-terminus	553 bp	1	1
AP 11	N-terminus	599 bp	2	1
AP 15	N-terminus	528 bp	1	2
AP 2	C-terminus	645 bp	1	3
AP 9	C-terminus	639 bp	0	1
AP13	C-terminus	638 bp	2	1

Table 1: Sequencing results after recombination of two alkaline phosphatase genes

2.8 Sequences

Figure 3 shows the N-terminal sequence alignment of parental ciap and hpap genes and their recombination variants (AP01, AP03, AP05, AP06, AP11 and AP15)

Figure 4 shows the C-terminal sequence alignment of parental parental ciap and hpap genes and their recombination variants (AP2, AP9 and AP13)

Figure 5 shows sequences of hpap, ciap, AP01, AP03, AP05, AP06, AP11, AP15, AP2, AP9 and AP13.

IDT Exhibit 1020.6488

Claims

1) Method for forming a polynucleotide sequence comprising the steps of

a) generating a nucleic acid fragment ladder by nucleic acid synthesis which is carried out in presence of a reaction mixture comprising at least one primer, an enzyme having nucleic acid-synthesizing activity and a mixture of nucleic acid chain-terminating and nucleic acid chain-elongating molecules and a template polynucleotide,

b) removing the chain-terminating molecules or changing it into non-terminating molecules,

c) reassembly of the polynucleotide by hybridizing these fragments to one another in presence of a thermostable enzyme having nucleic acid-synthesizing activity and conducting nucleic acid synthesis either
ii) in presence of a mixture of nucleic acid chain-terminating molecules and nucleic acid chain elongating molecules or
ii) in presence of a reaction mixture of nucleic acid chain elongating molecules which does not contain chain-terminating molecules
whereas in case of i) steps b) and c) will be repeated subsequently.

2) Method for forming a polynucleotide sequence according to claim 1 whereas the template polynucleotide in step a) and/or c) is a mixture consisting of different template polynucleotides.

3) Method for forming a polynucleotide sequence according to claim 1 or 2 wherein the nucleic acid chain-terminating molecule is a ddNTP and the nucleic acid elongating molecule is a dNTP.

4) Method for forming a polynucleotide sequence according to one of the claims 1-3 wherein the enzyme having nucleic-acid-synthesizing activity is a DNA polymerase and the template polynucleotide is DNA.

5) Method for forming a polynucleotide sequence according to one of the claims 1-4 wherein the enzyme having nucleic acid-synthesizing activity is a RNA-dependent DNA polymerase and the template polynucleotide is RNA.

6) A polynucleotide sequence obtained by a method according to any one of the claims 1-5.

7) Method for providing mutant polypeptides or proteins comprising the steps according to any one of the claims 1-5 and further comprising the steps of

- cloning procedures - as cloning the amplified product into a vector and expression of the encoded protein.
- screening or selection procedures.

8) Method for providing mutant polypeptides or proteins comprising the steps according to claim 7 additionally comprising an amplification step of reassembled product

9) Mutant polypeptide or protein obtained by a method according to claim 7 or 8.

10) Use of a method according to any of the claims 1 to 5 and 7 to 8 for optimizing biomolecules.

11) Use of a method according to any of the claims 1 to 5 and 7 to 8 for optimizing enzymes.

Fig.1

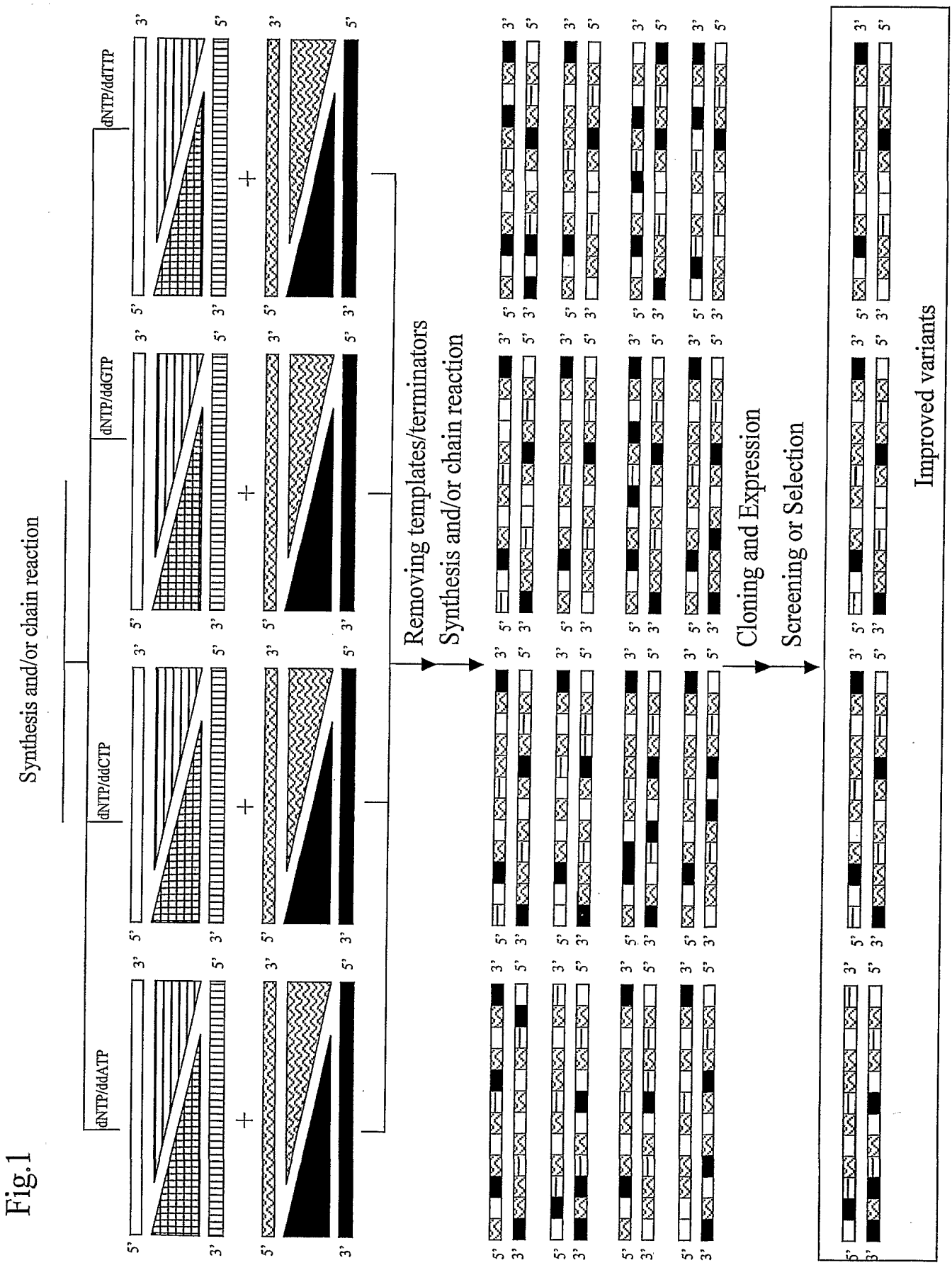

Synthesis and/or chain reaction
dNTP/ddATP
dNTP/ddCTP
dNTP/ddGTP
dNTP/ddTTP
Removing templates/terminators
Synthesis and/or chain reaction
Cloning and Expression
Screening or Selection
Improved variants

Figure 2:

Primer arrangement along human placental alkaline phosphatase (hpap) and calf intestinal alkaline phosphatase (ciap) genes

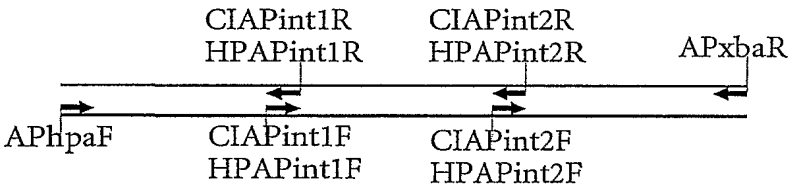


Figure 3 1/4

The N-terminal sequence alignment of parental ciap and hpap genes and their recombination variants (AP01, AP03, AP05, AP06, AP11 and AP15).

```
                10        20        30        40        50
CIAP  : CTTCGGCGTTCAGTAACACGCGTTAACGTGCTGAACAGCCGGGCATTTTT : 50
HPAP  : CTTCGGCGTTCAGTAACACGCGTTAACGTGCTGAACAGCCGGGCATTTTT : 50
AP01  : CTTCGGCGTTCAGTAACACGCGTTAACGTGCTGAACAGCCGGGCATTCTT : 50
AP03  : CTTCGGCGTTCAGTAACACGCGTTAACGTGCTGAACAGCCGGGCATTTTT : 50
AP05  : CTTCGGCGTTCAGTAACACGCGTTAACGTGCTGAACAGCCGGGCATTTTT : 50
AP06  : CTTCGGCGTTCAGTAACACGCGTTAACGTGCTGAACAGCCGGGCATTTTT : 50
AP11  : CTTCGGCGTTCAGTAACACGCGTTAACGTGCTGAACAGCCGGGCATTTTT : 50
AP15  : CTTCGGCGTTCAGTAACACGCGTTACCGTGCTGAACAGCCGGGCATTTTT : 50

                60        70        80        90       100
CIAP  : TTACGCTATACCCTACATAATAAAACCGGAGCTACCATGAATAAGAAGGT : 100
HPAP  : TTACGCTATACCCTACATAATAAAACCGGAGCTACCATGAATAAGAAGGT : 100
AP01  : TTACGCTATACCCTACATAATAAAACCGGAGCTACCATGAATAAGAAGGT : 100
AP03  : TTACGCTATACCCTACATAATAAAACCGGAGCTACCATGAATAAGAAGGT : 100
AP05  : TTACGCTATACCCTACATAATAAAACCGGAGCTACCATGAATAAGAAGGT : 100
AP06  : TTACGCTATACCCTACATAATAAAACCGGAGCTACCATGAATAAGAAGGT : 100
AP11  : TTACGCTATACCCTACATAATAAAACCGGAGCTACCATGAATAAGAAGGT : 100
AP15  : TTACGCTATACCCTACATAATAAAACCGGAGCTACCATGAATAAGAAGGT : 100

               110       120       130       140       150
CIAP  : ACTGACCCTTTCTGCCGTGATGGCAAGTCTGTTATTCGGGGCCCACGCGC : 150
HPAP  : ACTGACCCTTTCTGCCGTGATGGCAAGTCTGTTATTCGGCGCGCACGCGC : 150
AP01  : ACTGACCCTTTCTGCCGTGATGGCAAGTCTGTTATTCGGGGCCCACGCGC : 150
AP03  : ACTGACCCTTTCTGCCGTGATGGCAAGTCTGTTATTCGGCGCGCACGCGC : 150
AP05  : ACTGACCCTTTCTGCCGTGATGGCAAGTCTGTTATTCGGCGCGCACGCGC : 150
AP06  : ACTGACCCTTTCTGCCGTGATGGCAAGTCTGTTATTCGGCGCGCACGCGC : 150
AP11  : ATTGACCCTTTCTGCCGTGATGGCAAGTCTGTTATTCGGCGCGCACGCGC : 150
AP15  : ACTGACCCTTTCTGCCGTGATGGCAAGTCTGTTATTCGGGGCCCACGCGC : 150

               160       170       180       190       200
CIAP  : ACGCGGCGATCCCAGCTGAGGAGGAAAACCCCGCCTTCTGGAACCGCCAG : 200
HPAP  : ACGCGGCGATCCCAGTTGAGGAGGACAACCCCGACTTCTGGAACCGCCAG : 200
AP01  : ACGCGGCGATCCCAGTTGAGGAGGACAACCCCGACTTCTGGAACCGCCAG : 200
AP03  : ACGCGGCGATCCCAGTTGAGGAGGACAACCCCGACTTCTGGAACCGCCAG : 200
AP05  : ACGCGGCGATCCCAGTTGAGGAGGACAACCCCGACTTCTGGAACCGCCAG : 200
AP06  : ACGCGGCGATCCCAGTTGAGGAGGACAACCCCGACTTCTGGAACCGCCAG : 200
AP11  : ACGCGGCGATCCCAGTTGAGGAGGACAACCCCGACTTCTGGAACCGCCAG : 200
AP15  : ACGCGGCGATCCCAGCTGAGGAGGAAAACCCCGCCTTCTGGAACCGCCAG : 200
```

Figure 3 2/4

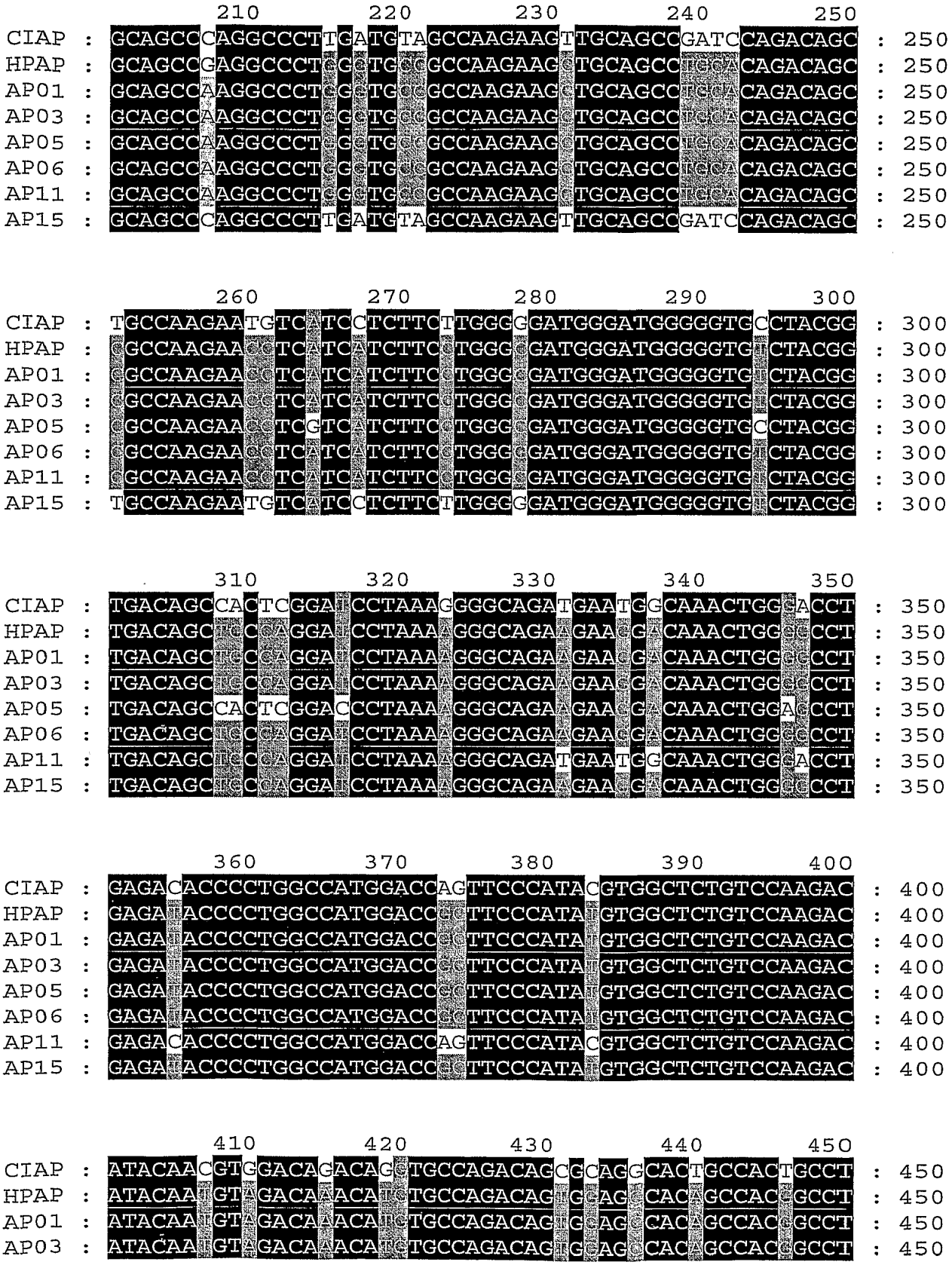

210 220 230 240 250
CIAP : GCAGCCCAGGCCCTTGATGTAGCCAAGAAGTTGCAGCCGATCCAGACAGC : 250
HPAP : GCAGCCGAGGCCCTGGCTGCCGCCAAGAAGCTGCAGCCTGCACAGACAGC : 250
AP01 : GCAGCCAAGGCCCTGGCTGCCGCCAAGAAGCTGCAGCCTGCACAGACAGC : 250
AP03 : GCAGCCAAGGCCCTGGCTGCCGCCAAGAAGCTGCAGCCTGCACAGACAGC : 250
AP05 : GCAGCCAAGGCCCTGGCTGCCGCCAAGAAGCTGCAGCCTGCACAGACAGC : 250
AP06 : GCAGCCAAGGCCCTGGCTGCCGCCAAGAAGCTGCAGCCTGCACAGACAGC : 250
AP11 : GCAGCCAAGGCCCTGGCTGCCGCCAAGAAGCTGCAGCCTGCACAGACAGC : 250
AP15 : GCAGCCCAGGCCCTTGATGTAGCCAAGAAGTTGCAGCCGATCCAGACAGC : 250
260 270 280 290 300
CIAP : TGCCAAGAATGTCATCCTCTTCTTGGGGGATGGGATGGGGGTGCCTACGG : 300
HPAP : GGCCAAGAACGTCATCATCTTCCTGGGCGATGGGATGGGGGTGTCTACGG : 300
AP01 : GGCCAAGAACGTCATCATCTTCCTGGGCGATGGGATGGGGGTGTCTACGG : 300
AP03 : GGCCAAGAACGTCATCATCTTCCTGGGCGATGGGATGGGGGTGTCTACGG : 300
AP05 : GGCCAAGAACGTCGTCATCTTCCTGGGCGATGGGATGGGGGTGCCTACGG : 300
AP06 : GGCCAAGAACGTCATCATCTTCCTGGGCGATGGGATGGGGGTGTCTACGG : 300
AP11 : GGCCAAGAACGTCATCATCTTCCTGGGCGATGGGATGGGGGTGTCTACGG : 300
AP15 : TGCCAAGAATGTCATCCTCTTCTTGGGGGATGGGATGGGGGTGTCTACGG : 300
310 320 330 340 350
CIAP : TGACAGCCACTCGGAACCTAAAGGGGCAGATGAATGGCAAACTGGCACCT : 350
HPAP : TGACAGCTGCCAGGAACCTAAAAGGGCAGAAGAACGACAAACTGGCGCCT : 350
AP01 : TGACAGCTGCCAGGAACCTAAAAGGGCAGAAGAACGACAAACTGGCGCCT : 350
AP03 : TGACAGCTGCCAGGAACCTAAAAGGGCAGAAGAACGACAAACTGGCGCCT : 350
AP05 : TGACAGCCACTCGGACCCTAAAAGGGCAGAAGAACGACAAACTGGACCCT : 350
AP06 : TGACAGCTGCCAGGAACCTAAAAGGGCAGAAGAACGACAAACTGGCGCCT : 350
AP11 : TGACAGCTGCCAGGAACCTAAAAGGGCAGATGAATGGCAAACTGGCACCT : 350
AP15 : TGACAGCTGCCAGGAACCTAAAAGGGCAGAAGAACGACAAACTGGCGCCT : 350
360 370 380 390 400
CIAP : GAGACACCCCTGGCCATGGACCAGTTCCCATACGTGGCTCTGTCCAAGAC : 400
HPAP : GAGATACCCCTGGCCATGGACCGCTTCCCATATGTGGCTCTGTCCAAGAC : 400
AP01 : GAGATACCCCTGGCCATGGACCGCTTCCCATATGTGGCTCTGTCCAAGAC : 400
AP03 : GAGATACCCCTGGCCATGGACCGCTTCCCATATGTGGCTCTGTCCAAGAC : 400
AP05 : GAGATACCCCTGGCCATGGACCGCTTCCCATATGTGGCTCTGTCCAAGAC : 400
AP06 : GAGATACCCCTGGCCATGGACCGCTTCCCATATGTGGCTCTGTCCAAGAC : 400
AP11 : GAGACACCCCTGGCCATGGACCAGTTCCCATACGTGGCTCTGTCCAAGAC : 400
AP15 : GAGATACCCCTGGCCATGGACCGCTTCCCATATGTGGCTCTGTCCAAGAC : 400
410 420 430 440 450
CIAP : ATACAACGTGGACAGACAGCTGCCAGACAGCGCAGGCACTGCCACTGCCT : 450
HPAP : ATACAATGTAGACAAACATGTGCCAGACAGTGCAGCCACAGCCACCGCCT : 450
AP01 : ATACAATGTAGACAAACATGTGCCAGACAGTGCAGCCACAGCCACCGCCT : 450
AP03 : ATACAATGTAGACAAACATGTGCCAGACAGTGCAGCCACAGCCACCGCCT : 450

Figure 3 3/4

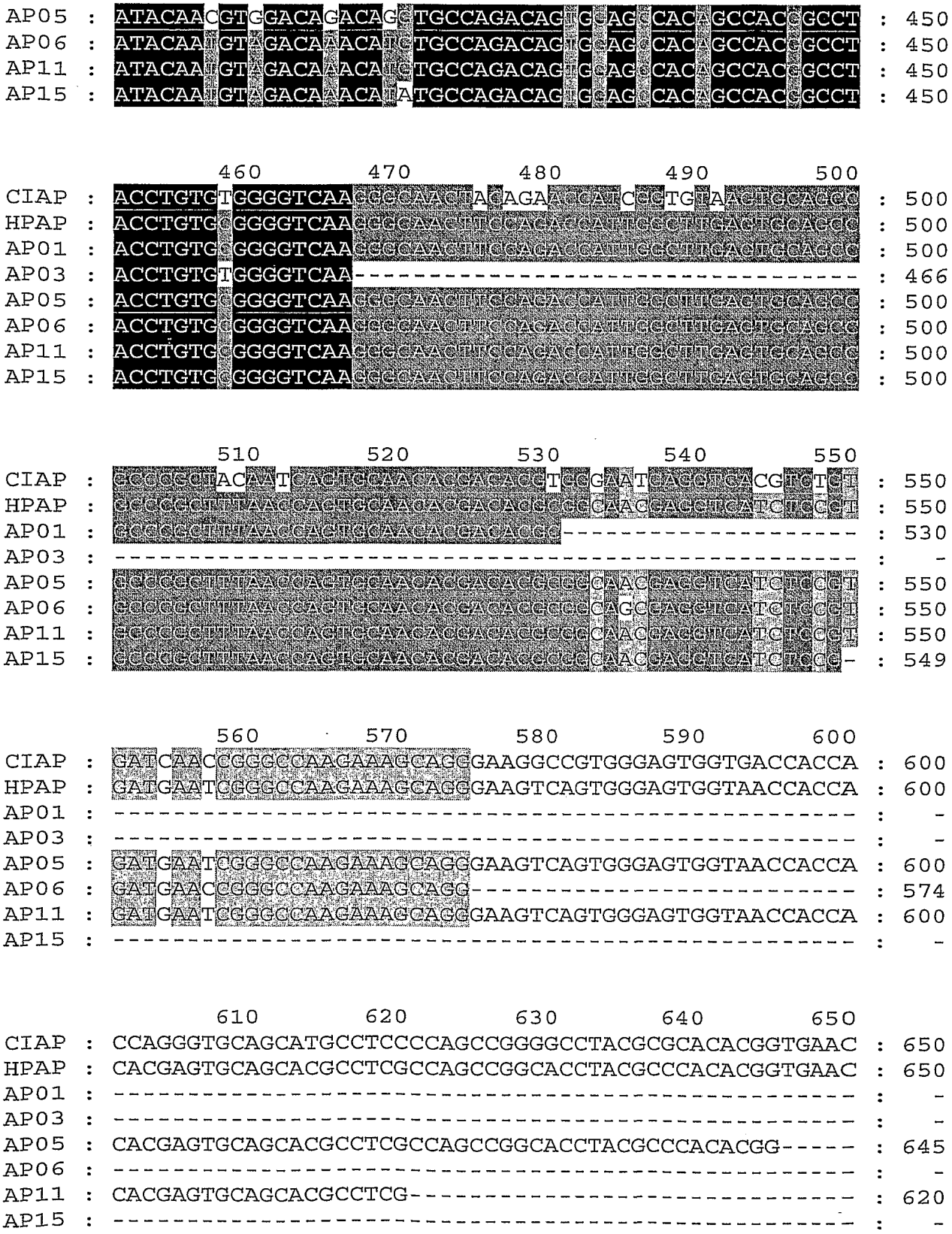

AP05 : ATACAACGTGGACAGACAGCTGCCAGACAGTGCAGCCACAGCCACCGCCT : 450
AP06 : ATACAATGTAGACAAACATCTGCCAGACAGTGCAGCCACAGCCACCGCCT : 450
AP11 : ATACAATGTAGACAAACATCTGCCAGACAGTGCAGCCACAGCCACCGCCT : 450
AP15 : ATACAATGTAGACAAACATATGCCAGACAGTGCAGCCACAGCCACCGCCT : 450
460 470 480 490 500
CIAP : ACCTGTGTGGGGTCAACCCCAACTACAGAACCATCCCTGTAACTCCACCC : 500
HPAP : ACCTGTGCGGGGTCAACCCCAACTTCCACACCATTCCCTTCACTCCACCC : 500
AP01 : ACCTGTGCGGGGTCAACCCCAACTTCCACACCATTCCCTTCACTCCACCC : 500
AP03 : ACCTGTGTGGGGTCAA---------- : 466
AP05 : ACCTGTGCGGGGTCAACCCCAACTTCCACACCATTCCCTTCACTCCACCC : 500
AP06 : ACCTGTGCGGGGTCAACCCCAACTTCCACACCATTCCCTTCACTCCACCC : 500
AP11 : ACCTGTGCGGGGTCAACCCCAACTTCCACACCATTCCCTTCACTCCACCC : 500
AP15 : ACCTGTGCGGGGTCAACCCCAACTTCCACACCATTCCCTTCACTCCACCC : 500
510 520 530 540 550
CIAP : CCCCCCCTACAATCACTCCAACAGCACACCTCCGAATCACGTCACGTGTCT : 550
HPAP : CCCCCCCTTTAACCACTCCAACACCACACCCCCCAACCACCTCATCTCCCT : 550
AP01 : CCCCCCCTTTAACCACTCCAACACCACACCC---------- : 530
AP03 : ---------- : -
AP05 : CCCCCCCTTTAACCACTCCAACACCACACCCCCCAACCACCTCATCTCCCT : 550
AP06 : CCCCCCCTTTAACCACTCCAACACCACACCCCCCAGCCACCTCATCTCCCT : 550
AP11 : CCCCCCCTTTAACCACTCCAACACCACACCCCCCAACCACCTCATCTCCCT : 550
AP15 : CCCCCCCTTTAACCACTCCAACACCACACCCCCCAACCACCTCATCTCCC- : 549
560 570 580 590 600
CIAP : GATCAACCGGGCCAAGAAAGCAGGGAAGGCCGTGGGAGTGGTGACCACCA : 600
HPAP : GATGAATCGGGCCAAGAAAGCAGGGAAGTCAGTGGGAGTGGTAACCACCA : 600
AP01 : ---------- : -
AP03 : ---------- : -
AP05 : GATGAATCGGGCCAAGAAAGCAGGGAAGTCAGTGGGAGTGGTAACCACCA : 600
AP06 : GATGAACCGGGCCAAGAAAGCAGG---------- : 574
AP11 : GATGAATCGGGCCAAGAAAGCAGGGAAGTCAGTGGGAGTGGTAACCACCA : 600
AP15 : ---------- : -
610 620 630 640 650
CIAP : CCAGGGTGCAGCATGCCTCCCCAGCCGGGGCCTACGCGCACACGGTGAAC : 650
HPAP : CACGAGTGCAGCACGCCTCGCCAGCCGGCACCTACGCCCACACGGTGAAC : 650
AP01 : ---------- : -
AP03 : ---------- : -
AP05 : CACGAGTGCAGCACGCCTCGCCAGCCGGCACCTACGCCCACACGG----- : 645
AP06 : ---------- : -
AP11 : CACGAGTGCAGCACGCCTCG---------- : 620
AP15 : ---------- : -

Figure 3 4/4

```
                660        670        680
CIAP  : CGAAACTGGTACTCAGACGCCGACCTGCCTGC : 682
HPAP  : CGCAACTGGTACTCGGACGCCGACGTGCCTGC : 682
AP01  : -------------------------------- :   -
AP03  : -------------------------------- :   -
AP05  : -------------------------------- :   -
AP06  : -------------------------------- :   -
AP11  : -------------------------------- :   -
AP15  : -------------------------------- :   -
```

CIAP: SEQ ID NO: 14
HPAP: SEQ ID NO: 13
AP01: SEQ ID NO: 15
AP03: SEQ ID NO: 16
AP05: SEQ ID NO: 17
AP06: SEQ ID NO: 18
AP11: SEQ ID NO: 19
AP15: SEQ ID NO: 20

IDT Exhibit 1020.6496

Figure 4 1/3

The C-terminal sequence alignment of parental parental ciap and hpap genes and their recombination variants (AP2, AP9 and AP13)

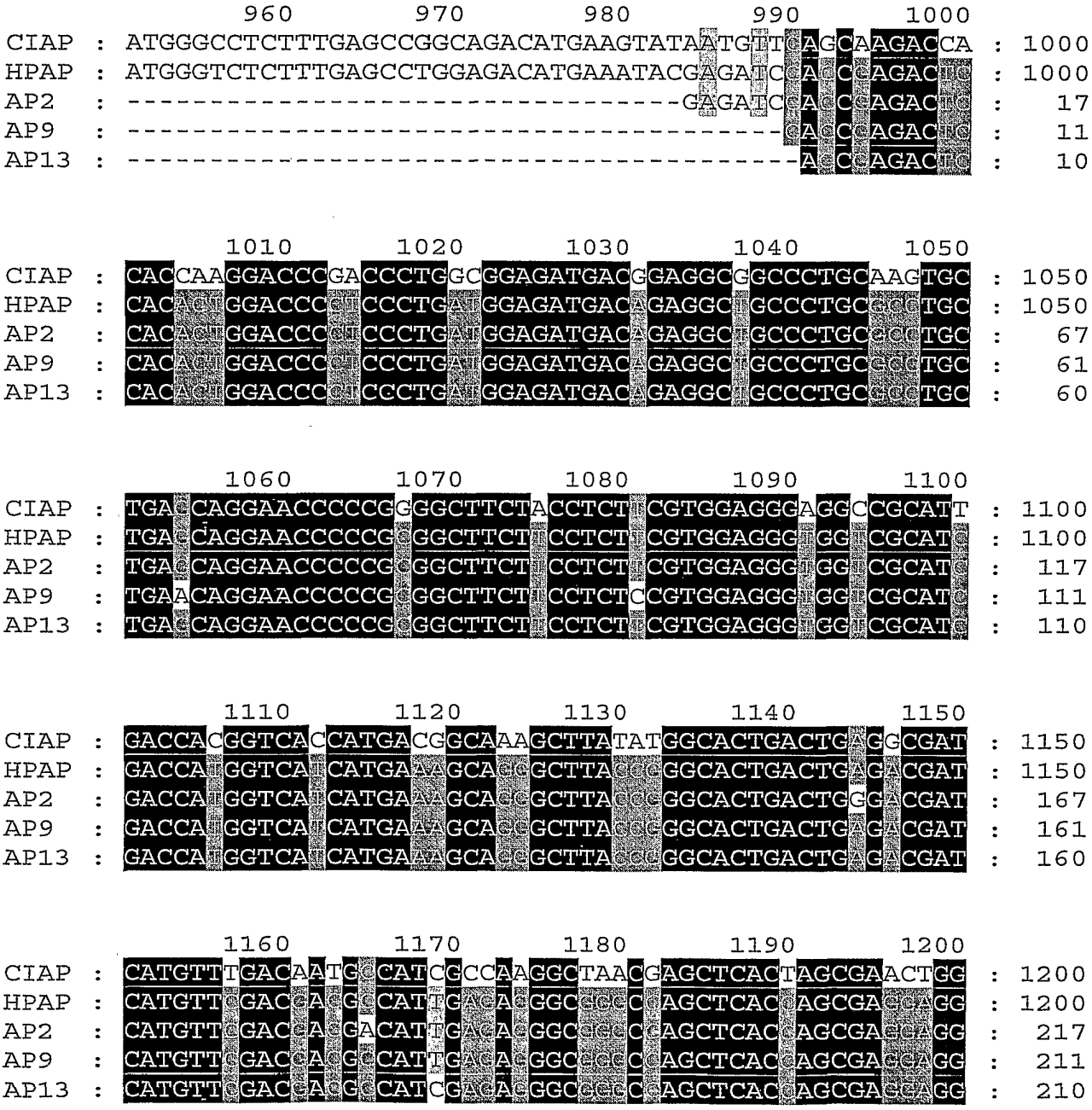


Figure 4 2/3

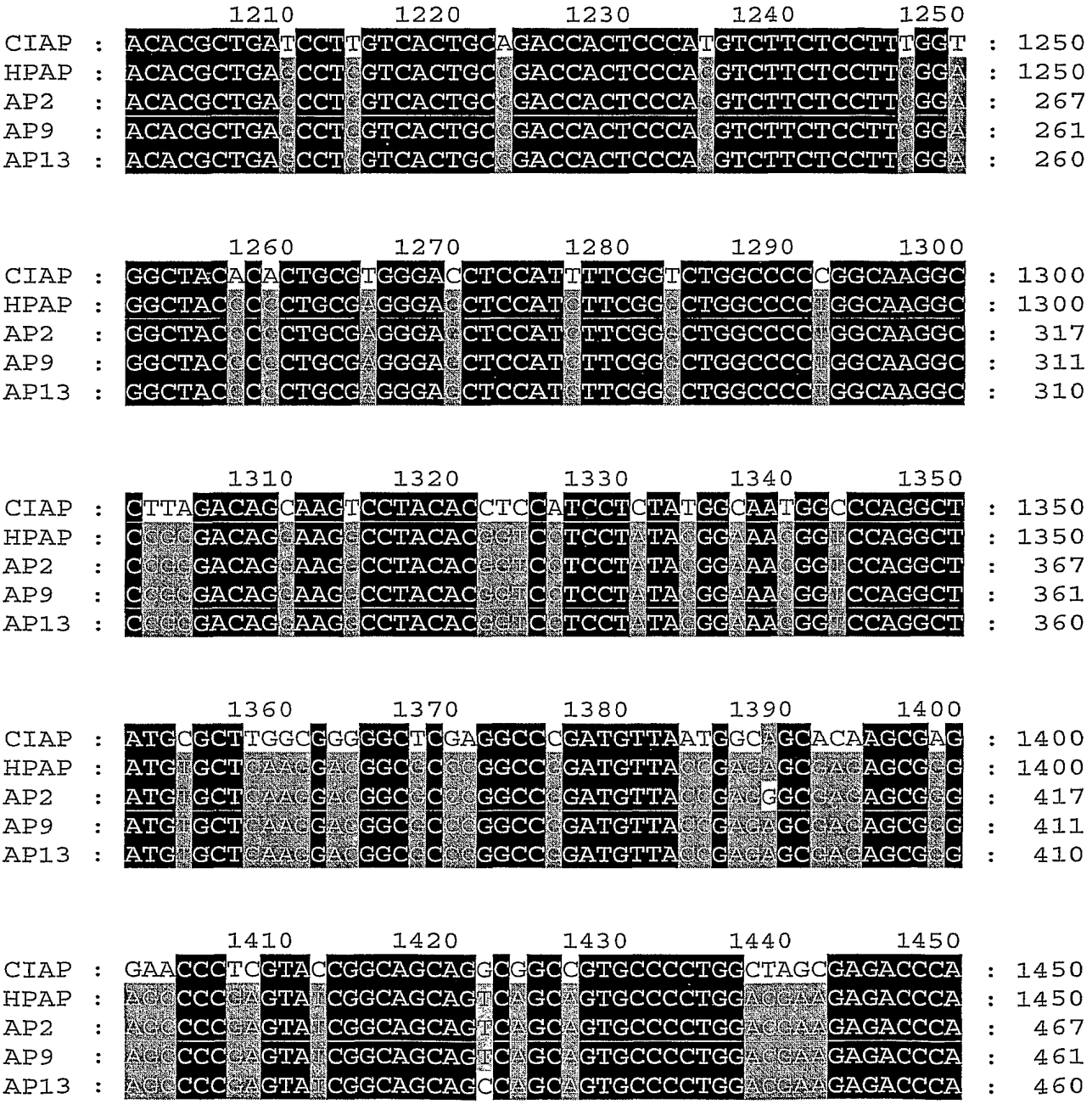
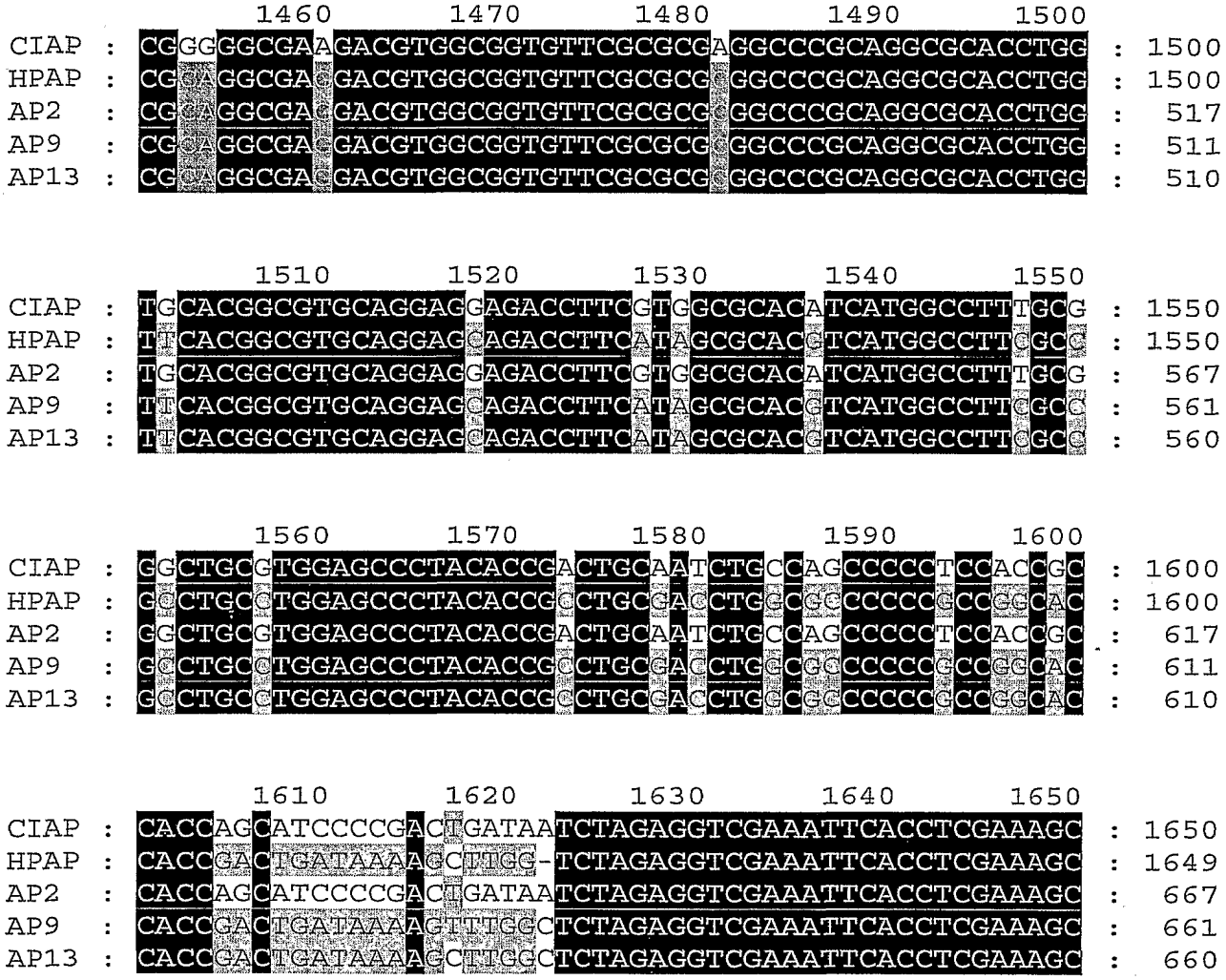

1210 1220 1230 1240 1250
CIAP : ACACGCTGATCCTTGTCACTGCAGACCACTCCCATGTCTTCTCCTTTGGT : 1250
HPAP : ACACGCTGACCCTCGTCACTGCCGACCACTCCCACGTCTTCTCCTTCGGA : 1250
AP2 : ACACGCTGACCCTCGTCACTGCCGACCACTCCCACGTCTTCTCCTTCGGA : 267
AP9 : ACACGCTGACCCTCGTCACTGCCGACCACTCCCACGTCTTCTCCTTCGGA : 261
AP13 : ACACGCTGACCCTCGTCACTGCCGACCACTCCCACGTCTTCTCCTTCGGA : 260
1260 1270 1280 1290 1300
CIAP : GGCTACACCACTGCGTGGGACCTCCATTTTCGGTCTGGCCCCCGGCAAGGC : 1300
HPAP : GGCTACCCCCTGCGAGGGACCTCCATCTTCGGCCTGGCCCCCGGCAAGGC : 1300
AP2 : GGCTACCCCCTGCGAGGGACCTCCATCTTCGGCCTGGCCCCCGGCAAGGC : 317
AP9 : GGCTACCCCCTGCGAGGGACCTCCATCTTCGGCCTGGCCCCCGGCAAGGC : 311
AP13 : GGCTACCCCCTGCGAGGGACCTCCATCTTCGGCCTGGCCCCCGGCAAGGC : 310
1310 1320 1330 1340 1350
CIAP : CTTAGACAGCAAGTCCTACACCTCCATCCTCTATGGCAATGGCCCAGGCT : 1350
HPAP : CCCCGACAGCAAGCCCTACACCGTCCTCCTATACGGAAACGGTCCAGGCT : 1350
AP2 : CCCCGACAGCAAGCCCTACACCGTCCTCCTATACGGAAACGGTCCAGGCT : 367
AP9 : CCCCGACAGCAAGCCCTACACCGTCCTCCTATACGGAAACGGTCCAGGCT : 361
AP13 : CCCCGACAGCAAGCCCTACACCGTCCTCCTATACGGAAACGGTCCAGGCT : 360
1360 1370 1380 1390 1400
CIAP : ATGCGCTTGGCGGGGGCTCGAGGCCCGATGTTAATGGCAGCACAAGCGAG : 1400
HPAP : ATGTGCTCAACGACGGCCCCGGCCCGATGTTACCGACAGCCAGAGCGCG : 1400
AP2 : ATGTGCTCAACGACGGCCCCGGCCCGATGTTACCGACGGCCAGAGCGCG : 417
AP9 : ATGTGCTCAACGACGGCCCCGGCCCGATGTTACCGACAGCCAGAGCGCG : 411
AP13 : ATGTGCTCAACGACGGCCCCGGCCCGATGTTACCGACAGCCAGAGCGCG : 410
1410 1420 1430 1440 1450
CIAP : GAACCCTCGTACCGGCAGCAGGCGGCCGTGCCCCTGGCTAGCGAGACCCA : 1450
HPAP : AGCCCCCAGTATCGGCAGCAGTCAGCAGTGCCCCTGGACGAAGAGACCCA : 1450
AP2 : AGCCCCCAGTATCGGCAGCAGTCAGCAGTGCCCCTGGACGAAGAGACCCA : 467
AP9 : AGCCCCCAGTATCGGCAGCAGTCAGCAGTGCCCCTGGACGAAGAGACCCA : 461
AP13 : AGCCCCCAGTATCGGCAGCAGCCAGCAGTGCCCCTGGACGAAGAGACCCA : 460

Figure 4 3/3



CIAP: SEQ ID NO: 22
HPAP: SEQ ID NO: 21
AP2: SEQ ID NO: 23
AP9: SEQ ID NO: 24
AP13: SEQ ID NO: 25

Figure 5: 1/3

Sequences of hpap, ciap, AP01, AP03, AP05, AP06, AP11, AP15, AP2, AP9 and AP13.

ciap (SEQ ID NO: 26)

CTTCGGCGTTCAGTAACACGCGTTAACGTGCTGAACAGCCGGGCATTTTTTTACGCTATACCCTACATAATAAAACCGGA
GCTACCATGAATAAGAAGGTACTGACCCTTTCTGCCGTGATGGCAAGTCTGTTATTCGGGGCCCACGCGCACGCGGCGAT
CCCAGCTGAGGAGGAAAACCCCGCCTTCTGGAACCGCCAGGCAGCCCAGGCCCTTGATGTAGCCAAGAAGTTGCAGCCGA
TCCAGACAGCTGCCAAGAATGTCATCCTCTTCTTGGGGGATGGGATGGGGGTGCCTACGGTGACAGCCACTCGGATCCTA
AAGGGGCAGATGAATGGCAAACTGGGACCTGAGACACCCCTGGCCATGGACCAGTTCCCATACGTGGCTCTGTCCAAGAC
ATACAACGTGGACAGACAGGTGCCAGACAGCGCAGGCACTGCCACTGCCTACCTGTGTGGGGTCAAGGGCAACTACAGAA
CCATCGGTGTAAGTGCAGCCGCCCGCTACAATCAGTGCAACACGACACGTGGGAATGAGGTCACGTCTGTGATCAACCGG
GCCAAGAAAGCAGGGAAGGCCGTGGGAGTGGTGACCACCACCAGGGTGCAGCATGCCTCCCCAGCCGGGGCCTACGCGCA
CACGGTGAACCGAAACTGGTACTCAGACGCCGACCTGCCTGCTGATGCACAGAAGAATGGCTGCCAGGACATCGCCGCAC
AGCTGGTCTACAACATGGATATTGACGTGATCCTGGGTGGAGGCCGAATGTACATGTTTCCTGAGGGGACCCCAGACCCT
GAATACCCAGATGATGCCAGTGTGAATGGAGTCCGGAAGGACAAGCAGAACCTGGTGCAGGAATGGCAGGCCAAGCACCA
GGGAGCCCAGTATGTGTGGAACCGCACTGCGCTCCTTCAGGCGGCCGATGACTCCAGTGTAACACACCTCATGGGCCTCT
TTGAGCCGGCAGACATGAAGTATAATGTTCAGCAAGACCACACCAAGGACCCGACCCTGGCGGAGATGACGGAGGCGGCC
CTGCAAGTGCTGAGCAGGAACCCCCGGGGCTTCTACCTCTTCGTGGAGGGAGGCCGCATTGACCACGGTCACCATGACGG
CAAAGCTTATATGGCACTGACTGAGGCGATCATGTTTGACAATGCCATCGCCAAGGCTAACGAGCTCACTAGCGAACTGG
ACACGCTGATCCTTGTCACTGCAGACCACTCCCATGTCTTCTCCTTTGGTGGCTACACACTGCGTGGGACCTCCATTTTC
GGTCTGGCCCCCGGCAAGGCCTTAGACAGCAAGTCCTACACCTCCATCCTCTATGGCAATGGCCCAGGCTATGCGCTTGG
CGGGGGCTCGAGGCCCGATGTTAATGGCAGCACAAGCGAGGAACCCTCGTACCGGCAGCAGGCGGCCGTGCCCCTGGCTA
GCGAGACCCACGGGGGCGAAGACGTGGCGGTGTTCGCGCGAGGCCCGCAGGCGCACCTGGTGCACGGCGTGCAGGAGGAG
ACCTTCGTGGCGCACATCATGGCCTTTGCGGGCTGCGTGGAGCCCTACACCGACTGCAATCTGCCAGCCCCCTCCACCGC
CACCAGCATCCCCGACTGATAATCTAGAGGTCGAAATTCACCTCGAAAGC

Hpap (SEQ ID NO: 27)

CTTCGGCGTTCAGTAACACGCGTTAACGTGCTGAACAGCCGGGCATTTTTTTACGCTATACCCTACATAATAAAACCGGA
GCTACCATGAATAAGAAGGTACTGACCCTTTCTGCCGTGATGGCAAGTCTGTTATTCGGCGCGCACGCGCACGCGGCGAT
CCCAGTTGAGGAGGAGAACCCGGACTTCTGGAACCGCGAGGCAGCCGAGGCCCTGGGTGCCGCCAAGAAGCTGCAGCCTG
CACAGACAGCCGCCAAGAACCTCATCATCTTCCTGGGCGATGGGATGGGGGTGTCTACGGTGACAGCTGCCAGGATCCTA
AAAGGGCAGAAGAAGGACAAACTGGGGCCTGAGATACCCCTGGCCATGGACCGCTTCCCATATGTGGCTCTGTCCAAGAC
ATACAATGTAGACAAACATGTGCCAGACAGTGGAGCCACAGCCACGGCCTACCTGTGCGGGGTCAAGGGCAACTTCCAGA
CCATTGGCTTGAGTGCAGCCGCCCGCTTTAACCAGTGCAACACGACACGCGGCAACGAGGTCATCTCCGTGATGAATCGG
GCCAAGAAAGCAGGGAAGTCAGTGGGAGTGGTAACCACCACACGAGTGCAGCACGCCTCGCCAGCCGGCACCTACGCCCA
CACGGTGAACCGCAACTGGTACTCGGACGCCGACGTGCCTGCCTCGGCCCGCCAGGAGGGGTGCCAGGACATCGCTACGC
AGCTCATCTCCAACATGGACATTGACGTGATCCTAGGTGGAGGCCGAAAGTACATGTTTCCCATGGGAACCCCAGACCCT
GAGTACCCAGATGACTACAGCCAAGGTGGGACCAGGCTGGACGGGAAGAATCTGGTGCAGGAATGGCTGGCGAAGCGCCA
GGGTGCCCGGTATGTGTGGAACCGCACTGAGCTCATGCAGGCTTCCCTGGACCCGTCTGTGACCCATCTCATGGGTCTCT
TTGAGCCTGGAGACATGAAATACGAGATCCACCGAGACTCCACACTGGACCCCTCCCTGATGGAGATGACAGAGGCTGCC
CTGCGCCTGCTGAGCAGGAACCCCCGCGGCTTCTTCCTCTTCGTGGAGGGTGGTCGCATCGACCATGGTCATCATGAAAG
CAGGGCTTACCGGGCACTGACTGAGACGATCATGTTCGACGACGCCATTGAGAGGGCGGGCCAGCTCACCAGCGAGGAGG
ACACGCTGAGCCTCGTCACTGCCGACCACTCCCACGTCTTCTCCTTCGGAGGCTACCCCCTGCGAGGGAGCTCCATCTTC
GGGCTGGCCCCTGGCAAGGCCCGGGACAGGAAGGCCTACACGGTCCTCCTATACGGAAACGGTCCAGGCTATGTGCTCAA
GGACGGCGCCCGGCCGGATGTTACCGAGAGCGAGAGCGGGAGCCCCGAGTATCGGCAGCAGTCAGCAGTGCCCCTGGACG

Figure 5: 2/3

AAGAGACCCACGCAGGCGAGGACGTGGCGGTGTTCGCGCGCGGCCCGCAGGCGCACCTGGTTCACGGCGTGCAGGAGCAG
ACCTTCATAGCGCACGTCATGGCCTTCGCCGCCTGCCTGGAGCCCTACACCGCCTGCGACCTGGCGCCCCCCGCCGGCAC
CACCGACTGATAAAAGCTTGGTCTAGAGGTCGAAATTCACCTCGAAAGC

AP01 (SEQ ID NO: 28)
CTTCGGCGTTCAGTAACACGCGTTAACGTGCTGAACAGCCGGGCATTCTTTTACGCTATACCCTACATAATAAAACCGGA
GCTACCATGAATAAGAAGGTACTGACCCTTTCTGCCGTGATGGCAAGTCTGTTATTCGGGGCCCACGCGCACGCGGCGAT
CCCAGTTGAGGAGGAGAACCCGGACTTCTGGAACCGCGAGGCAGCCAAGGCCCTGGGTGCCGCCAAGAAGCTGCAGCCTG
CACAGACAGCCGCCAAGAACCTCATCATCTTCCTGGGCGATGGGATGGGGGTGTCTACGGTGACAGCTGCCAGGATCCTA
AAAGGGCAGAAGAAGGACAAACTGGGGCCTGAGATACCCCTGGCCATGGACCGCTTCCCATATGTGGCTCTGTCCAAGAC
ATACAATGTAGACAAACATGTGCCAGACAGTGGAGCCACAGCCACGGCCTACCTGTGCGGGGTCAAGGGCAACTTCCAGA
CCATTGGCTTGAGTGCAGCCGCCCGCTTTAACCAGTGCAACACGACACGC

AP03 (SEQ ID NO: 29)
CTTCGGCGTTCAGTAACACGCGTTAACGTGCTGAACAGCCGGGCATTTTTTTACGCTATACCCTACATAATAAAACCGGA
GCTACCATGAATAAGAAGGTACTGACCCTTTCTGCCGTGATGGCAAGTCTGTTATTCGGCGCGCACGCGCACGCGGCGAT
CCCAGTTGAGGAGGAGAACCCGGACTTCTGGAACCGCGAGGCAGCCAAGGCCCTGGGTGCCGCCAAGAAGCTGCAGCCTG
CACAGACAGCCGCCAAGAACCTCATCATCTTCCTGGGCGATGGGATGGGGGTGTCTACGGTGACAGCTGCCAGGATCCTA
AAAGGGCAGAAGAAGGACAAACTGGGGCCTGAGATACCCCTGGCCATGGACCGCTTCCCATATGTGGCTCTGTCCAAGAC
ATACAATGTAGACAAACATGTGCCAGACAGTGGAGCCACAGCCACGGCCTACCTGTGTGGGGTCAA

AP05 (SEQ ID NO: 30)
CTTCGGCGTTCAGTAACACGCGTTAACGTGCTGAACAGCCGGGCATTTTTTTACGCTATACCCTACATAATAAAACCGGA
GCTACCATGAATAAGAAGGTACTGACCCTTTCTGCCGTGATGGCAAGTCTGTTATTCGGCGCGCACGCGCACGCGGCGAT
CCCAGTTGAGGAGGAGAACCCGGACTTCTGGAACCGCGAGGCAGCCAAGGCCCTGGGTGCCGCCAAGAAGCTGCAGCCTG
CACAGACAGCCGCCAAGAACCTCGTCATCTTCCTGGGCGATGGGATGGGGGTGCCTACGGTGACAGCCACTCGGACCCTA
AAAGGGCAGAAGAAGGACAAACTGGAGCCTGAGATACCCCTGGCCATGGACCGCTTCCCATATGTGGCTCTGTCCAAGAC
ATACAACGTGGACAGACAGGTGCCAGACAGTGGAGCCACAGCCACGGCCTACCTGTGCGGGGTCAAGGGCAACTTCCAGA
CCATTGGCTTGAGTGCAGCCGCCCGCTTTAACCAGTGCAACACGACACGCGGCAACGAGGTCATCTCCGTGATGAATCGG
GCCAAGAAAGCAGGGAAGTCAGTGGGAGTGGTAACCACCACACGAGTGCAGCACGCCTCGCCAGCCGGCACCTACGCCCA
CACGG

AP06 (SEQ ID NO: 31)
CTTCGGCGTTCAGTAACACGCGTTAACGTGCTGAACAGCCGGGCATTTTTTTACGCTATACCCTACATAATAAAACCGGA
GCTACCATGAATAAGAAGGTACTGACCCTTTCTGCCGTGATGGCAAGTCTGTTATTCGGCGCGCACGCGCACGCGGCGAT
CCCAGTTGAGGAGGAGAACCCGGACTTCTGGAACCGCGAGGCAGCCAAGGCCCTGGGTGCCGCCAAGAAGCTGCAGCCTG
CACAGACAGCCGCCAAGAACCTCATCATCTTCCTGGGCGATGGGATGGGGGTGTCTACGGTGACAGCTGCCAGGATCCTA
AAAGGGCAGAAGAAGGACAAACTGGGGCCTGAGATACCCCTGGCCATGGACCGCTTCCCATATGTGGCTCTGTCCAAGAC
ATACAATGTAGACAAACATGTGCCAGACAGTGGAGCCACAGCCACGGCCTACCTGTGCGGGGTCAAGGGCAACTTCCAGA
CCATTGGCTTGAGTGCAGCCGCCCGCTTTAACCAGTGCAACACGACACGCGGCAGCGAGGTCATCTCCGTGATGAACCGG
GCCAAGAAAGCAGG

AP11 (SEQ ID NO: 32)
CTTCGGCGTTCAGTAACACGCGTTAACGTGCTGAACAGCCGGGCATTTTTTTACGCTATACCCTACATAATAAAACCGGA
GCTACCATGAATAAGAAGGTATTGACCCTTTCTGCCGTGATGGCAAGTCTGTTATTCGGCGCGCACGCGCACGCGGCGAT
CCCAGTTGAGGAGGAGAACCCGGACTTCTGGAACCGCGAGGCAGCCAAGGCCCTGGGTGCCGCCAAGAAGCTGCAGCCTG
CACAGACAGCCGCCAAGAACCTCATCATCTTCCTGGGCGATGGGATGGGGGTGTCTACGGTGACAGCTGCCAGGATCCTA

Figure 5: 3/3

AAAGGGCAGATGAATGGCAAACTGGGACCTGAGACACCCCTGGCCATGGACCAGTTCCCATACGTGGCTCTGTCCAAGAC
ATACAATGTAGACAAACATGTGCCAGACAGTGGAGCCACAGCCACGGCCTACCTGTGCGGGGTCAAGGGCAACTTCCAGA
CCATTGGCTTGAGTGCAGCCGCCCGCTTTAACCAGTGCAACACGACACGCGGCAACGAGGTCATCTCCGTGATGAATCGG
GCCAAGAAAGCAGGGAAGTCAGTGGGAGTGGTAACCACCACACGAGTGCAGCACGCCTCG

AP15 (SEQ ID NO: 33)
CTTCGGCGTTCAGTAACACGCGTTACCGTGCTGAACAGCCGGGCATTTTTTTACGCTATACCCTACATAATAAAACCGGA
GCTACCATGAATAAGAAGGTACTGACCCTTTCTGCCGTGATGGCAAGTCTGTTATTCGGGGCCCACGCGCACGCGGCGAT
CCCAGCTGAGGAGGAAAACCCCGCCTTCTGGAACCGCCAGGCAGCCCAGGCCCTTGATGTAGCCAAGAAGTTGCAGCCGA
TCCAGACAGCTGCCAAGAATGTCATCCTCTTCTTGGGGGATGGGATGGGGGTGTCTACGGTGACAGCTGCCAGGATCCTA
AAAGGGCAGAAGAAGGACAAACTGGGGCCTGAGATACCCCTGGCCATGGACCGCTTCCCATATGTGGCTCTGTCCAAGAC
ATACAATGTAGACAAACATATGCCAGACAGTGGAGCCACAGCCACGGCCTACCTGTGCGGGGTCAAGGGCAACTTCCAGA
CCATTGGCTTGAGTGCAGCCGCCCGCTTTAACCAGTGCAACACGACACGCGGCAACGAGGTCATCTCCG

AP2 (SEQ ID NO: 34)
GAGATCCACCGAGACTCCACACTGGACCCCTCCCTGATGGAGATGACAGAGGCTGCCCTGCGCCTGCTGAGCAGGAACCC
CCGCGGCTTCTTCCTCTTCGTGGAGGGTGGTCGCATCGACCATGGTCATCATGAAAGCAGGGCTTACCGGGCACTGACTG
GGACGATCATGTTCGACGACGACATTGAGAGGGCGGGCCAGCTCACCAGCGAGGAGGACACGCTGAGCCTCGTCACTGCC
GACCACTCCCACGTCTTCTCCTTCGGAGGCTACCCCCTGCGAGGGAGCTCCATCTTCGGGCTGGCCCCTGGCAAGGCCCG
GGACAGGAAGGCCTACACGGTCCTCCTATACGGAAACGGTCCAGGCTATGTGCTCAAGGACGGCGCCCGGCCGGATGTTA
CCGAGGGCGAGAGCGGGAGCCCCGAGTATCGGCAGCAGTCAGCAGTGCCCCTGGACGAAGAGACCCACGCAGGCGAGGAC
GTGGCGGTGTTCGCGCGCGGCCCGCAGGCGCACCTGGTGCACGGCGTGCAGGAGGAGACCTTCGTGGCGCACATCATGGC
CTTTGCGGGCTGCGTGGAGCCCTACACCGACTGCAATCTGCCAGCCCCCTCCACCGCCACCAGCATCCCCGACTGATAAT
CTAGAGGTCGAAATTCACCTCGAAAGC

>AP9 (SEQ ID NO: 35)
CACCGAGACTCCACACTGGACCCCTCCCTGATGGAGATGACAGAGGCTGCCCTGCGCCTGCTGAACAGGAACCCCCGCGG
CTTCTTCCTCTCCGTGGAGGGTGGTCGCATCGACCATGGTCATCATGAAAGCAGGGCTTACCGGGCACTGACTGAGACGA
TCATGTTCGACGACGCCATTGAGAGGGCGGGCCAGCTCACCAGCGAGGAGGACACGCTGAGCCTCGTCACTGCCGACCAC
TCCCACGTCTTCTCCTTCGGAGGCTACCCCCTGCGAGGGAGCTCCATCTTCGGGCTGGCCCCTGGCAAGGCCCGGGACAG
GAAGGCCTACACGGTCCTCCTATACGGAAACGGTCCAGGCTATGTGCTCAAGGACGGCGCCCGGCCGGATGTTACCGAGA
GCGAGAGCGGGAGCCCCGAGTATCGGCAGCAGTCAGCAGTGCCCCTGGACGAAGAGACCCACGCAGGCGAGGACGTGGCG
GTGTTCGCGCGCGGCCCGCAGGCGCACCTGGTTCACGGCGTGCAGGAGCAGACCTTCATAGCGCACGTCATGGCCTTCGC
CGCCTGCCTGGAGCCCTACACCGCCTGCGACCTGGCGCCCCCCGCCGGCACCACCGACTGATAAAAGTTTGGCTCTAGAG
GTCGAAATTCACCTCGAAAGC

>AP13 (SEQ ID NO: 36)
ACCGAGACTCCACACTGGACCCCTCCCTGATGGAGATGACAGAGGCTGCCCTGCGCCTGCTGAGCAGGAACCCCCGCGGC
TTCTTCCTCTTCGTGGAGGGTGGTCGCATCGACCATGGTCATCATGAAAGCAGGGCTTACCGGGCACTGACTGAGACGAT
CATGTTCGACGACGCCATCGAGAGGGCGGGCCAGCTCACCAGCGAGGAGGACACGCTGAGCCTCGTCACTGCCGACCACT
CCCACGTCTTCTCCTTCGGAGGCTACCCCCTGCGAGGGAGCTCCATCTTCGGGCTGGCCCCTGGCAAGGCCCGGGACAGG
AAGGCCTACACGGTCCTCCTATACGGAAACGGTCCAGGCTATGTGCTCAAGGACGGCGCCCGGCCGGATGTTACCGAGAG
CGAGAGCGGGAGCCCCGAGTATCGGCAGCAGCCAGCAGTGCCCCTGGACGAAGAGACCCACGCAGGCGAGGACGTGGCGG
TGTTCGCGCGCGGCCCGCAGGCGCACCTGGTTCACGGCGTGCAGGAGCAGACCTTCATAGCGCACGTCATGGCCTTCGCC
GCCTGCCTGGAGCCCTACACCGCCTGCGACCTGGCGCCCCCCGCCGGCACCACCGACTGATAAAAGCTTGGCTCTAGAGG
TCGAAATTCACCTCGAAAGC

SEQUENCE LISTING

<110> Roche Diagnostics GmbH

<120> A walk-through technique for in vitro recombination of polynucleotide sequences

<130> 19061-Koe

<160> 36

<170> PatentIn version 3.1

<210> 1

<211> 28

<212> DNA

<213> Artificial primer

<400> 1
ggtagagcga gtctcgaggg ggagatgc 28

<210> 2

<211> 22

<212> DNA

<213> Artificial primer

<400> 2
agccggcgtg acgtgggtca gc 22

<210> 3

<211> 21

<212> DNA

<213> Artificial primer

<400> 3
cttcggcgtt cagtaacacg c 21

<210> 4

<211> 22

<212> DNA

<213> Artificial primer

<400> 4
gctttcgagg tgaatttcga cc 22

<210> 5

<211> 21

<212> DNA

<213> Artificial primer

<400> 5
ggtcacgtct gtgatcaacc g 21

<210> 6

<211> 21

<212> DNA

<213> Artificial primer

<400> 6
cggttgatca cagacgtgac c 21

<210> 7

<211> 23

<212> DNA

<213> Artificial primer

<400> 7
cgcaaagctt atatggcact gac 23

<210> 8

<211> 23

<212> DNA

<213> Artificial primer

<400> 8
gtcagtgcca tataagcttt gcc 23

<210> 9

<211> 23

<212> DNA

<213> Artificial primer

<400> 9
ccaagaaagc agggaagtca gtg 23

<210> 10

<211> 23

<212> DNA

<213> Artificial primer

<400> 10
cactgacttc cctgctttct tgg 23

<210> 11

<211> 22

<212> DNA

<213> Artificial primer

<400> 11
catgttcgac gacgccattg ag 22

<210> 12

<211> 22

<212> DNA

<213> Artificial primer

<400> 12
ctcaatggcg tcgtcgaaca tg 22

<210> 13

<211> 700

<212> DNA

<213> Human placental phosphatase

<400> 13
gttaacgtgc tgaacagccg ggcatttttt tacgctatac cctacataat aaaaccggag 60

ctaccatgaa taagaaggta ctgacccttt ctgccgtgat ggcaagtctg ttattcggcg 120

cgcacgcgca cgcggcgatc ccagttgagg aggagaaccc ggacttctgg aaccgcgagg 180

cagccgaggc cctgggtgcc gccaagaagc tgcagcctgc acagacagcc gccaagaacc 240

aagggcagaa gaaggacaaa ctggggcctg agataccccct ggccatggac cgcttcccat 360

atgtggctct gtccaagaca tacaatgtag acaaacatgt gccagacagt ggagccacag 420

ccacggccta cctgtgcggg gtcaagggca acttccagac cattggcttg agtgcagccg 480

cccgctttaa ccagtgcaac acgacacgcg gcaacgaggt catctccgtg atgaatcggg 540

ccaagaaagc agggaagtca gtgggagtgg taaccaccac acgagtgcag cacgcctcgc 600

cagccggcac ctacgcccac acggtgaacc gcaactggta ctcggacgcc gacgtgcctg 660

cctcggcccg ccaggagggg tgccaggaca tcgctacgca 700

<210> 14

<211> 700

<212> DNA

<213> Calf intestinal alkaline phosphatase

<400> 14

gttaacgtgc tgaacagccg ggcatttttt tacgctatac cctacataat aaaaccggag 60

ctaccatgaa taagaaggta ctgacccttt ctgccgtgat ggcaagtctg ttattcgggg 120

cccacgcgca cgcggcgatc ccagctgagg aggaaaaccc cgccttctgg aaccgccagg 180

cagcccaggc ccttgatgta gccaagaagt tgcagccgat ccagacagct gccaagaatg 240

tcatcctctt cttgggggat gggatggggg tgcctacggt gacagccact cggatcctaa 300

aggggcagat gaatggcaaa ctgggacctg agacacccct ggccatggac cagttcccat 360

acgtggctct gtccaagaca tacaacgtgg acagacaggt gccagacagc gcaggcactg 420

ccactgccta cctgtgtggg gtcaagggca actacagaac catcggtgta agtgcagccg 480

cccgctacaa tcagtgcaac acgacacgtg ggaatgaggt cacgtctgtg atcaaccggg 540

ccaagaaagc agggaaggcc gtgggagtgg tgaccaccac cagggtgcag catgcctccc 600

cagccggggc ctacgcgcac acggtgaacc gaaactggta ctcagacgcc gacctgcctg 660

ctgatgcaca gaagaatggc tgccaggaca tcgccgcaca 700

<210> 15

<211> 509

<212> DNA

<213> Human placental phosphatase and calf intestinal alkaline phosphatase

<400> 15
gttaacgtgc tgaacagccg ggcattcttt tacgctatac cctacataat aaaaccggag 60

ctaccatgaa taagaaggta ctgacccttt ctgccgtgat ggcaagtctg ttattcgggg 120

cccacgcgca cgcggcgatc ccagttgagg aggagaaccc ggacttctgg aaccgcgagg 180

cagccaaggc cctgggtgcc gccaagaagc tgcagcctgc acagacagcc gccaagaacc 240

tcatcatctt cctgggcgat gggatggggg tgtctacggt gacagctgcc aggatcctaa 300

aagggcagaa gaaggacaaa ctggggcctg agataccccc ggccatggac cgcttcccat 360

atgtggctct gtccaagaca tacaatgtag acaaacatgt gccagacagt ggagccacag 420

ccacggccta cctgtgcggg gtcaagggca acttccagac cattggcttg agtgcagccg 480

cccgctttaa ccagtgcaac acgacacgc 509

<210> 16

<211> 445

<212> DNA

<213> Human placental phosphatase and calf intestinal alkaline phosphatase

<400> 16
gttaacgtgc tgaacagccg ggcatttttt tacgctatac cctacataat aaaaccggag 60

ctaccatgaa taagaaggta ctgacccttt ctgccgtgat ggcaagtctg ttattcggcg 120

cgcacgcgca cgcggcgatc ccagttgagg aggagaaccc ggacttctgg aaccgcgagg 180

cagccaaggc cctgggtgcc gccaagaagc tgcagcctgc acagacagcc gccaagaacc 240

tcatcatctt cctgggcgat gggatggggg tgtctacggt gacagctgcc aggatcctaa 300

aagggcagaa gaaggacaaa ctggggcctg agataccccct ggccatggac cgcttcccat 360

atgtggctct gtccaagaca tacaatgtag acaaacatgt gccagacagt ggagccacag 420

ccacggccta cctgtgtggg gtcaa 445

<210> 17

<211> 624

<212> DNA

<213> Human placental phosphatase and calf intestinal alkaline phosphatase

<400> 17

gttaacgtgc tgaacagccg ggcatttttt tacgctatac cctacataat aaaaccggag 60

ctaccatgaa taagaaggta ctgacccttt ctgccgtgat gttaacgtgc tgaacagccg 120

ggcatttttt tacgctatac cctacataat aaaaccggag ctaccatgaa taagaaggta 180

ctgacccttt ctgccgtgat gccaagaagc tgcagcctgc acagacagcc gccaagaacc 240

tcgtcatctt cctgggcgat gggatggggg tgcctacggt gacagccact cggaccctaa 300

aagggcagaa gaaggacaaa ctggagcctg agataccccct ggccatggac cgcttcccat 360

atgtggctct gtccaagaca tacaacgtgg acagacaggt gccagacagt ggagccacag 420

ccacggccta cctgtgcggg gtcaagggca acttccagac cattggcttg agtgcagccg 480

cccgctttaa ccagtgcaac acgacacgcg gcaacgaggt catctccgtg atgaatcggg 540

ccaagaaagc agggaagtca gtgggagtgg taaccaccac acgagtgcag cacgcctcgc 600

cagccggcac ctacgcccac acgg 624

<210> 18

<211> 553

<212> DNA

<213> Human placental phosphatase and calf intestinal alkaline phosphatase

<400> 18
gttaacgtgc tgaacagccg ggcatttttt tacgctatac cctacataat aaaaccggag 60

ctaccatgaa taagaaggta ctgacccttt ctgccgtgat ggcaagtctg ttattcggcg 120

cgcacgcgca cgcggcgatc ccagttgagg aggagaaccc ggacttctgg aaccgcgagg 180

cagccaaggc cctgggtgcc gccaagaagc tgcagcctgc acagacagcc gccaagaacc 240

tcatcatctt cctgggcgat gggatggggg tgtctacggt gacagctgcc aggatcctaa 300

aagggcagaa gaaggacaaa ctggggcctg agataccccc ggccatggac cgcttcccat 360

atgtggctct gtccaagaca tacaatgtag acaaacatgt gccagacagt ggagccacag 420

ccacggccta cctgtgcggg gtcaagggca acttccagac cattggcttg agtgcagccg 480

cccgctttaa ccagtgcaac acgacacgcg gcagcgaggt catctccgtg atgaaccggg 540

ccaagaaagc agg 553

<210> 19

<211> 599

<212> DNA

<213> Human placental phosphatase and calf intestinal alkaline phosphatase

<400> 19
gttaacgtgc tgaacagccg ggcatttttt tacgctatac cctacataat aaaaccggag 60

ctaccatgaa taagaaggta ttgacccttt ctgccgtgat ggcaagtctg ttattcggcg 120

cgcacgcgca cgcggcgatc ccagttgagg aggagaaccc ggacttctgg aaccgcgagg 180

cagccaaggc cctgggtgcc gccaagaagc tgcagcctgc acagacagcc gccaagaacc 240

tcatcatctt cctgggcgat gggatggggg tgtctacggt gacagctgcc aggatcctaa 300

aagggcagat gaatggcaaa ctgggacctg agacacccct ggccatggac cagttcccat 360

acgtggctct gtccaagaca tacaatgtag acaaacatgt gccagacagt ggagccacag 420

ccacggccta cctgtgcggg gtcaagggca acttccagac cattggcttg agtgcagccg 480

cccgctttaa ccagtgcaac acgacacgcg gcaacgaggt catctccgtg atgaatcggg 540

ccaagaaagc agggaagtca gtgggagtgg taaccaccac acgagtgcag cacgcctcg 599

<210> 20

<211> 528

<212> DNA

<213> Human placental phosphatase and calf intestinal alkaline

<400> 20
gttaccgtgc tgaacagccg ggcatttttt tacgctatac cctacataat aaaaccggag 60

ctaccatgaa taagaaggta ctgacccttt ctgccgtgat ggcaagtctg ttattcgggg 120

cccacgcgca cgcggcgatc ccagctgagg aggaaaaccc cgccttctgg aaccgccagg 180

cagcccaggc ccttgatgta gccaagaagt tgcagccgat ccagacagct gccaagaatg 240

tcatcctctt cttgggggat gggatggggg tgtctacggt gacagctgcc aggatcctaa 300

aagggcagaa gaaggacaaa ctggggcctg agataccccc ggccatggac cgcttcccat 360

atgtggctct gtccaagaca tacaatgtag acaaacatat gccagacagt ggagccacag 420

ccacggccta cctgtgcggg gtcaagggca acttccagac cattggcttg agtgcagccg 480

cccgctttaa ccagtgcaac acgacacgcg gcaacgaggt catctccg 528

<210> 21

<211> 606

<212> DNA

<213> Human placental phosphatase

<400> 21
cttccctgga cccgtctgtg acccatctca tgggtctctt tgagcctgga gacatgaaat 60

acgagatcca ccgagactcc acactggacc cctccctgat ggagatgaca gaggctgccc 120

tgcgcctgct gagcaggaac ccccgcggct tcttcctctt cgtggagggt ggtcgcatcg 180

accatggtca tcatgaaagc agggcttacc gggcactgac tgagacgatc atgttcgacg 240

acgccattga gagggcgggc cagctcacca gcgaggagga cacgctgagc ctcgtcactg 300

ggtcctccta tacggaaacg gtccaggcta tgtgctcaag gacggcgccc ggccggatgt 360

taccgagagc gagagcggga gccccgagta tcggcagcag tcagcagtgc ccctggacga 420

agagacccac gcaggcgagg acgtggcggt gttcgcgcgc ggcccgcagg cgcacctggt 480

tcacggcgtg caggagcaga ccttcatagc gcacgtcatg gccttcgccg cctgcctgga 540

gccctacacc gcctgcgacc tggcgccccc cgccggcacc accgactgat aaaagcttgg 600

tctaga 606

<210> 22

<211> 707

<212> DNA

<213> Calf intestinal alkaline phosphatase

<400> 22

cggccgatga ctccagtgta acacacctca tgggcctctt tgagccggca gacatgaagt 60

ataatgttca gcaagaccac accaaggacc cgaccctggc ggagatgacg gaggcggccc 120

tgcaagtgct gagcaggaac ccccggggct tctacctctt cgtggaggga ggccgcattg 180

accacggtca ccatgacggc aaagcttata tggcactgac tgaggcgatc atgtttgaca 240

atgccatcgc caaggctaac gagctcacta gcgaactgga cacgctgatc cttgtcactg 300

cagaccactc ccatgtcttc tcctttggtg gctacacact gcgtgggacc tccattttcg 360

gtctggcccc cggcaaggcc ttagacagca agtcctacac ctccatcctc tatggcaatg 420

gcccaggcta tgcgcttggc gggggctcga ggcccgatgt taatggcagc acaagcgagg 480

aaccctcgta ccggcagcag gcggccgtgc ccctggctag cgagacccac gggggcgaag 540

acgtggcggt gttcgcgcga ggcccgcagg cgcacctggt gcacggcgtg caggaggaga 600

ccttcgtggc gcacatcatg gcctttgcgg gctgcgtgga gccctacacc gactgcaatc 660

tgccagcccc ctccaccgcc accagcatcc ccgactgata atctaga 707

<210> 23

<211> 763

<212> DNA

<213> Human placental phosphatase and calf intestinal alkaline phosphatase

<400> 23

gagatccacc gagactccac actggacccc tccctgatgg agatgacaga ggctgccctg 60

cgcctgctga gcaggaaccc ccgcggcttc ttcctcttcg tggagggtgg tcgcatcgac 120

catggtcatc atgaaagcag ggcttaccgg gcactgactg ggacgatcat gttcgacgac 180

gacattgaga gggcgggcca gctcaccagc gaggaggaca cgctgagcct cgtcactgcc 240

gaccactccc acgtcttctc cttcggaggc taccccctgc gagggagctc catcttcggg 300

ctggcccctg gcaaggcccg ggacaggaag gcctacacgg tcctcctata cggaaacggt 360

ccaggctatg tgctcaagga cggcgcccgg ccggatgtta ccgagggcga gagcgggagc 420

cccgagtatc ggcagcagtc agcagtgccc ctggacgaag agacccacgc aggcgaggac 480

gtggcggtgt tcgcgcgcgg cccgcaggcg cacctggtgc acggcgtgca ggaggagacc 540

ttcgtggcgc acatcatggc ctttgcgggc tgcgtggagc cctacaccga ctgcaatctg 600

ccagcccccc ccaccgccac cagcatcccc gactgataat ctagaggtcg aaattcacct 660

cgaaagcaag ggcgaattct gcagatatcc atcacactgg cggccgctcg agcatgcatc 720

tagagggccc aattcgccct atagtgagtc gtattacaat tca 763

<210> 24

<211> 757

<212> DNA

<213> Human placental phosphatase and calf intestinal alkaline phosphatase

<400> 24
caccgagact ccacactgga cccctccctg atggagatga cagaggctgc cctgcgcctg 60

ctgaacagga acccccgcgg cttcttcctc tccgtggagg gtggtcgcat cgaccatggt 120

catcatgaaa gcagggctta ccgggcactg actgagacga tcatgttcga cgacgccatt 180

gagagggcgg gccagctcac cagcgaggag gacacgctga gcctcgtcac tgccgaccac 240

tcccacgtct tctccttcgg aggctacccc ctgcgaggga gctccatctt cgggctggcc 300

cctggcaagg cccgggacag gaaggcctac acggtcctcc tatacggaaa cggtccaggc 360

tatgtgctca aggacggcgc ccggccggat gttaccgaga gcgagagcgg gagccccgag 420

tatcggcagc agtcagcagt gcccctggac gaagagaccc acgcaggcga ggacgtggcg 480

gtgttcgcgc gcggcccgca ggcgcacctg gttcacggcg tgcaggagca gaccttcata 540

gcgcacgtca tggccttcgc cgcctgcctg gagccctaca ccgcctgcga cctggcgccc 600

cccgccggca ccaccgactg ataaaagttt ggctctagag gtcgaaattc acctcgaaag 660

caagggcgaa ttctgcagat atccatcaca ctggcggccg ctcgagcatg catctagagg 720

gcccaattcg ccctatagtg agtcgtatta caattca 757

<210> 25

<211> 756

<212> DNA

<213> Human placental phosphatase and calf intestinal alkaline phosphatase

<400> 25
accgagactc cacactggac ccctccctga tggagatgac agaggctgcc ctgcgcctgc 60

tgagcaggaa cccccgcggc ttcttcctct tcgtggaggg tggtcgcatc gaccatggtc 120

atcatgaaag cagggcttac cgggcactga ctgagacgat catgttcgac gacgccatcg 180

agagggcggg ccagctcacc agcgaggagg acacgctgag cctcgtcact gccgaccact 240

cccacgtctt ctccttcgga ggctaccccc tgcgagggag ctccatcttc gggctggccc 300

ctggcaaggc ccgggacagg aaggcctaca cggtcctcct atacggaaac ggtccaggct 360

atgtgctcaa ggacggcgcc cggccggatg ttaccgagag cgagagcggg agccccgagt 420

atcggcagca gccagcagtg cccctggacg aagagaccca cgcaggcgag gacgtggcgg 480

tgttcgcgcg cggcccgcag gcgcacctgg ttcacggcgt gcaggagcag accttcatag 540

cgcacgtcat ggccttcgcc gcctgcctgg agccctacac cgcctgcgac ctggcgcccc 600

ccgccggcac caccgactga taaaagcttg gctctagagg tcgaaattca cctcgaaagc 660

aagggcgaat tctgcagata tccatcacac tggcggccgc tcgagcatgc atctagaggg 720

cccaattcgc cctatagtga gtcgtattac aattca 756

<210> 26

<211> 1650

<212> DNA

<213> Calf intestinal alkaline phosphatase

<400> 26

cttcggcgtt cagtaacacg cgttaacgtg ctgaacagcc gggcattttt ttacgctata 60

ccctacataa taaaaccgga gctaccatga ataagaaggt actgaccctt tctgccgtga 120

tggcaagtct gttattcggg gcccacgcgc acgcggcgat cccagctgag gaggaaaacc 180

ccgccttctg gaaccgccag gcagcccagg cccttgatgt agccaagaag ttgcagccga 240

tccagacagc tgccaagaat gtcatcctct tcttgggggа tgggatgggg gtgcctacgg 300

tgacagccac tcggatccta aaggggcaga tgaatggcaa actgggacct gagacacccc 360

tggccatgga ccagttccca tacgtggctc tgtccaagac atacaacgtg gacagacagg 420

tgccagacag cgcaggcact gccactgcct acctgtgtgg ggtcaagggc aactacagaa 480

ccatcggtgt aagtgcagcc gcccgctaca atcagtgcaa cacgacacgt gggaatgagg 540

tcacgtctgt gatcaaccgg gccaagaaag cagggaaggc cgtgggagtg gtgaccacca 600

ccagggtgca gcatgcctcc ccagccgggg cctacgcgca cacggtgaac cgaaactggt 660

actcagacgc cgacctgcct gctgatgcac agaagaatgg ctgccaggac atcgccgcac 720

agctggtcta caacatggat attgacgtga tcctgggtgg aggccgaatg tacatgtttc 780

ctgaggggac cccagaccct gaatacccag atgatgccag tgtgaatgga gtccggaagg 840

acaagcagaa cctggtgcag gaatggcagg ccaagcacca gggagcccag tatgtgtgga 900

accgcactgc gctccttcag gcggccgatg actccagtgt aacacacctc atgggcctct 960

ttgagccggc agacatgaag tataatgttc agcaagacca caccaaggac ccgaccctgg 1020

cggagatgac ggaggcggcc ctgcaagtgc tgagcaggaa cccccggggc ttctacctct 1080

tcgtggaggg aggccgcatt gaccacggtc accatgacgg caaagcttat atggcactga 1140

ctgaggcgat catgtttgac aatgccatcg ccaaggctaa cgagctcact agcgaactgg 1200

acacgctgat ccttgtcact gcagaccact cccatgtctt ctcctttggt ggctacacac 1260

tgcgtgggac ctccattttc ggtctggccc ccggcaaggc cttagacagc aagtcctaca 1320

cctccatcct ctatggcaat ggcccaggct atgcgcttgg cgggggctcg aggcccgatg 1380

ttaatggcag cacaagcgag gaaccctcgt accggcagca ggcggccgtg cccctggcta 1440

gcgagaccca cgggggcgaa gacgtggcgg tgttcgcgcg aggcccgcag gcgcacctgg 1500

tgcacggcgt gcaggaggag accttcgtgg cgcacatcat ggcctttgcg ggctgcgtgg 1560

agccctacac cgactgcaat ctgccagccc cctccaccgc caccagcatc cccgactgat 1620

aatctagagg tcgaaattca cctcgaaagc 1650

<210> 27

<211> 1649

<212> DNA

<213> Human placental phosphatase

<400> 27
cttcggcgtt cagtaacacg cgttaacgtg ctgaacagcc gggcattttt ttacgctata 60
ccctacataa taaaaccgga gctaccatga ataagaaggt actgaccctt tctgccgtga 120
tggcaagtct gttattcggc gcgcacgcgc acgcggcgat cccagttgag gaggagaacc 180
cggacttctg gaaccgcgag gcagccgagg ccctgggtgc cgccaagaag ctgcagcctg 240
cacagacagc cgccaagaac ctcatcatct tcctgggcga tgggatgggg gtgtctacgg 300
tgacagctgc caggatccta aaagggcaga agaaggacaa actggggcct gagataccc 360
tggccatgga ccgcttccca tatgtggctc tgtccaagac atacaatgta gacaaacatg 420
tgccagacag tggagccaca gccacggcct acctgtgcgg ggtcaagggc aacttccaga 480
ccattggctt gagtgcagcc gcccgcttta accagtgcaa cacgacacgc ggcaacgagg 540
tcatctccgt gatgaatcgg gccaagaaag cagggaagtc agtgggagtg gtaaccacca 600
cacgagtgca gcacgcctcg ccagccggca cctacgccca cacggtgaac cgcaactggt 660
actcggacgc cgacgtgcct gcctcggccc gccaggaggg gtgccaggac atcgctacgc 720
agctcatctc caacatggac attgacgtga tcctaggtgg aggccgaaag tacatgtttc 780
ccatgggaac cccagaccct gagtacccag atgactacag ccaaggtggg accaggctgg 840
acgggaagaa tctggtgcag gaatggctgg cgaagcgcca gggtgcccgg tatgtgtgga 900
accgcactga gctcatgcag gcttccctgg acccgtctgt gacccatctc atgggtctct 960
ttgagcctgg agacatgaaa tacgagatcc accgagactc cacactggac ccctccctga 1020
tggagatgac agaggctgcc ctgcgcctgc tgagcaggaa cccccgcggc ttcttcctct 1080
tcgtggaggg tggtcgcatc gaccatggtc atcatgaaag cagggcttac cgggcactga 1140
ctgagacgat catgttcgac gacgccattg agagggcggg ccagctcacc agcgaggagg 1200
acacgctgag cctcgtcact gccgaccact cccacgtctt ctccttcgga ggctacccc 1260
tgcgagggag ctccatcttc gggctggccc ctggcaaggc ccgggacagg aaggcctaca 1320
cggtcctcct atacggaaac ggtccaggct atgtgctcaa ggacggcgcc cggccggatg 1380
ttaccgagag cgagagcggg agccccgagt atcggcagca gtcagcagtg cccctggacg 1440

aagagaccca cgcaggcgag gacgtggcgg tgttcgcgcg cggcccgcag gcgcacctgg 1500

ttcacggcgt gcaggagcag accttcatag cgcacgtcat ggccttcgcc gcctgcctgg 1560

agccctacac cgcctgcgac ctggcgcccc ccgccggcac caccgactga taaaagcttg 1620

gtctagaggt cgaaattcac ctcgaaagc 1649

<210> 28

<211> 530

<212> DNA

<213> Calf intestinal alkaline phosphatase and human placental phosphatase

<400> 28

cttcggcgtt cagtaacacg cgttaacgtg ctgaacagcc gggcattctt ttacgctata 60

ccctacataa taaaaccgga gctaccatga ataagaaggt actgaccctt tctgccgtga 120

tggcaagtct gttattcggg gcccacgcgc acgcggcgat cccagttgag gaggagaacc 180

cggacttctg gaaccgcgag gcagccaagg ccctgggtgc cgccaagaag ctgcagcctg 240

cacagacagc cgccaagaac ctcatcatct tcctgggcga tgggatgggg gtgtctacgg 300

tgacagctgc caggatccta aaagggcaga agaaggacaa actggggcct gagataccc 360

tggccatgga ccgcttccca tatgtggctc tgtccaagac atacaatgta gacaaacatg 420

tgccagacag tggagccaca gccacggcct acctgtgcgg ggtcaagggc aacttccaga 480

ccattggctt gagtgcagcc gcccgcttta accagtgcaa cacgacacgc 530

<210> 29

<211> 466

<212> DNA

<213> Calf intestinal alkaline phosphatase and human placental phosphatase

<400> 29
cttcggcgtt cagtaacacg cgttaacgtg ctgaacagcc gggcattttt ttacgctata 60

ccctacataa taaaaccgga gctaccatga ataagaaggt actgaccctt tctgccgtga 120

tggcaagtct gttattcggc gcgcacgcgc acgcggcgat cccagttgag gaggagaacc 180

cggacttctg gaaccgcgag gcagccaagg ccctgggtgc cgccaagaag ctgcagcctg 240

cacagacagc cgccaagaac ctcatcatct tcctgggcga tgggatgggg gtgtctacgg 300

tgacagctgc caggatccta aaagggcaga agaaggacaa actggggcct gagataccc 360

tggccatgga ccgcttccca tatgtggctc tgtccaagac atacaatgta gacaaacatg 420

tgccagacag tggagccaca gccacggcct acctgtgtgg ggtcaa 466

<210> 30

<211> 645

<212> DNA

<213> Calf intestinal alkaline phosphatase and human placental phosphatase

<400> 30
cttcggcgtt cagtaacacg cgttaacgtg ctgaacagcc gggcattttt ttacgctata 60

ccctacataa taaaaccgga gctaccatga ataagaaggt actgaccctt tctgccgtga 120

tggcaagtct gttattcggc gcgcacgcgc acgcggcgat cccagttgag gaggagaacc 180

cggacttctg gaaccgcgag gcagccaagg ccctgggtgc cgccaagaag ctgcagcctg 240

cacagacagc cgccaagaac ctcgtcatct tcctgggcga tgggatgggg gtgcctacgg 300

tgacagccac tcggacccta aaagggcaga agaaggacaa actggagcct gagataccc 360

tggccatgga ccgcttccca tatgtggctc tgtccaagac atacaacgtg gacagacagg 420

tgccagacag tggagccaca gccacggcct acctgtgcgg ggtcaagggc aacttccaga 480

ccattggctt gagtgcagcc gcccgcttta accagtgcaa cacgacacgc ggcaacgagg 540

tcatctccgt gatgaatcgg gccaagaaag cagggaagtc agtgggagtg gtaaccacca 600

cacgagtgca gcacgcctcg ccagccggca cctacgccca cacgg 645

<210> 31

<211> 574

<212> DNA

<213> Calf intestinal alkaline phosphatase and human placental phosphatase

<400> 31
cttcggcgtt cagtaacacg cgttaacgtg ctgaacagcc gggcattttt ttacgctata 60

ccctacataa taaaaccgga gctaccatga ataagaaggt actgaccctt tctgccgtga 120

tggcaagtct gttattcggc gcgcacgcgc acgcggcgat cccagttgag gaggagaacc 180

cggacttctg gaaccgcgag gcagccaagg ccctgggtgc cgccaagaag ctgcagcctg 240

cacagacagc cgccaagaac ctcatcatct tcctgggcga tgggatgggg gtgtctacgg 300

tgacagctgc caggatccta aaagggcaga agaaggacaa actggggcct gagataccc 360

tggccatgga ccgcttccca tatgtggctc tgtccaagac atacaatgta gacaaacatg 420

tgccagacag tggagccaca gccacggcct acctgtgcgg ggtcaagggc aacttccaga 480

ccattggctt gagtgcagcc gcccgcttta accagtgcaa cacgacacgc ggcagcgagg 540

tcatctccgt gatgaaccgg gccaagaaag cagg 574

<210> 32

<211> 620

<212> DNA

<213> Calf intestinal alkaline phosphatase and human placental phosphatase

<400> 32
cttcggcgtt cagtaacacg cgttaacgtg ctgaacagcc gggcattttt ttacgctata 60

ccctacataa taaaaccgga gctaccatga ataagaaggt attgaccctt tctgccgtga 120

tggcaagtct gttattcggc gcgcacgcgc acgcggcgat cccagttgag gaggagaacc 180

cggacttctg gaaccgcgag gcagccaagg ccctgggtgc cgccaagaag ctgcagcctg 240

cacagacagc cgccaagaac ctcatcatct tcctgggcga tgggatgggg gtgtctacgg 300

tgacagctgc caggatccta aaagggcaga tgaatggcaa actgggacct gagacacccc 360

tggccatgga ccagttccca tacgtggctc tgtccaagac atacaatgta gacaaacatg 420

tgccagacag tggagccaca gccacggcct acctgtgcgg ggtcaagggc aacttccaga 480

ccattggctt gagtgcagcc gcccgcttta accagtgcaa cacgacacgc ggcaacgagg 540

tcatctccgt gatgaatcgg gccaagaaag cagggaagtc agtgggagtg gtaaccacca 600

cacgagtgca gcacgcctcg 620

<210> 33

<211> 549

<212> DNA

<213> Calf intestinal alkaline phosphatase and human placental phosphatase

<400> 33

cttcggcgtt cagtaacacg cgttaccgtg ctgaacagcc gggcattttt ttacgctata 60

ccctacataa taaaaccgga gctaccatga ataagaaggt actgaccctt tctgccgtga 120

tggcaagtct gttattcggg gcccacgcgc acgcggcgat cccagctgag gaggaaaacc 180

ccgccttctg gaaccgccag gcagcccagg cccttgatgt agccaagaag ttgcagccga 240

tccagacagc tgccaagaat gtcatcctct tcttggggga tgggatgggg gtgtctacgg 300

tgacagctgc caggatccta aaagggcaga agaaggacaa actggggcct gagatacccc 360

tggccatgga ccgcttccca tatgtggctc tgtccaagac atacaatgta gacaaacata 420

tgccagacag tggagccaca gccacggcct acctgtgcgg ggtcaagggc aacttccaga 480

ccattggctt gagtgcagcc gcccgcttta accagtgcaa cacgacacgc ggcaacgagg 540

tcatctccg 549

<210> 34

<211> 667

<212> DNA

<213> Calf intestinal alkaline phosphatase and human placental phosphatase

<400> 34
gagatccacc gagactccac actggacccc tccctgatgg agatgacaga ggctgccctg 60

cgcctgctga gcaggaaccc ccgcggcttc ttcctcttcg tggagggtgg tcgcatcgac 120

catggtcatc atgaaagcag ggcttaccgg gcactgactg ggacgatcat gttcgacgac 180

gacattgaga gggcgggcca gctcaccagc gaggaggaca cgctgagcct cgtcactgcc 240

gaccactccc acgtcttctc cttcggaggc taccccctgc gagggagctc catcttcggg 300

ctggcccctg gcaaggcccg ggacaggaag gcctacacgg tcctcctata cggaaacggt 360

ccaggctatg tgctcaagga cggcgcccgg ccggatgtta ccgagggcga gagcgggagc 420

cccgagtatc ggcagcagtc agcagtgccc ctggacgaag agacccacgc aggcgaggac 480

gtggcggtgt tcgcgcgcgg cccgcaggcg cacctggtgc acggcgtgca ggaggagacc 540

ttcgtggcgc acatcatggc ctttgcgggc tgcgtggagc cctacaccga ctgcaatctg 600

ccagcccccc ccaccgccac cagcatcccc gactgataat ctagaggtcg aaattcacct 660

cgaaagc 667

<210> 35

<211> 661

<212> DNA

<213> Calf intestinal alkaline phosphatase and human placental phosphatase

<400> 35
caccgagact ccacactgga cccctccctg atggagatga cagaggctgc cctgcgcctg 60

ctgaacagga acccccgcgg cttcttcctc tccgtggagg gtggtcgcat cgaccatggt 120

catcatgaaa gcagggctta ccgggcactg actgagacga tcatgttcga cgacgccatt 180

gagagggcgg gccagctcac cagcgaggag gacacgctga gcctcgtcac tgccgaccac 240

tcccacgtct tctccttcgg aggctacccc ctgcgaggga gctccatctt cgggctggcc 300

cctggcaagg cccgggacag gaaggcctac acggtcctcc tatacggaaa cggtccaggc 360

tatgtgctca aggacggcgc ccggccggat gttaccgaga gcgagagcgg gagccccgag 420

tatcggcagc agtcagcagt gcccctggac gaagagaccc acgcaggcga ggacgtggcg 480

gtgttcgcgc gcggcccgca ggcgcacctg gttcacggcg tgcaggagca gaccttcata 540

gcgcacgtca tggccttcgc cgcctgcctg gagccctaca ccgcctgcga cctggcgccc 600

cccgccggca ccaccgactg ataaaagttt ggctctagag gtcgaaattc acctcgaaag 660

c 661

<210> 36

<211> 660

<212> DNA

<213> Calf intestinal alkaline phosphatase and human placental phosphatase

<400> 36
accgagactc cacactggac ccctccctga tggagatgac agaggctgcc ctgcgcctgc 60

tgagcaggaa cccccgcggc ttcttcctct tcgtggaggg tggtcgcatc gaccatggtc 120

atcatgaaag cagggcttac cgggcactga ctgagacgat catgttcgac gacgccatcg 180

agagggcggg ccagctcacc agcgaggagg acacgctgag cctcgtcact gccgaccact 240

cccacgtctt ctccttcgga ggctacccca tgcgagggag ctccatcttc gggctggccc 300

ctggcaaggc ccgggacagg aaggcctaca cggtcctcct atacggaaac ggtccaggct 360

atgtgctcaa ggacggcgcc cggccggatg ttaccgagag cgagagcggg agccccgagt 420

atcggcagca gccagcagtg cccctggacg aagagaccca cgcaggcgag gacgtggcgg 480

tgttcgcgcg cggcccgcag gcgcacctgg ttcacggcgt gcaggagcag accttcatag 540

cgcacgtcat ggccttcgcc gcctgcctgg agccctacac cgcctgcgac ctggcgcccc 600

ccgccggcac caccgactga taaaagcttg gctctagagg tcgaaattca cctcgaaagc 660

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 January 2005 (13.01.2005)

PCT

(10) International Publication Number
WO 2005/003375 A2

(51) International Patent Classification[7]: C12Q

(21) International Application Number:
PCT/US2004/002570

(22) International Filing Date: 28 January 2004 (28.01.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/443,471	29 January 2003 (29.01.2003)	US
60/465,071	23 April 2003 (23.04.2003)	US
60/476,602	6 June 2003 (06.06.2003)	US
60/476,504	6 June 2003 (06.06.2003)	US
60/476,313	6 June 2003 (06.06.2003)	US
60/476,592	6 June 2003 (06.06.2003)	US
60/497,985	25 August 2003 (25.08.2003)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/443,471 (CON)
Filed on	29 January 2003 (29.01.2003)
US	60/465,071 (CON)
Filed on	23 April 2003 (23.04.2003)
US	60/476,602 (CON)
Filed on	6 June 2003 (06.06.2003)
US	60/476,504 (CON)
Filed on	6 June 2003 (06.06.2003)
US	60/476,313 (CON)
Filed on	6 June 2003 (06.06.2003)
US	60/476,592 (CON)
Filed on	6 June 2003 (06.06.2003)
US	60/497,985 (CON)
Filed on	25 August 2003 (25.08.2003)

(71) Applicant *(for all designated States except US)*: **454 CORPORATION** [US/US]; 20 Commercial Street, Branford, CT 06405 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **LEAMON, John, H.** [US/US]; 251 Nortontown Road, Guilford, CN 06437 (US). **LOHMAN, Kenton** [US/US]; 425 Schoolside Lane, Guilford, CN 06437 (US). **ROTHBERG, Jonathan, M.** [US/US]; 1701 Moose Hill Road, Gilford, CT 06437 (US). **WEINER, Michael, P.** [US/US]; 52 Leighton Trail, Guilford, CT 06437 (US).

(74) Agent: **ELRIFI, Ivor, R.**; Mintz, Levin, Cohn, Ferris, Glovsky, and Popeo, P., C., 666 Third Avenue, New York, NY 10017 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US (patent), UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, 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, IT, LU, MC, NL, 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*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHODS OF AMPLIFYING AND SEQUENCING NUCLEIC ACIDS

(57) Abstract: An apparatus and method for performing rapid DNA sequencing, such as genomic sequencing, is provided herein. The method includes the steps of preparing a sample DNA for genomic sequencing, amplifying the prepared DNA in a representative manner, and performing multiple sequencing reaction on the amplified DNA with only one primer hybridization step.

WO 2005/003375 A2

IDT Exhibit 1020.6525

METHODS OF AMPLIFYING AND SEQUENCING NUCLEIC ACIDS

FIELD OF THE INVENTION

This invention relates to a method and apparatus for determining the base sequences of DNA. More particularly, this invention relates to methods and an apparatus with which the base sequences of a genome can be amplified and determined automatically or semiautomatically.

BACKGROUND OF THE INVENTION

Development of rapid and sensitive nucleic acid sequencing methods utilizing automated DNA sequencers has revolutionized modern molecular biology. Analysis of entire genomes of plants, fungi, animals, bacteria, and viruses is now possible with a concerted effort by a series of machines and a team of technicians. However, the goal of rapid and automated or semiautomatic sequencing of a genome in a short time has not been possible. There continues to be technical problems for accurate sample preparation, amplification and sequencing.

One technical problem which hinders sequence analysis of genomes has been the inability of the investigator to rapidly prepare numerous nucleic acid sample encompassing a complete genome in a short period of time.

Another technical problem is the inability to representatively amplified a genome to a level that is compatible with the sensitivity of current sequencing methods. Modern economically feasible sequencing machines, while sensitive, still require in excess of one million copies of a DNA fragment for sequencing. Current methods for providing high copies of DNA sequencing involves variations of cloning or *in vitro* amplification which cannot amplify the number of individual clones (600,000 or more, and tens of millions for a human genome) necessary for sequencing a whole genome economically.

Yet another technical problem in the limitation of current sequencing methods which can perform, at most, one sequencing reaction per hybridization of oligonucleotide primer. The hybridization of sequencing primers is often the rate limiting step constricting the throughput of DNA sequencers.

In most cases, Polymerase Chain Reaction (PCR; Saiki, R.K., *et al.*, *Science* 1985, 230, 1350-1354; Mullis, K., *et al.*, *Cold Spring Harb. Symp. Quant. Biol.* 1986, 51 Pt 1, 263-273) plays an integral part in obtaining DNA sequence information, amplifying minute

IDT Exhibit 1020.6526

amounts of specific DNA to obtain concentrations sufficient for sequencing. Yet, scaling current PCR technology to meet the increasing demands of modern genetics is neither cost effective nor efficient, especially when the requirements for full genome sequencing are considered.

Efforts to maximize time and cost efficiencies have typically focused on two areas: decreasing the reaction volume required for amplifications and increasing the number of simultaneous amplifications performed. Miniaturization confers the advantage of lowered sample and reagent utilization, decreased amplification times and increased throughput scalability.

While conventional thermocyclers require relatively long cycling times due to thermal mass restrictions (Woolley, A.T., *et al.*, Anal. Chem. 1996, 68, 4081-4086), smaller reaction volumes can be cycled more rapidly. Continuous flow PCR devices have utilized etched microchannels in conjunction with fixed temperature zones to reduce reaction volumes to sub-microliter sample levels (Lagally, E.T., *et al.*, Analytical Chemistry 2001, 73, 565-570; Schneegas, I., *et al.*, *Lab on a Chip - The Royal Society of Chemistry* 2001, 1, 42-49).

Glass microcapillaries, heated by air (Kalinina, O., *et al.*, *Nucleic Acids Res.* 1997, 25, 1999-2004) or infrared light (Oda, R.P., *et al.*, *Anal. Chem.* 1998, 70, 4361-4368; Huhmer, A.F. and Landers, J.P., *Anal. Chem.* 2000, 72, 5507-5512), have also served as efficient vessels for nanoliter scale reactions. Similar reaction volumes have been attained with microfabricated silicon thermocyclers (Burns, M.A., *et al.*, *Proc. Natl. Acad. Sci. USA* 1996, 93, 5556-5561).

In many cases, these miniaturizations have reduced total PCR reaction times to less than 30 minutes for modified electric heating elements (Kopp, M.U., *et al.*, *Science* 1998, 280, 1046-1048; Chiou, J., Matsudaira, P., Sonin, A. and Ehrlich, D., *Anal. Chem.* 2001, 73, 2018-2021) and hot air cyclers (Kalinina, O., *et al.*, *Nucleic Acids Res.* 1997, 25, 1999-2004), and to 240 seconds for some infrared controlled reactions (Giordano, B.C., *et al.*, *Anal. Biochem.* 2001, 291, 124-132).

Certain technologies employ increased throughput and miniaturization simultaneously; as in the 1536-well system design by Sasaki *et al.* (Sasaki, N., *et al.*, *DNA Res.* 1997, 4, 387-391), which maintained reaction volumes under 1 µl. As another example, Nagai *et al.* (Nagai, H., *et al.*, *Biosens. Bioelectron.* 2001, 16, 1015-1019; Nagai, H., *et al.*, *Anal. Chem.* 2001, 73, 1043-1047) reported amplification of a single test fragment in ten thousand 86 pl reaction pits etched in a single silicon wafer. Unfortunately, recovery and

IDT Exhibit 1020.6527

utilization of the amplicon from these methods have proven problematic, requiring evaporation through selectively permeable membranes.

Despite these remarkable improvements in reactions volumes and cycle times, none of the previous strategies have provided the massively parallel amplification required to dramatically increase throughput to levels required for analysis of the entire human genome. DNA sequencers continue to be slower and more expensive than would be desired. In the pure research setting it is perhaps acceptable if a sequencer is slow and expensive. But when it is desired to use DNA sequencers in a clinical diagnostic setting such inefficient sequencing methods are prohibitive even for a well financed institution.The large-scale parallel sequencing of thousands of clonally amplified targets would greatly facilitate large-scale, whole genome library analysis without the time consuming sample preparation process and expensive, error-prone cloning processes. Successful high capacity, solid-phase, clonal DNA amplification can be used for numerous applications. Accordingly, it is clear that there exists a need for preparation of a genome or large template nucleic acids for sequencing, for amplification of the nucleic acid template, and for the sequencing of the amplified template nucleic acids without the constraint of one sequencing reaction per hybridization. Furthermore, there is a need for a system to connect these various technologies into a viable automatic or semiautomatic sequencing machine.

BRIEF SUMMARY OF THE INVENTION

This invention describes an integrated system, comprising novel methods and novel apparatus for (1) nucleic acid sample preparation, (2) nucleic acid amplification, and (3) DNA sequencing.

The invention provides a novel method for preparing a library of multiple DNA sequences, particularly derived from large template DNA or whole (or partial) genome DNA. Sequences of single stranded DNA are prepared from a sample of large template DNA or whole or partial DNA genomes through fragmentation, polishing, adaptor ligation, nick repair, and isolation of single stranded DNA. The method provides for generating a ssDNA library linked to solid supports comprising: (a) generating a library of ssDNA templates; (b) attaching the ssDNA templates to solid supports; and (c) isolating the solid supports on which one ssDNA template is attached.

The invention also provides for a method of amplifying each individual member of a DNA library in a single reaction tube, by, e.g., encapsulating a plurality of DNA samples

IDT Exhibit 1020.6528

individually in a microcapsule of an emulsion, performing amplification of the plurality of encapsulated nucleic acid samples simultaneously, and releasing said amplified plurality of DNA from the microcapsules for subsequent reactions. In one embodiment, single copies of the nucleic acid template species are hybridized to DNA capture beads, suspended in complete amplification solution and emulsified into microreactors (typically 100 to 200 microns in diameter), after which amplification (e.g., PCR) is used to clonally increase copy number of the initial template species to more than 1,000,000 copies of a single nucleic acid sequence, preferably between 2 and 20 million copies of a single nucleic acid. The amplification reaction, for example, may be performed simultaneously with at least 3,000 microreactors per microliter of reaction mix, and may be performed with over 300,000 microreactors in a single 100µl volume test tube (e.g., a PCR reaction tube). The present invention also provides for a method of enriching for those beads that contains a successful DNA amplification event (i.e., by removing beads that have no DNA attached thereto).

The invention also provides for a method of sequencing a nucleic acid from multiple primers with a single primer hybridization step. Two or more sequencing primers are hybridized to the template DNA to be sequenced. All the sequencing primers are then protected except for one. Sequencing (e.g., pyrophosphate sequencing) is performed again by elongating the unprotected primer. The elongation is either allowed to go to completion (with additional polymerase and dNTPs if necessary) or is terminated (by polymerase and ddNTPs). Chain completion and/or termination reagents are removed. Then one of the protected primers is unprotected and sequencing is performed by elongating the newly unprotected primer. This process is continued until all the sequencing primers are deprotected and sequenced. In a preferred embodiment, two primers (one protected and one unprotected) are used to sequence both ends of a double stranded nucleic acid.

The invention also provides an apparatus and methods for sequencing nucleic acids using a pyrophosphate based sequencing approach. The apparatus has a charge coupled device (CCD) camera, microfluidics chamber, sample cartridge holder, pump and flow valves. The apparatus uses chemiluminescence as the detection method, which for pyrophosphate sequencing has an inherently low background. In a preferred embodiment, the sample cartridge for sequencing is termed the 'PicoTiter plate,' and is formed from a commercial fiber optics faceplate, acid-etched to yield hundreds of thousands of very small wells, each well volume of 75 pL. The apparatus includes a novel reagent delivery cuvette adapted for use with the arrays described herein, to provide fluid reagents to the picotiter

IDT Exhibit 1020.6529

plate, and a reagent delivery means in communication with the reagent delivery cuvette. Photons from each well on the picotiter plate are channeled into specific pixels on the CCD camera to detect sequencing reactions.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 depicts a schematic representation of the entire process of library preparation including the steps of template DNA fragmentation (Figure 1A), end polishing (Figure 1B), adaptor ligation (Figure 1C), nick repair, strand extension and gel isolation (Figure 1D). Figure 1E depicts a schematic representation of the stages for amplification and sequencing of template DNA (Figure 1E). Figure 1F depicts a representative agarose gel containing a sample preparation of a 180-350 base pair adenovirus DNA library according to the methods of this invention. Figure 1G depicts a detailed schematic representation of library preparation, amplification, and sequencing.

Figure 2A depicts a schematic representation of the universal adaptor design according to the present invention. Each universal adaptor is generated from two complementary ssDNA oligonucleotides that are designed to contain a 20 bp nucleotide sequence for PCR priming, a 20 bp nucleotide sequence for sequence priming and a unique 4 bp discriminating sequence comprised of a non-repeating nucleotide sequence (*i.e.*, ACGT, CAGT, etc.). Figure 2B depicts a representative universal adaptor sequence pair for use with the invention. Adaptor A sense strand: SEQ ID NO:1; Adaptor A antisense strand: SEQ ID NO:2; Adaptor B sense strand: SEQ ID NO:3; Adaptor B antisense strand: SEQ ID NO:4. Figure 2C depicts a schematic representation of universal adaptor design for use with the invention.

Figure 3 depicts the strand displacement and extension of nicked double-stranded DNA fragments according to the present invention. Following the ligation of universal adaptors generated from synthetic oligonucleotides, double-stranded DNA fragments will be generated that contain two nicked regions following T4 DNA ligase treatment (Figure 3A). The addition of a strand displacing enzyme (*i.e.*, Bst DNA polymerase I) will bind nicks (Figure 3B), strand displace the nicked strand and complete nucleotide extension of the strand (Figure 3C) to produce non-nicked double-stranded DNA fragments (Figure

IDT Exhibit 1020.6530

3D).

Figure 4 depicts the isolation of directionally-ligated single-stranded DNA according to the present invention using streptavidin-coated beads. Following ligation with universal adaptors A and B (the two different adaptors are sometimes referred to as a "first" and "second" universal adaptor), double-stranded DNA will contain adaptors in four possible combinations: AA, BB, AB and BA. When universal adaptor B contains a 5'-biotin, magnetic streptavidin-coated solid supports are used to capture and isolate the AB, BA and BB populations (population AA is washed away). The BB population is retained on the beads as each end of the double-stranded DNA is attached to a bead and is not released. However, upon washing in the presence of a low salt buffer, only populations AB and BA will release a single-stranded DNA fragment that is complementary to the bound strand. Single-stranded DNA fragments are isolated from the supernatant and used as template for subsequent amplification and sequencing. This method is described below in more detail.

Figure 5 depicts a schematic of the structure of a DNA capture bead.

Figure 6 depicts a schematic of one embodiment of a bead emulsion amplification process.

Figure 7 depicts a schematic of an enrichment process to remove beads that do not have any DNA attached thereto.

Figures 8A-B depict a schematic representation of the double ended sequencing reaction according to the present invention.

Figure 9 depicts a double-ended sequencing demonstration on a pyrosequencing apparatus according to the invention.

Figures 10A-F depict an exemplary double ended sequencing process.

Figures 11A-D depict schematic illustrations of rolling circle-based amplification using an anchor primer.

Figure 12 depicts a drawing of a sequencing apparatus according to the present invention.

Figure 13 depicts a drawing of a reagent delivery / perfusion chamber according to the present invention.

Figure 14 depicts a micrograph of a cavitated fiber optic bundle, termed a PicoTiter Plate™, of the invention.

IDT Exhibit 1020.6531

Figure 15 depicts a micrograph of a picotiter plate carpeted with beads having DNA template immobilized thereon and sulfurylase and luciferase immobilized thereon.

Figure 16 depicts a schematic illustration of the reagent flow chamber and FORA (PicoTiter Plate™).

Figure 17 depicts a diagram of the analytical instrument of the present invention.

Figure 18 depicts a schematic illustration of microscopic parallel sequencing reactions within a PicoTiter Plate™.

Figure 19 depicts a micrograph of single well reactions.

Figure 20 depicts a PicoTiterPlate™ loading cartridge. "A" refers to a PicoTiterPlate™ with microwells facing into the cartridge, the distance between the open sides of the PicoTiterPlate™ wells and the wall of the loading cartridge is 0.3 mm; "B" refers to a silicon sealing gasket; "C" refers to an inlet port; "D" refers to an inlet loading tube; "E" refers to an outlet port and "F" refers to an outlet tube. The PicoTiterPlate™ is held in the cartridge with plastic clamps. The liquid is filled via the inlet loading tube D and enters the space between the open sides of the PicoTiterPlate™ wells and the wall of the laoding cartridge through the inlet port C. The area defined by the silicon sealing gasket B is filled and excess liquid leaves the cartridge via the outlet port E and the outlet tubing F.

Figure 21 depicts a PicoTiterPlate™ amplification chamber in exploded view. "A" refers to an amplification chamber lid with six retaining bolts; "B" refers to a closed cell foam insulation pad; "C" refers to a 25 mm by 75 mm standard glass microscope slide; "D" refers to a 0.25 mm thick silicon sheet; "E" refers to a PicoTiterPlate™; "F" refers to an amplification chamber base; "G" refers to a second 0.25 mm thick silicon sheet.

Figure 22 depicts a schematic diagram of solid phase PicoTiterPlate™ PCR. The cylindrical structures symbolize individual PicoTiterPlate™ wells. Gray spheres symbolize beads with immobilized primers. Forward "F" (red) and Reverse "R" (blue) primers are shown in 5' to 3' orientation as indicated by arrows. Synthesized sequence complementary to the Forward and Reverse primers are shown as dark red (F complement) and dark blue (R complement) bars. Single stranded template DNA is shown as solid gray line, newly

IDT Exhibit 1020.6532

synthesized DNA strands as dashed gray lines. Fluorescently labeled hybridization probes are shown as green bars.

Figures 23A-C depict fluorescent probes hybridization to bead-immobilized test DNA fragments. Figure 23A (upper left) and 23B (upper right) illustrate the specificity of a mixed population of probes hybridized to fragment A and fragment B immobilized on control beads, respectively. Fragment B beads exhibited the Alexa Fluor 647 signal (red), and the fragment A beads exhibited the Alexa Fluor 488 signal (green). Figure 23C (bottom panel) depicts probe fluorescence from DNA capture beads after PTPCR. Beads displayed homogenous fragment A and fragment B signals, as well as mixes of templates, shown as varying degrees of yellow.

Figure 24 depicts representative BioAnalyzer output from analysis of a single stranded DNA library.

Figure 25 depicts an insert flanked by PCR primers and sequencing primers.

Figure 26 depicts truncated product produced by PCR primer mismatch at cross-hybridization region (CHR).

Figure 27 depicts the calculation for primer candidates based on melting temperature.

Figures 28A-D depict the assembly for the nebulizer used for the methods of the invention. A tube cap was placed over the top of the nebulizer (Figure 7A) and the cap was secured with a nebulizer clamp assembly (Figure 7B). The bottom of the nebulizer was attached to the nitrogen supply (Figure 7C) and the entire device was wrapped in parafilm (Figure 7D).

Figure 29A depicts representative results for LabChip analysis of a single stranded DNA library following nebulization and polishing.

Figure 29B depicts representative size distribution results for an adaptor-ligated single stranded DNA library following nebulization, polishing, and gel purification.

Figure 30 depiction of jig used to hold tubes on the stir plate below vertical syringe pump. The jig was modified to hold three sets of bead emulsion amplification reaction mixtures. The syringe was loaded with the PCR reaction mixture and beads.

Figure 31 depiction of optimal placement of syringes in vertical syringe pump and orientation of emulsion tubes below syringe outlets.

Figure 32 depiction of optimal placement of syringe pump pusher block against syringe

IDT Exhibit 1020.6533

plungers, and optimal orientation of jig on the stir plate. Using this arrangement, the syringe contents were expelled into the agitated emulsion oil.

Figure 33 depiction of beads (see arrows) suspended in individual microreactors according to the methods of the invention.

Figure 34 depiction of double ended sequencing results showing that the sequence of both ends of a DNA template are determined. SEQ ID NO:44: atgcacatggttgacacagtggt; SEQ ID NO:45: atgcacatggttgacacagtgg; SEQ ID NO:46: atgccaccgacctagtctcaaactt.

Figure 35 illustrates the encapsulation of a bead comprising two oligonucleotide sequences for double stranded sequencing.

Figure 36 illustrates solution phase PCR and drive to bead procedure – a step in a preferred embodiment of double ended sequencing.

Figure 37 illustrates emulsion breaking and recovery of amplified template DNA on a bead - a step in a preferred embodiment of double ended sequencing.

Figure 38 depicts a schematic representation of a preferred method of double stranded sequencing.

Figure 39 illustrates the results of sequencing a *Staphylococcus aureus* genome.

Figure 40 illustrates the average read lengths in one experiment involving double ended sequencing.

Figure 41 illustrates the number of wells for each genome span in a double ended sequencing experiment.

Figure 42 illustrates a typical output and alignment string from a double ended sequencing procedure. Sequences shown in order, from top to bottom: SEQ ID NO:47-SEQ ID NO:60.

DETAILED DESCRIPTION OF THE INVENTION

A novel platform is described herein which permits simultaneous amplification of three hundred thousand discrete PCR reactions (PTPCR) in volumes as low as 39.5 picoliters. The pooled PTPCR products from the entire reaction can be recovered through a wash step and assayed via real-time PCR for the presence and abundance of specific templates. Of greater interest, it is shown herein that these PTPCR products can be driven to solid supports and detected by hybridization with two color fluorescent probes, allowing high capacity, solid-phase, clonal DNA amplification and large-scale parallel sequencing.

The present invention is directed to a method and apparatus for performing genomic

IDT Exhibit 1020.6534

sequencing which satisfies the objectives of (1) preparing a nucleic acid (e.g., a genome) in a rapid and efficient manner for sequencing, (2) amplifying the nucleic acid in a representative manner, and (3) performing multiple sequencing reactions with only one primer hybridization. The present invention is particularly suited for genotyping, detection and diagnosis from a small sample of nucleic acid in a cost efficient manner. Each of these objectives are listed below.

Definitions:

Unless otherwise defined, 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. Methods and materials similar or equivalent to those described herein can be used in the practice of the present invention, and exemplified suitable methods and materials are described below. For example, methods may be described which comprise more than two steps. In such methods, not all steps may be required to achieve a defined goal and the invention envisions the use of isolated steps to achieve these discrete goals. The disclosures of all publications, patent applications, patents and other references are incorporated *in toto* herein by reference. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

As used herein, the term "universal adaptor" refers to two complementary and annealed oligonucleotides that are designed to contain a nucleotide sequence for PCR priming and a nucleotide sequence for sequence priming. Optionally, the universal adaptor may further include a unique discriminating key sequence comprised of a non-repeating nucleotide sequence (*i.e.*, ACGT, CAGT, etc.). A set of universal adaptors comprises two unique and distinct double-stranded sequences that can be ligated to the ends of double-stranded DNA. Therefore, the same universal adaptor or different universal adaptors can be ligated to either end of the DNA molecule. When comprised in a larger DNA molecule that is single stranded or when present as an oligonucleotide, the universal adaptor may be referred to as a single stranded universal adaptor.

"Target DNA" shall mean a DNA whose sequence is to be determined by the methods and apparatus of the invention.

Binding pair shall mean a pair of molecules that interact by means of specific non-covalent interactions that depend on the three-dimensional structures of the molecules involved. Typical pairs of specific binding partners include antigen-antibody, hapten-antibody, hormone-receptor, nucleic acid strand-complementary nucleic acid strand,

IDT Exhibit 1020.6535

substrate-enzyme, substrate analog-enzyme, inhibitor-enzyme, carbohydrate-lectin, biotin-avidin, and virus-cellular receptor.

As used herein, the term "discriminating key sequence" refers to a sequence consisting of at least one of each of the four deoxyribonucleotides (*i.e.*, A, C, G, T). The same discriminating sequence can be used for an entire library of DNA fragments. Alternatively, different discriminating key sequences can be used to track libraries of DNA fragments derived from different organisms.

As used herein, the term "plurality of molecules" refers to DNA isolated from the same source, whereby different organisms may be prepared separately by the same method. In one embodiment, the plurality of DNA samples is derived from large segments of DNA, whole genome DNA, cDNA, viral DNA or from reverse transcripts of viral RNA. This DNA may be derived from any source, including mammal (i.e., human, nonhuman primate, rodent or canine), plant, bird, reptile, fish, fungus, bacteria or virus.

As used herein, the term "library" refers to a subset of smaller sized DNA species generated from a single DNA template, either segmented or whole genome.

As used herein, the term "unique", as in "unique PCR priming regions" refers to a sequence that does not exist or exists at an extremely low copy level within the DNA molecules to be amplified or sequenced.

As used herein, the term "compatible" refers to an end of double stranded DNA to which an adaptor molecule may be attached (i.e., blunt end or cohesive end).

As used herein, the term "fragmenting" refers to a process by which a larger molecule of DNA is converted into smaller pieces of DNA.

As used herein, "large template DNA" would be DNA of more than 25kb, preferably more than 500kb, more preferably more than 1MB, and most preferably 5MB or larger.

As used herein, the term "stringent hybridization conditions" refers to those conditions under which only complimentary sequences will hybridize to each other.

The invention described here is generally a system and methods for processing nucleic acids. The system and methods can be used to process nucleic acids in a multitude of ways that utilize sequencing of nucleic acids. Such sequencing can be performed to determine the identity of a sequence of nucleic acids, or for single nucleotide polymorphism detection in nucleic acid fragments, for nucleic acid expression profiling (comparing the nucleic acid expression profile between two or more states – e.g., comparing between

IDT Exhibit 1020.6536

diseased and normal tissue or comparing between untreated tissue and tissue treated with drug, enzymes, radiation or chemical treatment), for haplotying (comparing genes or variations in genes on each of the two alleles present in a human subject), for karyotyping (diagnostically comparing one or more genes in a test tissue – typically from an embryo/fetus prior to conception to detect birth defects -- with the same genes from "normal" karyotyped subjects), and for genotyping (comparing one or more genes in a first individual of a species with the same genes in other individuals of the same species).

The system has a number of components. These include (1) the nucleic acid template that is to be processed, (2) a picotiter plate for containing the nucleic acid template, (3) a flow chamber and fluid delivery means that permits flow of nucleic acid processing reagents over the nucleic acid template where the processing reagents generate light as the nucleic acid is processed, (4) a light capture means that detects light emitted as the nucleic acid is processed and that converts the captured light into data, and (5) data processing means that processes the data to yield meaningful information about the nucleic acid that has been processed. Each of these components of the system will be discussed in detail below.

1. NUCLEIC ACID TEMPLATE AND PREPARATION THEREOF

Nucleic Acid Templates

The nucleic acid templates that can be sequenced according to the invention, *e.g.*, a nucleic acid library, in general can include open circular or closed circular nucleic acid molecules. A "closed circle" is a covalently closed circular nucleic acid molecule, *e.g.*, a circular DNA or RNA molecule. An "open circle" is a linear single-stranded nucleic acid molecule having a 5' phosphate group and a 3' hydroxyl group.

In one embodiment, the single stranded nucleic acid contains at least 100 copies of a specific nucleic acid sequence, each copy covalently linked end to end. In some embodiments, the open circle is formed *in situ* from a linear double-stranded nucleic acid molecule. The ends of a given open circle nucleic acid molecule can be ligated by DNA ligase. Sequences at the 5' and 3' ends of the open circle molecule are complementary to two regions of adjacent nucleotides in a second nucleic acid molecule, *e.g.*, an adapter region of an anchor primer (sometimes referred to as an adapter), or to two regions that are nearly adjoining in a second DNA molecule. Thus, the ends of the open-circle molecule can be

IDT Exhibit 1020.6537

ligated using DNA ligase, or extended by DNA polymerase in a gap-filling reaction. Open circles are described in detail in Lizardi, U.S. Pat. No. 5,854,033, fully incorporated herein by reference. An open circle can be converted to a closed circle in the presence of a DNA ligase (for DNA) or RNA ligase following, *e.g.*, annealing of the open circle to an anchor primer.

If desired, nucleic acid templates can be provided as padlock probes. Padlock probes are linear oligonucleotides that include target-complementary sequences located at each end, and which are separated by a linker sequence. The linkers can be ligated to ends of members of a library of nucleic acid sequences that have been, *e.g.*, physically sheared or digested with restriction endonucleases. Upon hybridization to a target-sequence, the 5'- and 3'-terminal regions of these linear oligonucleotides are brought in juxtaposition. This juxtaposition allows the two probe segments (if properly hybridized) to be covalently-bound by enzymatic ligation (*e.g.*, with T4 DNA ligase), thus converting the probes to circularly-closed molecules which are catenated to the specific target sequences (see *e.g.*, Nilsson, *et al.*, 1994. *Science* 265: 2085-2088). The resulting probes are suitable for the simultaneous analysis of many gene sequences both due to their specificity and selectivity for gene sequence variants (see *e.g.*, Lizardi, *et al.*, 1998. *Nat. Genet.* 19: 225-232; Nilsson, *et al.*, 1997. *Nat. Genet.* 16: 252-255) and due to the fact that the resulting reaction products remain localized to the specific target sequences. Moreover, intramolecular ligation of many different probes is expected to be less susceptible to non-specific cross-reactivity than multiplex PCR-based methodologies where non-cognate pairs of primers can give rise to irrelevant amplification products (see *e.g.*, Landegren and Nilsson, 1997. *Ann. Med.* 29: 585-590).

A starting nucleic acid template library can be constructed comprising either single-stranded or double-stranded nucleic acid molecules, provided that the nucleic acid sequence includes a region that, if present in the library, is available for annealing, or can be made available for annealing, to an anchor primer sequence. For example, when used as a template for rolling circle amplification, a region of a double-stranded template needs to be at least transiently single-stranded in order to act as a template for extension of the anchor primer.

Library templates can include multiple elements, including, but not limited to, one or more regions that are complementary to the anchor primer. For example, the template libraries may include a region complementary to a sequencing primer, a control nucleotide region, and an insert sequence comprised of the sequencing template to be subsequently characterized. As is explained in more detail below, the control nucleotide region is used to calibrate the relationship between the amount of byproduct and the number of nucleotides

IDT Exhibit 1020.6538

incorporated. As utilized herein the term "complement" refers to nucleotide sequences that are able to hybridize to a specific nucleotide sequence to form a matched duplex.

In one embodiment, a library template includes: (*i*) two distinct regions that are complementary to the anchor primer, (*ii*) one region homologous to the sequencing primer, (*iii*) one optional control nucleotide region, (*iv*) an insert sequence of, *e.g.*, 30-500, 50-200, or 60-100 nucleotides, that is to be sequenced. The template can, of course, include two, three, or all four of these features.

The template nucleic acid can be constructed from any source of nucleic acid, *e.g.*, any cell, tissue, or organism, and can be generated by any art-recognized method. Suitable methods include, *e.g.*, sonication of genomic DNA and digestion with one or more restriction endonucleases (RE) to generate fragments of a desired range of lengths from an initial population of nucleic acid molecules. Preferably, one or more of the restriction enzymes have distinct four-base recognition sequences. Examples of such enzymes include, *e.g.*, Sau3A1, MspI, and TaqI. Preferably, the enzymes are used in conjunction with anchor primers having regions containing recognition sequences for the corresponding restriction enzymes. In some embodiments, one or both of the adapter regions of the anchor primers contain additional sequences adjoining known restriction enzyme recognition sequences, thereby allowing for capture or annealing to the anchor primer of specific restriction fragments of interest to the anchor primer. In other embodiments, the restriction enzyme is used with a type IIS restriction enzyme.

Alternatively, template libraries can be made by generating a complementary DNA (cDNA) library from RNA, *e.g.*, messenger RNA (mRNA). The cDNA library can, if desired, be further processed with restriction endonucleases to obtain a 3' end characteristic of a specific RNA, internal fragments, or fragments including the 3' end of the isolated RNA. Adapter regions in the anchor primer may be complementary to a sequence of interest that is thought to occur in the template library, *e.g.*, a known or suspected sequence polymorphism within a fragment generated by endonuclease digestion.

In one embodiment, an indexing oligonucleotide can be attached to members of a template library to allow for subsequent correlation of a template nucleic acid with a population of nucleic acids from which the template nucleic acid is derived. For example, one or more samples of a starting DNA population can be fragmented separately using any of the previously disclosed methods (*e.g.*, restriction digestion, sonication). An indexing oligonucleotide sequence specific for each sample is attached to, *e.g.*, ligated to, the termini

IDT Exhibit 1020.6539

of members of the fragmented population. The indexing oligonucleotide can act as a region for circularization, amplification and, optionally, sequencing, which permits it to be used to index, or code, a nucleic acid so as to identify the starting sample from which it is derived.

Distinct template libraries made with a plurality of distinguishable indexing primers can be mixed together for subsequent reactions. Determining the sequence of the member of the library allows for the identification of a sequence corresponding to the indexing oligonucleotide. Based on this information, the origin of any given fragment can be inferred.

The invention includes a sample preparation process that results in a solid or a mobile solid substrate array containing a plurality of anchor primers or adapaters covalently linked to template nucleic acids.

When the template nucleic acid is circular, formation of the covalently linked anchor primer and one or more copies of the target nucleic acid preferably occurs by annealing the anchor primer to a complementary region of a circular nucleic acid, and then extending the annealed anchor primer with a polymerase to result in formation of a nucleic acid containing one or more copies of a sequence complementary to the circular nucleic acid.

Attachment of the anchor primer to a solid or mobile solid substrate can occur before, during, or subsequent to extension of the annealed anchor primer. Thus, in one embodiment, one or more anchor primers are linked to the solid or a mobile solid substrate, after which the anchor primer is annealed to a target nucleic acid and extended in the presence of a polymerase. Alternatively, in a second embodiment, an anchor primer is first annealed to a target nucleic acid, and a 3'OH terminus of the annealed anchor primer is extended with a polymerase. The extended anchor primer is then linked to the solid or mobile solid substrate. By varying the sequence of anchor primers, it is possible to specifically amplify distinct target nucleic acids present in a population of nucleic acids.

Outlined below is a preferred embodiment for the preparation of template nucleic acids for amplification and sequencing reactions. The invention includes a method for preparing the sample DNA comprised of seven general steps: (a) fragmenting large template DNA or whole genomic DNA samples to generate a plurality of digested DNA fragments; (b) creating compatible ends on the plurality of digested DNA samples; (c) ligating a set of universal adaptor sequences onto the ends of fragmented DNA molecules to make a plurality of adaptor-ligated DNA molecules, wherein each universal adaptor sequence has a known and unique base sequence comprising a common PCR primer sequence, a common sequencing primer sequence and a discriminating four base key sequence and wherein one

IDT Exhibit 1020.6540

adaptor is attached to biotin; (d) separating and isolating the plurality of ligated DNA fragments; (e) removing any portion of the plurality of ligated DNA fragments; (f) nick repair and strand extension of the plurality of ligated DNA fragments; (g) attaching each of the ligated DNA fragments to a solid support; and (h) isolating populations comprising single-stranded adaptor-ligated DNA fragments for which there is a unique adaptor at each end (i.e., providing directionality).

The following discussion summarizes the basic steps involved in the methods of the invention. The steps are recited in a specific order, however, as would be known by one of skill in the art, the order of the steps may be manipulated to achieve the same result. Such manipulations are contemplated by the inventors. Further, some steps may be minimized as would also be known by one of skill in the art.

Fragmentation

In the practice of the methods of the present invention, the fragmentation of the DNA sample can be done by any means known to those of ordinary skill in the art. Preferably, the fragmenting is performed by enzymatic or mechanical means. The mechanical means may be sonication or pnysical shearing. The enzymatic means may be performed by digestion with nucleases (e.g., Deoxyribonuclease I (DNase I)) or one or more restriction endonucleases. In a preferred embodiment, the fragmentation results in ends for which the sequence is not known.

In a preferred embodiment, the enzymatic means is DNase I. DNase I is a versatile enzyme that nonspecifically cleaves double-stranded DNA (dsDNA) to release 5'-phosphorylated di-, tri-, and oligonucleotide products. DNase I has optimal activity in buffers containing Mn^{2+}, Mg^{2+} and Ca^{2+}, but no other salts. The purpose of the DNase I digestion step is to fragment a large DNA genome into smaller species comprising a library. The cleavage characteristics of DNase I will result in random digestion of template DNA (*i.e.*, no sequence bias) and in the predominance of blunt-ended dsDNA fragments when used in the presence of manganese-based buffers (Melgar, E. and D.A. Goldthwait. 1968. Deoxyribonucleic acid nucleases. II. The effects of metal on the mechanism of action of deoxyribonuclease I. *J. Biol. Chem.* 243: 4409). The range of digestion products generated following DNase I treatment of genomic templates is dependent on three factors: i) amount of enzyme used (units); ii) temperature of digestion (°C); and iii) incubation time (minutes). The DNase I digestion conditions outlined below have been optimized to yield genomic

IDT Exhibit 1020.6541

libraries with a size range from 50-700 base pairs (bp).

In a preferred embodiment, the DNase I digests large template DNA or whole genome DNA for 1-2 minutes to generate a population of polynucleotides. In another preferred embodiment, the DNase I digestion is performed at a temperature between 10°C-37°C. In yet another preferred embodiment, the digested DNA fragments are between 50 bp to 700 bp in length.

Polishing

Digestion of genomic DNA (gDNA) templates with DNase I in the presence of Mn^{2+} will yield fragments of DNA that are either blunt-ended or have protruding termini with one or two nucleotides in length. In a preferred embodiment, an increased number of blunt ends are created with *Pfu* DNA polymerase. In other embodiments, blunt ends can be created with less efficient DNA polymerases such as T4 DNA polymerase or Klenow DNA polymerase. Pfu "polishing" or blunt ending is used to increase the amount of blunt-ended species generated following genomic template digestion with DNase I. Use of *Pfu* DNA polymerase for fragment polishing will result in the fill-in of 5' overhangs. Additionally, *Pfu* DNA polymerase does not exhibit DNA extendase activity but does have 3'→ 5' exonuclease activity that will result in the removal of single and double nucleotide extensions to further increase the amount of blunt-ended DNA fragments available for adaptor ligation (Costa, G.L. and M.P. Weiner. 1994a. Protocols for cloning and analysis of blunt-ended PCR-generated DNA fragments. *PCR Methods Appl* 3(5):S95; Costa, G.L., A. Grafsky and M.P. Weiner. 1994b. Cloning and analysis of PCR-generated DNA fragments. *PCR Methods Appl* 3(6):338; Costa, G.L. and M.P. Weiner. 1994c. Polishing with T4 or Pfu polymerase increases the efficiency of cloning of PCR products. *Nucleic Acids Res.* 22(12):2423).

Adaptor Ligation

If the libraries of nucleic acids are to be attached to the solid substrate, then preferably the nucleic acid templates are annealed to anchor primer sequences using recognized techniques (see, *e.g.*, Hatch, *et al.*, 1999. *Genet. Anal. Biomol. Engineer.* 15: 35-40; Kool, U.S. Patent No. 5,714, 320 and Lizardi, U.S. Patent No. 5,854,033). In general, any procedure for annealing the anchor primers to the template nucleic acid sequences is suitable as long as it results in formation of specific, *i.e.*, perfect or nearly perfect, complementarity between the adapter region or regions in the anchor primer sequence and a sequence present

IDT Exhibit 1020.6542

in the template library.

In a preferred embodiment, following fragmentation and blunt ending of the DNA library, universal adaptor sequences are added to each DNA fragment. The universal adaptors are designed to include a set of unique PCR priming regions that are typically 20 bp in length located adjacent to a set of unique sequencing priming regions that are typically 20 bp in length optionally followed by a unique discriminating key sequence consisting of at least one of each of the four deoxyribonucleotides (i.e., A, C, G, T). In a preferred embodiment, the discriminating key sequence is 4 bases in length. In another embodiment, the discriminating key sequence may be combinations of 1-4 bases. In yet another embodiment, each unique universal adaptor is forty-four bp (44 bp) in length. In a preferred embodiment the universal adaptors are ligated, using T4 DNA ligase, onto each end of the DNA fragment to generate a total nucleotide addition of 88 bp to each DNA fragment. Different universal adaptors are designed specifically for each DNA library preparation and will therefore provide a unique identifier for each organism. The size and sequence of the universal adaptors may be modified as would be apparent to one of skill in the art.

For example, to prepare two distinct universal adaptors (i.e., "first" and "second"), single-stranded oligonucleotides may be ordered from a commercial vendor (*i.e.*, Integrated DNA Technologies, IA or Operon Technologies, CA). In one embodiment, the universal adaptor oligonucleotide sequences are modified during synthesis with two or three phosphorothioate linkages in place of phosphodiester linkages at both the 5' and 3' ends. Unmodified oligonucleotides are subject to rapid degradation by nucleases and are therefore of limited utility. Nucleases are enzymes that catalyze the hydrolytic cleavage of a polynucleotide chain by cleaving the phosphodiester linkage between nucleotide bases. Thus, one simple and widely used nuclease-resistant chemistry available for use in oligonucleotide applications is the phosphorothioate modification. In phosphorothioates, a sulfur atom replaces a non-bridging oxygen in the oligonucleotide backbone making it resistant to all forms of nuclease digestion (i.e. resistant to both endonuclease and exonuclease digestion). Each oligonucleotide is HPLC-purified to ensure there are no contaminating or spurious oligonucleotide sequences in the synthetic oligonucleotide preparation. The universal adaptors are designed to allow directional ligation to the blunt-ended, fragmented DNA. Each set of double-stranded universal adaptors are designed with a PCR priming region that contains noncomplementary 5' four-base overhangs that cannot ligate to the blunt-ended DNA fragment as well as prevent ligation with each other at these

IDT Exhibit 1020.6543

ends. Accordingly, binding can only occur between the 3' end of the adaptor and the 5' end of the DNA fragment or between the 3' end of the DNA fragment and the 5' end of the adaptor. Double-stranded universal adaptor sequences are generated by using single-stranded oligonucleotides that are designed with sequences that allow primarily complimentary oligonucleotides to anneal, and to prevent cross-hybridization between two non-complimentary oligonucleotides. In one embodiment, 95% of the universal adaptors are formed from the annealing of complimentary oligonucleotides. In a preferred embodiment, 97% of the universal adaptors are formed from the annealing of complimentary oligonucleotides. In a more preferred embodiment, 99% of the universal adaptors are formed from the annealing of complimentary oligonucleotides. In a most preferred embodiment, 100% of the universal adaptors are formed from the annealing of complimentary oligonucleotides.

One of the two adaptors can be linked to a support binding moiety. In a preferred embodiment, a 5' biotin is added to the first universal adaptor to allow subsequent isolation of ssDNA template and noncovalent coupling of the universal adaptor to the surface of a solid support that is saturated with a biotin-binding protein (i.e. streptavidin, neutravidin or avidin). Other linkages are well known in the art and may be used in place of biotin-streptavidin (for example antibody/antigen-epitope, receptor/ligand and oligonucleotide pairing or complimentarity) one embodiment, the solid support is a bead, preferably a polystyrene bead. In one preferred embodiment, the bead has a diameter of about 2.8 μm. As used herein, this bead is referred to as a "sample prep bead".

Each universal adaptor may be prepared by combining and annealing two ssDNA oligonucleotides, one containing the sense sequence and the second containing the antisense (complementary) sequence. Schematic representation of the universal adaptor design is outlined in Figure 2.

Isolation of Ligation Products

The universal adaptor ligation results in the formation of fragmented DNAs with adaptors on each end, unbound single adaptors, and adaptor dimers. In a preferred embodiment, agarose gel electrophoresis is used as a method to separate and isolate the adapted DNA library population from the unligated single adaptors and adaptor dimer populations. In other embodiments, the fragments may be separated by size exclusion chromatography or sucrose sedimentation. The procedure of DNase I digestion of DNA

IDT Exhibit 1020.6544

typically yields a library population that ranges from 50-700 bp. In a preferred embodiment, upon conducting agarose gel electrophoresis in the presence of a DNA marker, the addition of the 88 bp universal adaptor set will shift the DNA library population to a larger size and will result in a migration profile in the size range of approximately 130-800 bp; adaptor dimers will migrate at 88 bp; and adaptors not ligated will migrate at 44 bp. Therefore, numerous double-stranded DNA libraries in sizes ranging from 200-800 bp can be physically isolated from the agarose gel and purified using standard gel extraction techniques. In one embodiment, gel isolation of the adapted ligated DNA library will result in the recovery of a library population ranging in size from 200-400 bp. Other methods of distinguishing adaptor-ligated fragments are known to one of skill in the art.

Nick Repair

Because the DNA oligonucleotides used for the universal adaptors are not 5' phosphorylated, gaps will be present at the 3' junctions of the fragmented DNAs following ligase treatment (see Figure 3A). These two "gaps" or "nicks" can be filled in by using a DNA polymerase enzyme that can bind to, strand displace and extend the nicked DNA fragments. DNA polymerases that lack 3'→ 5' exonuclease activity but exhibit 5' → 3' exonuclease activity have the ability to recognize nicks, displace the nicked strands, and extend the strand in a manner that results in the repair of the nicks and in the formation of non-nicked double-stranded DNA (see Figure 3B and 3C) (Hamilton, S.C., J.W. Farchaus and M.C. Davis. 2001. DNA polymerases as engines for biotechnology. *BioTechniques* 31:370).

Several modifying enzymes are utilized for the nick repair step, including but not limited to polymerase, ligase and kinase. DNA polymerases that can be used for this application include, for example, *E. coli* DNA pol I, *Thermoanaerobacter thermohydrosulfuricus* pol I, and bacteriophage phi 29. In a preferred embodiment, the strand displacing enzyme *Bacillus stearothermophilus* pol I (Bst DNA polymerase I) is used to repair the nicked dsDNA and results in non-nicked dsDNA (see Figure 3D). In another preferred embodiment, the ligase is T4 and the kinase is polynucleotide kinase.

Isolation of Single-Stranded DNA

Following the generation of non-nicked dsDNA, ssDNAs comprising both the first and second adaptor molecules are to be isolated (desired populations are designated below

IDT Exhibit 1020.6545

with asterisks; "A" and "B" correspond to the first and second adaptors). Double-stranded DNA libraries will have adaptors bound in the following configurations:

Universal Adaptor A – DNA fragment – Universal Adaptor A
Universal Adaptor B – DNA fragment – Universal Adaptor A*
Universal Adaptor A – DNA fragment – Universal Adaptor B*
Universal Adaptor B – DNA fragment – Universal Adaptor B

Universal adaptors are designed such that only one universal adaptor has a 5' biotin moiety. For example, if universal adaptor B has a 5'biotin moiety, streptavidin-coated sample prep beads can be used to bind all double-stranded DNA library species with universal adaptor B. Genomic library populations that contain two universal adaptor A species will not contain a 5' biotin moiety and will not bind to streptavidin-containing sample prep beads and thus can be washed away. The only species that will remain attached to beads are those with universal adaptors A and B and those with two universal adaptor B sequences. DNA species with two universal adaptor B sequences (i.e., biotin moieties at each 5'end) will be bound to streptavidin-coated sample prep beads at each end, as each strand comprised in the double strand will be bound. Double-stranded DNA species with a universal adaptor A and a universal adaptor B will contain a single 5'biotin moiety and thus will be bound to streptavidin-coated beads at only one end. The sample prep beads are magnetic, therefore, the sample prep beads will remain coupled to a solid support when magnetized. Accordingly, in the presence of a low-salt ("melt" or denaturing) solution, only those DNA fragments that contain a single universal adaptor A and a single universal adaptor B sequence will release the complementary unbound strand. This single-stranded DNA population may be collected and quantitated by, for example, pyrophosphate sequencing, real-time quantitative PCR, agarose gel electrophoresis or capillary gel electrophoresis.

Attachment of Template to Beads

In one embodiment, ssDNA libraries that are created according to the methods of the invention are quantitated to calculate the number of molecules per unit volume. These molecules are annealed to a solid support (bead) that contain oligonucleotide capture primers that are complementary to the PCR priming regions of the universal adaptor ends of the ssDNA species. Beads are then transferred to an amplification protocol. Clonal populations

IDT Exhibit 1020.6546

of single species captured on DNA beads may then sequenced. In one embodiment, the solid support is a bead, preferably a sepharose bead. As used herein, this bead is referred to as a "DNA capture bead".

The beads used herein may be of any convenient size and fabricated from any number of known materials. Example of such materials include: inorganics, natural polymers, and synthetic polymers. Specific examples of these materials include: cellulose, cellulose derivatives, acrylic resins, glass; silica gels, polystyrene, gelatin, polyvinyl pyrrolidone, co-polymers of vinyl and acrylamide, polystyrene cross-linked with divinylbenzene or the like (see, Merrifield Biochemistry 1964, 3, 1385-1390), polyacrylamides, latex gels, polystyrene, dextran, rubber, silicon, plastics, nitrocellulose, celluloses, natural sponges, silica gels, glass, metals plastic, cellulose, cross-linked dextrans (e.g., Sephadex^{TM}) and agarose gel (Sepharose^{TM}) and solid phase supports known to those of skill in the art. In one embodiment, the diameter of the DNA capture bead is in the range of 20-70μm. In a preferred embodiment, the diameter of the DNA capture bead is in a range of 20-50μm. In a more preferred embodiment, the diameter of the DNA capture bead is about 30μm.

In one aspect, the invention includes a method for generating a library of solid supports comprising: (a) preparing a population of ssDNA templates according to the methods disclosed herein; (b) attaching each DNA template to a solid support such that there is one molecule of DNA per solid support; (c) amplifying the population of single-stranded templates such that the amplification generates a clonal population of each DNA fragment on each solid support; (d) sequencing clonal populations of beads.

In one embodiment, the solid support is a DNA capture bead. In another embodiment, the DNA is genomic DNA, cDNA or reverse transcripts of viral RNA. The DNA may be attached to the solid support, for example, via a biotin-streptavidin linkage, a covalent linkage or by complementary oligonucleotide hybridization. In one embodiment, each DNA template is ligated to a set of universal adaptors. In another embodiment, the universal adaptor pair comprises a common PCR primer sequence, a common sequencing primer sequence and a discriminating key sequence. Single-stranded DNAs are isolated that afford unique ends; single stranded molecules are then attached to a solid support and exposed to amplification techniques for clonal expansion of populations. The DNA may be amplified by PCR.

In another aspect, the invention provides a library of solid supports made by the methods described herein.

IDT Exhibit 1020.6547

The nucleic acid template (e.g., DNA template) prepared by this method may be used for many molecular biological procedures, such as linear extension, rolling circle amplification, PCR and sequencing. This method can be accomplished in a linkage reaction, for example, by using a high molar ratio of bead to DNA. Capture of single-stranded DNA molecules will follow a poisson distribution and will result in a subset of beads with no DNA attached and a subset of beads with two molecules of DNA attached. In a preferred embodiment, there would be one bead to one molecule of DNA. In addition, it is possible to include additional components in the adaptors that may be useful for additional manipulations of the isolated library.

2. NUCLEIC ACID TEMPLATE AMPLIFICATION

In order for the nucleic acid template to be sequenced according to the methods of this invention the copy number must be amplified to generate a sufficient number of copies of the template to produce a detectable signal by the light detection means. Any suitable nucleic acid amplification means may be used.

A number of *in vitro* nucleic acid amplification techniques have been described. These amplification methodologies may be differentiated into those methods: (*i*) which require temperature cycling - polymerase chain reaction (PCR) (see *e.g.,* Saiki, *et al.,* 1995. *Science* 230: 1350-1354), ligase chain reaction (see *e.g.,* Barany, 1991. *Proc. Natl. Acad. Sci. USA* 88: 189-193; Barringer, *et al.,* 1990. *Gene* 89: 117-122) and transcription-based amplification (see *e.g.,* Kwoh, *et al.,* 1989. *Proc. Natl. Acad. Sci. USA* 86: 1173-1177) and (*ii*) isothermal amplification systems - self-sustaining, sequence replication (see *e.g.,* Guatelli, *et al.,* 1990. *Proc. Natl. Acad. Sci. USA* 87: 1874-1878); the Qβ replicase system (see *e.g.,* Lizardi, *et al.,* 1988. *BioTechnology* 6: 1197-1202); strand displacement amplification Nucleic Acids Res. 1992 Apr 11;20(7):1691-6.; and the methods described in PNAS 1992 Jan 1;89(1):392-6; and NASBA J Virol Methods. 1991 Dec;35(3):273-86.

In one embodiment, isothermal amplification is used. Isothermal amplification also includes rolling circle-based amplification (RCA). RCA is discussed in, *e.g.,* Kool, U.S. Patent No. 5,714,320 and Lizardi, U.S. Patent No. 5,854,033; Hatch, *et al.,* 1999. *Genet. Anal. Biomol. Engineer.* 15: 35-40. The result of the RCA is a single DNA strand extended from the 3' terminus of the anchor primer (and thus is linked to the solid support matrix) and including a concatamer containing multiple copies of the circular template annealed to a primer sequence. Typically, 1,000 to 10,000 or more copies of circular templates, each

IDT Exhibit 1020.6548

having a size of, *e.g.*, approximately 30-500, 50-200, or 60-100 nucleotides size range, can be obtained with RCA.

The product of RCA amplification following annealing of a circular nucleic acid molecule to an anchor primer is shown schematically in Figure 11A. A circular template nucleic acid 102 is annealed to an anchor primer 104, which has been linked to a surface 106 at its 5' end and has a free 3' OH available for extension. The circular template nucleic acid 102 includes two adapter regions 108 and 110 which are complementary to regions of sequence in the anchor primer 104. Also included in the circular template nucleic acid 102 is an insert 112 and a region 114 homologous to a sequencing primer, which is used in the sequencing reactions described below.

Upon annealing, the free 3'-OH on the anchor primer 104 can be extended using sequences within the template nucleic acid 102. The anchor primer 102 can be extended along the template multiple times, with each iteration adding to the sequence extended from the anchor primer a sequence complementary to the circular template nucleic acid. Four iterations, or four rounds of rolling circle replication, are shown in Figure 11A as the extended anchor primer amplification product 114. Extension of the anchor primer results in an amplification product covalently or otherwise physically attached to the substrate 106. A number of *in vitro* nucleic acid amplification techniques may be utilized to extend the anchor primer sequence. The amplification is typically performed in the presence of a polymerase, *e.g.*, a DNA or RNA-directed DNA polymerase, and one, two, three, or four types of nucleotide triphosphates, and, optionally, auxiliary binding proteins. In general, any polymerase capable of extending a primed 3'-OH group can be used a long as it lacks a 3' to 5' exonuclease activity. Suitable polymerases include, *e.g.*, the DNA polymerases from *Bacillus stearothermophilus*, *Thermus acquaticus*, *Pyrococcus furiosis*, *Thermococcus litoralis*, and *Thermus thermophilus*, bacteriophage T4 and T7, and the *E. coli* DNA polymerase I Klenow fragment. Suitable RNA-directed DNA polymerases include, *e.g.*, the reverse transcriptase from the Avian Myeloblastosis Virus, the reverse transcriptase from the Moloney Murine Leukemia Virus, and the reverse transcriptase from the Human Immunodeficiency Virus-I.

Additional embodiments of circular templates and anchor primers are shown in more detail in Figure 11B-11D. Figure 11B illustrates an annealed open circle linear substrate that can serve, upon ligation, as a template for extension of an anchor primer. A template molecule having the sequence 5' – tcg tgt gag gtc tca gca tct tat gta tat tta ctt cta ttc tca gtt

IDT Exhibit 1020.6549

gcc taa gct gca gcc a – 3' (SEQ ID NO:5) is annealed to an anchor primer having a biotin linker at its 5' terminus and the sequence 5'-gac ctc aca cga tgg ctg cag ctt – 3' (SEQ ID NO:6). Annealing of the template results in juxtaposition of the 5' and 3' ends of the template molecule. The 3'OH of the anchor primer can be extended using the circular template.

The use of a circular template and an anchor primer for identification of single nucleotide polymorphisms is shown in Figure 11C. Shown is a generic anchor primer having the sequence 5' - gac ctc aca cga tgg ctg cag ctt – 3'(SEQ ID NO:7). The anchor primer anneals to an SNP probe having the sequence 5' – ttt ata tgt att cta cga ctc tgg agt gtg cta ccg acg tcg aat ccg ttg act ctt atc ttc a – 3' (SEQ ID NO:8). The SNP probe in turn hybridizes to a region of a SNP-containing region of a gene having the sequence 5' – cta gct cgt aca tat aaa tga aga taa gat cct g – 3' (SEQ ID NO:9). Hybridization of a nucleic acid sequence containing the polymorphism to the SNP probe complex allows for subsequent ligation and circularization of the SNP probe. The SNP probe is designed so that its 5' and 3' termini anneal to the genomic region so as to abut in the region of the polymorphic site, as is indicated in Figure 11C. The circularized SNP probe can be subsequently extended and sequenced using the methods described herein. A nucleic acid lacking the polymorphism does not hybridize so as to result in juxtaposition of the 5' and 3' termini of the SNP probe. In this case, the SNP probe cannot be ligated to form a circular substrate needed for subsequent extension.

Figure 11D illustrates the use of a gap oligonucleotide to along with a circular template molecule. An anchor primer having the sequence 5'–gac ctc aca cga gta gca tgg ctg cag ctt – 3' (SEQ ID NO:10) is attached to a surface through a biotin linker. A template molecule having the sequence 5' – tcg tgt gag gtc tca gca tct tat gta tat tta ctt cta ttc tca gtt gcc taa gct gca gcc a – 3' (SEQ ID NO:11) is annealed to the anchor primer to result in partially single stranded, or gapped region, in the anchor primer flanked by a double-stranded region. A gapping molecule having the sequence 5' – tgc tac – 3' then anneals to the anchor primer. Ligation of both ends of the gap oligonucleotide to the template molecule results in formation of a circular nucleic acid molecule that can act as a template for rolling circle amplification.

RCA can occur when the replication of the duplex molecule begins at the origin. Subsequently, a nick opens one of the strands, and the free 3'-terminal hydroxyl moiety generated by the nick is extended by the action of DNA polymerase. The newly synthesized strand eventually displaces the original parental DNA strand. This aforementioned type of

IDT Exhibit 1020.6550

replication is known as rolling-circle replication (RCR) because the point of replication may be envisaged as "rolling around" the circular template strand and, theoretically, it could continue to do so indefinitely. Additionally, because the newly synthesized DNA strand is covalently-bound to the original template, the displaced strand possesses the original genomic sequence (*e.g.*, gene or other sequence of interest) at its 5'-terminus. In RCR, the original genomic sequence is followed by any number of "replication units" complementary to the original template sequence, wherein each replication unit is synthesized by continuing revolutions of said original template sequence. Hence, each subsequent revolution displaces the DNA which is synthesized in the previous replication cycle.

Through the use of the RCA reaction, a strand may be generated which represents many tandem copies of the complement to the circularized molecule. For example, RCA has recently been utilized to obtain an isothermal cascade amplification reaction of circularized padlock probes *in vitro* in order to detect single-copy genes in human genomic DNA samples (see Lizardi, *et al.*, 1998. *Nat. Genet.* 19: 225-232). In addition, RCA has also been utilized to detect single DNA molecules in a solid phase-based assay, although difficulties arose when this technique was applied to *in situ* hybridization (see Lizardi, *et al.*, 1998. *Nat. Genet.* 19: 225-232).

If desired, RCA can be performed at elevated temperatures, *e.g.*, at temperatures greater than 37° C, 42° C, 45° C, 50° C, 60° C, or 70° C. In addition, RCA can be performed initially at a lower temperature, *e.g.*, room temperature, and then shifted to an elevated temperature. Elevated temperature RCA is preferably performed with thermostable nucleic acid polymerases and with primers that can anneal stably and with specificity at elevated temperatures.

RCA can also be performed with non-naturally occurring oligonucleotides, *e.g.*, peptide nucleic acids. Further, RCA can be performed in the presence of auxiliary proteins such as single-stranded binding proteins.

The development of a method of amplifying short DNA molecules which have been immobilized to a solid support, termed RCA has been recently described in the literature (see *e.g.*, Hatch, *et al.*, 1999. *Genet. Anal. Biomol. Engineer.* 15: 35-40; Zhang, *et al.*, 1998. *Gene* 211: 277-85; Baner, *et al.*, 1998. *Nucl. Acids Res.* 26: 5073-5078; Liu, *et al.*, 1995. *J. Am. Chem. Soc.* 118: 1587-1594; Fire and Xu, 1995. *Proc. Natl. Acad. Sci. USA* 92: 4641-4645; Nilsson, *et al.*, 1994. *Science* 265: 2085-2088). RCA targets specific DNA sequences through hybridization and a DNA ligase reaction. The circular product is then subsequently

IDT Exhibit 1020.6551

used as a template in a rolling circle replication reaction.

Other examples of isothermal amplification systems include, *e.g.*, (*i*) self-sustaining, sequence replication (see *e.g.*, Guatelli, *et al.*, 1990. *Proc. Natl. Acad. Sci. USA* 87: 1874-1878), (*ii*) the Qβ replicase system (see *e.g.*, Lizardi, *et al.*, 1988. *BioTechnology* 6: 1197-1202), and (*iii*) nucleic acid sequence-based amplification (NASBA™; see Kievits, *et al.*, 1991. *J. Virol. Methods* 35: 273-286).

PCR Amplification of Nucleic Acid Templates

In a preferred embodiment, polymerase chain reaction ("PCR") is used to generate additional copies of the template nucleic acids. The PCR amplification step may be performed prior to distribution of the nucleic acid templates onto the picotiter plate or may be performed after the nucleic acid templates have been distributed onto the picotiter plate.

Bead Emulsion PCR Amplification

In a preferred embodiment, a PCR amplification step is performed prior to distribution of the nucleic acid templates onto the picotiter plate.

In a particularly preferred embodiment, a novel amplification system, herein termed "bead emulsion amplification" is performed by attaching a template nucleic acid (e.g., DNA) to be amplified to a solid support, preferably in the form of a generally spherical bead. A library of single stranded template DNA prepared according to the sample prepration methods of this invention is an example of one suitable source of the starting nucleic acid template library to be attached to a bead for use in this amplification method.

The bead is linked to a large number of a single primer species (*i.e.*, primer B in Figure 6) that is complementary to a region of the template DNA. Template DNA annealed to the bead bound primer. The beads are suspended in aqueous reaction mixture and then encapsulated in a water-in-oil emulsion. The emulsion is composed of discrete aqueous phase microdroplets, approximately 60 to 200 um in diameter, enclosed by a thermostable oil phase. Each microdroplet contains, preferably, amplification reaction solution (i.e., the reagents necessary for nucleic acid amplification). An example of an amplification would be a PCR reaction mix (polymerase, salts, dNTPs) and a pair of PCR primers (primer A and primer B). See, Figure 6A. A subset of the microdroplet population also contains the DNA bead comprising the DNA template. This subset of microdroplet is the basis for the amplification. The microcapsules that are not within this subset have no template DNA and will not participate in amplification. In one embodiment, the amplification technique is PCR

IDT Exhibit 1020.6552

and the PCR primers are present in a 8:1 or 16:1 ratio (i.e., 8 or 16 of one primer to 1 of the second primer) to perform asymmetric PCR.

In this overview, the DNA is annealed to an oligonucleotide (primer B) which is immobilized to a bead. During thermocycling (Figure 6B), the bond between the single stranded DNA template and the immobilized B primer on the bead is broken, releasing the template into the surrounding microencapsulated solution. The amplification solution, in this case, the PCR solution, contains addition solution phase primer A and primer B. Solution phase B primers readily bind to the complementary b' region of the template as binding kinetics are more rapid for solution phase primers than for immobilized primers. In early phase PCR, both A and B strands amplify equally well (Figure 6C).

By midphase PCR (*i.e.,* between cycles 10 and 30) the B primers are depleted, halting exponential amplification. The reaction then enters asymmetric amplification and the amplicon population becomes dominated by A strands (Figure 6D). In late phase PCR (Figure 6E), after 30 to 40 cycles, asymmetric amplification increases the concentration of A strands in solution. Excess A strands begin to anneal to bead immobilized B primers. Thermostable polymerases then utilize the A strand as a template to synthesize an immobilized, bead bound B strand of the amplicon.

In final phase PCR (Figure 6F), continued thermal cycling forces additional annealing to bead bound primers. Solution phase amplification may be minimal at this stage but concentration of immobilized B strands increase. Then, the emulsion is broken and the immobilized product is rendered single stranded by denaturing (by heat, pH etc.) which removes the complimentary A strand. The A primers are annealed to the A' region of immobilized strand, and immobilized strand is loaded with sequencing enzymes, and any necessary accessory proteins. The beads are then sequenced using recognized pyrophosphate techniques (described, e.g., in US patent 6,274,320, 6258,568 and 6,210,891, incorporated *in toto* herein by reference).

IDT Exhibit 1020.6553

Template design

In a preferred embodiment, the DNA template to be amplified by bead emulsion amplification can be a population of DNA such as, for example, a genomic DNA library or a cDNA library. It is preferred that each member of the population have a common nucleic acid sequence at the first end and a common nucleic acid sequence at a second end. This can be accomplished, for example, by ligating a first adaptor DNA sequence to one end and a second adaptor DNA sequence to a second end of the DNA population. Many DNA and cDNA libraries, by nature of the cloning vector (e.g., Bluescript, Stratagene, La Jolla, CA) fit this description of having a common sequence at a first end and a second common sequence at a second end of each member DNA. The DNA template may be of any size amenable to *in vitro* amplification (including the preferred amplification techniques of PCR and asymmetric PCR). In a preferred embodiment, the DNA template is between about 150 to 750 bp in size, such as, for example about 250 bp in size.

Binding Nucleic Acid Template to Capture Beads

In a first step, a single stranded nucleic acid template to be amplified is attached to a capture bead. The nucleic acid template may be attached to the solid support capture bead in any manner known in the art. Numerous methods exist in the art for attaching DNA to a solid support such as the preferred microscopic bead. According to the present invention, covalent chemical attachment of the DNA to the bead can be accomplished by using standard coupling agents, such as water-soluble carbodiimide, to link the 5'-phosphate on the DNA to amine-coated capture beads through a phosphoamidate bond. Another alternative is to first couple specific oligonucleotide linkers to the bead using similar chemistry, and to then use DNA ligase to link the DNA to the linker on the bead. Other linkage chemistries to join the oligonucleotide to the beads include the use of N-hydroxysuccinamide (NHS) and its derivatives. In such a method, one end of the oligonucleotide may contain a reactive group (such as an amide group) which forms a covalent bond with the solid support, while the other end of the linker contains a second reactive group that can bond with the oligonucleotide to be immobilized. In a preferred embodiment, the oligonucleotide is bound to the DNA capture bead by covalent linkage. However, non-covalent linkages, such as chelation or antigen-antibody complexes, may also be used to join the oligonucleotide to the bead.

Oligonucleotide linkers can be employed which specifically hybridize to unique

IDT Exhibit 1020.6554

sequences at the end of the DNA fragment, such as the overlapping end from a restriction enzyme site or the "sticky ends" of bacteriophage lambda based cloning vectors, but blunt-end ligations can also be used beneficially. These methods are described in detail in US 5,674,743. It is preferred that any method used to immobilize the beads will continue to bind the immobilized oligonucleotide throughout the steps in the methods of the invention.

In one embodiment, each capture bead is designed to have a plurality of nucleic acid primers that recognize (i.e., are complementary to) a portion of the nucleic template, and the nucleic acid template is thus hybridized to the capture bead. In the methods described herein, clonal amplification of the template species is desired, so it is preferred that only one unique nucleic acid template is attached to any one capture bead.

The beads used herein may be of any convenient size and fabricated from any number of known materials. Example of such materials include: inorganics, natural polymers, and synthetic polymers. Specific examples of these materials include: cellulose, cellulose derivatives, acrylic resins, glass, silica gels, polystyrene, gelatin, polyvinyl pyrrolidone, co-polymers of vinyl and acrylamide, polystyrene cross-linked with divinylbenzene or the like (as described, *e.g.,* in Merrifield, Biochemistry 1964, 3, 1385-1390), polyacrylamides, latex gels, polystyrene, dextran, rubber, silicon, plastics, nitrocellulose, natural sponges, silica gels, control pore glass, metals, cross-linked dextrans (e.g., Sephadex^{TM}) agarose gel (Sepharose^{TM}), and solid phase supports known to those of skill in the art. In a preferred embodiment, the capture beads are Sepharose beads approximately 25 to 40 μm in diameter.

Emulsification

Capture beads with attached single strand template nucleic acid are emulsified as a heat stable water-in-oil emulsion. The emulsion may be formed according to any suitable method known in the art. One method of creating emulsion is described below but any method for making an emulsion may be used. These methods are known in the art and include adjuvant methods, counterflow methods, crosscurrent methods, rotating drum methods, and membrane methods. Furthermore, the size of the microcapsules may be adjusted by varying the flow rate and speed of the components. For example, in dropwise addition, the size of the drops and the total time of delivery may be varied. Preferably, the emulsion contains a density of bead "microreactors" at a density of about 3,000 beads per microliter.

The emulsion is preferably generated by suspending the template-attached beads in

IDT Exhibit 1020.6555

amplification solution. As used herein, the term "amplification solution" means the sufficient mixture of reagents that is necessary to perform amplification of template DNA. One example of an amplification solution, a PCR amplification solution, is provided in the Examples below – it will be appreciated that various modifications may be made to the PCR solution.

In one embodiment, the bead / amplification solution mixture is added dropwise into a spinning mixture of biocompatible oil (e.g., light mineral oil, Sigma) and allowed to emulsify. The oil used may be supplemented with one or more biocompatible emulsion stabilizers. These emulsion stabilizers may include Atlox 4912, Span 80, and other recognized and commercially available suitable stabilizers. Preferably, the droplets formed range in size from 5 micron to 500 microns, more preferably, from between about 50 to 300 microns, and most preferably, from 100 to 150 microns.

There is no limitation in the size of the microreactors. The microreactors should be sufficiently large to encompass sufficient amplification reagents for the degree of amplification required. However, the microreactors should be sufficiently small so that a population of microreactors, each containing a member of a DNA library, can be amplified by conventional laboratory equipment (e.g., PCR thermocycling equipment, test tubes, incubators and the like).

With the limitations described above, the optimal size of a microreactor may be between 100 to 200 microns in diameter. Microreactors of this size would allow amplification of a DNA library comprising about 600,000 members in a suspension of microreactors of less than 10 ml in volume. For example, if PCR was the chosen amplification method, 10 mls would fit in 96 tubes of a regular thermocycler with 96 tube capacity. In a preferred embodiment, the suspension of 600,000 microreactors would have a volume of less than 1 ml. A suspension of less than 1 ml may be amplified in about 10 tubes of a conventional PCR thermocycler. In a most preferred embodiment, the suspension of 600,000 microreactors would have a volume of less than 0.5 ml.

Amplification

After encapsulation, the template nucleic acid may be amplified by any suitable method of DNA amplification including transcription-based amplification systems (Kwoh D. et al., Proc. Natl. Acad Sci. (U.S.A.) 86:1173 (1989); Gingeras T. R. et al., PCT appl. WO 88/10315; Davey, C. et al., European Patent Application Publication No. 329,822; Miller, H.

IDT Exhibit 1020.6556

I. et al., PCT appl. WO 89/06700, and "race" (Frohman, M. A., In: PCR Protocols: A Guide to Methods and Applications, Academic Press, NY (1990)) and "one-sided PCR" (Ohara, O. et al., Proc. Natl. Acad. Sci. (U.S.A.) 86.5673-5677 (1989)). Still other less common methods such as "di-oligonucleotide" amplification, isothermal amplification (Walker, G. T. et al., Proc. Natl. Acad. Sci. (U.S.A.) 89:392-396 (1992)), and rolling circle amplification (reviewed in 5,714,320), may be used in the present invention.

In a preferred embodiment, DNA amplification is performed by PCR. PCR according to the present invention may be performed by encapsulating the target nucleic acid, bound to a bead, with a PCR solution comprising all the necessary reagents for PCR. Then, PCR may be accomplished by exposing the emulsion to any suitable thermocycling regimen known in the art. In a preferred embodiment, between 30 and 50 cycles, preferably about 40 cycles, of amplification are performed. It is desirable, but not necessary, that following the amplification procedure there be one or more hybridization and extension cycles following the cycles of amplification. In a preferred embodiment, between 10 and 30 cycles, preferably about 25 cycles, of hybridization and extension are performed (e.g., as described in the examples). Routinely, the template DNA is amplified until typically at least two million to fifty million copies, preferably about ten million to thirty million copies of the template DNA are immobilized per bead.

IDT Exhibit 1020.6557

Breaking the Emulsion and Bead Recovery

Following amplification of the template, the emulsion is "broken" (also referred to as "demulsification" in the art). There are many methods of breaking an emulsion (see, e.g., U.S. Patent 5,989,892 and references cited therein) and one of skill in the art would be able to select the proper method. In the present invention, one preferred method of breaking the emulsion is to add additional oil to cause the emulsion to separate into two phases. The oil phase is then removed, and a suitable organic solvent (e.g., hexanes) is added. After mixing, the oil/organic solvent phase is removed. This step may be repeated several times. Finally, the aqueous layers above the beads are removed. The beads are then washed with an organic solvent / annealing buffer mixture (e.g., one suitable annealing buffer is described in the examples), and then washed again in annealing buffer. Suitable organic solvents include alcohols such as methanol, ethanol and the like.

The amplified template-containing beads may then be resuspended in aqueous solution for use, for example, in a sequencing reaction according to known technologies. (*See*, Sanger, F. et al., Proc. Natl. Acad. Sci. U.S.A. 75, 5463-5467 (1977); Maxam, A. M. & Gilbert, W. Proc Natl Acad Sci USA 74, 560-564 (1977); Ronaghi, M. et al., Science 281, 363, 365 (1998); Lysov, I. et al., Dokl Akad Nauk SSSR 303, 1508-1511 (1988); Bains W. & Smith G. C. J.TheorBiol 135, 303-307(1988); Drnanac, R. et al., Genomics 4, 114-128 (1989); Khrapko, K. R. et al., FEBS Lett 256. 118-122 (1989); Pevzner P. A. J Biomol Struct Dyn 7, 63-73 (1989); Southern, E. M. et al., Genomics 13, 1008-1017 (1992).) If the beads are to be used in a pyrophosphate-based sequencing reaction (described, e.g., in US patent 6,274,320, 6258,568 and 6,210,891, and incorporated *in toto* herein by reference), then it is necessary to remove the second strand of the PCR product and anneal a sequencing primer to the single stranded template that is bound to the bead.

Briefly, the second strand is melted away using any number of commonly known methods such as NaOH, low ionic (*e.g.*, salt) strength, or heat processing. Following this melting step, the beads are pelleted and the supernatant is discarded. The beads are resuspended in an annealing buffer, the sequencing primer added, and annealed to the bead-attached single stranded template using a standard annealing cycle.

Purifying the beads

At this point, the amplified DNA on the bead may be sequenced either directly on the

IDT Exhibit 1020.6558

bead or in a different reaction vessel. In an embodiment of the present invention, the DNA is sequenced directly on the bead by transferring the bead to a reaction vessel and subjecting the DNA to a sequencing reaction (e.g., pyrophosphate or Sanger sequencing). Alternatively, the beads may be isolated and the DNA may be removed from each bead and sequenced. In either case, the sequencing steps may be performed on each individual bead. However, this method, while commercially viable and technically feasible, may not be most effective because many of the beads will be negative beads (a bead that does not have amplified DNA attached). Accordingly, the following optional process may be used for removing beads that contain no nucleic acid template prior to distribution onto the picotiter plate.

A high percentage of the beads may be "negative" (i.e., have no amplified nucleic acid template attached thereto) if the goal of the initial DNA attachment is to minimize beads with two different copies of DNA. For useful pyrophosphate sequencing, each bead should contain multiple copies of a single species of DNA. This requirement is most closely approached by maximizing the total number of beads with a single fragment of DNA bound (before amplification). This goal can be achieved by the observation of a mathematical model.

For the general case of "N" number of DNAs randomly distributed among M number of beads, the relative bead population containing any number of DNAs depends on the ratio of N/M. The fraction of beads containing N DNAs R(N) may be calculated using the Poisson distribution:

$R(N) = \exp\text{-}(N/M) \times (N/M)^N/N!$ (where X is the multiplication symbol)

The table below shows some calculated values for various N/M (the average DNA fragment to bead ratio) and N (the number of fragments actually bound to a bead).

IDT Exhibit 1020.6559

N/M	0.1	0.5	1	2
R(0)	0.9	0.61	0.37	0.13
R(1)	0.09	0.3	0.37	0.27
R(N>1)	0.005	0.09	0.26	0.59

In the table the top row denotes the various ratios of N/M. R(0) denotes the fraction of beads with no DNA, R(1) denotes the fraction of beads with one DNA attached (before amplification) and R(N>1) denotes the fraction of DNA with more than one DNA attached (before amplification).

The table indicates that the maximum fraction of beads containing a single DNA fragment is 0.37 (37%) and occurs at a fragment to bead ratio of one. In this mixture, about 63% of the beads is useless for sequencing because they have either no DNA or more than a single species of DNA. Additionally, controlling the fragment to bead ratio require complex calculations and variability could produce bead batches with a significantly smaller fraction of useable beads.

This inefficiency could be significantly ameliorated if beads containing amplicon (originating from the binding of at least one fragment) could be separated from those without amplicon (originating from beads with no bound fragments). An amplicon is defined as any nucleic acid molecules produced by an *in vitro* nucleic amplification technique. Binding would be done at low average fragment-to-bead ratios (N/M <1), minimizing the ratio of beads with more than one DNA bound. A separation step would remove most or all of the beads with no DNA leaving an enriched population of beads with one species of amplified DNA. These beads may be applied to any method of sequencing such as, for example, pyrophosphate sequencing. Because the fraction of beads with one amplicon (N=1) has been enriched, any method of sequencing would be more efficient.

As an example, with an average fragment to bead ratio of 0.1, 90% of the beads will have no amplicon, 9% of the beads would be useful with one amplicon, and 0.5% of the beads will have more than one amplicon. An enrichment process of the invention will remove the 90% of the zero amplicon beads leaving a population of beads where the sequenceable fraction (N=1) is:

1-(0.005/0.09)=94%.

Dilution of the fragment to bead mixture, along with separation of beads containing amplicon can yield an enrichment of 2.5 folds over the optimal unenriched method. 94%/37% (see table above N/M=1) = 2.5. An additional benefit of the enrichment

IDT Exhibit 1020.6560

procedure of the invention is that the ultimate fraction of sequenceable beads is relatively insensitive to variability in N/M. Thus, complex calculations to derive the optimal N/M ratio are either unnecessary or may be performed to a lower level of precision. This will ultimately make the procedure more suitable to performance by less trained personnel or automation. An additional benefit of the procedure is that the zero amplicon beads may be recycled and reused. While recycling is not necessary, it may reduce cost or the total bulk of reagents making the method of the invention more suitable for some purposes such as, for example, portable sampling, remote robotic sampling and the like. In addition, all the benefits of the procedure (i.e., less trained personnel, automation, recycling of reagents) will reduce the cost of the procedure. The procedure is described in more detail below.

The enrichment procedure may be used to treat beads that have been amplified in the bead emulsion method above. The amplification is designed so that each amplified molecule contains the same DNA sequence at its 3' end. The nucleotide sequence may be a 20 mer but may be any sequence from 15 bases or more such as 25 bases, 30 bases, 35 bases, or 40 bases or longer. Naturally, while longer oligonucleotide ends are functional, they are not necessary. This DNA sequence may be introduced at the end of an amplified DNA by one of skill in the art. For example, if PCR is used for amplification of the DNA, the sequence may be part of one member of the PCR primer pair.

A schematic of the enrichment process is illustrated in Figure 7. Here, the amplicon-bound bead mixed with 4 empty beads represents the fragment-diluted amplification bead mixture. In step 1, a biotinylated primer complementary to the 3' end of the amplicon is annealed to the amplicon. In step 2, DNA polymerase and the four natural deoxynucleotides triphosphates (dNTPs) are added to the bead mix and the biotinylated primer is extended. This extension is to enhance the bonding between the biotinylated primer and the bead-bound DNA. This step may be omitted if the biotinylated primer – DNA bond is strong (e.g., in a high ionic environment). In Step 3, streptavidin coated beads susceptible to attraction by a magnetic field (referred to herein as "magnetic streptavidin beads") are introduced to the bead mixtures. Magnetic beads are commercially available, for example, from Dynal (M290). The streptavidin capture moieties binds biotins hybridized to the amplicons, which then specifically fix the amplicon-bound beads to the magnetic streptavidin beads.

In step 5, a magnetic field (represented by a magnet) is applied near the reaction mixture, which causes all the "magnetic streptavidin beads/amplicon bound bead complexes" to be positioned along one side of the tube most proximal to the magnetic field. Magnetic

IDT Exhibit 1020.6561

beads without amplicon bound beads attached are also expected to be positioned along the same side. Beads without amplicons remain in solution. The bead mixture is washed and the beads not immobilized by the magnet (i.e., the empty beads) are removed and discarded. In step 6, the extended biotinylated primer strand is separated from the amplicon strand by "melting" – a step that can be accomplished, for example, by heat or a change in pH. The heat may be 60°C in low salt conditions (e.g., in a low ionic environment such as 0.1X SSC). The change in pH may be accomplished by the addition of NaOH. The mixture is then washed and the supernatant, containing the amplicon bound beads, is recovered while the now unbound magnetic beads are retained by a magnetic field. The resultant enriched beads may be used for DNA sequencing. It is noted that the primer on the DNA capture bead may be the same as the primer of step 2 above. In this case, annealing of the amplicon-primer complementary strands (with or without extension) is the source of target-capture affinity.

The biotin streptavidin pair could be replaced by a variety of capture-target pairs. Two categories are pairs whose binding can be subsequently cleaved and those which bind irreversibly, under conditions that are practically achievable. Cleavable pairs include thiol-thiol, Digoxigenin/ anti- Digoxigenin, -Captavidin™ if cleavage of the target-capture complex is desired.

As described above, step 2 is optional. If step 2 is omitted, it may not be necessary to separate the magnetic beads from the amplicon bound beads. The amplicon bound beads, with the magnetic beads attached, may be used directly for sequencing. If the sequencing were to be performed in a microwell, separation would not be necessary if the amplicon bound bead-magnetic bead complex can fit inside the microwell.

While the use of magnetic capture beads is convenient, capture moieties can be bound to other surfaces. For example, streptavidin could be chemically bound to a surface, such as, the inner surface of a tube. In this case, the amplified bead mixture may be flowed through. The amplicon bound beads will tend to be retained until "melting" while the empty beads will flow through. This arrangement may be particularly advantageous for automating the bead preparation process.

While the embodiments described above is particularly useful, other methods can be envisioned to separate beads. For example, the capture beads may be labeled with a fluorescent moiety which would make the target-capture bead complex fluorescent. The target capture bead complex may be separated by flow cytometry or fluorescence cell sorter. Using large capture beads would allow separation by filtering or other particle size separation

IDT Exhibit 1020.6562

techniques. Since both capture and target beads are capable of forming complexes with a number of other beads, it is possible to agglutinate a mass of cross-linked capture-target beads. The large size of the agglutinated mass would make separation possible by simply washing away the unagglutinated empty beads. The methods described are described in more detail, for example, in Bauer, J.; J. Chromatography B, 722 (1999) 55-69 and in Brody et al., Applied Physics Lett. 74 (1999) 144-146.

The DNA capture beads each containing multiple copies of a single species of nucleic acid template prepared according to the above method are then suitable for distribution onto the picotiter plate.

Nucleic Acid Amplification On The Picotiter Plate

In an alternative embodiment, the nucleic acid template is distributed onto the picotiter plate prior to amplification and then amplified *in situ* on the picotiter plate. This method is described in detail in the Examples.

IDT Exhibit 1020.6563

3. SEQUENCING THE NUCLEIC ACID TEMPLATE

Pyrophosphate sequencing is used according to the methods of this invention to sequence the nucleic acid template. This technique is based on the detection of released pyrophosphate (Ppi) during DNA synthesis. See, e.g., Hyman, 1988. A new method of sequencing DNA. Anal Biochem. 174:423-36; Ronaghi, 2001. Pyrosequencing sheds light on DNA sequencing. Genome Res. 11:3-11.

In a cascade of enzymatic reactions, visible light is generated proportional to the number of incorporated nucleotides. The cascade starts with a nucleic acid polymerization reaction in which inorganic Ppi is released with nucleotide incorporation by polymerase. The released Ppi is converted to ATP by ATP sulfurylase, which provides the energy to luciferase to oxidize luciferin and generates light. Because the added nucleotide is known, the sequence of the template can be determined. Solid-phase pyrophosphate sequencing utilizes immobilized DNA in a three-enzyme system (see Figures). To increase the signal-to-noise ratio, the natural dATP has been replaced by dATPαS. Typically dATPαS is a mixture of two isomers (Sp and Rp); the use of pure 2'-deoxyadenosine-5'-O'- (1-thiotriphosphate) Sp-isomer in pyrophosphate sequencing allows substantially longer reads, up to doubling of the read length.

4. APPARATUS FOR SEQUENCING NUCLEIC ACIDS

This invention provides an apparatus for sequencing nucleic acids, which generally comprises one or more reaction chambers for conducting a sequencing reaction, means for delivering reactants to and from the reaction chamber(s), and means for detecting a sequencing reaction event. In another embodiment, the apparatus includes a reagent delivery cuvette containing a plurality of cavities on a planar surface. In a preferred embodiment, the apparatus is connected to at least one computer for controlling the individual components of the apparatus and for storing and/or analyzing the information obtained from detection of the sequence reaction event.

The invention also provides one or more reaction chambers are arranged on an inert substrate material, also referred to herein as a "solid support", that allows for discrete localization of the nucleic acid template and of the reactants in a sequencing reaction in a defined space, as well as for detection of the sequencing reaction event. Thus, as used herein, the terms "reaction chamber" or "analyte reaction chamber" refer to a localized area on the

IDT Exhibit 1020.6564

substrate material that facilitates interaction of reactants, *e.g.*, in a nucleic acid sequencing reaction. As discussed more fully below, the sequencing reactions contemplated by the invention preferably occur on numerous individual nucleic acid samples in tandem, in particular simultaneously sequencing numerous nucleic acid samples derived from genomic and chromosomal nucleic acid templates (e.g., DNA).

The apparatus of the invention therefore preferably comprises a sufficient number of reaction chambers to carry out such numerous individual sequencing reactions. In one embodiment, there are at least 10,000 reaction chambers, preferably at least 50,000 reaction chambers, more preferably greater than 100,000 reaction chambers, even more preferably greater than 200,000 reaction chambers.

Since the number of simultaneous sequencing reactions is limited by the number of reaction chambers, the throughput may be increased by fabricating plates containing increasing densities of wells. The table below shows this progression for a 14 x 43 mm and 30 x 60 mm active areas, derived from 25 x 75 mm and 40 x 75 mm arrays, respectively.

Table: Development of higher well count arrays.

Pitch (um)	Well Diameter (um)	# of Wells (14 x 43 mm)	# of Wells (30 x 60 mm)
50	44	275K	800K
43	38	375K	1.2M
35	31	575K	1.6M
25	22	1.1M	3.2M

The reaction chambers on the array typically take the form of a cavity or well in the substrate material, having a width and depth, into which reactants can be deposited. Typically the nucleic acid template is distributed into the reaction chamber on one or more solid supports or beads; the reactants are in a medium which facilitates the reaction and which flows through the reaction chamber. When formed as cavities or wells, the chambers are preferably of sufficient dimension and order to allow for (i) the introduction of the necessary reactants into the chambers, (ii) reactions to take place within the chamber and (iii) inhibition of mixing of reactants between chambers. The shape of the well or cavity is preferably circular or cylindrical, but can be multisided so as to approximate a circular or cylindrical shape. In one preferred embodiment, the shape of the well or cavity is substantially hexagonal. The cavity can have a smooth wall surface. In an additional embodiment, the cavity can have at least one irregular wall surface. The cavities can have a planar bottom or a concave bottom.

IDT Exhibit 1020.6565

The reaction chambers can be spaced between 5μm and 200μm apart. Spacing is determined by measuring the center-to-center distance between two adjacent reaction chambers. Typically, the reaction chambers can be spaced between 10μm and 150μm apart, preferably between 20 μm and 100 μm apart, most preferably between 40 and 60 μm apart. In one embodiment, the reaction chambers have a width (diameter) in one dimension of between 0.3 μm and 100 μm, more preferably between 20 μm and 70 μm and most preferably about 30 and 50 μm. The depth of the reaction chambers are preferably between 10 μm and 100 μm, preferably between 20 μm and 60 μm. Alternatively, the reaction chambers may have a depth that is between 0.25 and 5 times the width in one dimension of the reaction chamber or, in another embodiment, between 0.3 and 1 times the width in one dimension of the reaction chamber.

In a preferred embodiment, the array is fashioned from a sliced fiber optic bundle (i.e., a bundle of fused fiber optic cables) and the reaction chambers are formed by etching one surface of the fiber optic reactor array. The cavities can also be formed in the substrate via etching, molding or micromachining.

Each cavity or reaction chamber typically has a depth of between 10 μm and 100 μm; alternatively, the depth is between 0.25 and 5 times the size of the width of the cavity, preferably between 0.3 and 1 times the size of the width of the cavity.

In one embodiment, the arrays described herein typically include a planar top surface and a planar bottom surface, which is optically conductive such that optical signals from the reaction chambers can be detected through the bottom planar surface. In these arrays, typically the distance between the top surface and the bottom surface is no greater than 10 cm, preferably no greater than 2 cm, and usually between 0.5 mm to 5 mm, most preferably about 2 mm.

In a particularly preferred embodiment, the solid support is termed a picotiterplate, with reaction chambers having a center to center spacing of about 43 μm to 50 μm, a well diameter of between 38 μm to 44 μm, and a well volume of between 10 to 150 pL, preferably between 20 to 90 pL, more preferably between 40 to 85 pL, and most preferably about 75 pL.

In one embodiment, each cavity or reaction chamber of the array contains reagents for analyzing a nucleic acid or protein. Typically those reaction chambers that contain a nucleic acid (not all reaction chambers in the array are required to) contain only a single species of nucleic acid (i.e., a single sequence that is of interest). There may be a single copy of this

IDT Exhibit 1020.6566

species of nucleic acid in any particular reaction chamber, or there may be multiple copies. It is generally preferred that a reaction chamber contain at least 100,000 copies of the nucleic acid template sequence, preferably at least 1,000,000 copies, and more preferably between 2,000,000 to 20,000,000 copies, and most preferably between 5,000,000 to 15,000,000 copies of the nucleic acid. The ordinarily skilled artisan will appreciate that changes in the number of copies of a nucleic acid species in any one reaction chamber will affect the number of photons generated in a pyrosequencing reaction, and can be routinely adjusted to provide more or less photon signal as is required. In one embodiment the nucleic acid species is amplified to provide the desired number of copies using PCR, RCA, ligase chain reaction, other isothermal amplification, or other conventional means of nucleic acid amplification. In one embodiment, the nucleic acid is single stranded.

Solid Support Material

Any material can be used as the solid support material, as long as the surface allows for stable attachment of the primers and detection of nucleic acid sequences. The solid support material can be planar or can be cavitated, e.g., in a cavitated terminus of a fiber optic or in a microwell etched, molded, or otherwise micromachined into the planar surface, *e.g.* using techniques commonly used in the construction of microelectromechanical systems. See *e.g.*, Rai-Choudhury, HANDBOOK OF MICROLITHOGRAPHY, MICROMACHINING, AND MICROFABRICATION, VOLUME I: MICROLITHOGRAPHY, Volume PM39, SPIE Press (1997); Madou, CRC Press (1997), Aoki, *Biotech. Histochem.* 67: 98-9 (1992); Kane *et al.*, Biomaterials. 20: 2363-76 (1999); Deng *et al.*, *Anal. Chem.* 72:3176-80 (2000); Zhu *et al.*, *Nat. Genet.* 26:283-9 (2000). In some embodiments, the solid support is optically transparent, *e.g.*, glass.

An array of attachment sites on an optically transparent solid support can be constructed using lithographic techniques commonly used in the construction of electronic integrated circuits as described in, *e.g.*, techniques for attachment described in U.S. Patent Nos. 5,143,854, 5,445,934, 5,744,305, and 5,800,992; Chee *et al.*, *Science* 274: 610-614 (1996); Fodor *et al.*, *Nature* 364: 555-556 (1993); Fodor *et al.*, *Science* 251: 767-773 (1991); Gushin, *et al.*, *Anal. Biochem.* 250: 203-211 (1997); Kinosita *et al.*, *Cell* 93: 21-24 (1998); Kato-Yamada *et al.*, *J. Biol. Chem.* 273: 19375-19377 (1998); and Yasuda *et al.*, *Cell* 93: 1117-1124 (1998). Photolithography and electron beam lithography sensitize the solid support or substrate with a linking group that allows attachment of a modified biomolecule (*e.g.*, proteins or nucleic acids). See *e.g.*, Service, *Science* 283: 27-28

IDT Exhibit 1020.6567

(1999); Rai-Choudhury, HANDBOOK OF MICROLITHOGRAPHY, MICROMACHINING, AND MICROFABRICATION, VOLUME I: MICROLITHOGRAPHY, Volume PM39, SPIE Press (1997). Alternatively, an array of sensitized sites can be generated using thin-film technology as described in Zasadzinski *et al.*, *Science* 263: 1726-1733 (1994).

The substrate material is preferably made of a material that facilitates detection of the reaction event. For example, in a typical sequencing reaction, binding of a dNTP to a sample nucleic acid to be sequenced can be monitored by detection of photons generated by enzyme action on phosphate liberated in the sequencing reaction. Thus, having the substrate material made of a transparent or light conductive material facilitates detection of the photons.

In some embodiments, the solid support can be coupled to a bundle of optical fibers that are used to detect and transmit the light product. The total number of optical fibers within the bundle may be varied so as to match the number of individual reaction chambers in the array utilized in the sequencing reaction. The number of optical fibers incorporated into the bundle is designed to match the resolution of a detection device so as to allow 1:1 imaging. The overall sizes of the bundles are chosen so as to optimize the usable area of the detection device while maintaining desirable reagent (flow) characteristics in the reaction chamber. Thus, for a 4096 x 4096 pixel CCD (charge-coupled device) array with 15 μm pixels, the fiber bundle is chosen to be approximately 60 mm x 60 mm or to have a diameter of approximately 90 mm. The desired number of optical fibers are initially fused into a bundle or optical fiber array, the terminus of which can then be cut and polished so as to form a "wafer" of the required thickness (*e.g.*, 1.5 mm). The resulting optical fiber wafers possess similar handling properties to that of a plane of glass. The individual fibers can be any size diameter (*e.g.*, 3μm to 100 μm).

In some embodiments two fiber optic bundles are used: a first bundle is attached directly to the detection device (also referred to herein as the *fiber bundle* or *connector*) and a second bundle is used as the reaction chamber substrate (the *wafer* or *substrate*). In this case the two are placed in direct contact, optionally with the use of optical coupling fluid, in order to image the reaction centers onto the detection device. If a CCD is used as the detection device, the wafer could be slightly larger in order to maximize the use of the CCD area, or slightly smaller in order to match the format of a typical microscope slide—25 mm x 75 mm. The diameters of the individual fibers within the bundles are chosen so as to maximize the probability that a single reaction will be imaged onto a single pixel in the detection device, within the constraints of the state of the art. Exemplary diameters are 6-8 μm for the fiber

IDT Exhibit 1020.6568

bundle and 6-50 μm for the wafer, though any diameter in the range 3-100 μm can be used. Fiber bundles can be obtained commercially from CCD camera manufacturers. For example, the wafer can be obtained from Incom, Inc. (Charlton, MA) and cut and polished from a large fusion of fiber optics, typically being 2 mm thick, though possibly being 0.5 to 5 mm thick. The wafer has handling properties similar to a pane of glass or a glass microscope slide.

Reaction chambers can be formed in the substrate made from fiber optic material. The surface of the optical fiber is cavitated by treating the termini of a bundle of fibers, *e.g.*, with acid, to form an indentation in the fiber optic material. Thus, in one embodiment cavities are formed from a fiber optic bundle, preferably cavities can be formed by etching one end of the fiber optic bundle. Each cavitated surface can form a reaction chamber. Such arrays are referred to herein as fiber optic reactor arrays or FORA. The indentation ranges in depth from approximately one-half the diameter of an individual optical fiber up to two to three times the diameter of the fiber. Cavities can be introduced into the termini of the fibers by placing one side of the optical fiber wafer into an acid bath for a variable amount of time. The amount of time can vary depending upon the overall depth of the reaction cavity desired (see *e.g.*, Walt, *et al.*, 1996. *Anal. Chem.* 70: 1888). A wide channel cavity can have uniform flow velocity dimensions of approximately 14mm x 43mm. Thus, with this approximate dimension and at approximately 4.82×10^{-4} cavities/um^2 density, the apparatus can have approximately 290,000 fluidically accessible cavities. Several methods are known in the art for attaching molecules (and detecting the attached molecules) in the cavities etched in the ends of fiber optic bundles. See, *e.g.*, Michael, *et al.*, *Anal. Chem.* 70: 1242-1248 (1998); Ferguson, *et al.*, *Nature Biotechnology* 14: 1681-1684 (1996); Healey and Walt, *Anal. Chem.* 69: 2213-2216 (1997). A pattern of reactive sites can also be created in the microwell, using photolithographic techniques similar to those used in the generation of a pattern of reaction pads on a planar support. See, Healey, *et al.*, *Science* 269: 1078-1080 (1995); Munkholm and Walt, *Anal. Chem.* 58: 1427-1430 (1986), and Bronk, *et al.*, *Anal. Chem.* 67: 2750-2757 (1995).

The opposing side of the optical fiber wafer (*i.e.*, the non-etched side) is typically highly polished so as to allow optical-coupling (*e.g.*, by immersion oil or other optical coupling fluids) to a second, optical fiber bundle. This second optical fiber bundle exactly matches the diameter of the optical wafer containing the reaction chambers, and serve to act as a conduit for the transmission of light product to the attached detection device, such as a CCD imaging system or camera.

IDT Exhibit 1020.6569

In one preferred embodiment, the fiber optic wafer is thoroughly cleaned, *e.g.* by serial washes in 15% H_2O_2/15%NH_4OH volume:volume in aqueous solution, then six deionized water rinses, then 0.5M EDTA, then six deionized water, then 15% H_2O_2/15%NH_4OH, then six deionized water (one-half hour incubations in each wash).

The surface of the fiber optic wafer is preferably coated to facilitate its use in the sequencing reactions. A coated surface is preferably optically transparent, allows for easy attachment of proteins and nucleic acids, and does not negatively affect the activity of immobilized proteins. In addition, the surface preferably minimizes non-specific absorption of macromolecules and increases the stability of linked macromolecules (*e.g.*, attached nucleic acids and proteins).

Suitable materials for coating the array include, *e.g.*, plastic (e.g. polystyrene). The plastic can be preferably spin-coated or sputtered (0.1 μm thickness). Other materials for coating the array include gold layers, *e.g.* 24 karat gold, 0.1 μm thickness, with adsorbed self-assembling monolayers of long chain thiol alkanes. Biotin is then coupled covalently to the surface and saturated with a biotin-binding protein (e.g. streptavidin or avidin).

Coating materials can additionally include those systems used to attach an anchor primer to a substrate. Organosilane reagents, which allow for direct covalent coupling of proteins via amino, sulfhydryl or carboxyl groups, can also be used to coat the array. Additional coating substances include photoreactive linkers, *e.g.* photobiotin, (Amos et al., "Biomaterial Surface Modification Using Photochemical Coupling Technology," in *Encyclopedic Handbook of Biomaterials and Bioengineering, Part A: Materials*, Wise et al. (eds.), New York, Marcel Dekker, pp. 895926, 1995).

Additional coating materials include hydrophilic polymer gels (polyacrylamide, polysaccharides), which preferably polymerize directly on the surface or polymer chains covalently attached post polymerization (Hjerten, *J. Chromatogr.* 347,191 (1985); Novotny, *Anal. Chem.* 62,2478 (1990), as well as pluronic polymers (triblock copolymers, e.g. PPO-PEO-PPO, also known as F-108), specifically adsorbed to either polystyrene or silanized glass surfaces (Ho et al., *Langmuir* 14:3889-94, 1998), as well as passively adsorbed layers of biotin-binding proteins. The surface can also be coated with an epoxide which allows the coupling of reagents via an amine linkage.

In addition, any of the above materials can be derivatized with one or more functional groups, commonly known in the art for the immobilization of enzymes and nucleotides, *e.g.* metal chelating groups (e.g. nitrilo triacetic acid, iminodiacetic acid, pentadentate chelator),

IDT Exhibit 1020.6570

which will bind 6xHis-tagged proteins and nucleic acids.

Surface coatings can be used that increase the number of available binding sites for subsequent treatments, *e.g.* attachment of enzymes (discussed later), beyond the theoretical binding capacity of a 2D surface.

In a preferred embodiment, the individual optical fibers utilized to generate the fused optical fiber bundle/wafer are larger in diameter (*i.e.*, 6 μm to 12 μm) than those utilized in the optical imaging system (*i.e.*, 3 μm). Thus, several of the optical imaging fibers can be utilized to image a single reaction site.

In a particularly preferred embodiment, the sample cartridge for nucleic acid templete sequencing, termed the 'PicoTiter plate' is formed from a commercial fiber optics faceplate, acid-etched to yield well structures. Each optic fiber core is about 44 microns in diameter, with a 2-3 micron cladding, each well formed by acid etching to form a reaction well volume of about 65 pL to 85 pL, most preferably about 75 pL. The use of etched wells on a fiber optics faceplate surface serves a threefold purpose; i) delayed diffusion of the luminescence from emitting light in a different region of the array, ii) isolation of reaction chambers ("test-tubes") that contain the amplified template molecules, and iii) very efficient, high numerical aperture optical coupling to the CCD. Finally, the larger the amount of sequencing template immobilized within a well, the more optical signal one is able to achieve.

Delivery Means

An example of the means for delivering reactants to the reaction chamber is the perfusion chamber of the present invention is illustrated in Figure 13. The perfusion chamber includes a sealed compartment with transparent upper and lower side. It is designed to allow flow of solution over the surface of the substrate surface and to allow for fast exchange of reagents. Thus, it is suitable for carrying out, for example, the pyrophosphate sequencing reactions. The shape and dimensions of the chamber can be adjusted to optimize reagent exchange to include bulk flow exchange, diffusive exchange, or both in either a laminar flow or a turbulent flow regime.

The perfusion chamber is preferably detached from the imaging system while it is being prepared and only placed on the imaging system when sequencing analysis is performed. In one embodiment, the solid support (*i.e.*, a DNA chip or glass slide) is held in place by a metal or plastic housing, which may be assembled and disassembled to allow replacement of said solid support. The lower side of the solid support of the perfusion chamber carries the reaction chamber array and, with a traditional optical-based focal

IDT Exhibit 1020.6571

system, a high numerical aperture objective lens is used to focus the image of the reaction center array onto the CCD imaging system.

Many samples can thereby be analyzed in parallel. Using the method of the invention, many nucleic acid templates may be analyzed in this was by allowing the solution containing the enzymes and one nucleotide to flow over the surface and then detecting the signal produced for each sample. This procedure can then be repeated. Alternatively, several different oligonucleotides complementary to the template may be distributed over the surface followed by hybridization of the template. Incorporation of deoxynucleotides or dideoxynucleotides may be monitored for each oligonucleotide by the signal produced using the various oligonucleotides as primer. By combining the signals from different areas of the surface, sequence-based analyses may be performed by four cycles of polymerase reactions using the various dideoxynucleotides.

When the support is in the form of a cavitated array, *e.g.*, in the termini of a picotiter plate or other array of microwells, suitable delivery means for reagents include flowing and washing and also, *e.g.*, flowing, spraying, electrospraying, ink jet delivery, stamping, ultrasonic atomization (Sonotek Corp., Milton, NY) and rolling. When spraying is used, reagents may be delivered to the picotiter plate in a homogeneous thin layer produced by industrial type spraying nozzles (Spraying Systems, Co., Wheaton, IL) or atomizers used in thin layer chromatography (TLC), such as CAMAG TLC Sprayer (Camag Scientific Inc., Wilmington, NC). These sprayers atomize reagents into aerosol spray particles in the size range of 0.3 to 10 μm.

Successive reagent delivery steps are preferably separated by wash steps using techniques commonly known in the art. These washes can be performed, *e.g.*, using the above described methods, including high-flow sprayers or by a liquid flow over the picotiter plate or microwell array surface. The washes can occur in any time period after the starting material has reacted with the reagent to form a product in each reaction chamber but before the reagent delivered to any one reaction chamber has diffused out of that reaction chamber into any other reaction chamber. In one embodiment, any one reaction chamber is independent of the product formed in any other reaction chamber, but is generated using one or more common reagents.

An embodiment of a complete apparatus is illustrated in Figure 12. The apparatus includes an inlet conduit 200 in communication with a detachable perfusion chamber 226. The inlet conduit 200 allows for entry of sequencing reagents via a plurality of tubes 202-

IDT Exhibit 1020.6572

212, which are each in communication with a plurality of sequencing dispensing reagent vessels 214-224.

Reagents are introduced through the conduit 200 into the perfusion chamber 226 using either a pressurized system or pumps to drive positive flow. Typically, the reagent flow rates are from 0.05 to 50 ml/minute (*e.g.*, 1 to 50 ml/minute) with volumes from 0.100 ml to continuous flow (for washing). Valves are under computer control to allow cycling of nucleotides and wash reagents. Sequencing reagents, *e.g.*, polymerase can be either pre-mixed with nucleotides or added in stream. A manifold brings all six tubes 202-212 together into one for feeding the perfusion chamber. Thus several reagent delivery ports allow access to the perfusion chamber. For example, one of the ports may be utilized to allow the input of the aqueous sequencing reagents, while another port allows these reagents (and any reaction products) to be withdrawn from the perfusion chamber.

In a preferred embodiment, one or more reagents are delivered to an array immobilized or attached to a population of mobile solid supports, *e.g.*, a bead or microsphere. The bead or microsphere need not be spherical, irregular shaped beads may be used. They are typically constructed from numerous substances, *e.g.*, plastic, glass or ceramic and bead sizes ranging from nanometers to millimeters depending on the width of the reaction chamber. Various bead chemistries can be used *e.g.*, methylstyrene, polystyrene, acrylic polymer, latex, paramagnetic, thoria sol, carbon graphite and titanium dioxide. The construction or chemistry of the bead can be chosen to facilitate the attachment of the desired reagent.

In another embodiment, the bioactive agents are synthesized first, and then covalently attached to the beads. As is appreciated by someone skilled in the art, this will be done depending on the composition of the bioactive agents and the beads. The functionalization of solid support surfaces such as certain polymers with chemically reactive groups such as thiols, amines, carboxyls, etc. is generally known in the art.

In a preferred embodiment, the nucleic acid template is delivered to the picotiter plate on beads. The luciferase and sulfurylase enzymes are likewise delivered to each well on beads (see Figure), as is the DNA polymerase. It is noted that the one or more of the nucleic acid template, luciferase and sulfurylase may be delivered each on separate beads, or together on the same bead. To allow sequencing DNA at raised temperatures, we have cloned and modified the thermostable sulfurylase from Bacillus steareothermophilus. We have also cloned and modified several luciferase enzymes for solid-phase enzyme activity, including P.

IDT Exhibit 1020.6573

pennsylvanica and P. pyralis. The P. pyralis luciferase is used in a preferred embodiment.

"Blank" beads may be used that have surface chemistries that facilitate the attachment of the desired functionality by the user. Additional examples of these surface chemistries for blank beads include, but are not limited to, amino groups including aliphatic and aromatic amines, carboxylic acids, aldehydes, amides, chloromethyl groups, hydrazide, hydroxyl groups, sulfonates and sulfates.

These functional groups can be used to add any number of different candidate agents to the beads, generally using known chemistries. For example, candidate agents containing carbohydrates may be attached to an amino-functionalized support; the aldehyde of the carbohydrate is made using standard techniques, and then the aldehyde is reacted with an amino group on the surface. In an alternative embodiment, a sulfhydryl linker may be used. There are a number of sulfhydryl reactive linkers known in the art such as SPDP, maleimides, α-haloacetyls, and pyridyl disulfides (see for example the 1994 Pierce Chemical Company catalog, technical section on cross-linkers, pages 155-200, incorporated here by reference) which can be used to attach cysteine containing proteinaceous agents to the support. Alternatively, an amino group on the candidate agent may be used for attachment to an amino group on the surface. For example, a large number of stable bifunctional groups are well known in the art, including homobifunctional and heterobifunctional linkers (see Pierce Catalog and Handbook, pages 155-200). In an additional embodiment, carboxyl groups (either from the surface or from the candidate agent) may be derivatized using well known linkers (see Pierce catalog). For example, carbodiimides activate carboxyl groups for attack by good nucleophiles such as amines (see Torchilin et al., *Critical Rev. Thereapeutic Drug Carrier Systems*, 7(4):275-308 (1991)). Proteinaceous candidate agents may also be attached using other techniques known in the art, for example for the attachment of antibodies to polymers; see Slinkin et al., *Bioconj. Chem.* 2:342-348 (1991); Torchilin et al., *supra*; Trubetskoy et al., *Bioconj. Chem.* 3:323-327 (1992); King et al., *Cancer Res.* 54:6176-6185 (1994); and Wilbur et al., *Bioconjugate Chem.* 5:220-235 (1994). It should be understood that the candidate agents may be attached in a variety of ways, including those listed above. Preferably, the manner of attachment does not significantly alter the functionality of the candidate agent; that is, the candidate agent should be attached in such a flexible manner as to allow its interaction with a target.

Specific techniques for immobilizing enzymes on beads are known in the prior art. In one case, NH_2 surface chemistry beads are used. Surface activation is achieved with a 2.5%

IDT Exhibit 1020.6574

glutaraldehyde in phosphate buffered saline (10 mM) providing a pH of 6.9 (138 mM NaCl, 2.7 mM KCl). This mixture is stirred on a stir bed for approximately 2 hours at room temperature. The beads are then rinsed with ultrapure water plus 0.01% Tween 20 (surfactant) –0.02%, and rinsed again with a pH 7.7 PBS plus 0.01% tween 20. Finally, the enzyme is added to the solution, preferably after being prefiltered using a 0.45 μm amicon micropure filter.

The population of mobile solid supports are disposed in the reaction chambers. In some embodiments, 5% to 20% of the reaction chambers can have a mobile solid support with at least one reagent immobilized thereon, 20% to 60% of the reaction chambers can have a mobile solid support with at least one reagent immobilized thereon or 50% to 100% of the reaction chambers can have a mobile solid support with at least one reagent immobilized thereon. Preferably, at least one reaction chamber has a mobile solid support having at least one reagent immobilized thereon and the reagent is suitable for use in a nucleic acid sequencing reaction.

In some embodiments, the reagent immobilized to the mobile solid support can be a polypeptide with sulfurylase activity, a polypeptide with luciferase activity or a chimeric polypeptide having both sulfurylase and luciferase activity. In one embodiment, it can be a ATP sulfurylase and luciferase fusion protein. Since the product of the sulfurylase reaction is consumed by luciferase, proximity between these two enzymes may be achieved by covalently linking the two enzymes in the form of a fusion protein. This invention would be useful not only in substrate channeling but also in reducing production costs and potentially doubling the number of binding sites on streptavidin-coated beads.

In another embodiment, the sulfurylase is a thermostable ATP sulfurylase. In a preferred embodiment, the thermostable sulfurylase is active at temperatures above ambient (to at least 50°C). In one embodiment, the ATP sulfurylase is from a thermophile. In an additional embodiment, the mobile solid support can have a first reagent and a second reagent immobilized thereon, the first reagent is a polypeptide with sulfurylase activity and the second reagent is a polypeptide with luciferase activity.

In another embodiment, the reagent immobilized to the mobile solid support can be a nucleic acid; preferably the nucleic acid is a single stranded concatamer. In a preferred embodiment, the nucleic acid can be used for sequencing a nucleic acid, *e.g.*, a pyrosequencing reaction.

The invention also provides a method for detecting or quantifying ATP activity using

IDT Exhibit 1020.6575

a mobile solid support; preferably the ATP can be detected or quantified as part of a nucleic acid sequencing reaction.

A picotiter plate that has been "carpeted" with mobile solid supports with either nucleic acid or reagent enzymes attached thereto is shown as Figure 15.

5. METHODS OF SEQUENCING NUCLEIC ACIDS

Pyrophosphate-based sequencing is then performed. The sample DNA sequence and the extension primer are then subjected to a polymerase reaction in the presence of a nucleotide triphosphate whereby the nucleotide triphosphate will only become incorporated and release pyrophosphate (PPi) if it is complementary to the base in the target position, the nucleotide triphosphate being added either to separate aliquots of sample-primer mixture or successively to the same sample-primer mixture. The release of PPi is then detected to indicate which nucleotide is incorporated.

In one embodiment, a region of the sequence product is determined by annealing a sequencing primer to a region of the template nucleic acid, and then contacting the sequencing primer with a DNA polymerase and a known nucleotide triphosphate, *i.e.*, dATP, dCTP, dGTP, dTTP, or an analog of one of these nucleotides. The sequence can be determined by detecting a sequence reaction byproduct, as is described below.

The sequence primer can be any length or base composition, as long as it is capable of specifically annealing to a region of the amplified nucleic acid template. No particular structure for the sequencing primer is required so long as it is able to specifically prime a region on the amplified template nucleic acid. Preferably, the sequencing primer is complementary to a region of the template that is between the sequence to be characterized and the sequence hybridizable to the anchor primer. The sequencing primer is extended with the DNA polymerase to form a sequence product. The extension is performed in the presence of one or more types of nucleotide triphosphates, and if desired, auxiliary binding proteins.

Incorporation of the dNTP is preferably determined by assaying for the presence of a sequencing byproduct. In a preferred embodiment, the nucleotide sequence of the sequencing product is determined by measuring inorganic pyrophosphate (PPi) liberated from a nucleotide triphosphate (dNTP) as the dNMP is incorporated into an extended sequence primer. This method of sequencing, termed Pyrosequencing™ technology (PyroSequencing AB, Stockholm, Sweden) can be performed in solution (liquid phase) or as a solid phase technique. PPi-based sequencing methods are described generally in, *e.g.*, WO9813523A1,

IDT Exhibit 1020.6576

Ronaghi, *et al.*, 1996. *Anal. Biochem.* 242: 84-89, Ronaghi, *et al.*, 1998. *Science* 281: 363-365 (1998) and USSN 2001/0024790. These disclosures of PPi sequencing are incorporated herein in their entirety, by reference. See also , e.g., US patents 6,210,891 and 6,258,568, each fully incorporated herein by reference.

Pyrophosphate released under these conditions can be detected enzymatically (*e.g.*, by the generation of light in the luciferase-luciferin reaction). Such methods enable a nucleotide to be identified in a given target position, and the DNA to be sequenced simply and rapidly while avoiding the need for electrophoresis and the use of potentially dangerous radiolabels.

PPi can be detected by a number of different methodologies, and various enzymatic methods have been previously described (see *e.g.*, Reeves, *et al.*, 1969. *Anal.* Biochem. 28: 282-287; Guillory, *et al.*, 1971. *Anal. Biochem.* 39: 170-180; Johnson, *et al.*, 1968. *Anal. Biochem.* 15: 273; Cook, *et al.*, 1978. *Anal. Biochem.* 91: 557-565; and Drake, *et al.*, 1979. *Anal. Biochem.* 94: 117-120).

PPi liberated as a result of incorporation of a dNTP by a polymerase can be converted to ATP using, *e.g.*, an ATP sulfurylase. This enzyme has been identified as being involved in sulfur metabolism. Sulfur, in both reduced and oxidized forms, is an essential mineral nutrient for plant and animal growth (see *e.g.*, Schmidt and Jager, 1992. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* 43: 325-349). In both plants and microorganisms, active uptake of sulfate is followed by reduction to sulfide. As sulfate has a very low oxidation/reduction potential relative to available cellular reductants, the primary step in assimilation requires its activation via an ATP-dependent reaction (see *e.g.*, Leyh, 1993. *Crit. Rev. Biochem. Mol. Biol.* 28: 515-542). ATP sulfurylase (ATP: sulfate adenylyltransferase; EC 2.7.7.4) catalyzes the initial reaction in the metabolism of inorganic sulfate (SO_4^{-2}); see *e.g.*, Robbins and Lipmann, 1958. *J. Biol. Chem.* 233: 686-690; Hawes and Nicholas, 1973. *Biochem. J.* 133: 541-550). In this reaction SO_4^{-2} is activated to adenosine 5'-phosphosulfate (APS).

ATP sulfurylase has been highly purified from several sources, such as *Saccharomyces cerevisiae* (see *e.g.*, Hawes and Nicholas, 1973. *Biochem. J.* 133: 541-550); *Penicillium chrysogenum* (see *e.g.*, Renosto, *et al.*, 1990. *J. Biol. Chem.* 265: 10300-10308); rat liver (see *e.g.*, Yu, *et al.*, 1989. *Arch. Biochem. Biophys.* 269: 165-174); and plants (see *e.g.*, Shaw and Anderson, 1972. *Biochem. J.* 127: 237-247; Osslund, *et al.*, 1982. *Plant Physiol.* 70: 39-45). Furthermore, ATP sulfurylase genes have been cloned from prokaryotes (see *e.g.*, Leyh, *et al.*, 1992. *J. Biol. Chem.* 267: 10405-10410; Schwedock and Long, 1989. *Mol. Plant Microbe Interaction* 2: 181-194; Laue and Nelson, 1994. *J. Bacteriol.* 176: 3723-

IDT Exhibit 1020.6577

3729); eukaryotes (see *e.g.*, Cherest, *et al.*, 1987. *Mol. Gen. Genet.* 210: 307-313; Mountain and Korch, 1991. *Yeast* 7: 873-880; Foster, *et al.*, 1994. *J. Biol. Chem.* 269: 19777-19786); plants (see *e.g.*, Leustek, *et al.*, 1994. *Plant Physiol.* 105: 897-90216); and animals (see *e.g.*, Li, *et al.*, 1995. *J. Biol. Chem.* 270: 29453-29459). The enzyme is a homo-oligomer or heterodimer, depending upon the specific source (see *e.g.*, Leyh and Suo, 1992. *J. Biol. Chem.* 267: 542-545).

In some embodiments, a thermostable sulfurylase is used. Thermostable sulfurylases can be obtained from, *e.g.*, *Archaeoglobus* or *Pyrococcus* spp. Sequences of thermostable sulfurylases are available at database Acc. No. 028606, Acc. No. Q9YCR4, and Acc. No. P56863.

ATP sulfurylase has been used for many different applications, for example, bioluminometric detection of ADP at high concentrations of ATP (see *e.g.*, Schultz, *et al.*, 1993. *Anal. Biochem.* 215: 302-304); continuous monitoring of DNA polymerase activity (see *e.g.*, Nyrbn, 1987. *Anal. Biochem.* 167: 235-238); and DNA sequencing (see *e.g.*, Ronaghi, *et al.*, 1996. *Anal. Biochem.* 242: 84-89; Ronaghi, *et al.*, 1998. *Science* 281: 363-365; Ronaghi, *et al.*, 1998. *Anal. Biochem.* 267: 65-71).

Several assays have been developed for detection of the forward ATP sulfurylase reaction. The colorimetric molybdolysis assay is based on phosphate detection (see *e.g.*, Wilson and Bandurski, 1958. *J. Biol. Chem.* 233: 975-981), whereas the continuous spectrophotometric molybdolysis assay is based upon the detection of NADH oxidation (see *e.g.*, Seubert, *et al.*, 1983. *Arch. Biochem. Biophys.* 225: 679-691; Seubert, *et al.*, 1985. *Arch. Biochem. Biophys.* 240: 509-523). The later assay requires the presence of several detection enzymes. In addition, several radioactive assays have also been described in the literature (see *e.g.*, Daley, *et al.*, 1986. *Anal. Biochem.* 157: 385-395). For example, one assay is based upon the detection of ^{32}PPi released from ^{32}P-labeled ATP (see *e.g.*, Seubert, *et al.*, 1985. *Arch. Biochem. Biophys.* 240: 509-523) and another on the incorporation of ^{35}S into [^{35}S]-labeled APS (this assay also requires purified APS kinase as a coupling enzyme; see *e.g.*, Seubert, *et al.*, 1983. *Arch. Biochem. Biophys.* 225: 679-691); and a third reaction depends upon the release of $^{35}SO_4^{-2}$ from [^{35}S]-labeled APS (see *e.g.*, Daley, *et al.*, 1986. *Anal. Biochem.* 157: 385-395).

For detection of the reversed ATP sulfurylase reaction a continuous spectrophotometric assay (see *e.g.*, Segel, *et al.*, 1987. *Methods Enzymol.* 143: 334-349); a bioluminometric assay (see *e.g.*, Balharry and Nicholas, 1971. *Anal. Biochem.* 40: 1-17); an

IDT Exhibit 1020.6578

$^{35}SO_4^{-2}$ release assay (see *e.g.,* Seubert, *et al.,* 1985. *Arch. Biochem. Biophys.* 240: 509-523); and a ^{32}PPi incorporation assay (see *e.g.,* Osslund, *et al.,* 1982. *Plant Physiol.* 70: 39-45) have been previously described.

ATP produced by an ATP sulfurylase can be hydrolyzed using enzymatic reactions to generate light. Light-emitting chemical reactions (*i.e.,* chemiluminescence) and biological reactions (*i.e.,* bioluminescence) are widely used in analytical biochemistry for sensitive measurements of various metabolites. In bioluminescent reactions, the chemical reaction that leads to the emission of light is enzyme-catalyzed. For example, the luciferin-luciferase system allows for specific assay of ATP and the bacterial luciferase-oxidoreductase system can be used for monitoring of NAD(P)H. Both systems have been extended to the analysis of numerous substances by means of coupled reactions involving the production or utilization of ATP or NAD(P)H (see *e.g.,* Kricka, 1991. Chemiluminescent and bioluminescent techniques. *Clin. Chem.* 37: 1472-1281).

The development of new reagents have made it possible to obtain stable light emission proportional to the concentrations of ATP (see *e.g.,* Lundin, 1982. Applications of firefly luciferase In; *Luminescent Assays* (Raven Press, New York) or NAD(P)H (see *e.g.,* Lovgren, *et al.,* Continuous monitoring of NADH-converting reactions by bacterial luminescence. *J. Appl. Biochem.* 4: 103-111). With such stable light emission reagents, it is possible to make endpoint assays and to calibrate each individual assay by addition of a known amount of ATP or NAD(P)H. In addition, a stable light-emitting system also allows continuous monitoring of ATP- or NAD(P)H-converting systems.

Suitable enzymes for converting ATP into light include luciferases, *e.g.,* insect luciferases. Luciferases produce light as an end-product of catalysis. The best known light-emitting enzyme is that of the firefly, *Photinus pyralis* (*Coleoptera*). The corresponding gene has been cloned and expressed in bacteria (see *e.g.,* de Wet, *et al.,* 1985. *Proc. Natl. Acad. Sci. USA* 80: 7870-7873) and plants (see *e.g.,* Ow, *et al.,* 1986. *Science* 234: 856-859), as well as in insect (see *e.g.,* Jha, *et al.,* 1990. *FEBS Lett.* 274: 24-26) and mammalian cells (see *e.g.,* de Wet, *et al.,* 1987. *Mol. Cell. Biol.* 7: 725-7373; Keller, *et al.,* 1987. *Proc. Natl. Acad. Sci. USA* 82: 3264-3268). In addition, a number of luciferase genes from the Jamaican click beetle, *Pyroplorus plagiophihalamus* (*Coleoptera*), have recently been cloned and partially characterized (see *e.g.,* Wood, *et al.,* 1989. *J. Biolumin. Chemilumin.* 4: 289-301; Wood, *et al.,* 1989. *Science* 244: 700-702). Distinct luciferases can sometimes produce light of different wavelengths, which may enable simultaneous monitoring of light emissions at

IDT Exhibit 1020.6579

different wavelengths. Accordingly, these aforementioned characteristics are unique, and add new dimensions with respect to the utilization of current reporter systems.

Firefly luciferase catalyzes bioluminescence in the presence of luciferin, adenosine 5'-triphosphate (ATP), magnesium ions, and oxygen, resulting in a quantum yield of 0.88 (see *e.g.,* McElroy and Selinger, 1960. *Arch. Biochem. Biophys.* 88: 136-145). The firefly luciferase bioluminescent reaction can be utilized as an assay for the detection of ATP with a detection limit of approximately $1x10^{-13}$ M (see *e.g.,* Leach, 1981. *J. Appl. Biochem.* 3: 473-517). In addition, the overall degree of sensitivity and convenience of the luciferase-mediated detection systems have created considerable interest in the development of firefly luciferase-based biosensors (see *e.g.,* Green and Kricka, 1984. *Talanta* 31: 173-176; Blum, *et al.,* 1989. *J. Biolumin. Chemilumin.* 4: 543-550).

Using the above-described enzymes, the sequence primer is exposed to a polymerase and a known dNTP. If the dNTP is incorporated onto the 3' end of the primer sequence, the dNTP is cleaved and a PPi molecule is liberated. The PPi is then converted to ATP with ATP sulfurylase. Preferably, the ATP sulfurylase is present at a sufficiently high concentration that the conversion of PPi proceeds with first-order kinetics with respect to PPi. In the presence of luciferase, the ATP is hydrolyzed to generate a photon. The reaction preferably has a sufficient concentration of luciferase present within the reaction mixture such that the reaction, ATP $\rightarrow$ ADP + PO_4^{3-} + photon (light), proceeds with first-order kinetics with respect to ATP. The photon can be measured using methods and apparatuses described below. In one embodiment, the PPi and a coupled sulfurylase/luciferase reaction is used to generate light for detection. In some embodiments, either or both the sulfurylase and luciferase are immobilized on one or more mobile solid supports disposed at each reaction site.

The present invention thus permits PPi release to be detected during the polymerase reaction giving a real-time signal. The sequencing reactions may be continuously monitored in real-time. A procedure for rapid detection of PPi release is thus enabled by the present invention. The reactions have been estimated to take place in less than 2 seconds (Nyren and Lundin, *supra*). The rate limiting step is the conversion of PPi to ATP by ATP sulfurylase, while the luciferase reaction is fast and has been estimated to take less than 0.2 seconds. Incorporation rates for polymerases have also been estimated by various methods and it has been found, for example, that in the case of Klenow polymerase, complete incorporation of one base may take less than 0.5 seconds. Thus, the estimated total time for incorporation of one base and detection by this enzymatic assay is approximately 3 seconds. It will be seen

IDT Exhibit 1020.6580

therefore that very fast reaction times are possible, enabling real-time detection. The reaction times could further be decreased by using a more thermostable luciferase.

For most applications it is desirable to use reagents free of contaminants like ATP and PPi. These contaminants may be removed by flowing the reagents through a pre-column containing apyrase and/-or pyrophosphatase bound to resin. Alternatively, the apyrase or pyrophosphatase can be bound to magnetic beads and used to remove contaminating ATP and PPi present in the reagents. In addition it is desirable to wash away diffusible sequencing reagents, *e.g.*, unincorporated dNTPs, with a wash buffer. Any wash buffer used in pyrophosphate sequencing can be used.

In some embodiments, the concentration of reactants in the sequencing reaction include 1 pmol DNA, 3 pmol polymerase, 40 pmol dNTP in 0.2 ml buffer. See Ronaghi, *et al.*, *Anal. Biochem.* 242: 84-89 (1996).

The sequencing reaction can be performed with each of four predetermined nucleotides, if desired. A "complete" cycle generally includes sequentially administering sequencing reagents for each of the nucleotides dATP, dGTP, dCTP and dTTP (or dUTP), in a predetermined order. Unincorporated dNTPs are washed away between each of the nucleotide additions. Alternatively, unincorporated dNTPs are degraded by apyrase (see below). The cycle is repeated as desired until the desired amount of sequence of the sequence product is obtained. In some embodiments, about 10-1000, 10-100, 10-75, 20-50, or about 30 nucleotides of sequence information is obtained from extension of one annealed sequencing primer.

In some embodiments, the nucleotide is modified to contain a disulfide-derivative of a hapten such as biotin. The addition of the modified nucleotide to the nascent primer annealed to the anchored substrate is analyzed by a post-polymerization step that includes i) sequentially binding of, in the example where the modification is a biotin, an avidin- or streptavidin-conjugated moiety linked to an enzyme molecule, ii) the washing away of excess avidin- or streptavidin-linked enzyme, iii) the flow of a suitable enzyme substrate under conditions amenable to enzyme activity, and iv) the detection of enzyme substrate reaction product or products. The hapten is removed in this embodiment through the addition of a reducing agent. Such methods enable a nucleotide to be identified in a given target position, and the DNA to be sequenced simply and rapidly while avoiding the need for electrophoresis and the use of potentially dangerous radiolabels.

A preferred enzyme for detecting the hapten is horse-radish peroxidase. If desired, a

IDT Exhibit 1020.6581

wash buffer, can be used between the addition of various reactants herein. Apyrase can be used to remove unreacted dNTP used to extend the sequencing primer. The wash buffer can optionally include apyrase.

Example haptens, e.g., biotin, digoxygenin, the fluorescent dye molecules cy3 and cy5, and fluorescein, are incorporated at various efficiencies into extended DNA molecules. The attachment of the hapten can occur through linkages via the sugar, the base, and via the phosphate moiety on the nucleotide. Example means for signal amplification include fluorescent, electrochemical and enzymatic. In a preferred embodiment using enzymatic amplification, the enzyme, e.g. alkaline phosphatase (AP), horse-radish peroxidase (HRP), beta-galactosidase, luciferase, can include those for which light-generating substrates are known, and the means for detection of these light-generating (chemiluminescent) substrates can include a CCD camera.

In a preferred mode, the modified base is added, detection occurs, and the hapten-conjugated moiety is removed or inactivated by use of either a cleaving or inactivating agent. For example, if the cleavable-linker is a disulfide, then the cleaving agent can be a reducing agent, for example dithiothreitol (DTT), beta-mercaptoethanol, etc. Other embodiments of inactivation include heat, cold, chemical denaturants, surfactants, hydrophobic reagents, and suicide inhibitors.

Luciferase can hydrolyze dATP directly with concomitant release of a photon. This results in a false positive signal because the hydrolysis occurs independent of incorporation of the dATP into the extended sequencing primer. To avoid this problem, a dATP analog can be used which is incorporated into DNA, *i.e.*, it is a substrate for a DNA polymerase, but is not a substrate for luciferase. One such analog is α-thio-dATP. Thus, use of α-thio-dATP avoids the spurious photon generation that can occur when dATP is hydrolyzed without being incorporated into a growing nucleic acid chain.

Typically, the PPi-based detection is calibrated by the measurement of the light released following the addition of control nucleotides to the sequencing reaction mixture immediately after the addition of the sequencing primer. This allows for normalization of the reaction conditions. Incorporation of two or more identical nucleotides in succession is revealed by a corresponding increase in the amount of light released. Thus, a two-fold increase in released light relative to control nucleotides reveals the incorporation of two successive dNTPs into the extended primer.

If desired, apyrase may be "washed" or "flowed" over the surface of the solid support

IDT Exhibit 1020.6582

so as to facilitate the degradation of any remaining, non-incorporated dNTPs within the sequencing reaction mixture. Apyrase also degrades the generated ATP and hence "turns off" the light generated from the reaction. Upon treatment with apyrase, any remaining reactants are washed away in preparation for the following dNTP incubation and photon detection steps. Alternatively, the apyrase may be bound to the solid or mobile solid support.

Double Ended Sequencing

In a preferred embodiment we provide a method for sequencing from both ends of a nucleic acid template. Traditionally, the sequencing of two ends of a double stranded DNA molecule would require at the very least the hybridization of primer, sequencing of one end, hybridization of a second primer, and sequencing of the other end. The alternative method is to separate the individual strands of the double stranded nucleic acid and individually sequence each strand. The present invention provides a third alternative that is more rapid and less labor intensive than the first two methods.

The present invention provides for a method of sequential sequencing of nucleic acids from multiple primers. References to DNA sequencing in this application are directed to sequencing using a polymerase wherein the sequence is determined as the nucleotide triphosphate (NTP) is incorporated into the growing chain of a sequencing primer. One example of this type of sequencing is the pyro-sequencing detection pyrophosphate method (see, *e.g.*, U.S. Patents 6,274,320, 6258,568 and 6,210,891, each of which is incorporated in total herein by reference.).

In one embodiment, the present invention provides for a method for sequencing two ends of a template double stranded nucleic acid. The double stranded DNA is comprised of two single stranded DNA; referred to herein as a first single stranded DNA and a second single stranded DNA. A first primer is hybridized to the first single stranded DNA and a second primer is hybridized to the second single stranded DNA. The first primer is unprotected while the second primer is protected. "Protection" and "protected" are defined in this disclosure as being the addition of a chemical group to reactive sites on the primer that prevents a primer from polymerization by DNA polymerase. Further, the addition of such chemical protecting groups should be reversible so that after reversion, the now deprotected primer is once again able to serve as a sequencing primer. The nucleic acid sequence is determined in one direction (*e.g.*, from one end of the template) by elongating the first primer with DNA polymerase using conventional methods such as pyrophosphate sequencing. The

IDT Exhibit 1020.6583

second primer is then deprotected, and the sequence is determined by elongating the second primer in the other direction (*e.g.*, from the other end of the template) using DNA polymerase and conventional methods such as pyrophosphate sequencing. The sequences of the first and second primers are specifically designed to hybridize to the two ends of the double stranded DNA or at any location along the template in this method.

In another embodiment, the present invention provides for a method of sequencing a nucleic acid from multiple primers. In this method a number of sequencing primers are hybridized to the template nucleic acid to be sequenced. All the sequencing primers are reversibly protected except for one. A protected primer is an oligonucleotide primer that cannot be extended with polymerase and dNTPs which are commonly used in DNA sequencing reactions. A reversibly protected primer is a protected primer which can be deprotected. All protected primers referred to in this invention are reversibly protected. After deprotection, a reversibly protected primer functions as a normal sequencing primer and is capable of participating in a normal sequencing reaction.

The present invention provides for a method of sequential sequencing a nucleic acid from multiple primers. The method comprises the following steps: First, one or more template nucleic acids to be sequenced are provided. Second, a plurality of sequencing primers are hybridized to the template nucleic acid or acids. The number of sequencing primers may be represented by the number n where n can be any positive number greater than 1. That number may be, for example, 2, 3, 4, 5, 6, 7, 8, 9, 10 or greater. Of the primers, n-1 number may be protected by a protection group. So, for example, if n is 2, 3, 4, 5, 6, 7, 8, 9 or 10, n-1 would be 1, 2, 3, 4, 5, 6, 7, 8, 9 respectively. The remaining primer (*e.g.*, n number primers – (n-1) number of protected primers = one remaining primer) is unprotected. Third, the unprotected primer is extended and the template DNA sequence is determined by conventional methods such as, for example, pyrophosphate sequencing. Fourth, after the sequencing of the first primer, one of the remaining protected primers is unprotected. Fifth, unprotected primer is extended and the template DNA sequence is determined by conventional methods such as, for example, pyrophosphate sequencing. Optionally, the method may be repeated until sequencing is performed on all the protected primers.

In another aspect, the present invention includes a method of sequential sequencing of a nucleic acid comprising the steps of: (a) hybridizing 2 or more sequencing primers to the nucleic acid wherein all the primers except for one are reversibly protected; (b) determining a sequence of one strand of the nucleic acid by polymerase elongation from the unprotected

IDT Exhibit 1020.6584

primer; (c) deprotecting one of the reversibly protected primers into an unprotected primer; (d) repeating steps (b) and (c) until all the reversibly protected primers are deprotected and used for determining a sequence. In one embodiment, this method comprises one additional step between steps (b) and (c), i.e., the step of terminating the elongation of the unprotected primer by contacting the unprotected primer with DNA polymerase and one or more of a nucleotide triphosphate or a dideoxy nucleotide triphosphate. In yet another embodiment, this method further comprises an additional step between said step (b) and (c), i.e., terminating the elongation of the unprotected primer by contacting the unprotected primer with DNA polymerase and a dideoxy nucleotide triphosphate from ddATP, ddTTP, ddCTP, ddGTP or a combination thereof.

In another aspect, this invention includes a method of sequencing a nucleic acid comprising: (a) hybridizing a first unprotected primer to a first strand of the nucleic acid; (b) hybridizing a second protected primer to a second strand; (c) exposing the first and second strands to polymerase, such that the first unprotected primer is extended along the first strand; (d) completing the extension of the first sequencing primer; (e) deprotecting the second sequencing primer; and (f) exposing the first and second strands to polymerase so that the second sequencing primer is extended along the second strand. In a preferred embodiment, completing comprises capping or terminating the elongation.

In another embodiment, the present invention provides for a method for sequencing two ends of a template double stranded nucleic acid that comprises a first and a second single stranded DNA. In this embodiment, a first primer is hybridized to the first single stranded DNA and a second primer is hybridized to the second single stranded DNA in the same step. The first primer is unprotected while the second primer is protected.

Following hybridization, the nucleic acid sequence is determined in one direction (*e.g.*, from one end of the template) by elongating the first primer with DNA polymerase using conventional methods such as pyrophosphate sequencing. In a preferred embodiment, the polymerase is devoid of 3' to 5' exonuclease activity. The second primer is then deprotected, and its sequence is determined by elongating the second primer in the other direction (*e.g.*, from the other end of the template) with DNA polymerase using conventional methods such as pyrophosphate sequencing. As described earlier, the sequences of the first primer and the second primer are designed to hybridize to the two ends of the double stranded DNA or at any location along the template. This technique is especially useful for the sequencing of many template DNAs that contain unique sequencing primer hybridization

IDT Exhibit 1020.6585

sites on its two ends. For example, many cloning vectors provide unique sequencing primer hybridization sites flanking the insert site to facilitate subsequent sequencing of any cloned sequence (*e.g.*, Bluescript, Stratagene, La Jolla, CA).

One benefit of this method of the present invention is that both primers may be hybridized in a single step. The benefits of this and other methods are especially useful in parallel sequencing systems where hybridizations are more involved than normal. Examples of parallel sequencing systems are disclosed in copending U.S. patent application serial number 10/104,280, the disclosure of which is incorporated in total herein.

The oligonucleotide primers of the present invention may be synthesized by conventional technology, *e.g.*, with a commercial oligonucleotide synthesizer and/or by ligating together subfragments that have been so synthesized.

In another embodiment of the invention, the length of the double stranded target nucleic acid may be determined. Methods of determining the length of a double stranded nucleic acid are known in the art. The length determination may be performed before or after the nucleic acid is sequenced. Known methods of nucleic acid molecule length determination include gel electrophoresis, pulsed field gel electrophoresis, mass spectroscopy and the like. Since a blunt ended double stranded nucleic acid is comprised of two single strands of identical lengths, the determination of the length of one strand of a nucleic acid is sufficient to determine the length of the corresponding double strand.

The sequence reaction according to the present invention also allows a determination of the template nucleic acid length. First, a complete sequence from one end of the nucleic acid to another end will allow the length to be determined. Second, the sequence determination of the two ends may overlap in the middle allowing the two sequences to be linked. The complete sequence may be determined and the length may be revealed. For example, if the template is 100 bps long, sequencing from one end may determine bases 1 to 75; sequencing from the other end may determine bases 25 to 100; there is thus a 51 base overlap in the middle from base 25 to base 75; and from this information, the complete sequence from 1 to 100 may be determined and the length, of 100 bases, may be revealed by the complete sequence.

Another method of the present invention is directed to a method comprising the following steps. First a plurality of sequencing primers, each with a different sequence, is hybridized to a DNA to be sequenced. The number of sequencing primers may be any value greater than one such as, for example, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more. All of these primers

IDT Exhibit 1020.6586

are reversibly protected except for one. The one unprotected primer is elongated in a sequencing reaction and a sequence is determined. Usually, when a primer is completely elongated, it cannot extend and will not affect subsequent sequencing from another primer. If desired, the sequenced primer may be terminated using excess polymerase and dNTP or using ddNTPs. If a termination step is taken, the termination reagents (dNTPs and ddNTPs) should be removed after the step. Then, one of the reversibly protected primers is unprotected and sequencing from the second primer proceeds. The steps of deprotecting a primer, sequencing from the deprotected primer, and optionally, terminating sequencing from the primer is repeated until all the protected primers are unprotected and used in sequencing.

The reversibly protected primers should be protected with different chemical groups. By choosing the appropriate method of deprotection, one primer may be deprotected without affecting the protection groups of the other primers. In a preferred embodiment, the protection group is PO_4. That is, the second primer is protected by PO_4 and deprotection is accomplished by T4 polynucleotide kinase (utilizing its 3'-phosphatase activity). In another preferred embodiment, the protection is a thio group or a phosphorothiol group.

The template nucleic acid may be a DNA, RNA, or peptide nucleic acid (PNA). While DNA is the preferred template, RNA and PNA may be converted to DNA by known techniques such as random primed PCR, reverse transcription, RT-PCR or a combination of these techniques. Further, the methods of the invention are useful for sequencing nucleic acids of unknown and known sequence. The sequencing of nucleic acid of known sequence would be useful, for example, for confirming the sequence of synthesized DNA or for confirming the identity of suspected pathogen with a known nucleic acid sequence. The nucleic acids may be a mixture of more than one population of nucleic acids. It is known that a sequencing primer with sufficient specificity (*e.g.*, 20 bases, 25 bases, 30 bases, 35 bases, 40 bases, 45 bases, or 50 bases) may be used to sequence a subset of sequences in a long nucleic acid or in a population of unrelated nucleic acids. Thus, for example, the template may be one sequence of 10 Kb or ten sequences of 1 Kb each. In a preferred embodiment, the template DNA is between 50 bp to 700 bp in length. The DNA can be single stranded or double stranded.

In the case where the template nucleic acid is single stranded, a number of primers may be hybridized to the template nucleic acid as shown below:

5'--primer 4 --3' 5'-primer 3 --3' 5'-primer 2 -3' 5'-primer 1-3'

3' ----------template nucleic acid ----------5'

IDT Exhibit 1020.6587

In this case, it is preferred that the initial unprotected primer would be the primer that hybridizes at the most 5' end of the template. See primer 1 in the above illustration. In this orientation, the elongation of primer 1 would not displace (by strand displacement) primer 2, 3, or 4. When sequencing from primer 1 is finished, primer 2 can be unprotected and nucleic acid sequencing can commence. The sequencing from primer 2 may displace primer 1 or the elongated version of primer one but would have no effect on the remaining protected primers (primers 3 and 4). Using this order, each primer may be used sequentially and a sequencing reaction from one primer would not affect the sequencing from a subsequent primer.

One feature of the invention is the ability to use multiple sequencing primers on one or more nucleic acids and the ability to sequence from multiple primers using only one hybridization step. In the hybridization step, all the sequencing primers (*e.g.*, the n number of sequencing primers) may be hybridized to the template nucleic acid(s) at the same time. In conventional sequencing, usually one hybridization step is required for sequencing from one primer. One feature of the invention is that the sequencing from n primers (as defined above) may be performed by a single hybridization step. This effectively eliminates n-1 hybridization step.

In a preferred embodiment, the sequences of the n number of primers are sufficiently different that the primers do not cross hybridize or self-hybridize. Cross hybridization refers to the hybridization of one primer to another primer because of sequence complementarity. One form of cross hybridization is commonly referred to as a "primer dimer." In the case of a primer dimer, the 3' ends of two primers are complementary and form a structure that when elongated, is approximately the sum of the length of the two primers. Self-hybridization refers to the situation where the 5' end of a primer is complementary to the 3' end of the primer. In that case, the primer has a tendency to self hybridize to form a hairpin-like structure.

A primer can interact or become associated specifically with the template molecule. By the terms "interact" or "associate", it is meant herein that two substances or compounds (*e.g.*, primer and template; chemical moiety and nucleotide) are bound (*e.g.*, attached, bound, hybridized, joined, annealed, covalently linked, or otherwise associated) to one another sufficiently that the intended assay can be conducted. By the terms "specific" or "specifically", it is meant herein that two components bind selectively to each other. The parameters required to achieve specific interactions can be determined routinely, *e.g.*, using conventional methods in the art.

IDT Exhibit 1020.6588

To gain more sensitivity or to help in the analysis of complex mixtures, the protected primers can be modified (*e.g.*, derivatized) with chemical moieties designed to give clear unique signals. For example, each protected primer can be derivatized with a different natural or synthetic amino acid attached through an amide bond to the oligonucleotide strand at one or more positions along the hybridizing portion of the strand. The chemical modification can be detected, of course, either after having been cleaved from the target nucleic acid, or while in association with the target nucleic acid. By allowing each protected target nucleic acid to be identified in a distinguishable manner, it is possible to assay (*e.g.*, to screen) for a large number of different target nucleic acids in a single assay. Many such assays can be performed rapidly and easily. Such an assay or set of assays can be conducted, therefore, with high throughput efficiency as defined herein.

In the methods of the invention, after a first primer is elongated and the sequence of the template DNA is determined, a second primer is deprotected and sequenced. There is no interference between the sequencing reaction of the first primer with the sequencing reaction of the second, now unprotected, primer because the first primer is completely elongated or terminated. Because the first primer is completely elongated, the sequencing from the second primer, using conventional methods such a pyrophosphate sequencing, will not be affected by the presence of the elongated first primer. The invention also provides a method of reducing any possible signal contamination from the first primer. Signal contamination refers to the incidences where the first primer is not completely elongated. In that case, the first primer will continue to elongate when a subsequent primer is deprotected and elongated. The elongation of both the first and second primers may interfere with the determination of DNA sequence.

In a preferred embodiment, the sequencing reaction (*e.g.*, the chain elongation reaction) from one primer is first terminated or completed before a sequencing reaction is started on a second primer. A chain elongation reaction of DNA can be terminated by contacting the template DNA with DNA polymerase and dideoxy nucleotide triphosphates (ddNTPs) such as ddATP, ddTTP, ddGTP and ddCTP. Following termination, the dideoxy nucleotide triphosphates may be removed by washing the reaction with a solution without ddNTPs. A second method of preventing further elongation of a primer is to add nucleotide triphosphates (dNTPs such as dATP, dTTP, dGTP and dCTP) and DNA polymerase to a reaction to completely extend any primer that is not completely extended. Following complete extension, the dNTPs and the polymerases are removed before the next primer is

IDT Exhibit 1020.6589

deprotected. By completing or terminating one primer before deprotecting another primer, the signal to noise ratio of the sequencing reaction (*e.g.*, pyrophosphate sequencing) can be improved.

The steps of (a) optionally terminating or completing the sequencing, (b) deprotecting a new primer, and (c) sequencing from the deprotected primer may be repeated until a sequence is determined from the elongation of each primer. In this method, the hybridization step comprises "n" number of primers and one unprotected primer. The unprotected primer is sequenced first and the steps of (a), (b) and (c) above may be repeated.

In a preferred embodiment, pyrophosphate sequencing is used for all sequencing conducted in accordance with the method of the present invention.

In another preferred embodiment, the double ended sequencing is performed according to the process outlined in Figure 10. This process may be divided into six steps: (1) creation of a capture bead (Figure 10A); (2) drive to bead (DTB) PCR amplification (Figure 10B); (3) SL reporter system preparation (Figure 10C); (4) sequencing of the first strand (Figure 10D); (5) preparation of the second strand (Figures 10E and 10F); and (6) analysis of each strand (Figure 10G). This exemplary process is outlined below.

In step 1, an N-hydroxysuccinimide (NHS)-activated capture bead (*e.g.*, Amersham Biosciences, Piscataway, NJ) is coupled to both a forward primer and a reverse primer. NHS coupling forms a chemically stable amide bond with ligands containing primary amino groups. The capture bead is also coupled to biotin (Figure 10A). The beads (i.e., solid nucleic acid capturing supports) used herein may be of any convenient size and fabricated from any number of known materials. Example of such materials include: inorganics, natural polymers, and synthetic polymers. Specific examples of these materials include: cellulose, cellulose derivatives, acrylic resins, glass; silica gels, polystyrene, gelatin, polyvinyl pyrrolidone, co-polymers of vinyl and acrylamide, polystyrene cross-linked with divinylbenzene or the like (see, Merrifield Biochemistry 1964, 3, 1385-1390), polyacrylamides, latex gels, polystyrene, dextran, rubber, silicon, plastics, nitrocellulose, celluloses, natural sponges, silica gels, glass, metals plastic, cellulose, cross-linked dextrans (e.g., Sephadex™) and agarose gel (Sepharose™) and solid phase supports known to those of skill in the art. In a preferred embodiment, the capture beads are Sepharose beads approximately 25 to 40 μM in diameter.

In step 2, template DNA which has hybridized to the forward and reverse primers is added, and the DNA is amplified through a PCR amplification strategy (Figure 10B). In one

IDT Exhibit 1020.6590

embodiment, the DNA is amplified by Emulsion Polymerase Chain Reaction, Drive to Bead Polymerase Chain Reaction, Rolling Circle Amplification or Loop-mediated Isothermal Amplification. In step 3, streptavidin is added followed by the addition of sulfurylase and luciferase which are coupled to the streptavidin (Figure 10C). The addition of auxiliary enzymes during a sequencing method has been disclosed in U.S.S.N. 10/104,280 and U.S.S.N. 10/127,906, which are incorporated herein in their entireties by reference. In one embodiment, the template DNA has a DNA adaptor ligated to both the 5' and 3' end. In a preferred embodiment, the DNA is coupled to the DNA capture bead by hybridization of one of the DNA adaptors to a complimentary sequence on the DNA capture bead.

In the first step, single stranded nucleic acid template to be amplified is attached to a capture bead. The nucleic acid template may be attached to the capture bead in any manner known in the art. Numerous methods exist in the art for attaching the DNA to a microscopic bead. Covalent chemical attachment of the DNA to the bead can be accomplished by using standard coupling agents, such as water-soluble carbodiimide, to link the 5'-phosphate on the DNA to amine-coated microspheres through a phosphoamidate bond. Another alternative is to first couple specific oligonucleotide linkers to the bead using similar chemistry, and to then use DNA ligase to link the DNA to the linker on the bead. Other linkage chemistries include the use of N-hydroxysuccinamide (NHS) and its derivatives, to join the oligonucleotide to the beads. In such a method, one end of the oligonucleotide may contain a reactive group (such as an amide group) which forms a covalent bond with the solid support, while the other end of the linker contains another reactive group which can bond with the oligonucleotide to be immobilized. In a preferred embodiment, the oligonucleotide is bound to the DNA capture bead by covalent linkage. However, non-covalent linkages, such as chelation or antigen-antibody complexes, may be used to join the oligonucleotide to the bead.

Oligonucleotide linkers can be employed which specifically hybridize to unique sequences at the end of the DNA fragment, such as the overlapping end from a restriction enzyme site or the "sticky ends" of bacteriophage lambda based cloning vectors, but blunt-end ligations can also be used beneficially. These methods are described in detail in US 5,674,743, the disclosure of which is incorporated *in toto* herein. It is preferred that any method used to immobilize the beads will continue to bind the immobilized oligonucleotide throughout the steps in the methods of the invention. In a preferred embodiment, the oligonucleotide is bound to the DNA capture bead by covalent linkage. However, non-covalent linkages, such as chelation or antigen-antibody complexes, may be used to join the

IDT Exhibit 1020.6591

oligonucleotide to the bead.

In step 4, the first strand of DNA is sequenced by depositing the capture beads onto a PicoTiter plate (PTP), and sequencing by a method known to one of ordinary skill in the art (*e.g.*, pyrophosphate sequencing) (Figure 10D). Following sequencing, a mixture of dNTPs and ddNTPs are added in order to "cap" or terminate the sequencing process (Figure 10E). In step 5, the second strand of nucleic acid is prepared by adding apyrase to remove the ddNTPs and polynucleotide kinase (PNK) to remove the 3' phosphate group from the blocked primer strand (Figure 10F). Polymerase is then added to prime the second strand followed by sequencing of the second strand according to a standard method known to one of ordinary skill in the art (Figure 10G). In step 7, the sequence of the both the first and second strand is analyzed such that a contiguous DNA sequence is determined.

Detection means

The solid support is optically linked to an imaging system 230, which includes a CCD system in association with conventional optics or a fiber optic bundle. In one embodiment the perfusion chamber substrate includes a fiber optic array wafer such that light generated near the aqueous interface is transmitted directly through the optical fibers to the exterior of the substrate or chamber. When the CCD system includes a fiber optic connector, imaging can be accomplished by placing the perfusion chamber substrate in direct contact with the connector. Alternatively, conventional optics can be used to image the light, *e.g.*, by using a 1-1 magnification high numerical aperture lens system, from the exterior of the fiber optic substrate directly onto the CCD sensor. When the substrate does not provide for fiber optic coupling, a lens system can also be used as described above, in which case either the substrate or the perfusion chamber cover is optically transparent. An exemplary CCD imaging system is described above.

The imaging system 230 is used to collect light from the reactors on the substrate surface. Light can be imaged, for example, onto a CCD using a high sensitivity low noise apparatus known in the art. For fiber-optic based imaging, it is preferable to incorporate the optical fibers directly into the cover slip or for a FORA to have the optical fibers that form the microwells also be the optical fibers that convey light to the detector.

The imaging system is linked to a computer control and data collection system 240. In general, any commonly available hardware and software package can be used. The computer control and data collection system is also linked to the conduit 200 to control reagent delivery.

IDT Exhibit 1020.6592

The photons generated by the pyrophosphate sequencing reaction are captured by the CCD only if they pass through a focusing device (*e.g.*, an optical lens or optical fiber) and are focused upon a CCD element. However, the emitted photons will escape equally in all directions. In order to maximize their subsequent "capture" and quantitation when utilizing a planar array (*e.g.*, a DNA chip), it is preferable to collect the photons as close as possible to the point at which they are generated, *e.g.* immediately at the planar solid support. This is accomplished by either: (*i*) utilizing optical immersion oil between the cover slip and a traditional optical lens or optical fiber bundle or, preferably, (*ii*) incorporating optical fibers directly into the cover slip itself. Similarly, when a thin, optically transparent planar surface is used, the optical fiber bundle can also be placed against its back surface, eliminating the need to "image" through the depth of the entire reaction/perfusion chamber.

The reaction event, *e.g.*, photons generated by luciferase, may be detected and quantified using a variety of detection apparatuses, *e.g.*, a photomultiplier tube, a CCD, CMOS, absorbance photometer, a luminometer, charge injection device (CID), or other solid state detector, as well as the apparatuses described herein. In a preferred embodiment, the quantitation of the emitted photons is accomplished by the use of a CCD camera fitted with a fused fiber optic bundle. In another preferred embodiment, the quantitation of the emitted photons is accomplished by the use of a CCD camera fitted with a microchannel plate intensifier. A back-thinned CCD can be used to increase sensitivity. CCD detectors are described in, *e.g.*, Bronks, *et al.*, 1995. *Anal. Chem.* 65: 2750-2757.

An exemplary CCD system is a Spectral Instruments, Inc. (Tucson, AZ) Series 600 4-port camera with a Lockheed-Martin LM485 CCD chip and a 1-1 fiber optic connector (bundle) with 6-8 μm individual fiber diameters. This system has 4096 x 4096, or greater than 16 million pixels and has a quantum efficiency ranging from 10% to > 40%. Thus, depending on wavelength, as much as 40% of the photons imaged onto the CCD sensor are converted to detectable electrons.

In other embodiments, a fluorescent moiety can be used as a label and the detection of a reaction event can be carried out using a confocal scanning microscope to scan the surface of an array with a laser or other techniques such as scanning near-field optical microscopy (SNOM) are available which are capable of smaller optical resolution, thereby allowing the use of "more dense" arrays. For example, using SNOM, individual polynucleotides may be distinguished when separated by a distance of less than 100 nm, *e.g.*, 10nm x 10nm. Additionally, scanning tunneling microscopy (Binning et al., *Helvetica Physica Acta*, 55:726-

IDT Exhibit 1020.6593

735, 1982) and atomic force microscopy (Hanswa et al., *Annu Rev Biophys Biomol Struct*, 23:115-139, 1994) can be used.

Haplotype Application

Virtually any sequencing application can be accomplished using the methods and apparatus of this invention. In one embodiment we contemplate haplotype mapping. Human gene diversity is an important factor in the variability of patient response to pharmaceuticals. The most precise measurement of this diversity is the haplotype, which is the organization of polymorphic variation as it is found on a chromosome. Recently, major government and academic genome researchers in the US, Canada and Europe have agreed that haplotypes are a powerful tool that can reduce the complexity of genetic information to a practical form. Haplotypes can be used in drug discovery to improve the outcome of target validation and drug screening studies and in drug development to improve the design and reliability of clinical trials. Haplotype markers can be used to predict the efficacy and safety of new and approved drugs and will serve as the foundation for a new paradigm of personalized medicine matching patients to the right drug at the right dose via guidance from a database of clinical marker-associations.

Numerous empirical studies have shown that nearby SNP alleles are often in linkage disequilibrium (LD) with each other, such that the state of one SNP allele is often highly correlated with the allele of another close SNP. These correlations exist because of the shared history of tightly linked SNP's, which are co-transmitted from generation to generation. Patterns of human sequence variation (haplotypes) thus represent ancestral DNA segments. Historical meioses have slowly dissociated alleles from neighboring alleles on ancestral chromosomes, except for tightly linked variants. The extent of linkage disequilibrium in founder populations with recent bottlenecks had been the object of numerous studies – particularly in the cloning of simple Mendelian disorders disorders such as cystic fibrosis (16), Huntington's disease (11), diastrophic dysplasia (DTD) (8). Whereas these cloning studies benefited from the large chromosomal segments showing LD spanning over large distances (often in the megabase range), very little empirical data was available until recently regarding LD across the human genome in the world population.

We focus on three recent examples of large-scale surveys of LD (and haplotypes): (see, e.g, Reich, D. E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P. C., Richter, D. J., Lavery, T., Kouyoumjian, R., Farhadian, S. F., Ward, R. & Lander, E. S. 2001. Linkage disequilibrium in the human genome. Nature 411, 199-204.26). We sampled 19

IDT Exhibit 1020.6594

chromosome regions for their SNP content. High frequency SNP's spanning intervals of 2 to 160 kb were first genotyped in a Caucasian samples. Over all regions, LD was detectable at distances of about 60kb, with a significant difference between regions, as the range was as short as 6kb at one locus and as long 155 kb in another. Not surprisingly, LD was significantly correlated with the estimated local recombination rates. Further analysis in a Nigerian sample provided evidence of shorter LD in this population – although the allelic combinations over short distances were similar to the Caucasian sample. Overall – this work provided evidence that large blocks of LD are common across the human genome, and that genome-wide LD mapping of disease genes will be feasible.

Kits

The invention also comprises kits for use in methods of the invention which could include one or more of the following components: (a) a test specific primer which hybridizes to sample DNA so that the target position is directly adjacent to the 3' end of the primer; (b) a polymerase; (c) detection enzyme means for identifying PPi release; (d) deoxynucleotides including, in place of dATP, a dATP analogue which is capable of acting as a substrate for a polymerase but incapable of acting as a substrate for a said PPi-detection enzyme; and (e) optionally dideoxynucleotides, optionally ddATP being replaced by a ddATP analogue which is capable of acting as a substrate for a polymerase but incapable of acting as a substrate for a said PPi-detection enzyme. If the kit is for use with initial PCR amplification then it could also include the following components: (i) a pair of primers for PCR, at least one primer having means permitting immobilization of said primer; (ii) a polymerase which is preferably heat stable, for example Taq1 polymerase; (iii) buffers for the PCR reaction; and (iv) deoxynucleotides. Where an enzyme label is used to evaluate PCR, the kit will advantageously contain a substrate for the enzyme and other components of a detection system.

One embodiment of the invention is directed to a method for sequencing nucleic acids. The method involves fragmenting large template nucleic acid molecules to generate a plurality of fragmented nucleic acids. Then the fragmented nucleic acids are delivered into aqueous microreactors in a water-in-oil emulsion such that a plurality of aqueous microreactors comprise a single copy of a fragmented nucleic acid, a single bead capable of binding to the fragmented nucleic acid, and amplification reaction solution containing reagents necessary to perform nucleic acid amplification. In the next step, the fragmented

IDT Exhibit 1020.6595

nucleic acids is amplified in the microreactors to form amplified copies of the nucleic acids and binding the amplified copies to beads in the microreactors. Next, the beads are delivered to an array of at least 10,000 reaction chambers on a planar surface, wherein a plurality of the reaction chambers comprise no more than a single bead. Finally, a sequencing reaction is performed simultaneously on a plurality of the reaction chambers.

Another embodiment of the invention is directed to an array comprising a planar surface with a plurality of cavities thereon, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 20 to 100 μm and each cavity has a width in at least one dimension of between 20 μm and 70 μm. Further, there are at least 10,000 reaction chambers in the array. Each reaction chambers may contain at least 100,000 copies of a single species of single stranded nucleic acid template.

Another embodiment of the invention is directed to an array comprising a planar top surface and a planar bottom surface wherein the planar top surface has at least 10,000 cavities thereon, each cavity forming an analyte reaction chamber, the planar bottom surface is optically conductive such that optical signals from the reaction chambers can be detected through the bottom planar surface, wherein the distance between the top surface and the bottom surface is no greater than 5 mm, wherein the reaction chambers have a center to center spacing of between 20 to 100μm and each chamber having a width in at least one dimension of between 20μm and 70μm. The distance between the top surface and the bottom surface, in one embodiment, is no greater than 2 mm.

Another embodiment of the invention is directed to an array means for carrying out separate parallel common reactions in an aqueous environment. The array means may comprise a substrate comprising at least 10,000 discrete reaction chambers containing a starting material that is capable of reacting with a reagent, each of the reaction chambers being dimensioned such that when one or more fluids containing at least one reagent is delivered into each reaction chamber, the diffusion time for the reagent to diffuse out of the well exceeds the time required for the starting material to react with the reagent to form a product.

Another embodiment of the invention is directed to a method for delivering a bioactive agent to an array. The method comprises dispersing over the array a plurality of mobile solid supports, each mobile solid support having at least one reagent immobilized thereon, wherein the reagent is suitable for use in a nucleic acid sequencing reaction, where the array comprises a planar surface with a plurality of reaction chambers disposed thereon.

IDT Exhibit 1020.6596

The the reaction chambers may have a center to center spacing of between 20 to 100 μm and each reaction chamber has a width in at least one dimension of between 20 μm and 70 μm.

Another embodiment of the invention is directed to an apparatus for simultaneously monitoring an array of reaction chambers for light indicating that a reaction is taking place at a particular site. The apparatus comprises (a) an array of reaction chambers formed from a planar substrate comprising a plurality of cavitated surfaces, each cavitated surface forming a reaction chamber adapted to contain analytes, and wherein the reaction chambers have a center to center spacing of between 20 to 100 μm, each reaction chamber having a volume of between 10 to 150 pL, the array comprising more than 10,000 discrete reaction chambers; (b) an optically sensitive device arranged so that in use the light from a particular reaction chamber will impinge upon a particular predetermined region of the optically sensitive device; (c) means for determining the light level impinging upon each of the predetermined regions; and (d) means to record the variation of the light level with time for each of the reaction chamber.

Another embodiment of the invention is directed to an analytic sensor, comprising (a) an array formed from a first bundle of optical fibers with a plurality of cavitated surfaces at one end thereof, each cavitated surface forming a reaction chamber adapted to contain analytes, and wherein the reaction chambers have a center to center spacing of between 20 to 100 μm, a width of 20 to 70 μm, the array comprising more than 10,000 discrete reaction chambers; (b) an enzymatic or fluorescent means for generating light in the reaction chambers; and (c) light detection means comprising a light capture means and a second fiber optic bundle for transmitting light to the light detecting means, the second fiber optic bundle being in optical contact with the array, such that light generated in an individual reaction chamber is captured by a separate fiber or groups of separate fibers of the second fiber optic bundle for transmission to the light capture means.

Another embodiment of the invention is directed to a method for carrying out separate parallel common reactions in an aqueous environment. The first step involves delivering a fluid containing at least one reagent to an array, wherein the array comprises a substrate comprising at least 10,000 discrete reaction chambers, each reaction chamber adapted to contain analytes, and wherein the reaction chambers have a volume of between 10 to 150 pL and containing a starting material that is capable of reacting with the reagent, each of the reaction chambers being dimensioned such that when the fluid is delivered into each reaction chamber, the diffusion time for the reagent to diffuse out of the well exceeds the time

IDT Exhibit 1020.6597

required for the starting material to react with the reagent to form a product. The second step involves washing the fluid from the array in the time period (i) after the starting material has reacted with the reagent to form a product in each reaction chamber but (ii) before the reagent delivered to any one reaction chamber has diffused out of that reaction chamber into any other reaction chamber.

Another embodiment of the invention is directed to a method for delivering nucleic acid sequencing enzymes to an array. The array having a planar surface with a plurality of cavities thereon, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 20 to 100 μm. The method involves the step of dispersing over the array a plurality of mobile solid supports having one or more nucleic acid sequencing enzymes immobilized thereon, such that a plurality of the reaction chambers contain at least one mobile solid support.

Another embodiment of the invention is directed to a method for delivering a plurality of nucleic acid templates to an array. The array may have a planar surface with a plurality of cavities thereon, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 20 to 100 μm and the array having at least 10,000 reaction chambers. The method comprise the step of dispersing over the array a plurality of mobile solid supports, each mobile solid support having no more than a single species of nucleic acid template immobilized thereon, the dispersion causing no more than one mobile solid support to be disposed within any one reaction chamber.

Another embodiment of the invention is directed to a method for sequencing a nucleic acid. The method comprises the step of providing a plurality of single-stranded nucleic acid templates disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 20 to 100 μm and the planar surface has at least 10,000 reaction chambers. The next step involves performing a pyrophosphate based sequencing reaction simultaneously on all reaction chambers by annealing an effective amount of a sequencing primer to the nucleic acid templates and extending the sequencing primer with a polymerase and a predetermined nucleotide triphosphate to yield a sequencing product and, if the predetermined nucleotide triphosphate is incorporated onto the 3' end of the sequencing primer, a sequencing reaction byproduct. The third step involves identifying the sequencing reaction byproduct, thereby determining the sequence of the nucleic acid in each reaction chamber.

Another embodiment of the invention is directed to a method of determining the base

IDT Exhibit 1020.6598

sequence of a plurality of nucleotides on an array. The first step involves providing at least 10,000 DNA templates, each separately disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 20 to 100 μm, and a volume of between 10 to 150 pL. The second step involves adding an activated nucleotide 5'-triphosphate precursor of one known nitrogenous base to a reaction mixture in each reaction chamber, each reaction mixture comprising a template-directed nucleotide polymerase and a single-stranded polynucleotide template hybridized to a complementary oligonucleotide primer strand at least one nucleotide residue shorter than the templates to form at least one unpaired nucleotide residue in each template at the 3'-end of the primer strand, under reaction conditions which allow incorporation of the activated nucleoside 5'-triphosphate precursor onto the 3'-end of the primer strands, provided the nitrogenous base of the activated nucleoside 5'-triphosphate precursor is complementary to the nitrogenous base of the unpaired nucleotide residue of the templates. The third step involves detecting whether or not the nucleoside 5'-triphosphate precursor was incorporated into the primer strands in which incorporation of the nucleoside 5'-triphosphate precursor indicates that the unpaired nucleotide residue of the template has a nitrogenous base composition that is complementary to that of the incorporated nucleoside 5'-triphosphate precursor. The fourth step involves sequentially repeating steps (b) and (c), wherein each sequential repetition adds and, detects the incorporation of one type of activated nucleoside 5'-triphosphate precursor of known nitrogenous base composition. The fifth step involves determining the base sequence of the unpaired nucleotide residues of the template in each reaction chamber from the sequence of incorporation of the nucleoside precursors.

Another embodiment of the invention is directed to a method of identifying the base in a target position in a DNA sequence of template DNA. The first step involves providing at least 10,000 separate DNA templates are separately disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 20 to 100 μm, the DNA being rendered single stranded either before or after being disposed in the reaction chambers. The second step involves providing an extension primer is provided which hybridizes to the immobilized single-stranded DNA at a position immediately adjacent to the target position. The immobilized single-stranded DNA is subjected to a polymerase reaction in the presence of a predetermined deoxynucleotide or dideoxynucleotide, wherein if the predetermined deoxynucleotide or dideoxynucleotide is incorporated onto the 3' end of the sequencing

IDT Exhibit 1020.6599

primer then a sequencing reaction byproduct is formed. The fourth step involves identifying the sequencing reaction byproduct, thereby determining the nucleotide complementary to the base at the target position in each of the 10,000 DNA templates.

Another embodiment of the invention is directed to an apparatus for analyzing a nucleic acid sequence. The apparatus comprises: (a) a reagent delivery cuvette, wherein the cuvette includes an array comprising a planar surface with a plurality of cavities thereon, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 20 to 100 μm, and there are in excess of 10,000 reaction chambers, and wherein the reagent delivery cuvette contains reagents for use in a sequencing reaction; (b) a reagent delivery means in communication with the reagent delivery cuvette; (c) an imaging system in communication with the reagent delivery chamber; and (d) a data collection system in communication with the imaging system.

Another embodiment of the invention is directed to an apparatus for determining the base sequence of a plurality of nucleotides on an array. The apparatus comprises: (a) a reagent cuvette containing a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein there are in excess of 10,000 reaction chambers, each having a center to center spacing of between 20 to 100 μm and a volume of between 10 to 150 pL; (b) reagent delivery means for simultaneously adding to each reaction chamber an activated nucleotide 5'-triphosphate precursor of one known nitrogenous base to a reaction mixture in each reaction chamber, each reaction mixture comprising a template-directed nucleotide polymerase and a single-stranded polynucleotide template hybridized to a complementary oligonucleotide primer strand at least one nucleotide residue shorter than the templates to form at least one unpaired nucleotide residue in each template at the 3'-end of the primer strand, under reaction conditions which allow incorporation of the activated nucleoside 5'-triphosphate precursor onto the 3'-end of the primer strands, provided the nitrogenous base of the activated nucleoside 5'-triphosphate precursor is complementary to the nitrogenous base of the unpaired nucleotide residue of the templates; (c) detection means for detecting in each reaction chamber whether or not the nucleoside 5'-triphosphate precursor was incorporated into the primer strands in which incorporation of the nucleoside 5'-triphosphate precursor indicates that the unpaired nucleotide residue of the template has a nitrogenous base composition that is complementary to that of the incorporated nucleoside 5'-triphosphate precursor; and (d) means for sequentially repeating (b) and (c), wherein each sequential repetition adds and, detects the incorporation of one type of activated nucleoside

IDT Exhibit 1020.6600

5'-triphosphate precursor of known nitrogenous base composition; and (e) data processing means for determining the base sequence of the unpaired nucleotide residues of the template simultaneously in each reaction chamber from the sequence of incorporation of the nucleoside precursors.

Another embodiment of the invention is directed to an apparatus for processing a plurality of analytes. The apparatus comprises: (a) a flow chamber having disposed therein a substrate comprising at least 50,000 cavitated surfaces on a fiber optic bundle, each cavitated surface forming a reaction chamber adapted to contain analytes, and wherein the reaction chambers have a center to center spacing of between 20 to 100 μm and a diameter of 20 to 70 μm; (b) fluid means for delivering processing reagents from one or more reservoirs to the flow chamber so that the analytes disposed in the reaction chambers are exposed to the reagents; and (c) detection means for simultaneously detecting a sequence of optical signals from each of the reaction chambers, each optical signal of the sequence being indicative of an interaction between a processing reagent and the analyte disposed in the reaction chamber, wherein the detection means is in communication with the cavitated surfaces.

Another embodiment of the invention is directed to a method for sequencing a nucleic acid. The first step involves providing a plurality of single-stranded nucleic acid templates in an array having at least 50,000 discrete reaction sites. The second step involves contacting the nucleic acid templates with reagents necessary to perform a pyrophosphate-based sequencing reaction coupled to light emission. The third step involves detecting the light emitted from a plurality of reaction sites on respective portions of an optically sensitive device. The forth step involves converting the light impinging upon each of the portions of the optically sensitive device into an electrical signal which is distinguishable from the signals from all of the other reaction sites. The fifth step involves determining the sequence of the nucleic acid templates based on light emission for each of the discrete reaction sites from the corresponding electrical signal.

Another embodiment of the invention is directed to a method for sequencing nucleic acids. The first step involves fragmenting large template nucleic acid molecules to generate a plurality of fragmented nucleic acids. The second step involves attaching one strand of a plurality of the fragmented nucleic acids individually to beads to generate single stranded nucleic acids attached individually to beads. The third step involves delivering a population of the single stranded fragmented nucleic acids attached individually to beads to an array of at least 10,000 reaction chambers on a planar surface, wherein a plurality of the wells

IDT Exhibit 1020.6601

comprise no more than a one bead with on single stranded fragmented nucleic acid. The fourth step involves performing a sequencing reaction simultaneously on a plurality of the reaction chambers. The sequencing reaction may have the steps of (a) annealing an effective amount of a sequencing primer to the single stranded fragmented nucleic acid templates and extending the sequencing primer with a polymerase and a predetermined nucleotide triphosphate to yield a sequencing product and, if the predetermined nucleotide triphosphate is incorporated onto the 3' end of the sequencing primer, a sequencing reaction byproduct; and (b) identifying the sequencing reaction byproduct, thereby determining the sequence of the nucleic acid in a plurality of the reaction chambers. Alternatively, the sequencing reaction may comprises the steps of: (a) hybridizing two or more sequencing primers to one or a plurality of single strands of the nucleic acid molecule wherein all the primers except for one are reversibly blocked primers; (b) incorporating at least one base onto the nucleic acid molecule by polymerase elongation from an unblocked primer; (c) preventing further elongation of the unblocked primer; (d) deblocking one of the reversibly blocked primers into an unblocked primer; and (e) repeating steps (b) to (d) until at least one of the reversibly blocked primers are deblocked and used for determining a sequence.

Other materials and methods may be found in the following copending US patent applications: USSN 60/443,471 filed January 29, 2003 and USSN 60/465.071 filed April 23, 2003. All patents, patent applications, and references cited in this disclosure are incorporated herein by reference.

IDT Exhibit 1020.6602

EXAMPLES

Example 1: Sample Preparation

DNA Sample:

The DNA should be of high quality and free from contaminants such as proteins, nucleases, lipids, and other chemicals (such as residual EDTA from preparation) and salts. It is preferred that genomic DNA should have a 260/280 ratio of 1.8 or higher. If it is desired to sequence the genome of only one organism, then the DNA should be quality checked to ensure that there is no contaminating DNA. For example: a preparation of human DNA may be checked by PCR to ensure that it is not contaminated by bacterial DNA molecules. Another method of checking for contamination is by restriction digestion patterns and especially restriction digestion followed by Southern Blot using suitable probes known to be specific for an organism (e.g., human or mouse) and a second probe known to be specific for a possible contaminating organism (e.g., *E. coli*). If it is desired, the DNA should originate from a single clone of the organism (e.g., a colony if from bacteria).

Step 1: DNase I Digestion

The purpose of the DNase I digestion step is to fragment a large stretch of DNA such as a whole genome or a large portion of a genome into smaller species. This population of smaller-sized DNA species generated from a single DNA template is referred to as a "library". Deoxyribonuclease I (DNase I) is an endonuclease which cleaves double-stranded template DNA. The cleavage characteristics of DNase I allow random digestion of template DNA (i.e., minimal sequence bias) and will result in the predominance of blunt-ended, double-stranded DNA fragments when used in the presence of manganese-based buffers (Melgar and Goldthwait 1968). The digestion of genomic templates by DNase I is dependent on three factors: i) quantity of enzyme used (units); ii) temperature of digestion (°C); and iii) incubation time (minutes). The DNase I digestion conditions outlined below were optimized to yield DNA libraries in a size range from 50-700 base pairs (bp).

1. DNA was obtained and prepared to a concentration of 0.3 mg/ml in Tris-HCl (10mM, pH 7-8). A total of 134 µl of DNA (15 µg) was needed for this preparation. It is recommended to not use DNA preparations diluted with buffers containing EDTA (i.e., TE, Tris/EDTA). The presence of EDTA is inhibitory to enzyme digestion with DNase I. If the DNA preparation contains EDTA, it is important that the DNA be "salted" out of solution and reconstituted with the appropriate Tris-HCl buffer (10 mM, pH 7-8) or nanopure H_2O (pH 7-8).

IDT Exhibit 1020.6603

2. In a 0.2 ml tube, DNase I Buffer, comprising 50 µl Tris pH 7.5 (1M), 10 µl $MnCl_2$ (1M), 1 µl BSA (100 mg/ml), and 39 µl water was prepared.

3. In a separate 0.2 ml tube, 15 µl of DNase I Buffer and 1.5 µl of DNase I (1U/ml) was added. The reaction tube was placed in a thermal cycler set to 15°C.

4. The 134 µl of DNA (0.3 mg/ml) was added to the DNase I reaction tube placed in the thermal cycler set at 15°C. The lid was closed and the sample was incubated for exactly 1 minute. Following incubation, 50 µl of 50 mM EDTA was added to stop the enzyme digestion.

5. The digested DNA was purified by using the QiaQuick PCR purification kit. The digestion reaction was then split into four aliquots, and four spin columns were used to purify each aliquot (37.5 µl per spin column). Each column was eluted with 30 µl elution buffer (EB) according to the manufacturer's protocol. The eluates were then combined to generate a final reaction volume of 120 µl.

6. One 3 µl aliquot of the digestion reaction was saved for analysis using a BioAnalzyer DNA 1000 LabChip.

Step 2: Pfu Polishing

Digestion of DNA templates with DNase I yields fragments of DNA that are primarily blunt-ended, however, some fragments will have ends that contain protruding termini that are one or two nucleotides in length. *Pfu* polishing is used to increase the amount of blunt-ended species by fill-in (i.e., "blunting") of 5' overhangs. Additionally, *Pfu* DNA polymerase has 3'→ 5' exonuclease activity that will result in the removal of single and double nucleotide extensions. *Pfu* polishing increases the amount of blunt-ended DNA fragments available for adaptor ligation (Costa 1994a, 1994b, 1994c). The following *Pfu* polishing protocol was used.

1. In a 0.2 ml tube, 115 µl purified, DNase I-digested DNA fragments, 15 µl 10X Cloned *Pfu* buffer, 5 µl dNTPs (10 mM), and 15 µl cloned *Pfu* DNA polymerase (2.5 U/µl) were added in order.

2. The polishing reaction components were mixed well and incubated at 72° C for 30 minutes.

3. Following incubation, the reaction tube was removed and placed on ice for 2 minutes.

4. The polishing reaction mixture was then split into four aliquots and

IDT Exhibit 1020.6604

purified using QiaQuick PCR purification columns (37.5 μL on each column). Each column was eluted with 30 μl buffer EB according to the manufacturer's protocol. The eluates were then combined to generate a final reaction volume of 120 μL.

5. One 3 μl aliquot of the final polishing reaction was saved for analysis using a BioAnalzyer DNA 1000 LabChip.

Step 3: Ligation of Universal Adaptors to Fragmented DNA Library

Following fragmentation and polishing of the genomic DNA library, primer sequences are added to the ends of each DNA fragment. These primer sequences are termed "Universal Adaptors" and are comprised of double-stranded oligonucleotides that contain specific priming regions that afford both PCR amplification and nucleotide sequencing. The Universal Adaptors are designed to include a set of unique PCR priming regions that are 20 base pairs in length located adjacent to a set of unique sequencing priming regions that are 20 base pairs in length, followed by a unique 4-base "key" consisting of one of each deoxyribonucleotide (i.e., A, C, G, T). Each unique Universal Adaptor (termed "Universal Adaptor A" and "Universal Adaptor B") is forty-four base pairs (44 bp) in length. Universal Adaptors are ligated, using T4 DNA ligase, onto each end of the DNA fragment to generate a total nucleotide addition of 88 bp to each DNA fragment. Different Universal Adaptors are designed specifically for each genomic DNA library preparation and will therefore provide a unique identifier for each organism.

To prepare a pair of Universal Adaptors, single-stranded oligonucleotides are designed in-house and are manufactured through a commercial vendor. Universal Adaptor DNA oligonucleotides are designed with two phosphorothioate linkages at each oligonucleotide end that serve to protect against nuclease activity (Samini, T.D., B. Jolles, and A. Laigle. 2001. Best minimally modified antisense oligonucleotides according to cell nuclease activity. *Antisense Nucleic Acid Drug Dev.* 11(3):129., the disclosure of which is incorporated *in toto* herein by reference.). Each oligonucleotide is HPLC-purified to ensure there are no contaminating or spurious DNA oligonucleotide sequences in the final prep.

The Universal Adaptors are designed to allow directional ligation to the blunt-ended, fragmented genomic DNA. For each Universal Adaptor pair, the PCR priming region contains a 5' four-base overhang and a blunt-ended 3' Key region. Directionality is achieved as the blunt-end side of the Universal Adaptor ligates to the blunt-ended DNA fragment while the 5' overhang of the adaptor cannot ligate to the blunt-ended DNA fragment. Additionally,

IDT Exhibit 1020.6605

a 5' biotin is added to the Universal Adaptor B to allow subsequent isolation of ssDNA template (Step 8). Each Universal Adaptor is prepared by annealing, in a single tube, the two single-stranded complementary DNA oligonucleotides (i.e., one oligo containing the sense sequence and the second oligo containing the antisense sequence). The following ligation protocol was used.

1. In a 0.2 ml tube, 39 µl nH_2O (molecular biology grade water), 25 µl digested, polished DNA Library, 100 µl 2X Quick Ligase Reaction Buffer, 20 µl MMP1 (10 pm/µl) adaptor set, 100:1 ratio, and 16 µl Quick Ligase were added in order. The ligation reaction was mixed well and incubated at RT for 20 minutes.

2. The ligation reaction was then removed and a 10-µl aliquot of the ligation reaction was purified for use on the BioAnalyzer. A single spin column from the Qiagen Min-Elute kit was used. The column was eluted with 10 µl EB according to the procedure per manufacturers' protocol. A 1-µl aliquot of the purified ligation reaction was loaded using a BioAnalyzer DNA 1000 LabChip. This purification step is recommended as the unpurified ligation reaction contains high amounts of salt and PEG that will inhibit the sample from running properly on the BioAnalyzer.

3. The remainder of the ligation reaction (190 µL) was used for gel isolation in Step 4.

Step 3a: Microcon Filtration and Adaptor Construction. Total preparation time was approximately 25 min.

The Universal Adaptor ligation reaction requires a 100-fold excess of adaptors. To aid in the removal of these excess adaptors, the double-stranded gDNA library is filtered through a Microcon YM-100 filter device. Microcon YM-100 membranes can be used to remove double stranded DNA smaller than 125 bp. Therefore, unbound adaptors (44 bp), as well as adaptor dimers (88 bp) can be removed from the ligated gDNA library population. The following filtration protocol was used:

1. The 190 µL of the ligation reaction from Step 4 was applied into an assembled Microcon YM-100 device.

2. The device was placed in a centrifuge and spun at 5000 x g for approximately 6 minutes, or until membrane was almost dry.

3. To wash, 200 µl of 1X TE was added.

4. Sample was spun at 5000 x g for an additional 9 minutes, or until membrane

IDT Exhibit 1020.6606

was almost dry.

5. To recover, the reservoir was inserted into a new vial and spun at 3000 x g for 3 minutes. The reservoir was discarded. The recovered volume was approximately 10 µl. Next, 80 µl TE was added.

The Adaptors (A and B) were HPLC-purified and modified with phosphorothioate linkages prior to use. For Adaptor "A" (10 µM), 10 µl of 100 µM Adaptor A (44 bp, sense) was mixed with 10 µl of 100 µM Adaptor A (40 bp, antisense), and 30 µl of 1X Annealing Buffer (V_f = 50 µl) were mixed. The primers were annealed using the ANNEAL program on the Sample Prep Labthermal cycler (see below). For Adaptor "B" (10 µM), 10 µl of 100 µM Adaptor B (40 bp, sense) was mixed with 10 µl of 100 µM Adaptor B (44 bp, antisense), and 30 µl of 1X Annealing Buffer (V_f = 50 µl). The primers were annealed using the ANNEAL program on the Sample Prep Lab thermal cycler. Adaptor sets could be stored at -20°C until use.

ANNEAL-A program for primer annealing:

1. Incubate at 95°C, 1 min;
2. Decrease temperature to 15°C, at 0.1°C/sec; and
3. Hold at 15°C.

There was no orientation required for the genomic DNA insert fragment and the adaptors. Fragments could be ligated at either end. Four single-stranded DNA oligonucleotides were included in the Universal Adaptor set. Each single-stranded oligonucleotide was synthesized at 1 micromole scale and HPLC-purified. Each single-stranded oligonucleotide included four phosphorothioate linkages at each end.

Step 4: Gel Electrophoresis and Extraction of Adapted DNA Library

The Universal Adaptor ligation protocol produces the following: 1) fragmented DNAs with adaptors on either end; 2) unbound single adaptors; or 3) the formation of adaptor dimers. Agarose gel electrophoresis is used as a method to separate and isolate the adapted DNA library population from the unligated, single adaptors and adaptor dimer populations. The procedure of DNase I digestion of genomic DNA yields a library population that ranges from 50-700 bp (Step 1). The addition of the 88-bp Universal Adaptor set will shift the population to a larger size and will result in a migration profile in the size range of approximately 130-800 bp. Adaptor dimers will migrate at 88 bp and adaptors unligated will migrate at 44 bp. Therefore, genomic DNA libraries in size ranges > 200 bp can be

IDT Exhibit 1020.6607

physically isolated from the agarose gel and purified using standard gel extraction techniques. Gel isolation of the adapted DNA library will result in the recovery of a library population in a size range that is ≥200 bp (size range of library can be varied depending on application). The following electrophoresis and extraction protocol was used.

1. A 2% agarose gel was prepared.

2. 10 µl of 10X Ready-Load Dye was added to the remaining 90 µl of the DNA ligation mixture.

3. The dye/ligation reaction mixture was loaded into the gel using four adjacent lanes (25 µl per lane).

4. 10 µl of the100 bp ladder (0.1 µg/µl) was loaded two lanes away from ligation reaction lanes.

5. The gel was run at 100V for 3 hours.

6. When the gel run was complete, the gel was removed from the gel box and transferred to a flat surface covered with plastic wrap. DNA bands were visualized using a hand-held long-wave UV light. Using a sterile, single-use scalpel, the fragment sizes of 200 – 400 bp were cut out from the agarose gel. Using this approach, libraries with any size range can be isolated. It is also possible to isolate more than one size range. Where the library size range is 200-900 bp, it is possible to isolate several size ranges from a single well (i.e., 200-400 bp and 500-700 bp).

7. The DNA embedded in the agarose gel was isolated using a Qiagen MinElute Gel Extraction kit following the manufacturer's instructions. Briefly, Buffer QG was added to cover the agarose in the tube. The agarose was allowed to completely dissolve. The color of the Buffer QG was maintained by adjusting the pH according to the Qiagen instructions to minimize sample loss. Two MinElute spin columns (Qiagen) were used for purification. The large volume of dissolved agarose required each column to be loaded several times. The columns were eluded with 10 µl of Buffer EB which was pre-warmed at 55°C. The eluates were pooled to produce 20 µl of gDNA library.

8. One 1 µL aliquot of each isolated DNA library was analyzed using a BioAnalyzer DNA 1000 LabChip to assess the exact distribution of the DNA library population.

Step 5: Strand Displacement and Extension of Nicked Double Stranded DNA Library

Because the DNA oligonucleotides used for the Universal Adaptors are not

IDT Exhibit 1020.6608

phosphorylated, gaps are present at the 3' junctions of the fragmented gDNAs. These two "gaps" or "nicks" can be filled in by using a strand displacing DNA polymerase. The polymerase recognizes nicks, displaces the nicked strands, and extends the strand in a manner that results in repair of nicks and in the formation of non-nicked double-stranded DNA. The strand displacing enzyme used is the large fragment of *Bst* DNA polymerase.

1. In a 0.2 ml tube, 19 µl gel-extracted DNA library, 40 µl nH_2O, 8 µl 10X ThermoPol Reaction Buffer, 8 µl BSA (1 mg/ml), 2 µl dNTPs (10 mM), and 3 µl *Bst* I Polymerase (8 U/µl) were added in order.

2. The samples were mixed well and placed in a thermal cycler and incubated using the Strand Displacement incubation program: "BST". BST program for stand displacement and extension of nicked double-stranded DNA:

    1. Incubate at 65° C, 30 minutes;
    2. Incubate at 80° C, 10 minutes;
    3. Incubate at 58° C, 10 minutes; and
    4. Hold at 14° C.

3. One 1 µL aliquot of the *Bst*-treated DNA library was run using a BioAnalyzer DNA 1000 LabChip.

Step 6: Preparation of Streptavidin Beads

Following the generation of unnicked double-stranded genomic DNA, it is necessary to isolate single-stranded genomic DNAs containing flanking Universal Adaptor sequences. This step outlines the binding of biotin-tagged double-stranded DNA to streptavidin beads. For preparing streptavidin beads, the following protocol was used.

1. 100 µl Dynal M-270 Streptavidin beads were washed two times with 200 µl of 1X Binding Buffer (1 M NaCl, 0.5 mM EDTA, 5 mM Tris, pH 7.5) by applying the magnetic beads to the MPC.

2. The beads were resuspended in 100 µl 2X Binding buffer, then the remaining 79 µl of the *Bst*-treated DNA sample (from Step 5) and 20 µl water was added.

3. The bead solution was mixed well and placed on a tube rotator at RT for 20 minutes. The bead mixtures were washed, using the MPC, two times with 100 µl of 1X Binding Buffer, then washed two times with nH_2O. Binding & Washing (B&W) Buffer (2X and 1X): 2X B&W buffer was prepared by mixing 10 mM Tris•HCl (pH 7.5), 1 mM EDTA, and 2 M NaCl. The reagents were combined as listed above and mixed thoroughly. The

IDT Exhibit 1020.6609

solution can be stored at RT for 6 months; 1X B&W buffer was prepared by mixing 2X B&W buffer with nH_2O, 1:1. The final concentrations were half the above, i.e., 5 mM Tris•HCl (pH 7.5), 0.5 mM EDTA, and 1 M NaCl.

Step 7: Isolation of single-stranded DNA Library using Streptavidin Beads

Following binding of the double-stranded gDNA library to streptavidin beads, it is preferred to isolate from the ligated pool only the single-stranded gDNAs containing Universal Adaptor A and Universal Adaptor B (desired populations are designated below with asterisks). Double-stranded genomic DNA fragment pools will have adaptors bound in the following possible configurations:

Universal Adaptor A- gDNA Fragment -Universal Adaptor A

Universal Adaptor B- gDNA Fragment -Universal Adaptor A*

Universal Adaptor A- gDNA Fragment -Universal Adaptor B*

Universal Adaptor B- gDNA Fragment -Universal Adaptor B

Because only the Universal Adaptor B has a 5' biotin moiety, magnetic streptavidin-containing beads can be used to bind all gDNA library species that possess the Universal Adaptor B. Genomic library populations that contain two Universal Adaptor A species (or nonligated species) do not bind to streptavidin-containing beads and are removed during the wash procedure. The species that remain bound to bead after washing include those with Universal Adaptors A and B or those with two Universal Adaptor B ends.

Genomic DNA species with two Universal Adaptor B sequences with two biotin molecules can bind to the streptavidin-containing beads at both ends. Species with A and B adaptors having only a single biotin molecule can bind to the beads only at the "B" end. To isolate the single-stranded population, the bead-bound double-stranded DNA is treated with a sodium hydroxide solution that serves to disrupt the hydrogen bonding between the complementary DNA strands. If the DNA fragment has biotin on each end (Universal Adaptor B ends), both resulting single strands remain bound to the beads. If the fragment has only a single biotin (Universal Adaptors A and B), then the complementary strand separates from the DNA-bead complex.

The resulting single-stranded genomic DNA library is collected from the solution phase and is quantitated, e.g., using pyrophosphate sequencing (PyroSequence) or by using a RNA Pico 6000 LabChip (Agilent, Palo Alto, CA). Single-stranded genomic DNA libraries are quantitated by calculating the number of molecules per unit volume. Single-stranded gDNA molecules are then annealed (at a half copy per bead to obtain one effective copy

IDT Exhibit 1020.6610

per bead) to 25-30 μm sepharose beads containing DNA capture primers (PCR primer B). The templates are then amplified using emulsion polymerase chain reaction protocols. Subsequent sequencing may be conducted using known techniques. For isolation of the single stranded library, the following protocol was used.

1. 250 μl Melt Solution (0.125 M NaOH, 0.1 M NaCl)was added to washed beads from Step 6 above.

2. The bead solution was mixed well and the bead mixture was incubated at room temperature for 10 minutes on a tube rotator.

3. A Dynal MPC (magnetic particle concentrator) was used, the pellet beads were carefully removed, and the supernatant was set aside. The 250-μl supernatant included the single-stranded DNA library.

4. In a separate tube, 1250 μl PB (from QiaQuick Purification kit) was added and the solution was neutralized by adding 9 μl of 20% acetic acid.

5. Using a Dynal MPC, beads from the 250-μl supernatant including the single-stranded gDNA library were pelleted and the supernatant was carefully removed and transferred to the freshly prepared PB/acetic acid solution.

6. The 1500 μl solution was purified using a single QiaQuick purification spin column (load sample through same column two times at 750 μl per load). The single-stranded DNA library was eluted with 50 μl EB.

Step 8a: Single-stranded gDNA Quantitation using Pyrophosphate Sequencing. Total preparation time was approximately 1 hr.

1. In a 0.2 ml tube, the following reagents were added in order:

25 μl single-stranded gDNA
1 μl MMP2B sequencing primer
14 μl Library Annealing Buffer
40 μl total

2. The DNA was allowed to anneal using the ANNEAL-S Program (see Appendix, below).

3. The samples were run on PSQ (pyrophosphate sequencing jig) to determine the number of picomoles of template in each sample (see below). Methods of sequencing can be found in U.S. Patent 6,274,320; U.S. Patent 4,863,849; U.S. Patent 6,210,891; and U.S.

IDT Exhibit 1020.6611

Patent 6,258,568, the disclosures of which are incorporated *in toto* herein by reference. Calculations were performed to determine the number of single-stranded gDNA template molecules per microliter. The remaining 25 μL of prepared single-stranded gDNA library was used for amplification and subsequent sequencing (approximately 1 x 10^6 reactions).

Step 8b: Single-stranded gDNA Quantitation using RNA Pico 6000 LabChip. Total preparation time was approximately 30 minutes.

1. The mRNA Pico assay option was selected on the BioAnalyzer (Software version 2.12).

2. An RNA Pico 6000 LabChip was prepared on the BioAnalyzer according to the manufacturers' guidelines.

3. An RNA LabChip ladder (RNA 6000 ladder) was prepared according to manufacturer's (Ambion) directions. Briefly, the RNA LabChip ladder, in solution, was heated to 70°C for 2 minutes. The solution was chilled on ice for 5 minutes to snap cool the ladder. The solution was briefly centrifuged to clear any condensate from tube walls. The RNA LabChip Ladder was stored on ice and used within one day.

4. The ssDNA library to be analyzed was run in triplicate, in adjacent lanes, using three 1 μl aliquots.

5. The BioAnalyzer software was used to calculate the concentration of each ssDNA library lane (see the Table below and Figure 24. The average of all three lanes was used to calculate the DNA concentration of the library using the procedure outlined below.

   a. The peak integration lower limit line (large dash in Figure 24) was moved immediately in front of the library peak (see below).

   b. The peak integration upper limit line (large dash in the Figure 24) was moved immediately after the library peak. In this way, the peak integration line connecting the lower and upper integration lines followed the slope of the background.

   c. The mouse arrow was used to determine the average size of the peak in bases (usually near the peaks highest point) or a defined peak was used as chosen by the software.

   d. The integrated value was used for the amount of material in the peak. The value obtained for picograms recovered was converted into molecules recovered (see Table, below). The library concentration was then determined (molecules per microliter).

IDT Exhibit 1020.6612

Table

1	2	3	4	5	6	7	8	9
	pg/μL (1)	pg/μL (2)	pg/μL (3)	Average pg/μL	Mean Size (bp) 1	Mean Size (bp) 2	Mean Size (bp) 3	Average Size (bp)
sample	1633	1639	1645	1639	435	435	432	434

10	11	12	13	14	15
Ave MW (g/mole) Ribonucleotide	Ave MW (g/mole)	Library g/μL	moles/g	moles/μL	molecules/μL
328.2	1.42E+05	1.64E-09	7.02E-06	1.15E-14	6.93E+09

As shown in the Table above, the concentration of Library 1 was calculated as 1639 pg/μl (Column 5) and the average fragment size was 434 nucleotides (Column 9). These values were obtained from the Agilent 2100 software as described in Steps (a)-(d), above. The average molecular weight (MW) of a ribonucleotide is 328.2 g/mole (Column 10). The MW of the average library fragment (1.42×10^5 g/mole, Column 11) was calculated by multiplying the average fragment length (434) by the average ribonucleotide (328.2). The quantitated library (1639 pg/μl) was converted to grams per microliter (1.64×10^{-9} g/μl, Column 12). The number of moles per microliter (1.15×10^{-14} moles/μl, Column 14) was calculated by dividing the grams per microliter (1.64×10^{-9} g/μl, Column 12) by the average molecular weight of the library fragments (1.42×10^5, Column 11). Finally, the number of molecules per microliter (6.93×10^9 molecules/μl, Column 15) was derived by multiplying the number of moles per microliter (1.15×10^{-14} moles/μl, Column 14) by Avogadro's number (6.02×10^{23} molecules/mole).

The final library concentration was expected to be greater than 1×10^8 molecules/μl. A more important factor for library quality was adaptor dimer concentration. In Figure 24, the height of the library peak was determined approximately 10 fold greater than the adaptor dimer peak (the first peak after the marker). A library of good quality is expected to have a peak height at least 2 fold greater than the dimer peak. It should be noted that the RNA Pico 6000 LabChip provided estimates within 500% accuracy of the single-stranded gDNA concentration. Thus, it was important to perform an initial sequencing run using a titration of template to determine the number of copies per bead (cpb) of input gDNA. The recommended input DNA is 2.5 cpb, 1 cpb, 0.5 cpb, and 0.1 cpb. This titration was easily checked using the 4slot bead loading chamber on a 14 x 43 PTP.

IDT Exhibit 1020.6613

Step 9: Dilution and Storage of Single-Stranded gDNA library

The single-stranded gDNA library was eluted and quantitated in Buffer EB. To prevent degradation, the single-stranded gDNA library was stored frozen at -20°C in the presence of EDTA. After quantitation, an equal volume of 10 mM TE was added to the library stock. All subsequent dilutions was in TE. The yield was as follows:

Remaining final volume of ssDNA library following PSQ analysis = 25 µl.

Remaining final volume of ssDNA library following LabChip analysis = 47 µl.

For the initial stock dilution, single-stranded gDNA library was diluted to 100 million molecules/µl in 1X Library-Grade Elution Buffer. Aliquots of single-stranded gDNA library were prepared for common use. For this, 200,000 molecules/µl were diluted in 1X Library-Grade Elution Buffer and 20 µl aliquots were measured. Single-use library aliquots were stored at –20°C.

Step 10: Emulsion Polymerase Chain Reaction

Where increased numbers of cpb were preferred, bead emulsion PCR was performed as described in U.S. Patent Application Serial No. 06/476,504 filed June 6, 2003, incorporated herein by reference in its entirety.

Reagent preparation

The Stop Solution (50 mM EDTA) included 100 µl of 0.5 M EDTA mixed with 900 µl of nH_2O to obtain 1.0 ml of 50 mM EDTA solution. For 10 mM dNTPs, (10 µl dCTP (100 mM), 10 µl dATP (100 mM), 10 µl dGTP (100 mM), and 10 µl dTTP (100 mM) were mixed with 60 µl molecular biology grade water. All four 100 mM nucleotide stocks were thawed on ice. Then, 10 µl of each nucleotide was combined with 60 µl of nH_2O to a final volume of 100 µl, and mixed thoroughly. Next, 1 ml aliquots were dispensed into 1.5 ml microcentrifuge tubes. The stock solutions could be stored at -20°C for one year.

The 10 X Annealing buffer included 200 mM Tris (pH 7.5) and 50 mM magnesium acetate. For this solution, 24.23 g Tris was added to 800 ml nH_2O and the mixture was adjusted to pH 7.5. To this solution, 10.72 g of magnesium acetate was added and dissolved completely. The solution was brought up to a final volume of 1000 ml and could be stored at 4°C for 1 month. The 10 X TE included 100 mM Tris•HCl (pH 7.5) and 50 mM EDTA. These reagents were added together and mixed thoroughly. The solution could be stored at

IDT Exhibit 1020.6614

room temperature for 6 months.

Example 2: Primer Design

As discussed above, the universal adaptors are designed to include: 1) a set of unique PCR priming regions that are typically 20 bp in length (located adjacent to (2)); 2) a set of unique sequencing priming regions that are typically 20 bp in length; and 3) optionally followed by a unique discriminating key sequence consisting of at least one of each of the four deoxyribonucleotides (i.e., A, C, G, T). The probability of cross-hybridization between primers and unintended regions of the genome of interest is increased as the genome size increases and length of a perfect match with the primer decreases. However, this potential interaction with a cross-hybridizing region (CHR) is not expected to produce problems for the reasons set forth below.

In a preferred embodiment of the present invention, the single-stranded DNA library is utilized for PCR amplification and subsequent sequencing. Sequencing methodology requires random digestion of a given genome into 150 to 500 base pair fragments, after which two unique bipartite primers (composed of both a PCR and sequencing region) are ligated onto the 5' and 3' ends of the fragments (Figure 25). Unlike typical PCR amplifications where an existing section of the genome is chosen as a priming site based on melting temperature (T_m), uniqueness of the priming sequence within the genome and proximity to the particular region or gene of interest, the disclosed process utilizes synthetic priming sites that necessitates careful *de novo* primer design.

Tetramer Selection:

Strategies for *de novo* primer design are found in the published literature regarding work conducted on molecular tags for hybridization experiments (see, Hensel, M. and D.W. Holden, Molecular genetic approaches for the study of virulence in both pathogenic bacteria and fungi. Microbiology, 1996. 142(Pt 5): p. 1049-58; Shoemaker, D.D., et al., Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. Nat Genet, 1996. 14(4): p. 450-6) and PCR/LDR (polymerase chain reaction/ligation detection reaction) hybridization primers (see, Gerry, N.P., et al., Universal DNA microarray method for multiplex detection of low abundance point mutations. Journal of Molecular Biology, 1999. 292: p. 251-262; Witowski, N.E., et al., Microarray-based detection of select cardiovascular disease markers. BioTechniques, 2000. 29(5): p. 936-944.).

IDT Exhibit 1020.6615

The PCR/LDR work was particularly relevant and focused on designing oligonucleotide "zipcodes", 24 base primers comprised of six specifically designed tetramers with a similar final T_m. (see, Gerry, N.P., et al., Universal DNA microarray method for multiplex detection of low abundance point mutations. Journal of Molecular Biology, 1999. 292: p. 251-262; U.S. Pat. No. 6,506,594). Tetrameric components were chosen based on the following criteria: each tetramer differed from the others by at least two bases, tetramers that induced self-pairing or hairpin formations were excluded, and palindromic (AGCT) or repetitive tetramers (TATA) were omitted as well. Thirty-six of the 256 (4^4) possible permutations met the necessary requirements and were then subjected to further restrictions required for acceptable PCR primer design (Table 1).

IDT Exhibit 1020.6616

Table 1

	TT	TC	TG	TA	CT	CC	CG	CA	GT	GC	GG	GA	AT	AC	AG	AA
TT	TTTT	TTTC	TTTG	TTTA	TTCT	TTCC	TTCG	TTCA	TTGT	TTGC	TTGG	TTGA	TTAT	TTAC	TTAG	*TTAA*
TC	TCTT	TCTC	TCTG	TCTA	TCCT	TCCC	TCCG	TCCA	TCGT	TCGC	TCGG	*TCGA*	TCAT	TCAC	TCAG	TCAA
TG	TGTT	TGTC	TGTG	TGTA	TGCT	TGCC	TGCG	*TGCA*	TGGT	TGGC	TGGG	TGGA	TGAT	TGAC	TGAG	TGAA
TA	TATT	TATC	TATG	TATA	TACT	TACC	TACG	TACA	TAGT	TAGC	TAGG	TAGA	*TAAT*	TAAC	TAAG	TAAA
CT	CTTT	CTTC	CTTG	CTTA	CTCT	CTCC	CTCG	CTCA	CTGT	CTGC	CTGG	CTGA	CTAT	CTAC	*CTAG*	CTAA
CC	CCTT	CCTC	CCTG	CCTA	CCCT	CCCC	CCCG	CCCA	CCGT	CCGC	CCGG	CCGA	CCAT	CCAC	CCAG	CCAA
CG	CGTT	CGTC	CGTG	CGTA	CGCT	CGCC	CGCG	CGCA	CGGT	CGGC	CGGG	CGGA	CGAT	CGAC	CGAG	CGAA
CA	CATT	CATC	*CATG*	CATA	CACT	CACC	CACG	CACA	CAGT	CAGC	CAGG	CAGA	CAAT	CAAC	CAAG	CAAA
GT	GTTT	GTTC	GTTG	GTTA	GTCT	GTCC	GTCG	GTCA	GTGT	GTGC	GTGG	GTGA	GTAT	*GTAC*	GTAG	GTAA
GC	GCTT	GCTC	GCTG	GCTA	GCCT	GCCC	GCCG	GCCA	GCGT	GCGC	GCGG	GCGA	GCAT	GCAC	GCAG	GCAA
GG	GGTT	GGTC	GGTG	GGTA	GGCT	GGCC	GGCG	GGCA	GGGT	GGGC	GGGG	GGGA	GGAT	GGAC	GGAG	GGAA
GA	GATT	*GATC*	GATG	GATA	GACT	GACC	GACG	GACA	GAGT	GAGC	GAGG	GAGA	GAAT	GAAC	GAAG	GAAA
AT	ATTT	ATTC	ATTG	*ATTA*	ATCT	ATCC	ATCG	ATCA	ATGT	ATGC	ATGG	ATGA	ATAT	ATAC	ATAG	ATAA
AC	ACTT	ACTC	ACTG	ACTA	ACCT	ACCC	ACCG	ACCA	*ACGT*	ACGC	ACGG	ACGA	ACAT	ACAC	ACAG	ACAA
AG	AGTT	AGTC	AGTG	AGTA	*AGCT*	AGCC	AGCG	AGCA	AGGT	AGGC	AGGG	AGGA	AGAT	AGAC	AGAG	AGAA
AA	*AATT*	AATC	AATG	AATA	AACT	AACC	AACG	AACA	AAGT	AAGC	AAGG	AAGA	AAAT	AAAC	AAAG	AAAA

The table shows a matrix demonstrating tetrameric primer component selection based on criteria outlined by Gerry et al. 1999. *J. Mol. Bio.* 292: 251-262. Each tetramer was required to differ from all others by at least two bases. The tetramers could not be palindromic or complimentary with any other tetramer. Thirty-six tetramers were selected (bold, underlined); italicized sequences signal palindromic tetramers that were excluded from consideration.

Primer Design:

The PCR primers were designed to meet specifications common to general primer design (see, Rubin, E. and A.A. Levy, A mathematical model and a computerized simulation of PCR using complex templates. Nucleic Acids Res, 1996. 24(18): p. 3538-45; Buck, G.A., et al., Design strategies and performance of custom DNA sequencing primers. Biotechniques, 1999. 27(3): p. 528-36), and the actual selection was conducted by a computer program, MMP. Primers were limited to a length of 20 bases (5 tetramers) for efficient synthesis of the total bipartite PCR/sequencing primer. Each primer contained a two base GC clamp on the 5' end, and a single GC clamp on the 3' end (Table 2), and all primers shared similar T_m (+/- 2°C) (Figure 27). No hairpinning within the primer (internal hairpin stem $\Delta G > -1.9$ kcal/mol) was permitted. Dimerization was also controlled; a 3 base maximum acceptable dimer was allowed, but it could occur in final six 3' bases, and the maximum allowable ΔG for a 3' dimer was -2.0 kcal/mol. Additionally, a penalty was applied to primers in which the 3' ends were too similar to others in the group, thus preventing cross- hybridization between

IDT Exhibit 1020.6617

one primer and the reverse complement of another.

Table 2

	1-pos	2-pos	3-pos	4-pos	5-pos
1	CCAT	TGAT	TGAT	TGAT	ATAC
2	CCTA	CTCA	CTCA	CTCA	AAAG
3	CGAA	TACA	TACA	TACA	TTAG
4	CGTT	AGCC	AGCC	AGCC	AATC
5	GCAA	GACC	GACC	GACC	TGTC
6	GCTT	TCCC	TCCC	TCCC	AGTG
7	GGAC	ATCG	ATCG	ATCG	CTTG
8	GGTA	CACG	CACG	CACG	GATG
9		TGCG	TGCG	TGCG	TCTG
10		ACCT	ACCT	ACCT	
11		GTCT	GTCT	GTCT	
12		AGGA	AGGA	AGGA	
13		TTGA	TTGA	TTGA	
14		CAGC	CAGC	CAGC	
15		GTGC	GTGC	GTGC	
16		ACGG	ACGG	ACGG	
17		CTGT	CTGT	CTGT	
18		GAGT	GAGT	GAGT	
19		TCGT	TCGT	TCGT	

Table 2 shows possible permutations of the 36 selected tetrads providing two 5' and a single 3' G/C clamp. The internal positions are composed of remaining tetrads. This results in 8 x 19 x 19 x 19 x 9 permutations, or 493,848 possible combinations. Figure 27 shows first pass, T_m based selection of acceptable primers, reducing field of 493,848 primers to 56,246 candidates with T_m of 64 to 66°C.

IDT Exhibit 1020.6618

Table 3

The probability of perfect sequence matches for primers increases with decreasing match length requirements and increasing size of the genome of interest.

Match Length	Perfect match probability (1/(4^length))	% chance for match in Adeno ~ 35K bases	% chance for match in NCBI bacterial database ~ 488M bases	% chance for match in Human ~3B bases
20	9.1E-13	0.00%	0.04%	0.27%
19	7.3E-12	0.00%	0.65%	4.32%
18	4.4E-11	0.00%	5.76%	34.37%
17	2.3E-10	0.00%	35.69%	99.17%
16	1.2E-09	0.02%	97.52%	> 100%
15	5.6E-09	0.12%	> 100%	> 100%
14	2.6E-08	0.64%	> 100%	> 100%
13	1.2E-07	3.29%	> 100%	> 100%
12	5.4E-07	15.68%	> 100%	> 100%
11	2.4E-06	58.16%	> 100%	> 100%
10	1.0E-05	99.35%	> 100%	> 100%
9	4.6E-05	99.77%	> 100%	> 100%
8	2.0E-04	> 100%	> 100%	> 100%
7	8.5E-04	> 100%	> 100%	> 100%
6	3.7E-03	> 100%	> 100%	> 100%
5	1.6E-02	> 100%	> 100%	> 100%
4	6.4E-02	> 100%	> 100%	> 100%
3	2.5E-01	> 100%	> 100%	> 100%
2	7.1E-01	> 100%	> 100%	> 100%
1	1.0E+00	> 100%	> 100%	> 100%

The possibility of complimentary regions occurring within the genome of interest was not a major concern in the primer design process despite the reported tolerance of PCR to mismatches in complex sample populations (see, e.g., Rubin, E. and A.A. Levy, A mathematical model and a computerized simulation of PCR using complex templates. Nucleic Acids Res, 1996. 24(18): p. 3538-45). Although the probability of finding a perfect match to a 20 base primer is extremely low (4^{20}) (Table 3), the probability of finding less non-consecutive matches increases significantly with the size of the genome of interest. As a result, the probability of finding a perfect match of at least 10 of 20 bases is 99.35% for an Adenovirus genome. The probability of finding a 16 base perfect match is 97% for the sequences in the NCBI database (approximately 100 times larger than the Adenovirus genome). The probability of finding a 17 base perfect match to a 20 base primer is 99% for the sequences in the human genome (3 billion bases).

The high probability of primer cross-hybridization to regions of the genome is less problematic than one might expect due to the random DNA digestion used to produce the template fragments. Thus, the effects of a cross-hybridizing region (CHR) are fairly benign. It is unlikely that a CHR would be able to successfully compete with the perfect match between the PCR primers in solution and the template. In addition, any primers that include

IDT Exhibit 1020.6619

mismatches at their 3' end would be at a significant competitive disadvantage. Even if a CHR should out compete the intended PCR primer, it would produce a truncated PCR product, without a downstream site for the sequencing primer. If the truncated product could be driven to the capture bead and immobilized, one of two situations would result. If the CHR out-competed the solution–phase primer, then the immobilized product would lack a sequencing primer binding site, and would result in an empty PicoTiter plate (PTP) well. If the CHR out-competed the bead-bound primer, the sequencing primer would still be present, and the only effect would be a shorter insert. Neither result would unduly compromise the sequencing quality. Given the large amount of genomic material used in the sample preparation process (currently 25 μg, containing 5.29×10^{16} copies of the 35 Kb Adenovirus genome), oversampling can be used to provide fragments that lack the complete CHR, and allow standard PCR amplification of the region in question.

Example 3: Sample Preparation by Nebulization

Preparation of DNA by Nebulization

The purpose of the Nebulization step is to fragment a large stretch of DNA such as a whole genome or a large portion of a genome into smaller molecular species that are amenable to DNA sequencing. This population of smaller-sized DNA species generated from a single DNA template is referred to as a library. Nebulization shears double-stranded template DNA into fragments ranging from 50 to 900 base pairs. The sheared library contains single-stranded ends that are end-repaired by a combination of T4 DNA polymerase, *E. coli* DNA polymerase I (Klenow fragment), and T4 polynucleotide kinase. Both T4 and Klenow DNA polymerases are used to "fill-in" 3' recessed ends (5' overhangs) of DNA via their 5'-3' polymerase activity. The single-stranded 3'-5' exonuclease activity of T4 and Klenow polymerases will remove 3' overhang ends and the kinase activity of T4 polynucleotide kinase will add phosphates to 5' hydroxyl termini.

The sample was prepared as follows:

1. 15 μg of gDNA (genomic DNA) was obtained and adjusted to a final volume of 100 μl in 10 mM TE (10 mM Tris, 0.1 mM EDTA, pH 7.6; see reagent list at the end of section). The DNA was analyzed for contamination by measuring the O.D. $_{260/280}$ ratio, which was 1.8 or higher. The final gDNA concentration was expected to be approximately 300 μg/ml.

2. 1600 μl of ice-cold Nebulization Buffer (see end of section) was added to the gDNA.

IDT Exhibit 1020.6620

3. The reaction mixture was placed in an ice-cold nebulizer (CIS-US, Bedford, MA).

4. The cap from a 15 ml snap cap falcon tube was placed over the top of the nebulizer (Figure 28A).

5. The cap was secured with a clean Nebulizer Clamp assembly, consisting of the fitted cover (for the falcon tube lid) and two rubber O-rings (Figure 28B).

6. The bottom of the nebulizer was attached to a nitrogen supply and the entire device was wrapped in parafilm (Figures 28C and 28D).

7. While maintaining nebulizer upright (as shown in Figure 28D), 50 psi (pounds per square inch) of nitrogen was applied for 5 minutes. The bottom of the nebulizer was tapped on a hard surface every few seconds to force condensed liquid to the bottom.

8. Nitrogen was turned off after 5 minutes. After the pressure had normalized (30 seconds), the nitrogen source was remove from the nebulizer.

9. The parafilm was removed and the nebulizer top was unscrewed. The sample was removed and transferred to a 1.5 ml microcentrifuge tube.

10. The nebulizer top was reinstalled and the nebulizer was centrifuged at 500 rpm for 5 minutes.

11. The remainder of the sample in the nebulizer was collected. Total recovery was about 700 µl.

12. The recovered sample was purified using a QIAquick column (Qiagen Inc., Valencia, CA) according to manufacturer's directions. The large volume required the column to be loaded several times. The sample was eluted with 30 µl of Buffer EB (10 mM Tris HCl, pH 8.5;supplied in Qiagen kit) which was pre-warmed at 55°C.

13. The sample was quantitated by UV spectroscopy (2 µl in 198 µl water for 1:100 dilution).

Enzymatic Polishing

Nebulization of DNA templates yields many fragments of DNA with frayed ends. These ends are made blunt and ready for ligation to adaptor fragments by using three enzymes, T4 DNA polymerase, *E. coli* DNA polymerase (Klenow fragment) and T4 polynucleotide kinase.

The sample was prepared as follows:

1. In a 0.2 ml tube the following reagents were added in order:

28 µl purified, nebulized gDNA fragments

IDT Exhibit 1020.6621

5 µl water
5 µl 10 X T4 DNA polymerase buffer
5 µl BSA (1mg/ml)
2 µl dNTPs (10 mM)
5 µl T4 DNA polymerase (3 units/µl)
50 µl final volume

2. The solution of step 1 was mixed well and incubated at 25°C for 10 minutes in a MJ thermocycler (any accurate incubator may be used).

3. 1.25 µl *E. coli* DNA polymerase (Klenow fragment) (5 units/ml) was added.

4. The reaction was mixed well and incubated in the MJ thermocycler for 10 minutes at 25°C and for an additional 2 hrs at 16°C.

5. The treated DNA was purified using a QiaQuick column and eluted with 30 µl of Buffer EB (10 mM Tris HCl, pH 8.5) which was pre-warmed at 55°C.

6. The following reagents were combined in a 0.2 ml tube:

30 µl Qiagen purified, polished, nebulized gDNA fragments
5 µl water
5 µl 10 X T4 PNK buffer
5 µl ATP (10 mM)
5 µl T4 PNK (10 units/ml)
50 µl final volume

7. The solution was mixed and placed in a MJ thermal cycler using the T4 PNK program for incubation at 37°C for 30 minutes, 65°C for 20 minutes, followed by storage at 14°C.

8. The sample was purified using a QiaQuick column and eluted in 30 µl of Buffer EB which was pre-warmed at 55°C.

9. A 2 µl aliquot of the final polishing reaction was held for analysis using a BioAnalyzer DNA 1000 LabChip (see below).

Ligation of Adaptors

The procedure for ligating the adaptors was performed as follows:

1. In a 0.2 ml tube the following reagents were added in order:

20.6 µl molecular biology grade water
28 µl digested, polished gDNA Library

IDT Exhibit 1020.6622

60 µl 2 X Quick Ligase Reaction Buffer
1.8 µl MMP (200 pmol/µl) Universal Adaptor set
9.6 µl Quick Ligase
120 µl total

The above reaction was designed for 5 µg and was scaled depending on the amount of gDNA used.

2. The reagents were mixed well and incubated at 25°C for 20 minutes. The tube was on ice until the gel was prepared for agarose gel electrophoresis.

Gel Electrophoresis and Extraction of Adapted gDNA Library

Nebulization of genomic DNA yields a library population that ranges from 50-900 bp. The addition of the 88-bp Universal Adaptor set will shift the population to a larger size and will result in a migration profile with a larger size range (approximately 130-980 bp). Adaptor dimers will migrate at 88 bp and adaptors not ligated will migrate at 44 bp. Therefore, genomic DNA libraries isolated in size ranges ≥250 bp can be physically isolated from the agarose gel and purified using standard gel extraction techniques. Gel isolation of the adapted gDNA library will result in the recovery of a library population in a size range that is ≥250 bp (size range of library can be varied depending on application). The library size range after ligation of adapters is 130 to 980 bp. It should be noted that the procedure may be adapted for isolation of any band size range, such as, for example, 130 to 200 bp, 200 to 400 bp, 250 to 500 bp, 300 to 600 bp, 500 to 700 bp and the like by cutting different regions of the gel. The procedure described below was used to isolated fragments of 250 bp to 500 bp.

A 150 ml agarose gel was prepared to include 2% agarose, 1X TBE, and 4.5 µl ethidium bromide (10 mg/ml stock). The ligated DNA was mixed with 10X Ready Load Dye and loaded onto the gel. In addition, 10 µl of a100-bp ladder (0.1 µg/µl) was loaded on two lanes away from the ligation reaction flanking the sample. The gel was electrophoresed at 100 V for 3 hours. When the gel run was complete, the gel was removed from the gel box, transferred to a GelDoc, and covered with plastic wrap. The DNA bands were visualized using the Prep UV light. A sterile, single-use scalpel, was used to cut out a library population from the agarose gel with fragment sizes of 250 – 500 bp. This process was done as quickly as possible to prevent nicking of DNA. The gel slices were placed in a 15 ml falcon tube. The agarose-embedded gDNA library was isolated using a Qiagen MinElute Gel Extraction

IDT Exhibit 1020.6623

kit. Aliquots of each isolated gDNA library were analyzed using a BioAnalyzer DNA 1000 LabChip to assess the exact distribution of the gDNA library population.

Strand Displacement and Extension of the gDNA Library and Isolation of the Single Stranded gDNA Library Using Streptavidin Beads

Strand displacement and extension of nicked double-stranded gDNA library was performed as described in Example 1, with the exception that the *Bst*-treated samples were incubated in the thermal cycler at 65° C for 30 minutes and placed on ice until needed. Streptavidin beads were prepared as described in Example 1, except that the final wash was performed using two washes with 200 µl 1X Binding buffer and two washes with 200 µl nH_2O. Single-stranded gDNA library was isolated using streptavidin beads as follows. Water from the washed beads was removed and 250 µl of Melt Solution (see below) was added. The bead suspension was mixed well and incubated at room temperature for 10 minutes on a tube rotator. In a separate tube, 1250 µl of PB (from the QiaQuick Purification kit) and 9 µl of 20% acetic acid were mixed. The beads in 250 µl Melt Solution were pelleted using a Dynal MPC and the supernatant was carefully removed and transferred to the freshly prepared PB/acetic acid solution. DNA from the 1500 µl solution was purified using a single MinElute purification spin column. This was performed by loading the sample through the same column twice at 750 µl per load. The single stranded gDNA library was eluted with 15 µl of Buffer EB which was pre-warmed at 55°C.

Single Strand gDNA Quantitation and Storage

Single-stranded gDNA was quantitated using RNA Pico 6000 LabChip as described in Example 1. In some cases, the single stranded library was quantitated by a second assay to ensure the initial Agilent 2100 quantitation was performed accurately. For this purpose, RiboGreen quantitation was performed as described (ssDNA Quantitation by Fluorometry) to confirm the Agilent 2100 quantitation. If the two estimates differed by more than 3 fold, each analysis was repeated. If the quantitation showed greater than a 3 fold difference between the two procedures, a broader range of template to bead was used.

Dilution and storage of the single stranded gDNA library was performed as described in Example 1. The yield was as follows:

Remaining final volume of ssDNA library following LabChip analysis = 12 µl.

Remaining final volume of ssDNA library following RiboGreen analysis = 9 µl.

Final volume of ssDNA library after the addition of TE = 18 µl.

IDT Exhibit 1020.6624

An equal volume of TE was added to single-stranded gDNA library stock. Single-stranded gDNA library to 1 x 10^8 molecules/µl in Buffer TE. Stock was diluted (1/500) to 200,000 molecules/µl in TE and 20 µl aliquots were prepared.

Library Fragment Size Distribution After Nebulization

Typical results from Agilent 2100 DNA 1000 LabChip analysis of 1 µl of the material following Nebulization and polishing are shown in Figure 29A. The size range distribution of the majority of the product was expected to fall around 50 to 900 base pairs. The mean size (top of peak) was expected to be approximately 450 bp. Typical results from gel purification of adaptor ligated library fragments are shown in Figure 29B.

Reagents

Unless otherwise specified, the reagents listed in the Examples represent standard reagents that are commercially available. For example, Klenow, T4 DNA polymerase, T4 DNA polymerase buffer, T4 PNK, T4 PNK buffer, Quick T4 DNA Ligase, Quick Ligation Buffer, *Bst* DNA polymerase (Large Fragment) and ThermoPol reaction buffer are available from New England Biolabs (Beverly, MA). dNTP mix is available from Pierce (Rockford, IL). Agarose, UltraPure TBE, BlueJuice gel loading buffer and Ready-Load 100bp DNA ladder may be purchased from Invitrogen (Carlsbad, CA). Ethidium Bromide and 2-Propanol may be purchased from Fisher (Hampton, NH). RNA Ladder may be purchased from Ambion (Austin, TX). Other reagents are either commonly known and/or are listed below:

Melt Solution:

Ingredient	Quantity Required	Vendor	Stock Number
NaCl (5 M)	200 µl	Invitrogen	24740-011
NaOH (10 N)	125 µl	Fisher	SS255-1
molecular biology grade water	9.675 ml	Eppendorf	0032-006-205

The Melt Solution included 100 mM NaCl, and 125 mM NaOH. The listed reagents were combined and mixed thoroughly. The solution could be stored at RT for six months.

Binding & Washing (B&W) Buffer (2X and 1X):

Ingredient	Quantity Required	Vendor	Stock Number
UltraPure Tris-HCl (pH 7.5, 1 M)	250 µl	Invitrogen	15567-027
EDTA (0.5 M)	50 µl	Invitrogen	15575-020
NaCl (5 M)	10 ml	Invitrogen	24740-011

IDT Exhibit 1020.6625

Ingredient	Quantity Required	Vendor	Stock Number
molecular biology grade water	14.7 ml	Eppendorf	0032-006-205

The 2X B&W buffer included final concentrations of 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, and 2 M NaCl. The listed reagents were combined by combined and mixed thoroughly. The solution could be stored at RT for 6 months. The 1X B&W buffer was prepared by mixing 2X B&W buffer with picopure H_2O, 1:1. The final concentrations was half of that listed the above, i.e., 5 mM Tris-HCl (pH 7.5), 0.5 mM EDTA, and 1 M NaCl.

Other buffers included the following. 1X T4 DNA Polymerase Buffer: 50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl2, 1 mM dithiothreitol (pH 7.9 @ 25°C). TE: 10 mM Tris, 1 mM EDTA.

Special Reagent preparation:

TE (10 mM):

Ingredient	Quantity Required	Vendor	Stock Number
TE (1M)	1 ml	Fisher	BP1338-1
molecular biology grade water	99 ml	Eppendorf	0032-006-205

Reagents were mixed and the solution could be stored RT for six months.

Nebulization Buffer:

Ingredient	Quantity Required	Vendor	Stock Number
Glycerol	53.1 ml	Sigma	G5516
molecular biology grade water	42.1 ml	Eppendorf	0032-006-205
UltraPure Tris-HCl (pH 7.5, 1M)	3.7 ml	Invitrogen	15567-027
EDTA (0.5M)	1.1 ml	Sigma	M-10228

All reagents were added (glycerol was added last) to a Stericup and mixed well. The solution was labeled and could be stored at RT for six months.

ATP (10 mM):

Ingredient	Quantity Required	Vendor	Stock Number
ATP (100 mM)	10 µl	Roche	1140965
molecular biology grade water	90 µl	Eppendorf	0032-006-205

The reagents were mixed and the solution could be stored at -20°C for six months.

BSA (1 mg/ml):

Ingredient	Quantity Required	Vendor	Stock Number
BSA (10 mg/ml)	10 µl	NEB	M0203 kit
Molecular Biology Grade water	90 µl	Eppendorf	0032-006-205

IDT Exhibit 1020.6626

The reagents were mixed and the solution could be stored at 4°C for six months.

Library Annealing buffer, 10X:

Ingredient	Quantity Req.	Vendor	Stock No.
UltraPure Tris-HCl (pH 7.5, 1 M)	200 ml	Invitrogen	15567-027
Magnesium acetate, enzyme grade (1 M)	10.72 g	Fisher	BP-215-500
Molecular Biology Grade water	~ 1 L	Eppendorf	0032-006-205

The 10 X Annealing Buffer included 200 mM Tris (pH 7.5) and 50 mM magnesium acetate. For this buffer, 200 ml of Tris was added to 500 ml picopure H_2O. Next, 10.72 g of magnesium acetate was added to the solution and dissolved completely. The solution was adjusted to a final volume of 1000 ml. The solution could be stored at 4°C for six months. To avoid the potential for contamination of libraries, the buffer was aliquotted for single or short-term usage.

Adaptors:

Adaptor "A" (400 µM):

Ingredient	Quantity Req.	Vendor	Stock No.
Adaptor A (sense; HPLC-purified, phosphorothioate linkages, 44 bp, 1000 pmol/µl)	10.0 µl	IDT	custom
Adaptor A (antisense; HPLC-purified, Phosphorothioate linkages, 40 bp, 1000 pmol/µl)	10.0 µl	IDT	custom
Annealing buffer (10X)	2.5 µl	454 Corp.	previous table
molecular biology grade water	2.5 µl	Eppendorf	0032-006-205

For this solution, 10 µl of 1000 pmol/µl Adaptor A (44 bp, sense) was mixed with 10 µl of 1000 pmol/µl Adaptor A (40 bp, antisense), 2.5 µl of 10X Library Annealing Buffer, and 2.5 µl of water (V_f = 25 µl). The adaptors were annealed using the ANNEAL-A program (see Appendix, below) on the Sample Prep Lab thermal cycler. More details on adaptor design are provided in the Appendix.

Adaptor "B" (400 µM):

Ingredient	Quantity Req.	Vendor	Stock No.
Adaptor B (sense; HPLC-purified, phosphorothioate linkages, 40 bp, 1000 pmol/µl))	10 µl	IDT	Custom
Adaptor B (anti; HPLC-purified, phosphorothioate linkages, 5'Biotinylated, 44 bp, 1000 pmol/µl)	10 µl	IDT	Custom

IDT Exhibit 1020.6627

Ingredient	Quantity Req.	Vendor	Stock No.
Annealing buffer (10X)	2.5 µl	454 Corp.	previous table
molecular biology grade water	2.5 µl	Eppendorf	0032-006-205

For this solution, 10 µl of 1000 pmol/µl Adaptor B (40 bp, sense) was mixed with 10 µl of 1000 pmol/µl Adaptor B (44 bp, anti), 2.5 µl of 10 X Library Annealing Buffer, and 2.5 µl of water (V_f = 25 µl). The adaptors were annealed using the ANNEAL-A program (see Appendix) on the Sample Prep Lab thermal cycler. After annealing, adaptor "A" and adaptor "B" (V_f = 50 µl) were combined. Adaptor sets could be stored at -20°C until use.

20% Acetic Acid:

Ingredient	Quantity Required	Vendor	Stock Number
acetic acid, glacial	2 ml	Fisher	A35-500
molecular biology grade water	8 ml	Eppendorf	0032-006-205

For this solution, glacial acetic acid was added to the water.. The solution could be stored at RT for six months.

Adaptor Annealing Program:

ANNEAL-A program for primer annealing:

1. Incubate at 95°C, 1 min;
2. Reduce temperature to 15°C at 0.1°C/sec; and
3. Hold at 14°C.

T4 Polymerase/Klenow POLISH program for end repair:

1. Incubate at 25°C, 10 minutes;
2. Incubate at 16°C, 2 hours; and
3. Hold at 4°C.

T4 PNK Program for end repair:

1. Incubate at 37°C, 30 minutes;
2. Incubate at 65°C, 20 minutes; and
3. Hold at 14°C.

BST program for stand displacement and extension of nicked double-stranded gDNA:

1. Incubate at 65° C, 30 minutes; and

IDT Exhibit 1020.6628

2. Hold at 14° C.

Step 9: Dilution and Storage of Single-Stranded DNA library

Single-stranded DNA library in EB buffer: remaining final volume = 25 µl.

Initial Stock dilution was made as follows. Using Pyrosequencing (Pyrosequencing AB, Uppsala, Sweden) results, single-stranded DNA library was diluted to 100M molecules/µL in 1X Annealing Buffer (usually this was a 1:50 dilution).

Aliquots of single-stranded DNA Library were made for common use by diluting 200,000 molecules/µL in 1X Annealing Buffer and preparing 30 µL aliquots. Store at –20°C. Samples were utilized in emulsion PCR.

Reagent preparation :

Stop Solution (50 mM EDTA): 100 µl of 0.5 M EDTA was mixed with 900 µl of nH_2O to make 1.0 ml of 50 mM EDTA solution.

Solution of 10 mM dNTPs included 10 µl dCTP (100 mM), 10 µl dATP (100 mM), 10 µl dGTP (100 mM), and 10 µl dTTP (100 mM), 60 µl Molecular Biology Grade water, (nH_2O). All four 100 mM nucleotide stocks were thawed on ice. 10 µl of each nucleotide was combined with 60 µl of nH_2O to a final volume of 100 µl, and mixed thoroughly. 1 ml aliquots were dispensed into 1.5 ml microcentrifuge tubes, and stored at -20°C, no longer than one year.

Annealing buffer, 10X: 10 X Annealing buffer included 200 mM Tris (pH 7.5) and 50 mM magnesium acetate. For this solution, 24.23 g Tris was added to 800 ml nH_2O and adjusted to pH 7.5. To this, 10.72 g magnesium acetate was added and dissolved completely. The solution was brought up to a final volume of 1000 ml. The solution was able be stored at 4°C for 1 month.

10x TE: 10 X TE included 100 mM Tris•HCl (pH 7.5), and 50 mM EDTA. These reagents were added together and mixed thoroughly. The solution could be stored at room temperature for 6 months.

Example 4: Bead Emulsion PCR

The following procedures, including capture of the template DNA, DNA

IDT Exhibit 1020.6629

amplification, and recovery of the beads bound to amplified template, can be performed in a single tube. The emulsion format ensures the physical separation of the beads into 100-200 μm "microreactors" within this single tube, thus allowing for clonal amplification of the various templates. Immobilization of the amplification product is achieved through extension of the template along the oligonucleotides bound to the DNA capture beads. Typical, the copy number of the immobilized template ranges from 10 to 30 million copies per bead. The DNA capture beads affixed with multiple copies of a single species of nucleic acid template are ready for distribution onto PTPs.

The 300,000 75-picoliter wells etched in the PTP surface provide a unique array for the sequencing of short DNA templates in a massively parallel, efficient and cost-effective manner. However, this requires fairly large quantities (millions of copies) of clonal templates in each reaction well. The methods of the invention allow the user to clonally amplify single-stranded genomic template species thorough PCR reactions conducted in standard tubes or microtiter plates. Single copies of the template species may be mixed with capture beads, resuspended into complete PCR amplification solution, and emulsified into microreactors (100 to 200 μm in diameter), after which PCR amplification generates 10^7 –fold amplification of the initial template species. This procedure is much simpler and more cost-effective than previous methods.

Binding Nucleic Acid Template to Capture Beads

This example describes preparation of a population of beads that preferably have only one unique nucleic acid template attached thereto. Successful clonal amplification depends on the delivery of a controlled number of template species (0.5 to 1) to each bead. Delivery of excess species can result in PCR amplification of a mixed template population, preventing generation of meaningful sequence data while a deficiency of species will result in fewer wells containing template for sequencing. This can reduce the extent of genome coverage provided by the sequencing phase. As a result, it is preferred that the template concentration be accurately determined through replicated quantitation, and that the binding protocol be followed as outlined below.

Template Quality Control

The success of the Emulsion PCR reaction is related to the quality of the template species. Regardless of the care and detail paid to the amplification phase, poor quality

IDT Exhibit 1020.6630

templates will impede successful amplification and the generation of meaningful sequence data. To prevent unnecessary loss of time and money, it is important to check the quality of the template material before initiating the Emulsion PCR phase of the process. Preferably, the library should pass two quality control steps before it is used in Emulsion PCR. Its concentration and the distribution of products it contains should be determined. Ideally, the library should appear as a heterogeneous population of fragments with little or no visible adapter dimers (e.g., ~90 bases). Also, amplification with PCR primers should result in a product smear ranging, for example, from 300 to 500 bp. Absence of amplification product may reflect failure to properly ligate the adaptors to the template, while the presence of a single band of any size may reflect contamination of the template.

Preparation of the PCR solution

The main consideration for this phase is to prevent contamination of the PCR reaction mixture with stray amplicons. Contamination of the PCR reactions with a residual amplicon is one of the critical issues that can cause failure of a sequencing run. To reduce the possibility of contamination, proper lab technique should be followed, and reaction mixture preparation should be conducted in a clean room in a UV-treated laminar flow hood.

PCR Reaction Mix:

For 200 µl PCR reaction mixture (enough for amplifying 600,000 beads), the following reagents were combined in a 0.2 ml PCR tube:

Table 4

	Stock	Final	Microliters
HIFI Buffer	10 X	1 X	20
treated nucleotides	10 mM	1 mM	20
Mg	50 mM	2 mM	8
BSA	10 %	0.1 %	2
Tween 80	1 %	0.01 %	2
Ppase	2 U	0.003 U	0.333333
Primer MMP1a	100 µM	0.625 µM	1.25
Primer MMP1b	10 µM	0.078 µM	1.56
Taq polymerase	5 U	0.2 U	8
Water			136.6
Total			200

The tube was vortexed thoroughly and stored on ice until the beads are annealed with template.

IDT Exhibit 1020.6631

DNA Capture Beads:

1. 600,000 DNA capture beads were transferred from the stock tube to a 1.5 ml microfuge tube. The exact amount used will depend on bead concentration of formalized reagent.

2. The beads were pelleted in a benchtop mini centrifuge and supernatant was removed.

3. Steps 4-11 were performed in a PCR Clean Room.

4. The beads were washed with 1 mL of 1X Annealing Buffer.

5. The capture beads were pelleted in the microcentrifuge. The tube was turned 180° and spun again.

6. All but approximately 10 µl of the supernatant was removed from the tube containing the beads. The beads were not disturbed.

7. 1 mL of 1X Annealing Buffer was added and this mixture was incubated for 1 minute. The beads were then pelleted as in step 5.

8. All but approximately 100 µL of the material from the tube was removed.

9. The remaining beads and solution were transferred to a PCR tube.

10. The 1.5 mL tube was washed with 150 µL of 1X Annealing Buffer by pipetting up and down several times. This was added to the PCR tube containing the beads.

11. The beads were pelleted as in step 5 and all but 10 µL of supernatant was removed, taking care to not disturb the bead pellet.

12. An aliquot of quantitated single-stranded template DNA (sstDNA) was removed. The final concentration was 200,000-sst DNA molecules/µl.

13. 3 µl of the diluted sstDNA was added to PCR tube containing the beads. This was equivalent to 600,000 copies of sstDNA.

14. The tube was vortexed gently to mix contents.

15. The sstDNA was annealed to the capture beads in a PCR thermocycler with the program 80Anneal stored in the EPCR folder on the MJ Thermocycler, using the following protocol:

- 5 minutes at 65°C;
- Decrease by 0.1°C /sec to 60°C;
- Hold at 60°C for 1 minute;
- Decrease by 0.1°C /sec to 50°C;

IDT Exhibit 1020.6632

- Hold at 50°C for 1 minute;
- Decrease by 0.1°C/sec to 40°C;
- Hold at 40°C for 1 minute;
- Decrease by 0.1°C /sec to 20°C; and
- Hold at 10°C until ready for next step.

In most cases, beads were used for amplification immediately after template binding. If beads were not used immediately, they should were stored in the template solution at 4°C until needed. After storage, the beads were treated as follows.

16. As in step 6, the beads were removed from the thermocycler, centrifuged, and annealing buffer was removed without disturbing the beads.

17. The beads were stored in an ice bucket until emulsification (Example 2).

18. The capture beads included, on average, 0.5 to 1 copies of sstDNA bound to each bead, and were ready for emulsification.

Example 5: Emulsification

A PCR solution suitable for use in this step is described below. For 200 µl PCR reaction mix (enough for amplifying 600K beads), the following were added to a 0.2 ml PCR tube:

	Stock	Final	Microliters
HIFI Buffer	10 X	1 X	20
treated Nukes	10 mM	1 mM	20
Mg	50 mM	2 mM	8
BSA	10 %	0.1 %	2
Tween 80	1 %	0.01 %	2
Ppase	2 U	0.003 U	0.333333
Primer MMP1a	100 µM	0.625 µM	1.25
Primer MMP1b	10 µM	0.078 µM	1.56
Taq	5 U	0.2 U	8
Water			136. 6
Total			200

This example describes how to create a heat-stable water-in-oil emulsion containing about 3,000 PCR microreactors per microliter. Outlined below is a protocol for preparing the emulsion.

1. 200 µl of PCR solution was added to the 600,000 beads (both components from Example 1).

2. The solution was pipetted up and down several times to resuspend the beads.

IDT Exhibit 1020.6633

3. The PCR-bead mixture was allowed to incubate at room temperature for 2 minutes to equilibrate the beads with PCR solution.

4. 400 µl of Emulsion Oil was added to a UV-irradiated 2 ml microfuge tube.

5. An "amplicon-free" 1/4" stir magnetic stir bar was added to the tube of Emulsion Oil.

An amplicon-free stir bar was prepared as follows. A large stir bar was used to hold a 1/4" stir bar. The stir bar was then:

- Washed with DNA-Off (drip or spray);
- Rinsed with picopure water;
- Dried with a Kimwipe edge; and
- UV irradiated for 5 minutes.

6. The magnetic insert of a Dynal MPC-S tube holder was removed. The tube of Emulsion Oil was placed in the tube holder . The tube was set in the center of a stir plate set at 600 rpm.

7. The tube was vortexed extensively to resuspend the beads. This ensured that there was minimal clumping of beads.

8. Using a P-200 pipette, the PCR-bead mixture was added drop-wise to the spinning oil at a rate of about one drop every 2 seconds, allowing each drop to sink to the level of the magnetic stir bar and become emulsified before adding the next drop. The solution turned into a homogeneous milky white liquid with a viscosity similar to mayonnaise.

9. Once the entire PCR-bead mixture was been added, the microfuge tube was flicked a few times to mix any oil at the surface with the milky emulsion.

10. Stirring was continued for another 5 minutes.

11. Steps 9 and 10 were repeated.

12. The stir bar was removed from the emulsified material by dragging it out of the tube with a larger stir bar.

13. 10 µL of the emulsion was removed and placed on a microscope slide. The emulsion was covered with a cover slip and the emulsion was inspected at 50X magnification (10X ocular and 5X objective lens). A "good" emulsion was expected to include primarily single beads in isolated droplets (microreactors) of PCR solution in oil.

14. A suitable emulsion oil mixture with emulsion stabilizers was made as follows. The components for the emulsion mixture are shown in Table 5.

IDT Exhibit 1020.6634

Table 5

Ingredient	Quantity Required	Source	Ref. Number
Sigma Light Mineral Oil	94.5 g	Sigma	M-5904
Atlox 4912	1 g	Uniqema	NA
Span 80	4.5 g	Uniqema	NA

The emulsion oil mixture was made by prewarming the Atlox 4912 to 60°C in a water bath. Then, 4.5 grams of Span 80 was added to 94.5 grams of mineral oil to form a mixture. Then, one gram of the prewarmed Atlox 4912 was added to the mixture. The solutions were placed in a closed container and mixed by shaking and inversion. Any sign that the Atlox was settling or solidifying was remedied by warming the mixture to 60°C, followed by additional shaking.

Example 6: Amplification

This example describes amplification of the template DNA in the bead – emulsion mixture. According to this protocol of the invention, the DNA amplification phase of the processtakes 3 to 4 hours. After the amplification is complete, the emulsion may be left on the thermocycler for up to 12 hours before beginning the process of isolating the beads. PCR thermocycling was performed by placing 50 to 100 µl of the emulsified reaction mixture into individual PCR reaction chambers (i.e., PCR tubes). PCR was performed as follows:

1. The emulsion was transferred in 50-100 µL amounts into approximately 10 separate PCR tubes or a 96-well plate using a single pipette tip. For this step, the water-in-oil emulsion was highly viscous.

2. The plate was sealed, or the PCR tube lids were closed, and the containers were placed into in a MJ thermocycler with or without a 96-well plate adaptor.

3. The PCR thermocycler was programmed to run the following program:

- 1 cycle (4 minutes at 94° C) – Hotstart Initiation;
- 40 cycles (30 seconds at 94°C, 30 seconds at 58°C, 90 seconds at 68°C);
- 25 cycles (30 seconds at 94°C, 6 minutes at 58°C); and
- Storage at 14°C.

4. After completion of the PCR reaction, the amplified material was removed in order to proceed with breaking the emulsion and bead recovery.

Example 7: Breaking the Emulsion and Bead Recovery

IDT Exhibit 1020.6635

This example describes how to break the emulsion and recover the beads with amplified template thereon. Preferably, the post-PCR emulsion should remain intact. The lower phase of the emulsion should, by visual inspection, remain a milky white suspension. If the solution is clear, the emulsion may have partially resolved into its aqueous and oil phases, and it is likely that many of the beads will have a mixture of templates. If the emulsion has broken in one or two of the tubes, these samples should not be combined with the others. If the emulsion has broken in all of the tubes, the procedure should not be continued.

1. All PCR reactions from the original 600 µl sample were combined into a single 1.5 ml microfuge tube using a single pipette tip. As indicated above, the emulsion was quite viscous. In some cases, pipetting was repeated several times for each tube. As much material as possible was transferred to the 1.5 ml tube.

2. The remaining emulsified material was recovered from each PCR tube by adding 50 µl of Sigma Mineral Oil into each sample. Using a single pipette tip, each tube was pipetted up and down a few times to resuspend the remaining material.

3. This material was added to the 1.5 ml tube containing the bulk of the emulsified material.

4. The sample was vortexed for 30 seconds.

5. The sample was spun for 20 minutes in the tabletop microfuge tube at 13.2K rpm in the Eppendorf microcentrifuge.

6. The emulsion separated into two phases with a large white interface. As much of the top, clear oil phase as possible was removed. The cloudy material was left in the tube. Often a white layer separated the oil and aqueous layers. Beads were often observed pelleted at the bottom of the tube.

7. The aqueous layer above the beads was removed and saved for analysis (gel analysis, Agilent 2100, and Taqman). If an interface of white material persisted above the aqueous layer, 20 microliters of the underlying aqueous layer was removed. This was performed by penetrating the interface material with a pipette tip and withdrawing the solution from underneath.

8. In the PTP Fabrication and Surface Chemistry Room Fume Hood, 1 ml of Hexanes was added to the remainder of the emulsion.

9. The sample was vortexed for 1 minute and spun at full speed for 1minute.

10. In the PTP Fabrication and Surface Chemistry Room Fume Hood, the top,

IDT Exhibit 1020.6636

oil/hexane phase was removed and placed into the organic waste container. .

11. 1 ml of 1X Annealing Buffer was added in 80% Ethanol to the remaining aqueous phase, interface, and beads.

12. The sample was vortexed for 1 minute or until the white substance dissolved.

13. The sample was centrifuged for 1 minute at high speed. The tube was rotated 180 degrees, and spun again for 1 minute. The supernatant was removed without disturbing the bead pellet.

14. The beads were washed with 1 ml of 1X Annealing Buffer containing 0.1% Tween 20 and this step was repeated.

Example 8: Single Strand Removal and Primer Annealing

If the beads are to be used in a pyrophosphate-based sequencing reaction, then it is necessary to remove the second strand of the PCR product and anneal a sequencing primer to the single stranded template that is bound to the bead. This example describes a protocol for accomplishing that.

1. The beads were washed with 1 ml of water, and spun twice for 1 minute. The tube was rotated 180° between spins. After spinning, the aqueous phase was removed.

2. The beads were washed with 1 ml of 1 mM EDTA. The tube was spun as in step 1 and the aqueous phase was removed.

3. 1 ml of 0.125 M NaOH was added and the sample was incubated for 8 minutes.

4. The sample was vortexed briefly and placed in a microcentrifuge.

5. After 6 minutes, the beads were pelleted as in step 1 and as much solution as possible was removed.

6. At the completion of the 8 minute NaOH incubation, 1 ml of 1X Annealing Buffer was added.

7. The sample was briefly vortexed, and the beads were pelleted as in step 1. As much supernatant as possible was removed, and another 1 ml of 1X Annealing buffer was added.

8. The sample was briefly vortexed, the beads were pelleted as in step 1, and 800 µl of 1X Annealing Buffer was removed.

9. The beads were transferred to a 0.2 ml PCR tube.

10. The beads were transferred and as much Annealing Buffer as possible was

IDT Exhibit 1020.6637

removed, without disturbing the beads.

11. 100 µl of 1X Annealing Buffer was added.

12. 4 µl of 100 µM sequencing primer was added. The sample was vortexed just prior to annealing.

13. Annealing was performed in a MJ thermocycler using the "80Anneal" program.

14. The beads were washed three times with 200 µl of 1X Annealing Buffer and resuspended with 100 µl of 1X Annealing Buffer.

15. The beads were counted in a Hausser Hemacytometer. Typically, 300,000 to 500,000 beads were recovered (3,000-5,000 beads/µL).

16. Beads were stored at 4°C and could be used for sequencing for 1 week.

Example 9: Optional Enrichment Step

The beads may be enriched for amplicon containing bead using the following procedure. Enrichment is not necessary but it could be used to make subsequent molecular biology techniques, such as DNA sequencing, more efficient.

Fifty microliters of 10 µM (total 500 pmoles) of biotin-sequencing primer was added to the Sepharose beads containing amplicons from Example 5. The beads were placed in a thermocycler. The primer was annealed to the DNA on the bead by the thermocycler annealing program of Example 2.

After annealing, the sepharose beads were washed three times with Annealing Buffer containing 0.1% Tween 20. The beads, now containing ssDNA fragments annealed with biotin-sequencing primers, were concentrated by centrifugation and resuspended in 200 µl of BST binding buffer. Ten microliters of 50,000 unit/ml Bst-polymerase was added to the resuspended beads and the vessel holding the beads was placed on a rotator for five minutes. Two microliters of 10mM dNTP mixture (i.e., 2.5 µl each of 10 mM dATP, dGTP, dCTP and dTTP) was added and the mixture was incubated for an additional 10 minutes at room temperature. The beads were washed three times with annealing buffer containing 0.1% Tween 20 and resuspended in the original volume of annealing buffer.

Fifty microliters of Dynal Streptavidin beads (Dynal Biotech Inc., Lake Success, NY; M270 or MyOne™ beads at 10 mg/ml) was washed three times with Annealing Buffer containing 0.1% Tween 20 and resuspended in the original volume in Annealing Buffer containing 0.1% Tween 20. Then the Dynal bead mixture was added to the resuspended

IDT Exhibit 1020.6638

sepharose beads. The mixture was vortexed and placed in a rotator for 10 minutes at room temperature.

The beads were collected on the bottom of the test tube by centrifugation at 2300 g (500 rpm for Eppendorf Centrifuge 5415D). The beads were resuspended in the original volume of Annealing Buffer containing 0.1% Tween 20. The mixture, in a test tube, was placed in a magnetic separator (Dynal). The beads were washed three times with Annealing Buffer containing 0.1% Tween 20 and resuspended in the original volume in the same buffer. The beads without amplicons were removed by wash steps, as previously described. Only Sepharose beads containing the appropriated DNA fragments were retained.

The magnetic beads were separated from the sepharose beads by addition of 500 µl of 0.125 M NaOH. The mixture was vortexed and the magnetic beads were removed by magnetic separation. The Sepharose beads remaining in solution was transferred to another tube and washed with 400 µl of 50 mM Tris Acetate until the pH was stabilized at 7.6.

Example 10: Nucleic Acid Sequencing Using Bead Emulsion PCR

The following experiment was performed to test the efficacy of the bead emulsion PCR. For this protocol, 600,000 Sepharose beads, with an average diameter of 25-35 µm (as supplied my the manufacturer) were covalently attached to capture primers at a ratio of 30-50 million copies per bead. The beads with covalently attached capture primers were mixed with 1.2 million copies of single stranded Adenovirus Library. The library constructs included a sequence that was complimentary to the capture primer on the beads.

The adenovirus library was annealed to the beads using the procedure described in Example 1. Then, the beads were resuspended in complete PCR solution. The PCR Solution and beads were emulsified in 2 volumes of spinning emulsification oil using the same procedure described in Example 2. The emulsified (encapsulated) beads were subjected to amplification by PCR as outlined in Example 3. The emulsion was broken as outlined in Example 4. DNA on beads was rendered single stranded, sequencing primer was annealed using the procedure of Example 5.

Next, 70,000 beads were sequenced simultaneously by pyrophosphate sequencing using a pyrophosphate sequencer from 454 Life Sciences (New Haven, CT) (see co-pending application of Lohman et al., filed concurrently herewith entitled Methods of Amplifying and Sequencing Nucleic Acids" USSN 60/476,592 filed June 6, 2003). Multiple batches of 70,000 beads were sequenced and the data were listed in Table 6, below.

IDT Exhibit 1020.6639

Table 6

Alignment Error Tolerance	Alignments				Coverage	Inferred Read Error
	None	Single	Multiple	Unique		
0%	47916	1560		1110	54.98%	0.00%
5%	46026	3450		2357	83.16%	1.88%
10%	43474	6001	1	3742	95.64%	4.36%

This table shows the results obtained from BLAST analysis comparing the sequences obtained from the pyrophosphate sequencer against Adenovirus sequence. The first column shows the error tolerance used in the BLAST program. The last column shows the real error as determined by direct comparison to the known sequence.

BEAD EMULSION PCR FOR DOUBLE ENDED SEQUENCING

Example 11: Template Quality Control

As indicated previously, the success of the Emulsion PCR reaction was found to be related to the quality of the single stranded template species. Accordingly, the quality of the template material was assessed with two separate quality controls before initiating the Emulsion PCR protocol. First, an aliquot of the single-stranded template was run on the 2100 BioAnalyzer (Agilient). An RNA Pico Chip was used to verify that the sample included a heterogeneous population of fragments, ranging in size from approximately 200 to 500 bases. Second, the library was quantitated using the RiboGreen fluorescence assay on a Bio-Tek FL600 plate fluorometer. Samples determined to have DNA concentrations below 5 ng/μl were deemed too dilute for use.

Example 12: DNA Capture Bead Synthesis

Packed beads from a 1 mL N-hydroxysuccinimide ester (NHS)-activated Sepharose HP affinity column (Amersham Biosciences, Piscataway, NJ) were removed from the column. The 30 –25 μm size beads were selected by serial passage through 30 and 25 μm pore filter mesh sections (Sefar America, Depew, NY, USA). Beads that passed through the first filter, but were retained by the second were collected and activated as described in the product literature (Amersham Pharmacia Protocol # 71700600AP). Two different amine-labeled HEG (hexaethyleneglycol) long capture primers were obtained, corresponding to the 5' end of the sense and antisense strand of the template to be amplified, (5'-Amine-3 HEG

IDT Exhibit 1020.6640

spacers gcttacctgaccgacctctgcctatcccctgttgcgtgtc-3'; SEQ ID NO:12; and 5'-Amine-3 HEG spacers ccattccccagctcgtcttgccatctgttccctccctgtc-3'; SEQ ID NO:13) (IDT Technologies, Coralville, IA, USA). The primers were designed to capture of both strands of the amplification products to allow double ended sequencing, i.e., sequencing the first and second strands of the amplification products. The capture primers were dissolved in 20 mM phosphate buffer, pH 8.0, to obtain a final concentration of 1mM. Three microliters of each primer were bound to the sieved 30 – 25 μm beads. The beads were then stored in a bead storage buffer (50 mM Tris, 0.02% Tween and 0.02% sodium azide, pH 8). The beads were quantitated with a hemacytometer (Hausser Scientific, Horsham, PA, USA) and stored at 4°C until needed.

Example 13: PCR Reaction Mix Preparation and Formulation

As with any single molecule amplification technique, contamination of the reactions with foreign or residual amplicon from other experiments could interfere with a sequencing run. To reduce the possibility of contamination, the PCR reaction mix was prepared in a in a UV-treated laminar flow hood located in a PCR clean room. For each 600,000 bead emulsion PCR reaction, the following reagents were mixed in a 1.5 ml tube: 225 μl of reaction mixture (1X Platinum HiFi Buffer (Invitrogen)), 1 mM dNTPs, 2.5 mM $MgSO_4$ (Invitrogen), 0.1% BSA, 0.01% Tween, 0.003 U/μl thermostable PPi-ase (NEB), 0.125 μM forward primer (5'-gcttacctgaccgacctctg-3'; SEQ ID NO:14) and 0.125 μM reverse primer (5'-ccattccccagctcgtcttg-3'; SEQ ID NO:15) (IDT Technologies, Coralville, IA, USA) and 0.2 U/μl Platinum Hi-Fi Taq Polymerase (Invitrogen). Twenty-five microliters of the reaction mixture was removed and stored in an individual 200 μl PCR tube for use as a negative control. Both the reaction mixture and negative controls were stored on ice until needed.

Example 14: Binding Template Species to DNA Capture Beads

Successful clonal DNA amplification for sequencing relates to the delivery of a controlled number of template species to each bead. For the experiments described herein below, the typical target template concentration was determined to be 0.5 template copies per capture bead. At this concentration, Poisson distribution dictates that 61% of the beads have no associated template, 30% have one species of template, and 9% have two or more template species. Delivery of excess species can result in the binding and subsequent amplification of a mixed population (2 or more species) on a single bead, preventing the

IDT Exhibit 1020.6641

generation of meaningful sequence data. However, delivery of too few species will result in fewer wells containing template (one species per bead), reducing the extent of sequencing coverage. Consequently, it was deemed that the single-stranded library template concentration was important.

Template nucleic acid molecules were annealed to complimentary primers on the DNA capture beads by the following method, conducted in a UV-treated laminar flow hood. Six hundred thousand DNA capture beads suspended in bead storage buffer (see Example 9, above) were transferred to a 200 µl PCR tube. The tube was centrifuged in a benchtop mini centrifuge for 10 seconds, rotated 180°, and spun for an additional 10 seconds to ensure even pellet formation. The supernatant was removed, and the beads were washed with 200 µl of Annealing Buffer (20 mM Tris, pH 7.5 and 5 mM magnesium acetate). The tube was vortexed for 5 seconds to resuspend the beads, and the beads were pelleted as before. All but approximately 10 µl of the supernatant above the beads was removed, and an additional 200 µl of Annealing Buffer was added. The beads were again vortexed for 5 seconds, allowed to sit for 1 minute, and then pelleted as before. All but 10 µl of supernatant was discarded.

Next, 1.5 µl of 300,000 molecules/µl template library was added to the beads. The tube was vortexed for 5 seconds to mix the contents, and the templates were annealed to the beads in a controlled denaturation/annealing program preformed in an MJ thermocycler. The program allowed incubation for 5 minutes at 80°C, followed by a decrease by 0.1°C/sec to 70°C, incubation for 1 minute at 70°C, decrease by 0.1°C/sec to 60°C, hold at 60°C for 1 minute, decrease by 0.1°C/sec to 50°C, hold at 50°C for 1 minute, decrease by 0.1°C/sec to 20°C, hold at 20°C. Following completion of the annealing process, the beads were removed from the thermocycler, centrifuged as before, and the Annealing Buffer was carefully decanted. The capture beads included on average 0.5 copy of single stranded template DNA bound to each bead, and were stored on ice until needed.

Example 15: Emulsification

The emulsification process creates a heat-stable water-in-oil emulsion containing 10,000 discrete PCR microreactors per microliter. This serves as a matrix for single molecule, clonal amplification of the individual molecules of the target library. The reaction mixture and DNA capture beads for a single reaction were emulsified in the following manner. In a UV-treated laminar flow hood, 200 µl of PCR solution (from Example 10) was added to the tube containing the 600,000 DNA capture beads (from Example 11). The beads

IDT Exhibit 1020.6642

were resuspended through repeated pipetting. After this, the PCR-bead mixture was incubated at room temperature for at least 2 minutes, allowing the beads to equilibrate with the PCR solution. At the same time, 450 µl of Emulsion Oil (4.5 % (w:w) Span 80, 1% (w:w) Atlox 4912 (Uniqema, Delaware) in light mineral oil (Sigma)) was aliquotted into a flat-topped 2 ml centrifuge tube (Dot Scientific) containing a sterile ¼ inch magnetic stir bar (Fischer). This tube was then placed in a custom-made plastic tube holding jig, which was then centered on a Fisher Isotemp digital stirring hotplate (Fisher Scientific) set to 450 RPM.

The PCR-bead solution was vortexed for 15 seconds to resuspend the beads. The solution was then drawn into a 1 ml disposable plastic syringe (Benton-Dickenson) affixed with a plastic safety syringe needle (Henry Schein). The syringe was placed into a syringe pump (Cole-Parmer) modified with an aluminum base unit orienting the pump vertically rather than horizontally (Figure 30). The tube with the emulsion oil was aligned on the stir plate so that it was centered below the plastic syringe needle and the magnetic stir bar was spinning properly. The syringe pump was set to dispense 0.6 ml at 5.5 ml/hr. The PCR-bead solution was added to the emulsion oil in a dropwise fashion. Care was taken to ensure that the droplets did not contact the side of the tube as they fell into the spinning oil.

Once the emulsion was formed, great care was taken to minimize agitation of the emulsion during both the emulsification process and the post-emulsification aliquotting steps. It was found that vortexing, rapid pipetting, or excessive mixing could cause the emulsion to break, destroying the discrete microreactors. In forming the emulsion, the two solutions turned into a homogeneous milky white mixture with the viscosity of mayonnaise. The contents of the syringe were emptied into the spinning oil. Then, the emulsion tube was removed from the holding jig, and gently flicked with a forefinger until any residual oil layer at the top of the emulsion disappeared. The tube was replaced in the holding jig, and stirred with the magnetic stir bar for an additional minute. The stir bar was removed from the emulsion by running a magnetic retrieval tool along the outside of the tube, and the stir bar was discarded.

Twenty microliters of the emulsion was taken from the middle of the tube using a P100 pipettor and placed on a microscope slide. The larger pipette tips were used to minimize shear forces. The emulsion was inspected at 50X magnification to ensure that it was comprised predominantly of single beads in 30 to 150 micron diameter microreactors of PCR solution in oil (Figure 33). After visual examination, the emulsions were immediately amplified.

IDT Exhibit 1020.6643

Example 16: Amplification

The emulsion was aliquotted into 7-8 separate PCR tubes. Each tube included approximately 75 µl of the emulsion. The tubes were sealed and placed in a MJ thermocycler along with the 25 µl negative control described above. The following cycle times were used: 1 cycle of incubation for 4 minutes at 94°C (Hotstart Initiation), 30 cycles of incubation for 30 seconds at 94°C, and 150 seconds at 68°C (Amplification), and 40 cycles of incubation for 30 seconds at 94°C, and 360 seconds at 68°C (Hybridization and Extension). After completion of the PCR program, the tubes were removed and the emulsions were broken immediately or the reactions were stored at 10°C for up to 16 hours prior to initiating the breaking process.

Example 17: Breaking the emulsion and bead recovery

Following amplification, the emulstifications were examined for breakage (separation of the oil and water phases). Unbroken emulsions were combined into a single 1.5 ml microcentrifuge tube, while the occasional broken emulsion was discarded. As the emulsion samples were quite viscous, significant amounts remained in each PCR tube. The emulsion remaining in the tubes was recovered by adding 75 µl of mineral oil into each PCR tube and pipetting the mixture. This mixture was added to the 1.5 ml tube containing the bulk of the emulsified material. The 1.5 ml tube was then vortexed for 30 seconds. After this, the tube was centrifuged for 20 minutes in the benchtop microcentrifuge at 13.2K rpm (full speed).

After centrifugation, the emulsion separated into two phases with a large white interface. The clear, upper oil phase was discarded, while the cloudy interface material was left in the tube. In a chemical fume hood, 1 ml hexanes was added to the lower phase and interface layer. The mixture was vortexed for 1 minute and centrifuged at full speed for 1 minute in a benchtop microcentrifuge. The top, oil/hexane phase was removed and discarded. After this, 1 ml of 80% Ethanol/1X Annealing Buffer was added to the remaining aqueous phase, interface, and beads. This mixture was vortexed for 1 minute or until the white material from the interface was dissolved. The sample was then centrifuged in a benchtop microcentrifuge for 1 minute at full speed. The tube was rotated 180 degrees, and spun again for an additional minute. The supernatant was then carefully removed without disturbing the bead pellet.

The white bead pellet was washed twice with 1 ml Annealing Buffer containing 0.1%

IDT Exhibit 1020.6644

Tween 20. The wash solution was discarded and the beads were pelleted after each wash as described above. The pellet was washed with 1 ml Picopure water. The beads were pelleted with the centrifuge-rotate-centrifuge method used previously. The aqueous phase was carefully removed. The beads were then washed with 1 ml of 1 mM EDTA as before, except that the beads were briefly vortexed at a medium setting for 2 seconds prior to pelleting and supernatant removal.

Amplified DNA, immobilized on the capture beads, was treated to obtain single stranded DNA. The second strand was removed by incubation in a basic melt solution. One ml of Melt Solution (0.125 M NaOH, 0.2 M NaCl) was subsequently added to the beads. The pellet was resuspended by vortexing at a medium setting for 2 seconds, and the tube placed in a Thermolyne LabQuake tube roller for 3 minutes. The beads were then pelleted as above, and the supernatant was carefully removed and discarded. The residual Melt solution was neutralized by the addition of 1 ml Annealing Buffer. After this, the beads were vortexed at medium speed for 2 seconds. The beads were pelleted, and the supernatant was removed as before. The Annealing Buffer wash was repeated, except that only 800 µl of the Annealing Buffer was removed after centrifugation. The beads and remaining Annealing Buffer were transferred to a 0.2 ml PCR tube. The beads were used immediately or stored at 4°C for up to 48 hours before continuing on to the enrichment process.

Example 18: Optional Bead Enrichment

The bead mass included beads with amplified, immobilized DNA strands, and empty or null beads. As mentioned previously, it was calculated that 61% of the beads lacked template DNA during the amplification process. Enrichment was used to selectively isolate beads with template DNA, thereby maximizing sequencing efficiency. The enrichment process is described in detail below.

The single stranded beads from Example 14 were pelleted with the centrifuge-rotate-centrifuge method, and as much supernatant as possible was removed without disturbing the beads. Fifteen microliters of Annealing Buffer were added to the beads, followed by 2 µl of 100 µM biotinylated, 40 base enrichment primer (5'-Biotin- tetra-ethyleneglycol spacers ccattccccagctcgtcttgccatctgttccctccctgtctcag-3'; SEQ ID NO:16). The primer was complimentary to the combined amplification and sequencing sites (each 20 bases in length) on the 3' end of the bead-immobilized template. The solution was mixed by vortexing at a medium setting for 2 seconds, and the enrichment primers were annealed to the immobilized

IDT Exhibit 1020.6645

DNA strands using a controlled denaturation/annealing program in an MJ thermocycler. The program consisted of the following cycle times and temperatures: incubation for 30 seconds at 65° C, decrease by 0.1 °C/sec to 58°C, incubation for 90 seconds at 58° C, and hold at 10° C.

While the primers were annealing, Dynal MyOne™ streptavidin beads were resuspend by gentle swirling. Next, 20 µl of the MyOne™ beads were added to a 1.5 ml microcentrifuge tube containing 1 ml of Enhancing fluid (2 M NaCl, 10 mM Tris-HCl, 1 mM EDTA, pH 7.5). The MyOne bead mixture was vortexed for 5 seconds, and the tube was placed in a Dynal MPC-S magnet. The paramagnetic beads were pelleted against the side of the microcentrifuge tube. The supernatant was carefully removed and discarded without disturbing the MyOne™ beads. The tube was removed from the magnet, and 100 µl of enhancing fluid was added. The tube was vortexed for 3 seconds to resuspend the beads, and stored on ice until needed.

Upon completion of the annealing program, 100 µl of annealing buffer was added to the PCR tube containing the DNA capture beads and enrichment primer. The tube vortexed for 5 seconds, and the contents were transferred to a fresh 1.5 ml microcentrifuge tube. The PCR tube in which the enrichment primer was annealed to the capture beads was washed once with 200 µl of annealing buffer, and the wash solution was added to the 1.5 ml tube. The beads were washed three times with 1 ml of annealing buffer, vortexed for 2 seconds, and pelleted as before. The supernatant was carefully removed. After the third wash, the beads were washed twice with 1 ml of ice cold Enhancing fluid. The beads were vortexed, pelleted, and the supernatant was removed as before. The beads were resuspended in 150 µl ice cold Enhancing fluid and the bead solution was added to the washed MyOne™ beads.

The bead mixture was vortexed for 3 seconds and incubated at room temperature for 3 minutes on a LabQuake tube roller. The streptavidin-coated MyOne™ beads were bound to the biotinylated enrichment primers annealed to immobilized templates on the DNA capture beads. The beads were then centrifuged at 2,000 RPM for 3 minutes, after which the beads were vortexed with 2 second pulses until resuspended. The resuspended beads were placed on ice for 5 minutes. Following this, 500 µl of cold Enhancing fluid was added to the beads and the tube was inserted into a Dynal MPC-S magnet. The beads were left undisturbed for 60 seconds to allow pelleting against the magnet. After this, the supernatant with excess MyOne™ and null DNA capture beads was carefully removed and discarded.

The tube was removed from the MPC-S magnet, and 1 ml of cold enhancing fluid

IDT Exhibit 1020.6646

added to the beads. The beads were resuspended with gentle finger flicking. It was important not to vortex the beads at this time, as forceful mixing could break the link between the MyOne™ and DNA capture beads. The beads were returned to the magnet, and the supernatant removed. This wash was repeated three additional times to ensure removal of all null capture beads. To remove the annealed enrichment primers and MyOne™ beads, the DNA capture beads were resuspended in 400 µl of melting solution, vortexed for 5 seconds, and pelleted with the magnet. The supernatant with the enriched beads was transferred to a separate 1.5 ml microcentrifuge tube. For maximum recovery of the enriched beads, a second 400 µl aliquot of melting solution was added to the tube containing the MyOne™ beads. The beads were vortexed and pelleted as before. The supernatant from the second wash was removed and combined with the first bolus of enriched beads. The tube of spent MyOne™ beads was discarded.

The microcentrifuge tube of enriched DNA capture beads was placed on the Dynal MPC-S magnet to pellet any residual MyOne™ beads. The enriched beads in the supernatant were transferred to a second 1.5 ml microcentrifuge tube and centrifuged. The supernatant was removed, and the beads were washed 3 times with1 ml of annealing buffer to neutralize the residual melting solution. After the third wash, 800 µl of the supernatant was removed, and the remaining beads and solution were transferred to a 0.2 ml PCR tube. The enriched beads were centrifuged at 2,000 RPM for 3 minutes and the supernatant decanted. Next, 20 µl of annealing buffer and 3 µl of two different 100 µM sequencing primers (5'-ccatctgttccctccctgtc-3'; SEQ ID NO:17; and 5'-cctatcccctgttgcgtgtc-3' phosphate; SEQ ID NO:18) were added. The tube was vortexed for 5 seconds, and placed in an MJ thermocycler for the following 4-stage annealing program: incubation for 5 minutes at 65°C, decrease by 0.1°C/sec to 50°C, incubation for 1 minute at 50°C, decrease by 0.1°C/sec to 40°C, hold at 40°C for 1 minute, decrease by 0.1 °C /sec to 15°C, and hold at 15°C.

Upon completion of the annealing program, the beads were removed from thermocycler and pelleted by centrifugation for 10 seconds. The tube was rotated 180°, and spun for an additional 10 seconds. The supernatant was decanted and discarded, and 200 µl of annealing buffer was added to the tube. The beads were resuspended with a 5 second vortex, and pelleted as before. The supernatant was removed, and the beads resuspended in 100 µl annealing buffer. At this point, the beads were quantitated with a Multisizer 3 Coulter Counter (Beckman Coulter). Beads were stored at 4°C and were stable for at least 1 week.

IDT Exhibit 1020.6647

Example 19: Double Strand Sequencing

For double strand sequencing, two different sequencing primers are used; an unmodified primer MMP7A and a 3' phosphorylated primer MMP2Bp. There are multiple steps in the process. This process is shown schematically in Figure 38.

1. First Strand Sequencing. Sequencing of the first strand involves extension of the unmodified primer by a DNA polymerase through sequential addition of nucleotides for a predetermined number of cycles.

2. CAPPING: The first strand sequencing was terminated by flowing a Capping Buffer containing 25 mM Tricine, 5 mM Mangesium acetate, 1 mM DTT, 0.4 mg/ml PVP, 0.1 mg/ml BSA, 0.01% Tween and 2 μM of each dideoxynucleotides and 2 μM of each deoxynucleotide.

3. CLEAN: The residual deoxynucleotides and dideoxynucleotides was removed by flowing in Apyrase Buffer containing 25 mM Tricine, 5 mM Magnesium acetate, 1 mM DTT, 0.4 mg/ml PVP, 0.1 mg/ml BSA, 0.01% Tween and 8.5 units/L of Apyrase.

4. CUTTING: The second blocked primer was unblocked by removing the phosphate group from the 3' end of the modified 3' phosphorylated primer by flowing a Cutting buffer containing 5 units/ml of Calf intestinal phosphatases.

5. CONTINUE: The second unblocked primer was activated by addition of polymerase by flowing 1000 units/ml of DNA polymerases to capture all the available primer sites.

6. Second Strand Sequencing: Sequencing of the second strand by a DNA polymerase through sequential addition of nucleotides for a predetermined number of cycles.

Using the methods described above, the genomic DNA of *Staphylococcus aureus* was sequenced. The results are presented in Figure 39. A total of 31,785 reads were obtained based on 15770 reads of the first strand and 16015 reads of the second strand. Of these, a total of 11,799 reads were paired and 8187 reads were unpaired obtaining a total coverage of 38%.

Read lengths ranged from 60 to 130 with an average of 95 +/- 9 bases (Figure 40). The distribution of genome span and the number of wells of each genome span is shown in Figure 41. Representative alignment strings, from this genomic sequencing, are shown in Figure 42.

IDT Exhibit 1020.6648

Example 20: Template PCR

30 micron NHS Sepharose beads were coupled with 1 mM of each of the following primers:

MMP1A: cgtttcccctgtgtgccttg (SEQ ID NO:19)

MMP1B: ccatctgttgcgtgcgtgtc (SEQ ID NO:20)

Drive-to-bead PCR was performed in a tube on the MJ thermocycler by adding 50 µl of washed primer-coupled beads to a PCR master mix at a one-to-one volume-to-volume ratio. The PCR master mixture included:

1X PCR buffer;

1 mM of each dNTP;

0.625 µM primer MMP1A;

0.625 µM primer MMP1B;

1 µl of 1 unit/µl Hi Fi Taq (Invitrogen, San Diego, CA); and

~ 5-10 ng Template DNA (the DNA to be sequenced).

The PCR reaction was performed by programming the MJ thermocycler for the following: incubation at 94°C for 3 minutes; 39 cycles of incubation at 94°C for 30 seconds, 58°C for 30 seconds, 68°C for 30 seconds; followed by incubation at 94°C for 30 seconds and 58°C for 10 minutes; 10 cycles of incubation at 94°C for 30 seconds, 58°C for 30 seconds, 68°C for 30 seconds; and storage at 10°C.

Example 21: Template DNA Preparation and Annealing Sequencing Primer

The beads from Example 1 were washed two times with distilled water; washed once with 1 mM EDTA, and incubated with 0.125 M NaOH for 5 minutes. This removed the DNA strands not linked to the beads. Then, the beads were washed once with 50 mM Tris Acetate buffer, and twice with Annealing Buffer: 200mM Tris-Acetate, 50mM Mg Acetate, pH 7.5. Next, 500 pmoles of Sequencing Primer MMP7A (ccatctgttccctccctgtc; SEQ ID NO:21) and MMP2B-phos (***cctatcccctgttgcgtgtc***; SEQ ID NO:22) were added to the beads. The primers were annealed with the following program on the MJ thermocycler: incubation at 60°C for 5 minutes; temperature drop of 0.1 degree per second to 50°C; incubation at 50°C for 5 minutes; temperature drop of 0.1 degree per second to 4°C; incubation at 40°C for 5 minutes; temperature drop of 0.1 degree per second to 10°C. The template was then sequenced using standard pyrophosphate sequencing.

IDT Exhibit 1020.6649

Example 22: Sequencing and Stopping of the First Strand

The beads were spun into a 55 μm PicoTiter plate (PTP) at 3000 rpm for 10 minutes. The PTP was placed on a rig and run using de novo sequencing for a predetermined number of cycles. The sequencing was stopped by capping the first strand. The first strand was capped by adding 100 μl of 1X AB (50 mM Mg Acetate, 250 mM Tricine), 1000 unit/ml BST polymerase, 0.4 mg/ml single strand DNA binding protein, 1 mM DTT, 0.4 mg/ml PVP (Polyvinyl Pyrolidone), 10 uM of each ddNTP, and 2.5 μM of each dNTP. Apyrase was then flowed over in order to remove excess nucleotides by adding 1X AB, 0.4 mg/ml PVP, 1 mM DTT, 0.1 mg/ml BSA, 0.125 units/ml apyrase, incubated for 20 minutes.

Example 23: Preparation of Second Strand for Sequencing

The second strand was unblocked by adding 100 μl of 1X AB, 0.1 unit per ml poly nucleotide kinase, 5 mM DTT. The resultant template was sequenced using standard pyrophosphate sequencing (described, e.g., in US patent 6,274,320, 6258,568 and 6,210,891, incorporated herein by reference). The results of the sequencing method can be seen in Figure 10F where a fragment of 174 bp was sequenced on both ends using pyrophosphate sequencing and the methods described in these examples.

Example 24: Sequence Analysis of Nucleic Acid on a Picotiter Plate

The picotiter plate containing amplified nucleic acids as described in Example 2 is placed in a perfusion chamber. Then sulfurylase, apyrase, and luciferase are delivered to the picotiter plate.

The sequencing primer primes DNA synthesis extending into the insert suspected of having a polymorphism, as shown in Figures 11A-11D. The sequencing primer is first extended by delivering into the perfusion chamber, in succession, a wash solution, a DNA polymerase, and one of dTTP, dGTP, dCTP, or α thio dATP (a dATP analog). The sulfurylase, luciferase, and apyrase, attached to the termini convert any PPi liberated as part of the sequencing reaction to detectable light. The apyrase present degrades any unreacted dNTP. Light is typically allowed to collect for 3 seconds (although 1-100, *e.g.*, 2-10 seconds is also suitable) by a CCD camera linked to the fiber imaging bundle, after which additional wash solution is added to the perfusion chamber to remove excess nucleotides and byproducts. The next nucleotide is then added, along with polymerase, thereby repeating the cycle.

IDT Exhibit 1020.6650

During the wash the collected light image is transferred from the CCD camera to a computer. Light emission is analyzed by the computer and used to determine whether the corresponding dNTP has been incorporated into the extended sequence primer. Addition of dNTPs and pyrophosphate sequencing reagents is repeated until the sequence of the insert region containing the suspected polymorphism is obtained.

Example 25: On Picotiter Plate PCR Amplification

Picotiter plate Preparation: In a further embodiment, the single stranded library attached to beads are distributed directly onto the picotiter plate and then the nucleic acid template on each bead is amplified (using PCR or other known amplification technology) to generate sufficient copy number of the template that will generate detectable signal in the pyrophosphate-based sequencing methods disclosed herein.

Example 26: Sequence Analysis of Nucleic Acid on a PTP

Reagents used for sequence analysis and as controls were the four nucleotides and 0.1 μM Pyrophosphate (PPi) were made in substrate solution. Substrate solution refers to a mixture of 300 μM Luciferin and 4 μM adenosine 5'-phosphosulfate, APS, which are the substrates for the cascade of reactions involving PPi, Luciferase and Sulfurylase. The substrate was made in assay buffer. The concentration of PPi used to test the enzymes and determine the background levels of reagents passing through the chamber was 0.1 μM. The concentration of the nucleotides, dTTP, dGTP, dCTP was 6.5 μM and that of αdATP was 50 μM. Each of the nucleotides was mixed with DNA polymerase, Klenow at a concentration of 100 U/mL.

The PTP was placed in the flow chamber of the embodied instrument, and the flow chamber was attached to the faceplate of the CCD camera. The PTP was washed by flowing substrate (3 ml per min, 2 min) through the chamber. After this, a sequence of reagents was flown through the chamber by the pump connected to an actuator, which was programmed to switch positions, which had tubes inserted in the different reagents. The sequence of reagents, flow rates, and flow times were determined. The camera was set up in a fast acquisition mode, with exposure time = 2.5 s.

The signal output from the pad was determined as the average of counts on all the pixels within the pad. The frame number was equivalent tothe time passed during the experiment. Graphing was used to represent the flow of the different reagents.

IDT Exhibit 1020.6651

Example 27: Plate-based Platform for Picoliterscale PCR Reactions

Materials and methods

Unless otherwise indicated, all common laboratory chemicals were purchased either from Sigma (Sigma-Aldrich Corporation, St. Louis, MI) or Fisher (Fisher Scientific, Pittsburgh, PA).

The PicoTiterPlates™ (25 x 75 x 2 mm) were manufactured by anisotropic etching of fiber optic faceplates in a manner similar to that previously described (Pantano, P. and Walt, D.R., *Chemistry of Materials* 1996, 8, 2832-2835). Plates were etched in three different microwell depths, 26, 50 and 76 μm. Microwell center-to-center pitch was 50 μm, and well diameters ranged between 39 and 44 μm (See Figure 14), with a calculated well density of 480 wells/mm^2.

Solid-phase immobilization of oligonucleotide primers: Packed beads from a 1 ml NHS-activated Sepharose HP affinity column (Amersham Biosciences, Piscataway, NJ) were removed from the column and activated according to the manufacturer's instructions (Amersham Pharmacia Protocol # 71700600AP). Twenty-five microliters of a 1 mM amine-labeled HEG capture primer (5'-Amine-3 hexaethyleneglycol spacers ccatctgttgcgtgcgtgtc – 3'; SEQ ID NO:23) (IDT Technologies, Coralville, IA) in 20 mM phosphate buffer pH 8.0 were bound to the beads. After this, 36 to 25 μm beads were selected by serial passage through 36 and 25 μm pore filter mesh sections (Sefar America, Depew, NY). DNA capture beads that passed through the first filter, but were retained by the second were collected in bead storage buffer (50 mM Tris, 0.02% Tween, 0.02% Sodium Azide, pH 8), quantitated with a hemacytometer (Hausser Scientific, Horsham, PA) and stored at 4° C until needed.

Generation of test DNA fragments: Amplification test fragments were derived from a commercially available adenovirus serotype 5 vector, pAdEasy (Stratagene, La Jolla, CA). Fragments were amplified using bipartite PCR primers, the 5' end of which contained a 20 base amplification region, and a 20 base 3' section, complementary to a specific region of the adenovirus genome. Using these primers, two fragments were amplified from the 12933-13070 and 5659-5767 position of the adenovirus genome and assigned labels Fragment A and Fragment B, respectively.

The sequences for the forward and reverse primers for Fragment A was as follows. A slash (/) denotes the separation between the two regions of the primer: forward (5'-cgtttcccctgtgtgccttg / catcttgtccactaggctct-3'; SEQ ID NO:24-SEQ ID NO:25), and reverse (5'-ccatctgttgcgtgcgtgtc / accagcactcgcaccacc-3'; SEQ ID NO:26-SEQ ID NO:27). The

IDT Exhibit 1020.6652

primers for the Fragment B included: forward (5'-cgtttcccctgtgtgccttg / tacctctccgcgtaggcg-3'; SEQ ID NO:28-SEQ ID NO:29), and reverse (5'-ccatctgttgcgtgcgtgtc / ccccggacgagacgcag-3'; SEQ ID NO:30-SEQ ID NO:31).

Reaction conditions included 50 mM KCl, 10 mM Tris-HCl (pH 9.0), 0.1% Triton X-100, 2.5 mM $MgCl_2$, 0.2 mM dNTP, 1 μM each forward and reverse primer, 0.1 U/μl Taq (Promega, Madison, WI) and 50 nmol template DNA. Both templates were amplified with a PCR program that include 35 cycles of incubation at 94°C for 30 seconds, 56°C for 30 seconds, and 72°C for 90 seconds. With PCR primers, the total length of the amplified fragments was 178 bp for Fragment A and 148 bp for Fragment B.

To generate fluorescent probes, biotinylated double stranded fluorescent probes were prepared by PCR amplification from the pAdEasy vector as described above. However, the primer sequences were changed to prevent hybridization between the test fragment and probe primer regions. In addition, the reverse primers for both fragments utilized a 5'biotin followed by 3X hexaethyleneglycol spacers to permit product immobilization to beads prior to elution of the single stranded probe.

The sequence for the forward primer for the fluorescent Fragment A probe was as follows. A slash (/) denotes the separation between the two regions of the primer (5'-atctctgcctactaaccatgaag / catcttgtccactaggctct-3'; SEQ ID NO:32-SEQ ID NO:33). The sequence for the reverse primer was 5'-biotin –3X hexaethyleneglycol spacers – gtttctctccagcctctcaccga / accagcactcgcaccacc-3'; SEQ ID NO:34-SEQ ID NO:35. The primers for the Fragment B were as follows: forward (5'-atctctgcctactaaccatgaag / tacctctccgcgtaggcg-3'; SEQ ID NO:36-SEQ ID NO:37), and reverse (5'-biotin–3x hexaethyleneglycol spacers – gtttctctccagcctctcaccga / ccccggacgagacgcag-3'; SEQ ID NO:38-SEQ ID NO:39).

Fluorescent moieties were incorporated through the nucleotide mixture. This included 0.2 mM dATP/dGTP/dCTP, 0.15 mM TTP and 0.05 mM Alexa Fluor 488 – dUTP (Molecular Probes, Eugene, OR) for Fragment A. Alternately, 0.2 mM dATP/dGTP/TTP, 0.15 mM dCTP and 0.05 mM Alexa Fluor 647 - dCTP (Molecular Probes, Eugene, OR) was used for amplifying Fragment B. The fluorescent products were purified with a QIAquick PCR Purification Kit (Qiagen, Valencia, CA). The biotinylated DNA was subsequently bound to 100 μl (approximately 8.1 million) Streptavidin Sepharose High Performance beads (Amersham Biosciences) in 1X binding wash (5 mM Tris HCl pH 7.5, 1 M NaCl, 0.5 mM EDTA, 0.05% Tween-20) for 2 hours at room temperature. After incubation, the beads were

IDT Exhibit 1020.6653

washed three times in TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0) and incubated with 250 µl melt solution (0.125 N NaOH/0.1 M NaCl) for 2 minutes, releasing the single stranded probe from the beads.

Beads were pelleted with brief centrifugation in a benchtop centrifuge and the supernatant was neutralized in 1.25 ml buffer PB (Qiagen) with 1.9 µl glacial acetic acid. This mixture was repurified on a QiaQuick column (Qiagen), and the concentration of the purified probe was determined by TaqMan quantification using the BioRad iCycler (BioRad, Hercules, CA).

Solution-phase PTPCR was performed as follows. The PCR reaction mixture was loaded into individual wells of a single 14 mm x 43 mm PicoTiterPlate™. For this, 500 µl of PCR reaction mixture (1X Platinum HiFi Buffer (Invitrogen, Carlsbad, CA), 2.5 mM $MgSO_4$, 0.5% BSA, 1mM dNTPs (MBI Fermentas, Hanover, MD), 1 µM forward (5' – cgtttcccctgtgtgccttg – 3'; SEQ ID NO:40) and reverse (5' – ccatctgttgcgtgcgtgtc – 3'; SEQ ID NO:41) primers, 0.05% Tween-80, 1 U/µl Platinum High Fidelity DNA Polymerase (Invitrogen), 0.003 U/µl Thermostable Pyrophosphatase (USB, Cleveland, OH), and a calculated 5 copies of Fragment B template per well) were combined in a 1.5 ml microcentrifuge tube. The tube was vortexed thoroughly and stored on ice until the PicoTiterPlate™ loading cartridge was assembled.

The in-house PicoTiterPlate™ loading cartridge was attached to the PicoTiterPlate™ with two plastic clips, seating the silicon cartridge gasket firmly on the PicoTiterPlate™ surface (see Figure 20). The PCR reaction mix was drawn into a 1 ml disposable syringe, and the mouth of the syringe inserted into the input tube of the loading cartridge. The loading cartridge was placed on end, so that the input port was oriented at the bottom of cartridge, and the PCR mix was slowly loaded into the chamber. While loading, inspected through the transparent back of the PicoTiterPlate™ to ensure even, bubble-free delivery.

After loading, the PCR mixture was allowed to incubate for 5 minutes, at which time the reaction mixture was withdrawn from the PicoTiterPlate™ loading cartridge. The PicoTiterPlate™ was removed from the loading cartridge, and immediately placed in the amplification chamber (see Figure 21). The PicoTiterPlate™ surface was covered with a 0.25 mm thick Silpad A-2000 silicon sheet (The Bergquist Company, Chanhassen, MN). On top of this was placed a 25 mm x 75 mm standard glass microscope slide (Fisher). A closed cell foam insulation pad (Wicks Aircraft Supply, Highland, IL) was placed on top of the microscope slide. An aluminum lid was attached to the base of the chamber by six 25 mm

IDT Exhibit 1020.6654

bolts sealed the amplification chamber.

Once sealed, the amplification chamber was placed on a Thermocycler MJ PTC 225 Tetrad (MJ Research, Waltham, MA) equipped with Flat Block Alpha Units. The amplification program included incubation for 3 minutes at 94°C (Hotstart Initiation) followed by 40 cycles of incubation for 12 seconds at 94°C, 12 seconds at 58°C, 12 seconds at 68°C, with a 10° C final hold. After completion of the PCR program, the PicoTiterPlate™ was removed from the amplification chamber, and the loading cartridge was reattached. A disposable syringe was used to fill the cartridge chamber with 1 ml of H_2O, and allowed to incubate for 20 minutes 10° C at room temperature.

After the incubation was completed, the recovery solution was withdrawn from the loading cartridge and transferred to a 1.5 ml microcentrifuge tube. PCR product was quantitated using an iCycler RealTime PCR unit (BioRad) and FAM-labeled reporter probes (Epoch Biosciences, Bothell, WA). The TaqMan Universal PCR MasterMix (Applied Biosystems, Foster City, CA) was combined with 0.3 μM forward and reverse primers, 0.15 μM FAM-labeled probe, and 27 μl of the reaction mix added to each well of a 96 well PCR plate.

Purified fragments were used to create a standard curve (six standards ranging from 1 x 10^9 to 1 x 10^4 molecules per well), which was run in triplicate. The PCR amplification was run with the following parameters: incubation for 5 minutes at 94°C (hotstart initiation), 60 cycles of incubation for 15 seconds at 94°C, 45 seconds at 68°C, with a final hold at 4°C. Data was analyzed using the iCycler Optical Systems Software Version 2.3 (BioRad), and the PCR yield was quantitated using the iCycler data and Microsoft Excel (Microsoft, Redmond, WA).

Solid-phase PTPCR was performed similarly to solution phase PTPCR, except that DNA capture beads were loaded into the PicoTiterPlate™ wells prior to amplification by centrifugation as described below. In addition, the PCR mixture was loaded into the microwells after the bead deposition was completed. To facilitate retention of the capture beads during wash steps, the solid phase experiments utilized 50 μm deep PicoTiterPlate™s. The PicoTiterPlate™ was placed in an in-house built plexiglass bead loading jig. This was similar to the PicoTiterPlate™ loading jig described in Figure 20, except that the PicoTiterPlate™ was sandwiched between a bottom Plexiglas plate and a jig top plate, containing inlet and outlet ports, and sealed via a silicon gasket with plastic screws.

Template DNA was preannealed to the DNA capture beads at 5 template copies per

IDT Exhibit 1020.6655

bead by incubation at 80°C for 3 minutes, after which beads were allowed to cool to room temperature for 15 minutes. The beads were then spun into the PicoTiterPlate™ wells prior to loading the PCR reaction mixture. Bead Loading Buffer (450 µl; 1X Platinum HiFi PCR buffer (Invitrogen), 0.02% Tween-80) containing one hundred thousand Sepharose DNA capture beads (approximately 1 bead per 3 PicoTiterPlate™ wells) were injected by pipette into the jig through one of the inlet ports. Each inlet hole was then sealed with a circular adhesive pad (3M VHS, St. Paul, MN). The jig held the PicoTiterPlate™ with its wells facing up and covered with the bead suspension. This was centrifuged at 2000 rpm for 5 minutes at room temperature in an Allegra 6 centrifuge (Beckman Coulter, Fullerton, CA) using a Microtiter Rotor.

After centrifugation, the PicoTiterPlate™ was removed from the jig. The PCR reaction mix was loaded onto the PicoTiterPlate™ as described for solution phase PCR. However, the solid-phase PCR mixture omitted template since the template was preannealed to the DNA capture beads. The solid-phase PCR amplification program included additional hybridization/extension cycles to compensate for the slower kinetics of the immobilized primer. The program included incubation for 3 minutes at 94°C for hotstart initiation, 40 cycles of incubation for 12 seconds at 94°C, 12 seconds at 58°C, 12 seconds at 68°C, followed by 10 cycles of incubation for 12 seconds at 94°C, 10 minutes at 68°C for hybridization and extension, with a 10°C final hold.

Upon completion of the PCR program, the PicoTiterPlate™ was removed from the amplification chamber, and washed with 1 ml H_2O as described for solution phase PCR. The PicoTiterPlate™ was then prepared for hybridization detection of immobilized PCR product.

Hybridization was performed with fluorescently labeled probes as follows. After PTPCR was complete, the strand complementary to the immobilized strand was removed. For this, the whole PicoTiterPlate™ was incubated in 0.125 M NaOH for 8 minutes at room temperature. This solution was neutralized by two 5 minute washes in 50 ml of 20 mM Tris-acetate pH 7.5. The PicoTiterPlate™ was then placed in a custom-made 800 µl hybridization chamber, and blocked with hybridization buffer (3.5X SSC, 3.0% SDS, 20X SSC buffer is 3 M NaCl; 0.3 M Na_3-citrate) at 65°C for 30 minutes. The contents of the chamber were replaced with fresh hybridization buffer containing the probes: 20 nM fluorescent Fragment A (Alexa-488) and Fragment B (Alexa-647). The probes were allowed to hybridize to their targets. Incubation was carried out at 65°C for 4 hours while shaking at 200 RPM on an orbital shaker (Barnstead International, Dubuque, IA).

IDT Exhibit 1020.6656

After hybridization, the PicoTiterPlate™ was washed with 2X SSC, 0.1% SDS for 15 minutes at 37°C, followed by a 15 minute wash in 1x SSC at 37°C, with two final 15 minute washes in 0.2x SSC at 37°C. Following post-hybridization washing, the PicoTiterPlates ™ were air dried and placed in a FLA-8000 Fluorescent Image Analyzer (Fujifilm Medical Systems USA, Stamford, CT) and scanned at the 635 and 473 nm wavelength. The resulting 16-bit tiff images were imported into Genepix 4.0 (Axon Instruments, Union City, CA). A block of 100 analysis features was drawn over the area of interest and the 635 and 473 fluorescence intensities were recorded for each feature. Data was then exported to Microsoft Excel for further analysis.

Control beads were prepared as follows. Biotinylated test templates A and B were prepared by PCR amplification from the pAdEasy vector, purified, immobilized on Streptavidin Sepharose High Performance beads and strand separated as described under "Preparation of Fluorescent Probes". However, fluorescently labeled dNTPs were ommitted in the PCR reaction. Pelleted beads were washed 3 times with TE buffer and stored at 4°C in TE until deposition onto the PicoTiterPlate™.

Results

Solution-phase amplification was demonstrated by loading PicoTiterPlates with PCR master mix containing a calculated 5 template copies per PicoTiterPlate™ well. Reactions were run in duplicate in PicoTiterPlates with 26, 50 and 76 µm deep wells. Forty cycles of PTPCR amplification were performed as described in Material and Methods. Additives were incorporated to prevent the deleterious surface effects routinely reported with silica reaction vessels (Kalinina, O., *et al.*, *Nucleic Acids Res.* 1997, 25, 1999-2004;Wittwer, C.T. and Garling, D.J., *Biotechniques* 1991, 10, 76-83; Taylor, T.B., *et al.*, *Nucleic Acids Res.* 1997, 25, 3164-3168).

The inclusion of 0.5% BSA and 0.05% Tween-80 in the reaction mix was not only effective at reducing surface effects, it also facilitated amplification. Reducing the relative concentrations of either reagent had a negative effect on amplification. In addition, due to the polymerase-inactivating properties of silica surfaces (Taylor, T.B., *et al.*, *Nucleic Acids Res.*1997, 25, 3164-3168; Shoffner, M.A., Cheng, J., Hvichia, G.E., Kricka, L.J. and Wilding, P., *Nucleic Acids Res.* 1996, 24, 375-379), elevated Taq concentrations proved beneficial. Concentrations above 1 U/µl were optimum for enhancing amplicon yield.

Following PTPCR, the solution from each PicoTiterPlate™ was recovered and

IDT Exhibit 1020.6657

triplicate samples of each solution were quantified by TaqMan assay. A standard curve of diluted template (linear from 1 x 10^9 to 10^4 molecules, r^2 =0.995) was used to determine the concentration of the amplified product. The number of molecules amplified per well was obtained by dividing the amount of amplified product by the total number of wells in a PicoTiterPlate™ (372,380). The amount of amplification per well was calculated by dividing this number by the initial template concentration per well. PTPCR amplification was successful in all of the PicoTiterPlate™, with yields ranging from 2.36 x 10^6 fold in the 39.5 pl wells to 1.28 x 10^9 fold in the 50 pl wells (See Table below).

PicoTiterPlate Depth [μm]	Well Volume [pl]	Average Fold Amplification N=6	Fold Amplification SD	Final Product Conc. [M]
26	39.5	2.36E+06	1.02E+06	4.96E-07
50	76.0	1.28E+09	1.03E+09	1.40E-04
76	115.6	9.10E+08	4.95E+08	6.54E-05

The table shows PicoTiterPlate™ PCR amplification as determined by TaqMan Assay. Values reflect triplicate measurements taken from duplicate PicoTiterPlates. (N=6); SD = standard deviation.

Yield was influenced by well volume. The concentration of final product obtained for the 50 μm deep wells (1.4 x 10^{-4} M) was significantly greater (p value for ANOVA = 0.023) than that obtained in the 76 μm (6.54 x 10^{-5} M) deep wells, both were two orders of magnitude greater than the yield achieved in the 26 μm deep wells (4.96 x 10^{-7} M). The 50 μm deep microwell yield represented the optimal balancing of the costs and benefits associated with low-volume PCR. In this case, maximum elevation of the effective concentrations and low thermal mass of the reagents were obtained, but the surface to volume ratio was still low enough to prevent detrimental surface effects from significantly reducing amplification efficiency.

The final concentration of PTPCR product obtained in each of the different well depths (4.96 x 10^{-7} to 1.4 x 10^{-4} M) exceeded the10^{-8} M concentration typically reported as the maximum achievable before the PCR plateau effect occurs (Sardelli, A., *Amplifications* 1993, 9, 1-5). The higher effective concentration of primers and template molecules resulting from the low microwell volume increased the overall reaction efficiency and postponed the onset of the plateau phase until a higher molar yield was achieved. Alternatively, this effect was caused by the high concentration of Taq used in the PTPCR reactions, as elevated polymerase concentration has also been shown effective in delaying the plateau effect

IDT Exhibit 1020.6658

(Kainz, P., *Biochim. Biophys. Acta* 2000, 1494, 23-27; Collins, F.S., *et al.*, *Science* 2003, 300, 286-290). The amplification efficiency over 40 cycles was 44.3, 68.9 and 67.5% for the 26, 50 and 76 μm deep wells respectively, providing a high final concentration of amplicons. The greatest yield was observed in the 50 μm deep wells. It should be recognized, however, that cycle number optimization was not conducted; similar amplification yields could likely have been achieved with far fewer cycles, thereby increasing the efficiency of the PTPCR amplification.

The experimental strategy for clonal solid phase PTPCR, starting with a single effective copy of a single stranded DNA fragment, and finishing with a specific bead-immobilized DNA amplicon detected by fluorescent probe hybridization, is depicted in Figure 22 and described in detail below:

Stage 1: Each PicoTiterPlate™ well contains PCR reaction mix consisting of a single stranded template molecule (either single stranded and annealed to the DNA capture beads, as shown here, or free-floating in solution), Forward "F" (red) and Reverse "R" (blue) primers in solution, as well as R primers attached to a DNA capture bead. Solution phase primers are present in an 8:1 molar ratio, with the F primer in excess. Arrows indicate the 5' -> 3' DNA orientation.

Stage 2: The initial thermal cycle denatures the DNA template, allowing R primers in solution to bind to the complementary region on the template molecule. Thermostable polymerases initiate elongation at the primer site (dashed line), and in subsequent cycles, solution-phase exponential amplification ensues. Bead immobilized primers are not assumed to be major contributors to the amplification at this stage.

Stage 3: Early Phase PCR. During early exponential amplification (1 to 10 cycles) both F and R primers amplify the template equally, despite an excess of F primers in solution.

Stage 4: Mid Phase PCR. Between cycles 10 and 30, the R primers are depleted, halting exponential amplification. The reaction then enters an asymmetric amplification phase, and the amplicon population becomes increasingly dominated by F strands.

Stage 5: Late Phase PCR. After 30 to 40 cycles, asymmetric amplification continues to increase the concentration of F strands in solution. Excess F strands, without R strand complements, begin to anneal to bead-immobilized R primers. Thermostable polymerases utilize the F strand as a template to synthesize an immobilized R strand of the amplicon.

Stage 6: Final Phase PCR. Continued thermal cycling forces additional annealing to bead-bound primers. Solution phase amplification may be minimal at this stage, but

IDT Exhibit 1020.6659

concentration of immobilized R strands continues to increase.

Stage 7: The non-immobilized, F strand, complementary to the immobilized R strand, is removed by alkali denaturation. The DNA capture beads are now populated by single stranded R strands of the amplicon.

Stage 8: Fluorescently labeled probes (green bars) complementary to the R strand are annealed to the immobilized strand. Probes specific for particular strand sequences are labeled with unique fluorophores, resulting in a range of homogenous and heterogeneous fluorescent signals depending on the number of discrete templates amplified within a given PicoTiterPlate™ well.

Initially, fluorescently labeled probe specificity was confirmed by binding biotinylated Fragment A or Fragment B test DNA fragments to streptavidin Sepharose beads, loading the beads into a 50 μm deep PicoTiterPlate™ by centrifugation and hybridizing a mixed population of fluorescently labeled probes for the Fragment A and Fragment B fragments. No mixed signals or nonspecific hybridizations were observed; the beads with the Fragment A product displayed the 488 nm signal, while the Fragment B beads exhibited the 635 nm signal (See Figures 23A and 23B). Close examination of Figures 23A and 23B reveals a few Fragment A beads in the Fragment B pad and vice versa. Given the purity of the signal displayed by these nomadic beads, it is likely that they are either the product of some cross contamination during the loading process, or were washed from one pad to the other during subsequent wash steps.

As indicated in Figure 23C, the fluorescent probes detected successful solid phase PTPCR amplification of both Fragment A and Fragment B templates. The signals generated by the hybridized probe depended on the relative efficiency of dye incorporation within the probes, the sensitivity of the reactions to unequal amounts of template DNA, as well as the total and relative amounts of amplified product present on each bead. In addition, it is likely that the amount of template generated and retained on the DNA capture beads varied from well to well, and the number of capture primers bound to each bead is also likely to vary due to bead size distribution. As a result, the non-normalized ratios generated by the probe hybridization should be seen as semi-qualitative rather than quantitative data. Nevertheless, the fluorescent signals generated by the hybridized probes ranged from a homogeneous Fragment B signal (red) to an equally homogenous Fragment A signal (green), with heterogeneous mixes of the two signals (degrees of yellow) evident as well.

Due to the probe specificity displayed by the controls, as well as the sizeable number

IDT Exhibit 1020.6660

of homogenous red and green beads on the PicoTiterPlate™, it is unlikely that nonspecific probe hybridization caused the heterogeneous signals. The close proximity of homogenous beads of either template suggests it is unlikely that the heterogeneous beads resulted from amplicon leakage between wells during amplification; if intra-well cross-talk were responsible, one would expect to see heterogeneous beads located between homogenous beads of either template, and a generally patchy distribution of homogenous signals. Rather, it is likely that template molecules disassociated from their original bead and reannealed to new beads in the PicoTiterPlate™ loading mix prior to being spun into the microwells, or were washed from one bead to another as the PCR mix was applied to the PicoTiterPlate™. Regardless of the cause of the mixed template beads, the hybridization results show that PCR amplification in the PicoTiterPlate™ microwells can drive sufficient product to the DNA capture beads to enable fluorescent probe hybridization and detection.

Discussion

The results in this example demonstrate that PicoTiterPlate™-based PCR alleviates many factors associated with the DNA amplification process, such as high costs of reagents, large numbers of reactions, and lengthy reaction times, delivering another "evolutionary jump" in PCR technology. The microwells on a single PicoTiterPlate™ can function as up to 370,000 discrete reaction vessels achieving high yield (2.3×10^6 to 1.2×10^9 fold) amplification even at reaction volumes as low as 39.5 picoliters. As a result, throughput is increased, and the total reagent cost for PTPCR is reduced; the reaction volume contained in an entire 26 or 76 μm deep PicoTiterPlate™ is 15.3 and 43 μl, respectively. Increases in the size of the PicoTiterPlate™ can further increase the maximal throughput. For example, increasing the PicoTiterPlate™ dimensions to 40 mm x 75 mm provides approximately 1.4×10^6 discrete reaction vessels, and a PicoTiterPlate™ possessing the same perimeter dimensions as a commercially available 96-well PCR plate (85.47 mm x 127.81 mm) could contain as many as 5.24×10^6 wells.

Solution phase PCR amplifications, regardless of the number and volume in which they are conducted, are of limited utility unless the product can be recovered easily and efficiently. Previous efforsts in parallel PCR (Nagai, H., *et al.*, *Anal. Chem.* 2001, 73, 1043-1047) required evaporation of the liquid reaction mixture, leaving the amplicon dried to the walls of the microreactor, after which it could be recovered for further manipulations. The methodology disclosed herein avoids the problems of product recovery by including solid

IDT Exhibit 1020.6661

phase amplification, immobilizing the PCR product to a DNA capture bead. Thus, the product of a PicoTiterPlate™ microwell reaction is not 370,000 wells containing solution-phase PCR product, but up to 370,000 beads bound with immobilized PCR product. These PCR products are suitable for numerous solid-phase methods of nucleic acid interrogation including the potential capacity to support a massively parallel approach to sequencing whole genomes containing up to hundreds of millions of bases. The simplicity of the disclosed method would drastically reduce costs for sequencing and other applications now requiring robotics to maintain large-scale cloning and PCR.

The disclosures of one or more embodiments of the invention are set forth in the accompanying description. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are now described. Other features, objects, and advantages of the invention will be apparent from the description and from the claims. In the specification and the appended claims, the singular forms include plural referents unless the context clearly dictates otherwise. 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. Unless expressly stated otherwise, the techniques employed or contemplated herein are standard methodologies well known to one of ordinary skill in the art. The examples of embodiments are for illustration purposes only. All patents and publications cited in this specification are incorporated by reference.

Example 28: Rig Sequencing Method

Step 1: Preparation of pAdEasy PCR DNA Beads

This proceedure was used for a 384-well plate PCR of Adenovirus clones. Streptavidin-Sepharose beads (12 mls) were prepared for binding PCR fragments by washing once with 2 M NaCl solution and resuspending in 288 mls of 2 M NaCl. The washed beads were transferred to fifteen 96-well plates at 200 µl of bead suspension/well. The PCR products (25 µl) were transferred to a 384-deep-well plate using a Tecan TeMo robot. To bind DNA to solid supports, 25 µl of bead suspension (15,000 beads) were added to each well of every 384-deep well plate using a Tecan TeMo robot and mixed. The final concentration of NaCl in the binding reaction was 1 M. The binding reaction was incubated with shaking at room temperature for 3 hr on a shaker. The contents of the microtiter plates were pooled by inverting the 384-well plates on to a reservoir and centrifuging at 1000g in a Beckman

IDT Exhibit 1020.6662

Allegra benchtop centrifuge. The pooled beads were transferred into a 50 ml Falcon tube, centrifuged at 1000g, and the supernatant was removed.

Approximately a million beads (mobile solid support) were washed once with 100 µl of 2 M NaCl followed by two washes with distilled water (100 µl each). The washed beads were incubated in 300 µl melting reagent (0.1 M NaCl and 0.125 M NaOH) for 10 minutes in a rotator to remove the non-biotinylated DNA strand. The tube was centrifuged at maximum speed to pellet the beads and the melt solution was removed and discarded. The beads were washed with 100 µl of melt solution followed by three more washes with 1X Annealing buffer. After the washes, the beads were resuspended in 25 µl of 1X Annealing buffer.

Primer P2 (500 pmoles) was added to the bead mixture and mixed. The bead mixture, in tubes, was placed into an automated incubator (PCR thermocycler in this case) with the following temperature profiles: incubation at 60°C for 5 minutes, decrease 0.1°C/second down to 50°C, incubation at 50°C for 5 minutes, decrease 0.1°C/second down to 40°C, incubation at 40°C for 5 minutes, decrease 0.1°C/second down to 4°C, incubation at 4°C forever.

After annealing, the beads were washed carefully and resuspended in 200 µl of *Bst* DNA polymerase binding solution. Then, 10 µl aliquots (50,000 beads) of the bead suspension were processed for sequencing on the instrument described below.

Step 2: Preparation of Control DNA Beads

Six control DNA sequences TF 2, 7, 9, 10, 12 and 15 were cloned into pBluescript II KS + vector and plasmid DNA was used as template for PCR with one biotinylated primer for solid-phase immobilization of the amplicons.

The following reagents were added to a 1.7 ml tube to create a PCR mix.

10 X HIFI buffer	100 µl
10 mM dNTP mix	100 µl
50 mM $MgSO_4$	60 µl
5'-Bio-3HEG-MMP1B	10 µl
MMP1A	10 µl
HIFI *Taq* Polymerase	10 µl
Mol. Bio. Grade Water	690 µl

IDT Exhibit 1020.6663

Twenty microliters of plasmid template DNA was added and the mix was aliquoted by 50 µl into 0.2 ml PCR tubes. The following program was used for thermocycling: incubation at 94°C for 4 minutes; 39 cycles of incubation at 94°C for 15 seconds, 58°C for 30 seconds, 68°C for 90 seconds, and 68°C for 120 seconds; hold at 10°C.

Amplified DNA for each test fragment was purified using the Qiagen MinElute PCR Clean-Up Kit as per manufacturer's instructions. The purity and yield of each of the test fragments DNA was assessed using the Agilent 2100 Bioanalyzer and DNA 500 reagent kit and chip. Biotinylated PCR products were immobilized onto Sepharose streptavidin beads at 10 million DNA copies/bead.

Beads were washed once with 2 M NaCl solution. This was done by adding 100 µl, vortexing briefly to resuspend the beads, centrifuging for 1 minute at maximum speed to pellet the beads, and then removing the supernatant. This was followed by a second wash with 2 M NaCl. The beads were then resuspended in 30 µl of 2 M NaCl. PCR product was added to beads. The mixture was vortexed to resuspend the beads in solution and then placed in a rack, on a titer plate shaker, at speed 7, for 1 hour at room temperature.

The non-biotinylated second strand was removed by incubation with the alkaline melt solution (0.1 M NaOH / 0.15 M NaCl) for 10 minutes in an overhead rotator at room temperature. This was followed by washing the beads once with 100 µl of melt solution and three times with 100 µl of 1X annealing buffer (50 mM Tris-Acetate, pH 7.5; 5 mM $MgCl_2$). Sequencing primer was annealed to the immobilized single-stranded DNA by centrifugation for one minute at maximum speed. The supernatant was removed and the beads were resuspended in 25 µl of 1X annealing buffer. Next, 5 µl of sequencing primer MMP7A (100 pmol/µl) was added to the bead suspension and the following temperature profile was used to hybridize the sequencing primer:

- Incubation at 60°C for 5 minutes;
- Decrease 0.1°C/second down to 50°C;
- Incubation 50°C for 5 minutes;
- Decrease 0.1°C/second down to 40°C;
- Incubation at 40°C for 5 minutes;
- Decrease 0.1°C/second down to 4°C; and
- Hold at 4°C.

Beads were washed twice with 100 µl of 1X annealing buffer and then resuspended to

IDT Exhibit 1020.6664

a final volume of 200 µl with 1X annealing buffer and stored in 10 µl aliquots in labeled tube strips in a 4°C refrigerator.

Step 3: Sequencing Chemistry

Sepharose beads with immobilized single stranded DNA templates and annealed sequencing primer were incubated with *E. coli* single strand binding protein (Amersham Biosciences) (5 µl of 2.5 µg/µl ssb stock solution per 50,000 beads) and 500 U (10 µl of 50U/µl) of Bst DNA polymerase (NEB) in 200 µl of Bst polymerase binding solution (25 mM Tricine pH 7.8; 5 mM magnesium acetate; 1 mM DTT; 0.4 mg/ml PVP MW 360,000) for 30 minutes at room temperature on a rotator. After this, the DNA beads were mixed with the SL beads and deposited into the wells of the PicoTiter Plate as follows. Reagents required for a sequencing run on a 454 instrument included 1) substrate wash solution; 2) apyrase containing wash solution; 3) 100 nM inorganic pyrophosphate calibration standard; 4) individual nucleotide triphosphate solutions.

All solutions were prepared in the sulfurylase–luciferase assay buffer with enzyme substrates (25 mM Tricine pH 7.8; 5 mM magnesium acetate; 0.4 mg/ml PVP MW 360,000; 0.01% Tween 20; 300 µM D-luciferin; 4 µM APS). The substrate wash solution was identical to the luciferase assay buffer. The apyrase containing wash solution was based on the luciferase assay buffer, except no enzyme substrates (APS and D-luciferin) were added and this wash contained apyrase (Sigma St. Lous, MO; Pyrosequencing AB, Pyrosequencing, Inc. Westborough, MA) in the final concentration of 8.5 U / l.

Sodium pyrophosphate (PP_i) standard was prepared by adding sodium pyrophosphate tetrabasic decahydrate (Sigma, St. Louis, MO) to the luciferase assay buffer to a final concentration 100 nM. Nucleotide triphosphates (dCTP, dGTP, TTP; minimum diphosphate grade) (Amersham Biosciences AB, Uppsala, Sweden) were diluted to final concentration of 6.5 µM in the luciferase assay buffer. Deoxyadenosine triphosphate analog, 2'-Deoxyadenosine–5'–O–(1- thiotriphosphate), Sp-isomer (Sp-dATP-α-S, Biolog Life Science Institute, Bremen, Germany) was diluted to final concentration of 50 µM in the luciferase assay buffer.

Step 4: Cloning His6-BCCP-sulfurylase and His6-BCCP-luciferase

Bacillus stearothermophilus (*Bst*) ATP sulfurylase (E.C. 2.7.7.4) and firefly (*Photinus pyralis)* luciferase (E.C. 1.13.12.7) were cloned into *Nhe* I-*BamH* I digested pRSET-A vector

IDT Exhibit 1020.6665

(Invitrogen). The coding sequence of the BCCP (biotin carboxyl carrier protein) gene (Alix,J.H., DNA 8 (10), 779-789 (1989); Muramatsu, S. and Mizuno, T., Nucleic Acids Res. 17 (10), 3982 (1989), Jackowski, S. and Alix, J.H., J. Bacteriol. 172 (7), 3842-3848 (1990); Li, S.J. and Cronan, J.E. Jr., J. Biol. Chem. 267 (2), 855-863 (1992), Genbank accession number M80458) was used to design PCR primers to amplify the fragment corresponding to amino acids 87-165 of the BCCP protein. The forward primer was 5'-ctagctagcatggaagcgccagcagca-3'; SEQ ID NO:42 and the reverse primer was 5'-ccgggatccctcgatgacgaccagcggc-3'; SEQ ID NO:43. The PCR cocktail was prepared as Mix 1 and Mix 2, 25 µl each. Mix 1 included 75 pmoles of primers, 100 ng of *E. coli* genomic DNA and 5 µmoles of dNTPs. Mix 2 included 1 unit of Fidelity Expand DNA polymerase (Boehringer Mannheim/Roche Diagnostics Corporation, Indianapolis, IN, Cat. No. 1 732 641) and 5 µl of 10X Fidelity Expand buffer (Boehringer Mannheim/Roche Diagnostics Corporation, Indianapolis, IN). To allow PCR hot-start, Mix 1 and Mix 2 were heated separately for 20 seconds at 96°C before they were pooled. The pooled reaction was cycled as follows: incubation at 96°C for 3 min, 10 cycles of incubation at 96°C for 30 sec, 55°C for 1 min, and 68°C for 2 min, then 20 cycles of incubation at 96°C for 30s ec, 60°C for 1 min, and 68°C for 2 min, followed by a polishing step of incubation at 72°C for 7 min. After PCR, a single 250 bp fragment was obtained. The BCCP fragment was digested with *Nhe* I and *BamH* I and subcloned into *Nhe* I-*BamH* I digested pRSET-A.

Step 5: Expression of Sulfurylase and Luciferase.

The *Bst* ATP sulfurylase and *P. pyralis* luciferase open reading frames were amplified by PCR with primers that contain *Pst* I/*Hind* III and *BamH* I/*Xho* I sites (the first enzyme was at the 5' end and the second enzyme was at the 3'end), respectively. This produced an N-terminal fusion of 6X His and BCCP domain to ATP sulfurylase and luciferase. The enzymes were expressed in *E. coli* using biotin-supplemented growth media to allow for *in-vivo* biotinylation via the BCCP domain. The enzymes were purified to near homogeneity using a combination of IMAC and a size-exclusion column chromatography. Purification was assessed by electrophoresis using the Agilent 2100 Bioanalyzer on Protein 200 Plus chips.

Step 6: Solid Phase Immobilization of Luciferase and Sulfurylase

The enzymes were immobilized onto Dynal M-280 streptavidin coated magnetic

IDT Exhibit 1020.6666

microparticles (Dynal, Oslo, Norway) and Bangs microspheres (300 nm) by incubation of 1:3 mixture of ATP sulfurylase and luciferase, respectively. The binding was performed by mixing 50 μg of ATP sulfurylase and 150 μg of luciferase with 1 mg of Dynal M-280 beads or 0.6 mg of Bangs microspheres in TAGE buffer (25 mM Tris-Acetate pH 7.8, 200 mM ammonium sulfate, 15% v/v glycerol and 30% v/v ethylene glycol). The mixture was incubated for 1 hour at 4°C on a rotator. After binding, the beads could be stored at -20 °C in the enzyme solution for 3 months. Before use, beads were washed thoroughly in luciferase assay buffer containing 0.1 mg/ml bovine serum albumin (Sigma, St Louis, MO). Immobilized enzyme activity was assayed using a luminometer (Turner, Sunnyvale, California). Washed beads were stored on ice until deposition onto a PTP slide.

Step 7: PicoTiterPlates™ (PTPs)

The PicoTiterPlates™ (25 x 75 x 2 mm) were manufactured by anisotropic etching of fiber optic face plates in a manner similar to that described in literature. Plates were etched in three different microwell depths, 26, 50 and 76 mm. Microwell center-to-center pitch was 50 μm, and well diameters ranged between 39 and 44 μm with a calculated well density of 480 wells/mm^2.

Step 8: PTP Loading

Sepharose beads carrying DNA templates and Dynal M-280 / Bangs 0.3 μm bead mixture with immobilized sulfurylase and luciferase enzymes were deposited into individual wells of a PicoTiter plate using a centrifugation-based method. The procedure employed an in-house polycarbonate fixture (jig) which included a bottom plate (with slide positioning pegs), an elastomer sealing gasket, and a top plate with two loading ports. The PTP slide was placed onto the bottom plate with the etched side facing up and the top plate with sealing gasket in place was clamped on top of the PTP slide. The whole assembly was tightened with four plastic screws in order to provide a water-tight seal. The sealing gasket was designed to form a mask for bead deposition, resulting in one hexagonal area (14 x 43 mm) covering roughly 270,000 PTP wells.

Beads were deposited in ordered layers. The PTP was removed from incubating in Bead Wash Buffer. Layer 1, a mix of DNA and enzyme beads, was deposited. After centrifuging, Layer 1 supernatant was aspirated off the PTP and Layer 2, Dynal enzyme beads, was deposited.

IDT Exhibit 1020.6667

A bead suspension was prepared by mixing 150,000 DNA carrying Sepharose beads in 120 µl of the ssb/*Bst* pol binding mix (see above) with 270 µl of Dynal-SL and Bangs-SL beads (both at 10 mg/ml) in a total volume of 500 µl of the luciferase assay buffer containing 0.1 mg/ml bovine serum albumin. The bead slurry was vortexed and flowed into the bead deposition jig through pipetting ports. Care was taken to aviod introducing air bubbles. The jig/PTP assembly was centrifuged at 2000 rpm for 8 minutes in a Beckman Allegra 6 centrifuge equipped with a 4-position plate swing-out rotor. After centrifugation, the supernatant was carefully removed from the jig chamber using a pipette. A second layer of only Dynal-SL beads was deposited. This layer included 125 µl of Dynal-SL (at 10 mg/ml) and 375 µl Bead Wash Buffer in a 1.5 ml tube (2.5 mg/ml Dynal beads). The Dynal bead mixture was pipetted into the PTP main active area and centrifuged for 8 minutes at 2000 rpm. Layer 2 mixture was aspirated and the PTP was placed back into Bead Wash Buffer (luciferase assay buffer with 0.1 mg/ml bovine serum albumin and 8.5 U/l apyrase) until ready to load onto the Sequencer.

Step 9: Sequencing Instrument

The in-house sequencing instrument included three major assemblies: a fluidics subsystem, a PTP cartridge/flow chamber, and an imaging subsystem. The fluidics subsystem included reagent reservoirs, reagents inlet lines, a multi-valve manifold, and a peristaltic pump. It allowed for reagent delivery into the flow chamber, one reagent at a time, at a pre-programmed flow rate and duration. The PTP cartridge/flow chamber was designed in such a way that after attaching a PTP, there would be 300 µm space between the PTP top (etched side) and the chamber ceiling. It included means for temperature control of the reagents and PTP, as well as a light-tight housing. The polished side of a PTP was exposed at the back side of the PTP cartridge and was placed directly in contact with the imaging system. The imaging system comprised a CCD camera with a 1-1 imaging fiber bundle, as well as cryogenic cooling system for the camera, and camera control electronics. The camera used was a Spectral Instruments (Tucson, AZ) series 600 camera with a Fairchild Imaging LM485 CCD (16 million pixels, 15 µm pixel size). This was bonded directly to the imaging fiber bundle with 6 µm fiber pitch. The camera was cooled to –70° C and operated in a frame transfer mode. In this way, the center portion of the CCD was used for imaging while the outer portion of the CCD was used for image storage and read-out. The read-out occurred through 4 ports at each corner of the CCD. The data acquisition rate was set to 1 frame per

IDT Exhibit 1020.6668

30 seconds. The frame-transfer shift time was approximately 0.25 seconds. All camera images were stored in a UTIFF 16 format on a computer hard drive (IBM eServer xSeries 335, IBM, White Plains, NY).

Step 10: Sequencing Run Conditions

The cyclical delivery of sequencing reagents into the PTP wells and washing of the sequencing reaction byproducts from the wells was achieved by a pre-programmed operation of the fluidics system. The program was written in a form of a Microsoft Excel script, specifying the reagent name (Wash, dATPαS, dCTP, dGTP, dTTP, PP_i standard), flow rate and duration of each script step. Flow rate was set at 3 ml/min for all reagents and the linear velocity within the flow chamber was approximately. An initial wash step (5 minutes) was followed by a PP_i standard flow (2 min), followed by 21 or 42 cycles of (Wash – C – Wash – A – Wash – G – Wash – T), where each nucleotide flow was 0.5 minute and wash steps were 2 minutes. After all cycles of nucleotide additions and washes, a second PP_i standard flow (2 min) was delivered, followed by a final 5 minutes wash step. The total run time was 4 hours. Reagent volumes required to complete this run script were as follows: 300 ml each wash solution, 50 ml of each nucleotide solution, 20 ml of PP_i standard solution. During the run, all reagents were kept at room temperature. Because the flow chamber and flow chamber inlet tubing were maintained at 3 °C, all reagents entering the flow chamber were at 30°C.

References

Hamilton, S.C., J.W. Farchaus and M.C. Davis. 2001. DNA polymerases as engines for biotechnology. *BioTechniques* 31:370.

QiaQuick Spin Handbook (QIAGEN, 2001): hypertext transfer protocol://world wide web.qiagen.com/literature/handbooks/qqspin/1016893HBQQSpin_PCR_mc_prot.pdf.

Quick Ligation Kit (NEB): hypertext transfer protocol://world wide web.neb.com/neb/products/mod_enzymes/M2200.html.

MinElute kit (QIAGEN): hypertext transfer protocol://world wide web.qiagen.com/literature/handbooks/minelute/1016839_HBMinElute_Prot_Gel.pdf.

Biomagnetic Techniques in Molecular Biology, Technical Handbook, 3rd edition (Dynal, 1998): hypertext transfer protocol://world wide web.dynal.no/kunder/dynal/DynalPub36.nsf/cb927fbab127a0ad4125683b004b011c/4908f5b1a665858a41256adf005779f2/$FILE/Dynabeads M-280 Streptavidin.pdf.

IDT Exhibit 1020.6669

Bio Analyzer User Manual (Agilent): hypertext transfer protocol://world wide web.chem.agilent.com/temp/rad31B29/00033620.pdf

BioAnalyzer DNA and RNA LabChip Usage (Agilent): hypertext transfer protocol://world wide web.agilent.com/chem/labonachip

BioAnalyzer RNA 6000 Ladder (Ambion): hypertext transfer protocol://world wide web.ambion.com/techlib/spec/sp_7152.pdf

IDT Exhibit 1020.6670

CLAIMS

WE CLAIM:

1. A method for sequencing nucleic acids comprising:

(a) fragmenting large template nucleic acid molecules to generate a plurality of fragmented nucleic acids;

(b) delivering the fragmented nucleic acids into aqueous microreactors in a water-in-oil emulsion such that a plurality of aqueous microreactors comprise a single copy of a fragmented nucleic acid, a single bead capable of binding to the fragmented nucleic acid, and amplification reaction solution containing reagents necessary to perform nucleic acid amplification;

(c) amplifying the fragmented nucleic acids in the microreactors to form amplified copies of said nucleic acids and binding the amplified copies to beads in the microreactors;

(d) delivering the beads to an array of at least 10,000 reaction chambers on a planar surface, wherein a plurality of the reaction chambers comprise no more than a single bead; and

(e) performing a sequencing reaction simultaneously on a plurality of the reaction chambers.

2. The method of claim 1 wherein the reaction chambers have a center to center spacing of 20 to 100 μm.

3. The method of claim 1 wherein the fragmented nucleic acids are 30 – 500 bases.

4. The method of claim 1 wherein a plurality of the beads bind at least 10,000 amplified copies.

5. The method of claim 1 wherein step (c) is accomplished using polymerase chain reaction.

6. The method of claim 1 wherein the sequencing reaction is a pyrophosphate-based sequencing reaction.

7. The method of claim 1 wherein the sequencing reaction comprises the steps of:

(a) annealing an effective amount of a sequencing primer to the amplified copies of the nucleic acid and extending the sequencing primer with a polymerase and a predetermined nucleotide triphosphate to yield a

IDT Exhibit 1020.6671

sequencing product and, if the predetermined nucleotide triphosphate is incorporated onto a 3' end of said sequencing primer, a sequencing reaction byproduct; and

(b) identifying the sequencing reaction byproduct, thereby determining the sequence of the nucleic acid in a plurality of the reaction chambers.

8. The method of claim 1 wherein the sequencing reaction comprises the steps of:

(a) hybridizing two or more sequencing primers to one or a plurality of single strands of the nucleic acid molecule wherein all the primers except for one are reversibly blocked primers;

(b) incorporating at least one base onto the nucleic acid molecule by polymerase elongation from an unblocked primer;

(c) preventing further elongation of said unblocked primer;

(d) deblocking one of the reversibly blocked primers into an unblocked prime and

(e) repeating steps (b) to (d) until at least one of the reversibly blocked primers are deblocked and used for determining a sequence.

9. The method of claim 1 wherein the reaction chambers are cavities formed by etching one end of a fiber optic bundle.

10. An array comprising a planar surface with a plurality of cavities thereon, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 20 to 100 μm and each cavity has a width in at least one dimension of between 20 μm and 70 μm, and wherein there are at least 10,000 reaction chambers.

11. The array of claim 10 wherein a plurality of reaction chambers contains at least 100,000 copies of a single species of single stranded nucleic acid template.

12. The array of claim 11 wherein the single stranded nucleic acid templates are immobilized on mobile solid supports disposed in the reaction chambers.

13. The array of claim 10 wherein the center to center spacing is between 40 to 60 μm.

IDT Exhibit 1020.6672

14. The array of claim 10 wherein each cavity has a depth of between 20 μm and 60 μm.

15. An array comprising a planar top surface and a planar bottom surface wherein the planar top surface has at least 10,000 cavities thereon, each cavity forming an analyte reaction chamber, the planar bottom surface is optically conductive such that optical signals from the reaction chambers can be detected through the bottom planar surface, wherein the distance between the top surface and the bottom surface is no greater than 5 mm, wherein the reaction chambers have a center to center spacing of between 20 to 100μm and each chamber having a width in at least one dimension of between 20μm and 70μm.

16. The array of claim 15 wherein the distance between the top surface and the bottom surface is no greater than 2 mm.

17. The array of claim 10 or 15 wherein the number of cavities is greater than 50,000.

18. The array of claim 10 or 15 wherein the number of cavities is greater than 100,000.

19. The array of claim 10 or 15 wherein the shape of each reaction chamber is substantially hexagonal.

20. The array of claim 10 or 15 wherein each cavity has at least one irregular wall surface.

21. The array of claim 10 or 15 wherein the array is formed in a fused fiber optic bundle.

22. The array of claim 10 or 15 wherein each cavity has a smooth wall surface.

23. The array of claim 10 or 15 wherein the cavities are formed by etching one end of the fiber optic bundle.

24. The array of claim 10 or 15 wherein each cavity contains reagents for analyzing a nucleic acid or protein.

25. The array of claim 10 or 15 further comprising a second surface spaced apart from the planar array and in opposing contact therewith such that a flow chamber is formed over the array.

26. An array means for carrying out separate parallel common reactions in an aqueous environment, wherein the array means comprises a substrate comprising at least 10,000 discrete reaction chambers containing a starting material that is capable of reacting with a

IDT Exhibit 1020.6673

reagent, each of the reaction chambers being dimensioned such that when one or more fluids containing at least one reagent is delivered into each reaction chamber, the diffusion time for the reagent to diffuse out of the well exceeds the time required for the starting material to react with the reagent to form a product.

27. The array of claim 26 wherein each cavity contains reagents for analyzing a nucleic acid or protein.

28. The array of claim 26 further comprising a population of mobile solid supports disposed in the reaction chambers, each mobile solid support having one or more bioactive agents attached thereto.

29. The array of claim 26 wherein the cavities are formed in the substrate via etching, molding or micromaching.

30. The array of claim 17 wherein the substrate is a fiber optic bundle.

31. The array of claims 10, 15 or 26 wherein at least 5% to 20% of the reaction chambers contain at least one mobile solid support having at least one reagent immobilized thereon.

32. The array of claims 10, 15 or 26 wherein at least 20% to 60% of the reaction chambers have at least one mobile solid support having at least one reagent immobilized thereon.

33. The array of claims 10, 15 or 26 wherein at least 50% to 100% of the reaction chambers have at least one mobile solid support having at least one reagent immobilized thereon.

34. The array of claim 31 wherein the reagent immobilized on the mobile solid support is a polypeptide with sulfurylase activity.

35. The array of claims 31 wherein the reagent immobilized on the mobile solid support is a polypeptide with luciferase activity.

36. The array of claims 31 wherein the mobile solid support has both sulfurylase and luciferase immobilized.

37. The array of claim 31 wherein a plurality of reaction chambers contains at least 100,000 copies of a single species of single stranded nucleic acid template.

38. The array of claim 31 wherein the single stranded nucleic acid templates are immobilized

IDT Exhibit 1020.6674

on mobile solid supports disposed in the reaction chambers.

39. The array of claim 10, 15 or 26 wherein the nucleic acid is suitable for use in a pyrosequencing reaction.

40. A method for delivering a bioactive agent to an array, comprising dispersing over the array a plurality of mobile solid supports, each mobile solid support having at least one reagent immobilized thereon, wherein the reagent is suitable for use in a nucleic acid sequencing reaction, where the array comprises a planar surface with a plurality of reaction chambers disposed thereon, wherein the reaction chambers have a center to center spacing of between 20 to 100 μm and each reaction chamber has a width in at least one dimension of between 20 μm and 70 μm.

41. An apparatus for simultaneously monitoring an array of reaction chambers for light indicating that a reaction is taking place at a particular site, the apparatus comprising:

(a) an array of reaction chambers formed from a planar substrate comprising a plurality of cavitated surfaces, each cavitated surface forming a reaction chamber adapted to contain analytes, and wherein the reaction chambers have a center to center spacing of between 20 to 100 μm, each reaction chamber having a volume of between 10 to 150 pL, the array comprising more than 10,000 discrete reaction chambers;

(b) an optically sensitive device arranged so that in use the light from a particular reaction chamber will impinge upon a particular predetermined region of said optically sensitive device;

(c) means for determining the light level impinging upon each of said predetermined regions and

(d) means to record the variation of said light level with time for each of said reaction chamber.

42. An analytic sensor, comprising:

(a) an array formed from a first bundle of optical fibers with a plurality of cavitated surfaces at one end thereof, each cavitated surface forming a reaction chamber adapted to contain analytes, and wherein the reaction chambers have a center to center spacing of between 20 to 100 μm, a width of 20 to 70 μm, the array comprising more than 10,000 discrete reaction chambers;

IDT Exhibit 1020.6675

(b) an enzymatic or fluorescent means for generating light in the reaction chambers;

(c) light detection means comprising a light capture means and a second fiber optic bundle for transmitting light to the light detecting means, the second fiber optic bundle being in optical contact with the array, such that light generated in an individual reaction chamber is captured by a separate fiber or groups of separate fibers of the second fiber optic bundle for transmission to the light capture means.

43. The sensor of claim 42 wherein said sensor is suitable for use in a biochemical assay.

44. The sensor of claim 42 wherein said sensor is suitable for use in a cell-based assay.

45. The sensor of claim 42 wherein the light capture means is a CCD camera.

46. The sensor of claim 42 wherein the reaction chambers contain one or more mobile solid supports with a bioactive agent immobilized thereon.

47. A method for carrying out separate parallel common reactions in an aqueous environment, comprising:

(a) delivering a fluid containing at least one reagent to an array, wherein the array comprises a substrate comprising at least 10,000 discrete reaction chambers, each reaction chamber adapted to contain analytes, and wherein the reaction chambers have a volume of between 10 to 150 pL and containing a starting material that is capable of reacting with the reagent, each of the reaction chambers being dimensioned such that when the fluid is delivered into each reaction chamber, the diffusion time for the reagent to diffuse out of the well exceeds the time required for the starting material to react with the reagent to form a product; and

(b) washing the fluid from the array in the time period (i) after the starting material has reacted with the reagent to form a product in each reaction chamber but (ii) before the reagent delivered to any one reaction chamber has diffused out of that reaction chamber into any other reaction chamber.

48. The method of claim 47 wherein the product formed in any one reaction chamber is independent of the product formed in any other reaction chamber, but is generated using one or more common reagents

49. The method of claim 47 wherein the starting material is a nucleic acid sequence and at least one reagent in the fluid is a nucleotide or nucleotide analog

IDT Exhibit 1020.6676

50. The method of claim 47 wherein the fluid additionally comprises a polymerase capable of reacting the nucleic acid sequence and the nucleotide or nucleotide analog

51. The method of claim 47 additionally comprising repeating steps (a) and (b) sequentially.

52. A method for delivering nucleic acid sequencing enzymes to an array, said array having a planar surface with a plurality of cavities thereon, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 20 to 100 μm; the method comprising dispersing over the array a plurality of mobile solid supports having one or more nucleic acid sequencing enzymes immobilized thereon, such that a plurality of the reaction chambers contain at least one mobile solid support.

53. The method of claim 52 wherein one of the nucleic acid sequencing enzymes is a polypeptide having sulfurylase activity, luciferase activity or both.

54. A method for delivering a plurality of nucleic acid templates to an array, said array having a planar surface with a plurality of cavities thereon, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 20 to 100 μm and the array having at least 10,000 reaction chambers; the method comprising dispersing over the array a plurality of mobile solid supports, each mobile solid support having no more than a single species of nucleic acid template immobilized thereon, the dispersion causing no more than one mobile solid support to be disposed within any one reaction chamber.

55. The method of claim 54 wherein the nucleic acid sequence is a single stranded nucleic acid.

56. The method of claim 54 wherein at least 100,000 copies of a single species of nucleic acid template are immobilized on a plurality of the mobile solid supports.

57. The method of claim 54 wherein each single species of nucleic acid template is amplified on a picotiter plate to produce at least 2,000,000 copies per well of said nucleic acid template after being disposed in the reaction chamber.

58. The method of claim 57 wherein the nucleic acid sequence is amplified using an amplification technology selected from the group consisting of polymerase chain reaction, ligase chain reaction and isothermal DNA amplification.

IDT Exhibit 1020.6677

59. A method for sequencing a nucleic acid, the method comprising:

(a) providing a plurality of single-stranded nucleic acid templates disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 20 to 100 μm and the planar surface has at least 10,000 reaction chambers;

(b) performing a pyrophosphate based sequencing reaction simultaneously on all reaction chambers by annealing an effective amount of a sequencing primer to the nucleic acid templates and extending the sequencing primer with a polymerase and a predetermined nucleotide triphosphate to yield a sequencing product and, if the predetermined nucleotide triphosphate is incorporated onto the 3' end of said sequencing primer, a sequencing reaction byproduct; and

(c) identifying the sequencing reaction byproduct, thereby determining the sequence of the nucleic acid in each reaction chamber.

60. The method of claim 59, wherein the sequencing reaction byproduct is PPi and a coupled sulfurylase/luciferase reaction is used to generate light for detection.

61. The method of claim 60, wherein either or both of the sulfurylase and luciferase are immobilized on one or more mobile solid supports disposed at each reaction site.

62. A method of determining the base sequence of a plurality of nucleotides on an array, the method comprising:

(a) providing at least 10,000 DNA templates, each separately disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 20 to 100 μm, and a volume of between 10 to 150 pL; wherein

(b) adding an activated nucleotide 5'-triphosphate precursor of one known nitrogenous base to a reaction mixture in each reaction chamber, each reaction mixture comprising a template-directed nucleotide polymerase and a single-stranded polynucleotide template hybridized to a complementary oligonucleotide primer strand at least one nucleotide residue shorter than the templates to form at least one unpaired nucleotide residue in each template at the 3'-end of the primer strand, under reaction conditions which allow incorporation of the activated nucleoside 5'-triphosphate precursor onto the 3'-end of the primer strands,

IDT Exhibit 1020.6678

provided the nitrogenous base of the activated nucleoside 5'-triphosphate precursor is complementary to the nitrogenous base of the unpaired nucleotide residue of the templates;

(c) detecting whether or not the nucleoside 5'-triphosphate precursor was incorporated into the primer strands in which incorporation of the nucleoside 5'-triphosphate precursor indicates that the unpaired nucleotide residue of the template has a nitrogenous base composition that is complementary to that of the incorporated nucleoside 5'-triphosphate precursor; and

(d) sequentially repeating steps (b) and (c), wherein each sequential repetition adds and, detects the incorporation of one type of activated nucleoside 5'-triphosphate precursor of known nitrogenous base composition; and

(e) determining the base sequence of the unpaired nucleotide residues of the template in each reaction chamber from the sequence of incorporation of said nucleoside precursors.

63. A method of identifying the base in a target position in a DNA sequence of template DNA, wherein:

(a) at least 10,000 separate DNA templates are separately disposed within a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 20 to 100 μm, said DNA being rendered single stranded either before or after being disposed in the reaction chambers,

(b) an extension primer is provided which hybridizes to said immobilized single-stranded DNA at a position immediately adjacent to said target position;

(c) said immobilized single-stranded DNA is subjected to a polymerase reaction in the presence of a predetermined deoxynucleotide or dideoxynucleotide, wherein if the predetermined deoxynucleotide or dideoxynucleotide is incorporated onto the 3' end of said sequencing primer then a sequencing reaction byproduct is formed; and

(d) identifying the sequencing reaction byproduct, thereby determining the nucleotide complementary to the base at said target position in each of the 10,000 DNA templates.

64. The method of claim 63 wherein in place of deoxy- or dideoxy adenosine triphosphate (ATP) a dATP or ddATP analogue is used which is capable of acting as a substrate for a polymerase but incapable of acting as a substrate for said PPi—detection enzyme.

65. An apparatus for analyzing a nucleic acid sequence, the apparatus comprising:

IDT Exhibit 1020.6679

(a) a reagent delivery cuvette, wherein the cuvette includes an array comprising a planar surface with a plurality of cavities thereon, each cavity forming an analyte reaction chamber, wherein the reaction chambers have a center to center spacing of between 20 to 100 μm, and there are in excess of 10,000 reaction chambers, and wherein the reagent delivery cuvette contains reagents for use in a sequencing reaction;

(b) a reagent delivery means in communication with the reagent delivery cuvette;

(c) an imaging system in communication with the reagent delivery chamber; and

(d) a data collection system in communication with the imaging system.

66. An apparatus for determining the base sequence of a plurality of nucleotides on an array, the apparatus comprising:

(a) a reagent cuvette containing a plurality of cavities on a planar surface, each cavity forming an analyte reaction chamber, wherein there are in excess of 10,000 reaction chambers, each having a center to center spacing of between 20 to 100 μm and a volume of between 10 to 150 pL;

(b) reagent delivery means for simultaneously adding to each reaction chamber an activated nucleotide 5'-triphosphate precursor of one known nitrogenous base to a reaction mixture in each reaction chamber, each reaction mixture comprising a template-directed nucleotide polymerase and a single-stranded polynucleotide template hybridized to a complementary oligonucleotide primer strand at least one nucleotide residue shorter than the templates to form at least one unpaired nucleotide residue in each template at the 3'-end of the primer strand, under reaction conditions which allow incorporation of the activated nucleoside 5'-triphosphate precursor onto the 3'-end of the primer strands, provided the nitrogenous base of the activated nucleoside 5'-triphosphate precursor is complementary to the nitrogenous base of the unpaired nucleotide residue of the templates;

(c) detection means for detecting in each reaction chamber whether or not the nucleoside 5'-triphosphate precursor was incorporated into the primer strands in which incorporation of the nucleoside 5'-triphosphate precursor indicates that the unpaired nucleotide residue of the template has a nitrogenous base composition that is complementary to that of the incorporated nucleoside 5'-triphosphate precursor; and

(d) means for sequentially repeating steps (b) and (c), wherein each sequential repetition adds and, detects the incorporation of one type of activated nucleoside

IDT Exhibit 1020.6680

5'-triphosphate precursor of known nitrogenous base composition; and

(e) data processing means for determining the base sequence of the unpaired nucleotide residues of the template simultaneously in each reaction chamber from the sequence of incorporation of said nucleoside precursors.

67. An apparatus for processing a plurality of analytes, the apparatus comprising:

(a) a flow chamber having disposed therein a substrate comprising at least 50,000 cavitated surfaces on a fiber optic bundle, each cavitated surface forming a reaction chamber adapted to contain analytes, and wherein the reaction chambers have a center to center spacing of between 20 to 100 μm and a diameter of 20 to 70 μm;

(b) fluid means for delivering processing reagents from one or more reservoirs to the flow chamber so that the analytes disposed in the reaction chambers are exposed to the reagents; and

(c) detection means for simultaneously detecting a sequence of optical signals from each of the reaction chambers, each optical signal of the sequence being indicative of an interaction between a processing reagent and the analyte disposed in the reaction chamber, wherein the detection means is in communication with the cavitated surfaces.

68. The apparatus of claim 67 wherein the detection means is a CCD camera.

69. The apparatus of claim 67 wherein the analyte is nucleic acid.

70. The apparatus of claim 67 wherein the analytes are immobilized on one or more mobile solid supports that are disposed in the reaction chamber.

71. The apparatus of claim 67 wherein the processing reagents are immobilized on one or more mobile solid supports.

72. A method for sequencing a nucleic acid, the method comprising:

(a) providing a plurality of single-stranded nucleic acid templates in an array having at least 50,000 discrete reaction sites;

(b) contacting the nucleic acid templates with reagents necessary to perform a pyrophosphate-based sequencing reaction coupled to light emission;

(c) detecting the light emitted from a plurality of reaction sites on respective portions of an optically sensitive device;

IDT Exhibit 1020.6681

(d) converting the light impinging upon each of said portions of said optically sensitive device into an electrical signal which is distinguishable from the signals from all of said other reaction sites;

(e) determining the sequence of the nucleic acid templates based on light emission for each of said discrete reaction sites from the corresponding electrical signal.

73. The method of claim 1 further comprising the steps of:

(a) uniquely tagging fragmented nucleic acids from different biological sources libraries to create libraries of fragmented nucleic acids with different detectable sequence tags;

(b) sequencing said fragmented nucleic acids and detecting said detectable sequence tag from each said tagged nucleic acid fragment.

74. The method of claim 1 wherein the libraries are delivered individually or wherein the libraries are mixed and delivered simultaneously.

75. The method of claim 1 wherein said detectable sequence tag comprises an oligonucleotide of between 2 and 50 bases.

76. A method for sequencing nucleic acids comprising:

(a) fragmenting large template nucleic acid molecules to generate a plurality of fragmented nucleic acids;

(b) attaching one strand of a plurality of the fragmented nucleic acids individually to beads to generate single stranded nucleic acids attached individually to beads;

(c) delivering a population of the single stranded fragmented nucleic acids attached individually to beads to an array of at least 10,000 reaction chambers on a planar surface, wherein a plurality of the wells comprise no more than a one bead with on single stranded fragmented nucleic acid;

(d) performing a sequencing reaction simultaneously on a plurality of the reaction chambers.

77. The method of claim 76 wherein the reaction chambers have a center to center spacing of between 20 to 100 μm.

78. The method of claim 76 wherein the fragmented nucleic acids are between 30 – 500 bases.

IDT Exhibit 1020.6682

79. The method of claim 76 wherein the fragmented nucleic acids are amplified in the reaction chambers prior to step (d).

80. The method of claim 76 wherein the amplifying step is accomplished using polymerase chain reaction.

81. The method of claim 76 wherein the sequencing reaction is a pyrophosphate-based sequencing reaction.

82. The method of claim 76 wherein the sequencing reaction comprises the steps of:

(f) annealing an effective amount of a sequencing primer to the single stranded fragmented nucleic acid templates and extending the sequencing primer with a polymerase and a predetermined nucleotide triphosphate to yield a sequencing product and, if the predetermined nucleotide triphosphate is incorporated onto the 3' end of said sequencing primer, a sequencing reaction byproduct; and

(g) identifying the sequencing reaction byproduct, thereby determining the sequence of the nucleic acid in a plurality of the reaction chambers.

83. The method of claim 76 wherein the sequencing reaction comprises the steps of:

(a) hybridizing two or more sequencing primers to one or a plurality of single strands of the nucleic acid molecule wherein all the primers except for one are reversibly blocked primers;

(b) incorporating at least one base onto the nucleic acid molecule by polymerase elongation from an unblocked primer;

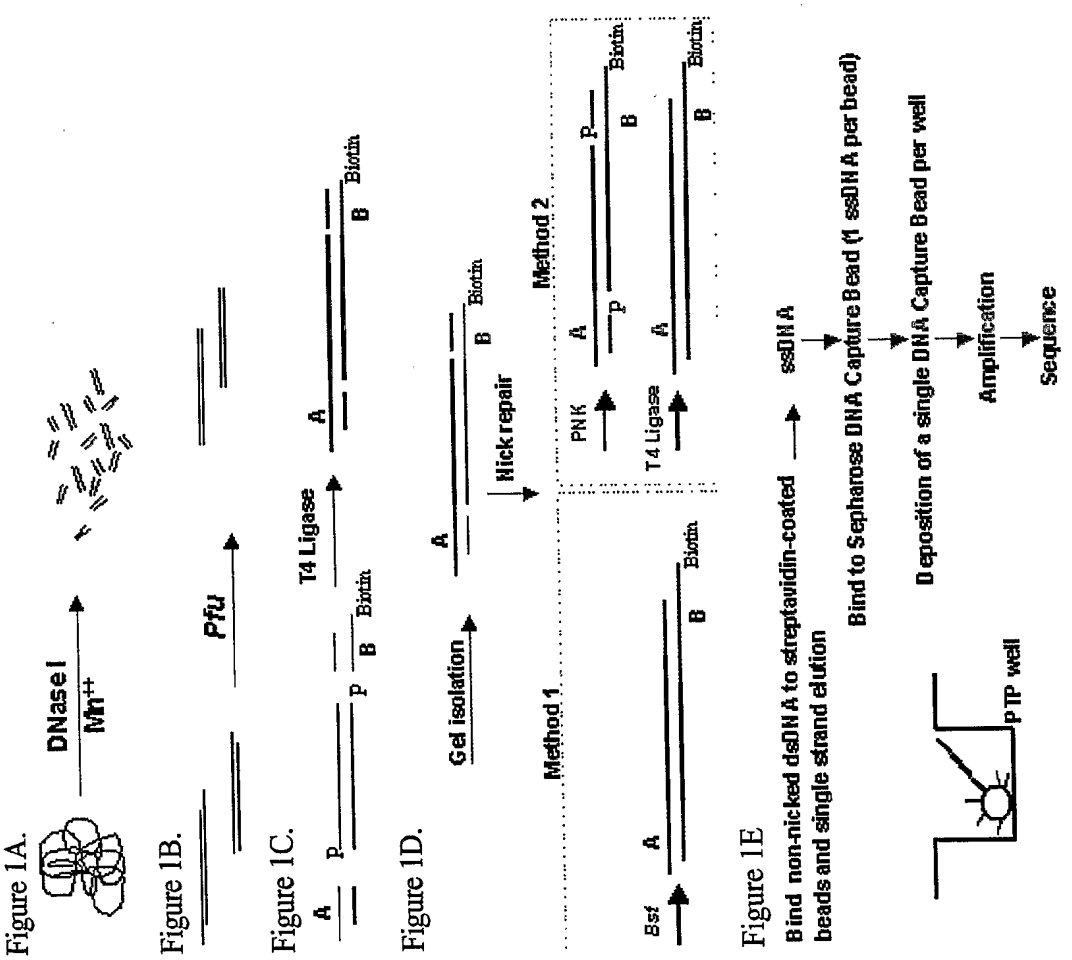
(c) preventing further elongation of said unblocked primer;

(d) deblocking one of the reversibly blocked primers into an unblocked primer; and

(e) repeating steps (b) to (d) until at least one of the reversibly blocked primers are deblocked and used for determining a sequence.

84. The method of claim 76 wherein the reaction chambers are cavities formed by etching one end of a fiber optic bundle.

IDT Exhibit 1020.6683


Figure 1A.
DNaseI
Mn++
Figure 1B.
Pfu
Figure 1C.
T4 Ligase
A
P
B
Biotin
Figure 1D.
Gel isolation
Nick repair
Method 1
Method 2
PNK
T4 Ligase
Bst
Figure 1E
Bind non-nicked dsDNA to streptavidin-coated beads and single strand elution
ssDNA
Bind to Sepharose DNA Capture Bead (1 ssDNA per bead)
Deposition of a single DNA Capture Bead per well
Amplification
Sequence
PTP well

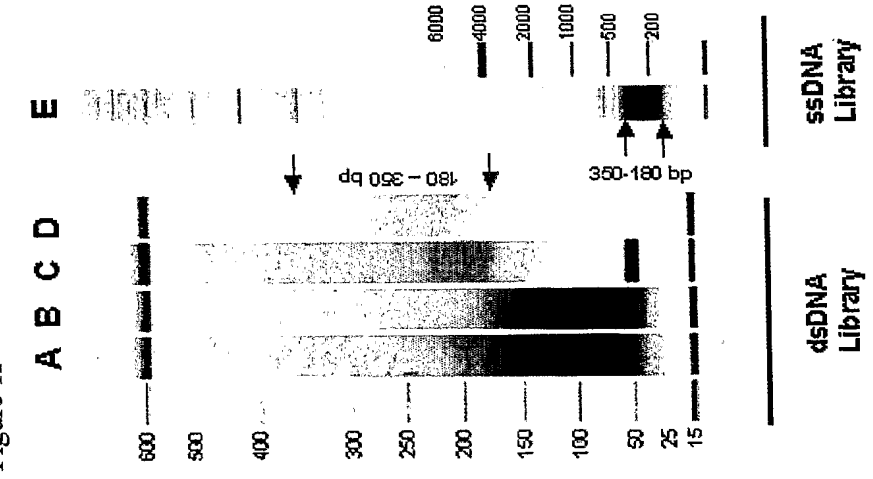

E
A B C D
6000
4000
2000
1000
500
200
180 – 350 bp
350-180 bp
ssDNA Library
dsDNA Library
600
500
400
300
250
200
150
100
50
25
15

Figure 1F

Figure 1G

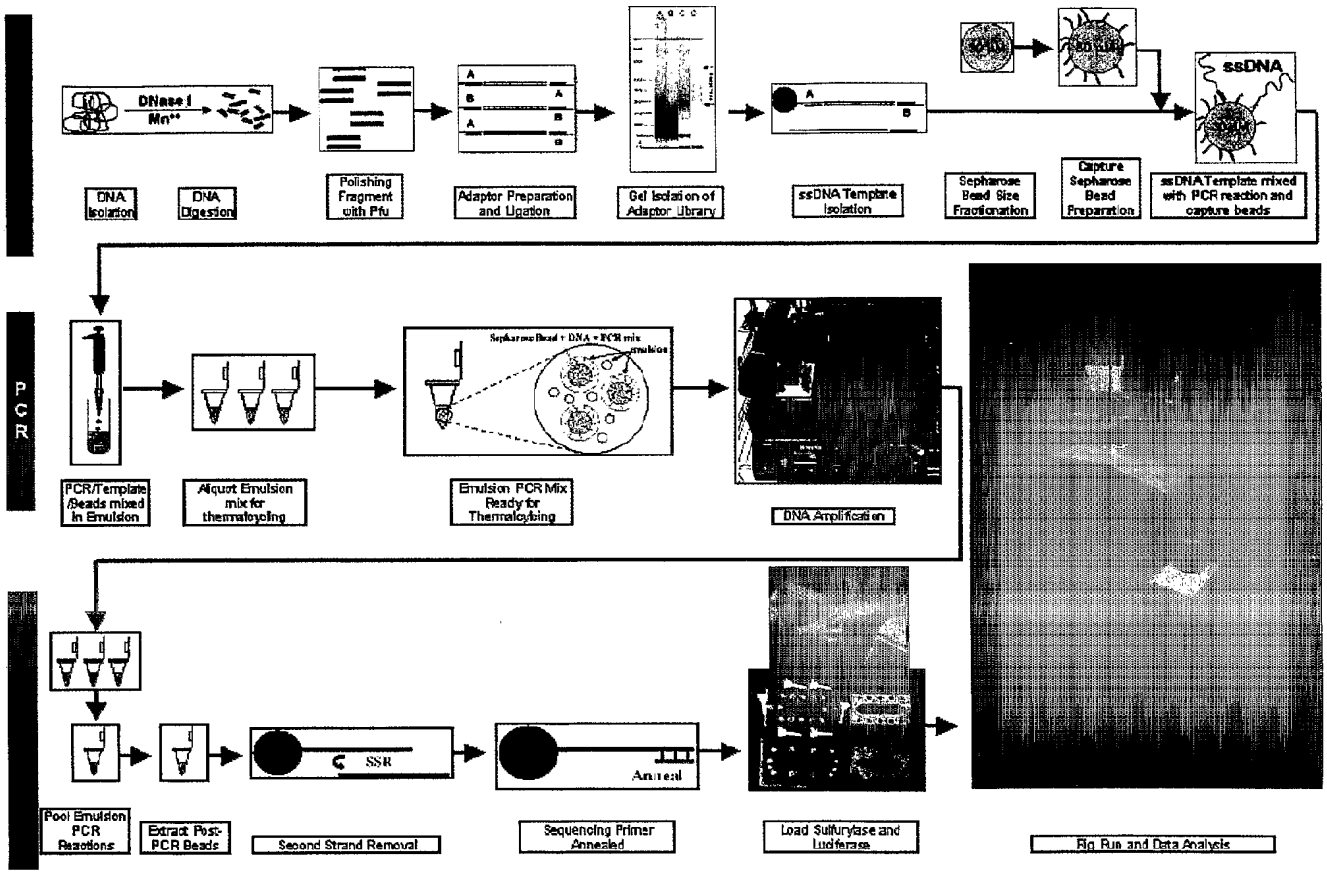

DNase I
Mn++
DNA Isolation
DNA Digestion
Polishing Fragment with Pfu
Adaptor Preparation and Ligation
Gel Isolation of Adaptor Library
ssDNA Template Isolation
Sepharose Bead Size Fractionation
Capture Sepharose Bead Preparation
ssDNA
ssDNA Template mixed with PCR reaction and capture beads
PCR
PCR/Template /Beads mixed in Emulsion
Aliquot Emulsion mix for thermalcycling
Emulsion PCR Mix Ready for Thermalcycling
DNA Amplification
SSR
Anneal
Pool Emulsion PCR Reactions
Extract Post-PCR Beads
Second Strand Removal
Sequencing Primer Annealed
Load Sulfurylase and Luciferase
Big Run and Data Analysis

Figure 2A

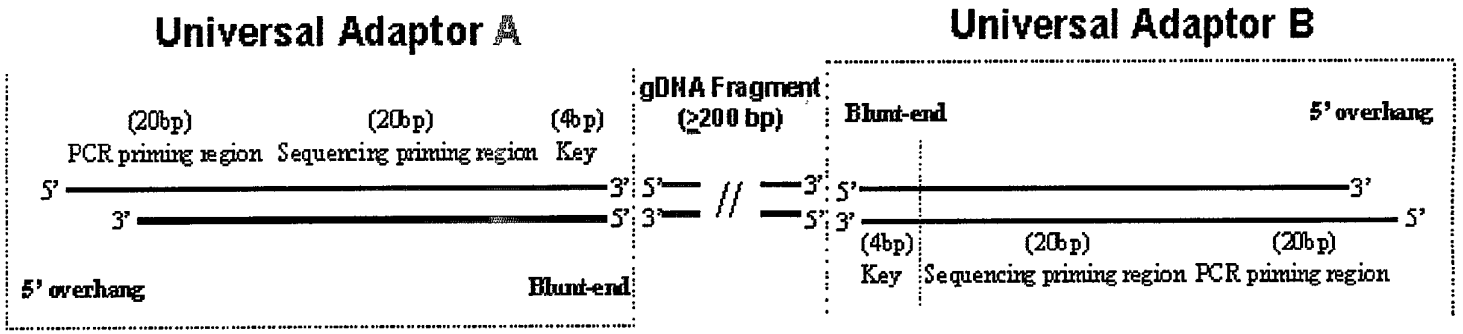


Figure 2B

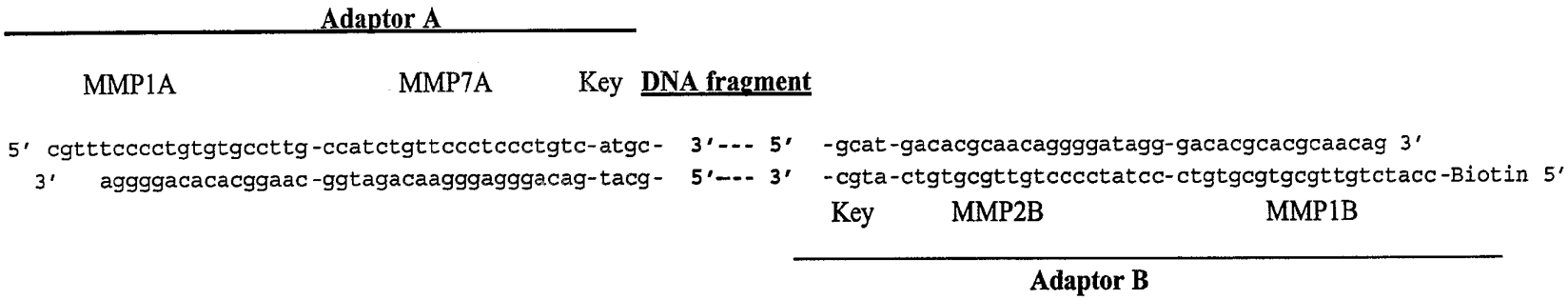


Figure 2C

Sense Strand 5' [PCR primer (20 bp)----- Seq Primer (20 bp)- Key (4 bp)] DNA Fragment [key (4 bp)-Seq Primer (20 bp)-PCR primer (20 bp)] 3'
Anti Strand 3' [PCR primer (20 bp)- Seq Primer (20 bp)- Key (4 bp)] DNA Fragment [key (4 bp)-Seq Primer (20 bp)-----PCR primer (20 bp)] 5' **(BEAD)**

IDT Exhibit 1020.6687

Figure 3

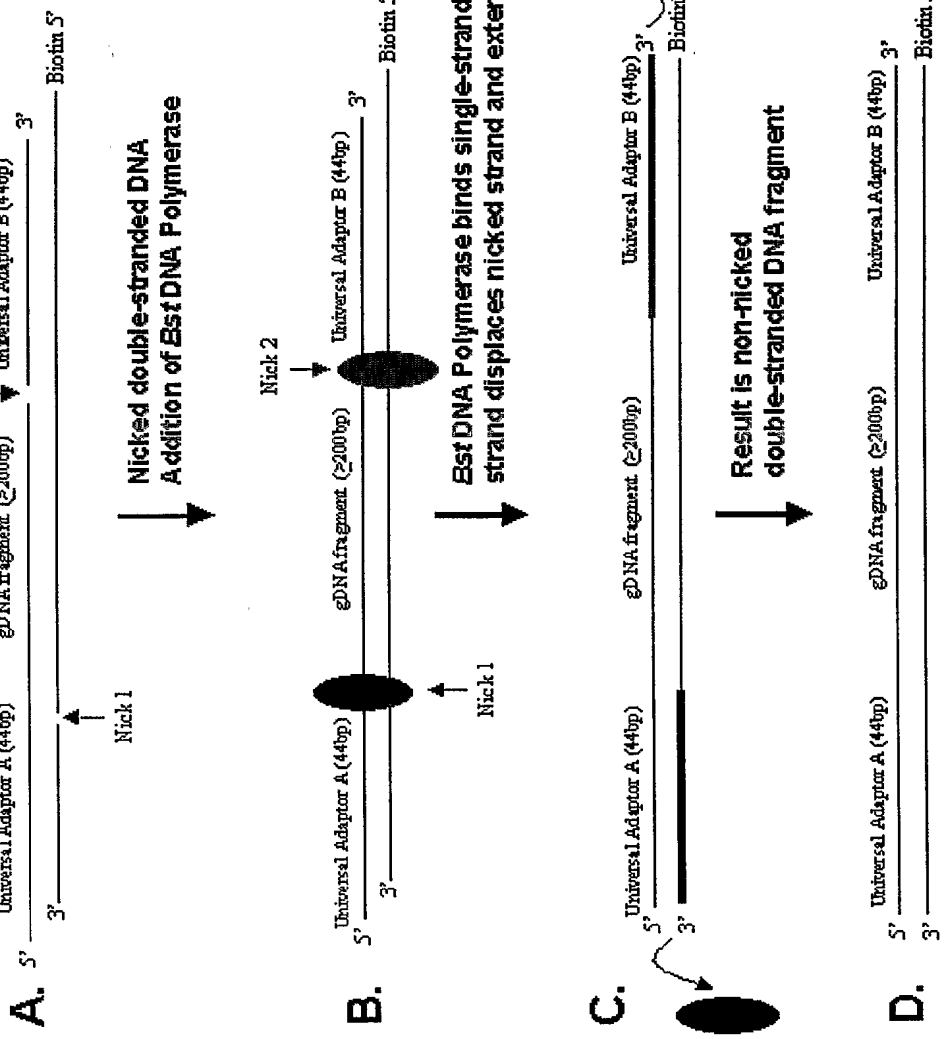

A.
Nick 2
Universal Adaptor B (44bp)
Biotin 5'
gDNA fragment (≥200bp)
3'
5'
Universal Adaptor A (44bp)
Nick 1
Nicked double-stranded DNA
Addition of Bst DNA Polymerase
B.
Nick 2
Universal Adaptor B (44bp)
Biotin 5'
gDNA fragment (≥200bp)
3'
5'
Universal Adaptor A (44bp)
3'
Nick 1
Bst DNA Polymerase binds single-stranded gaps,
strand displaces nicked strand and extends fragment
C.
Universal Adaptor B (44bp)
3'
Biotin 5'
gDNA fragment (≥200bp)
5'
Universal Adaptor A (44bp)
3'
Result is non-nicked
double-stranded DNA fragment
D.
Universal Adaptor B (44bp)
3'
Biotin 5'
gDNA fragment (≥200bp)
5'
Universal Adaptor A (44bp)
3'

Figure 4

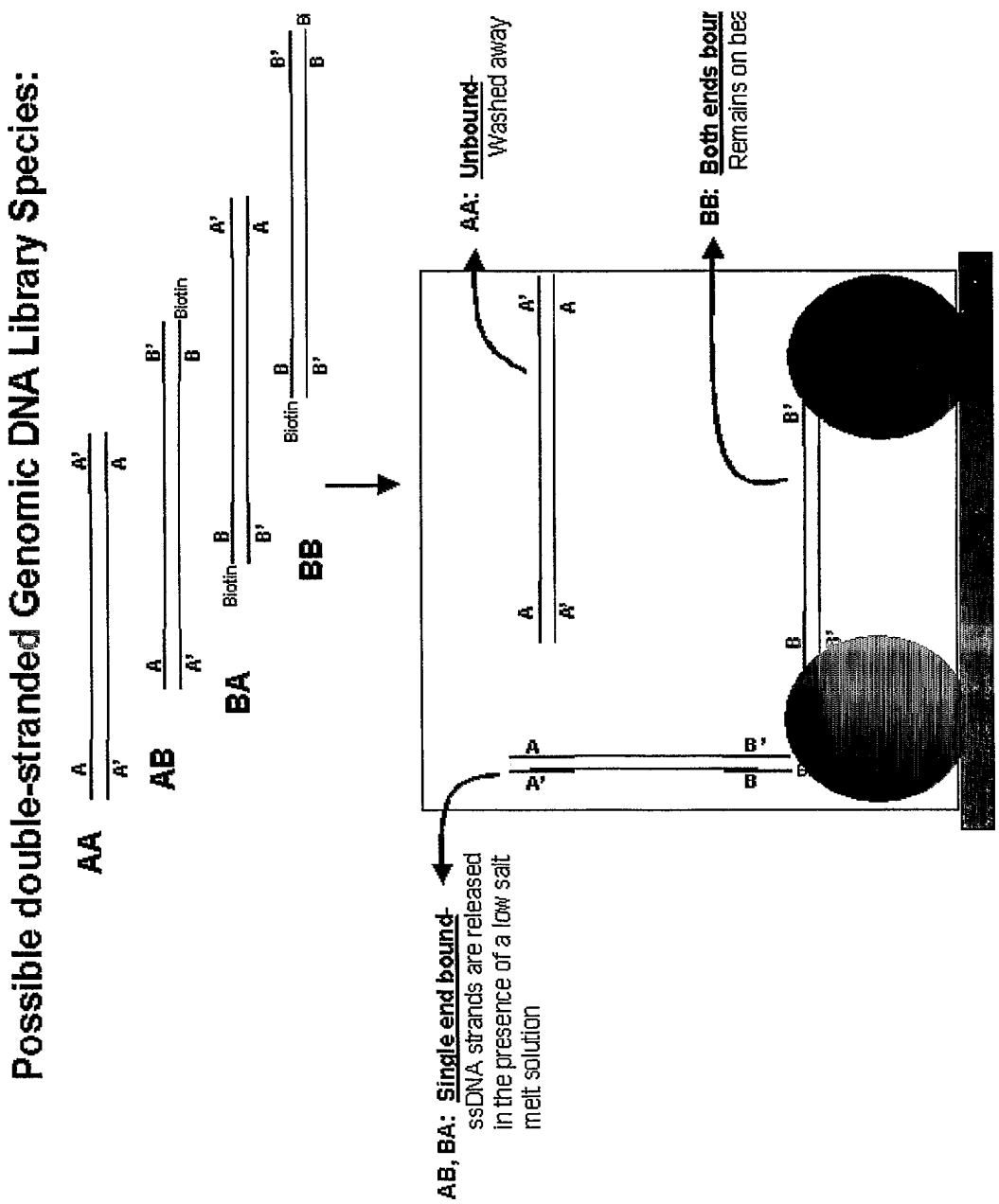

Possible double-stranded Genomic DNA Library Species:
AA
AB
BA
BB
Biotin
AA: Unbound
Washed away
BB: Both ends bour
Remains on bea
AB, BA: Single end bound
ssDNA strands are released
in the presence of a low salt
melt solution

Figure 5

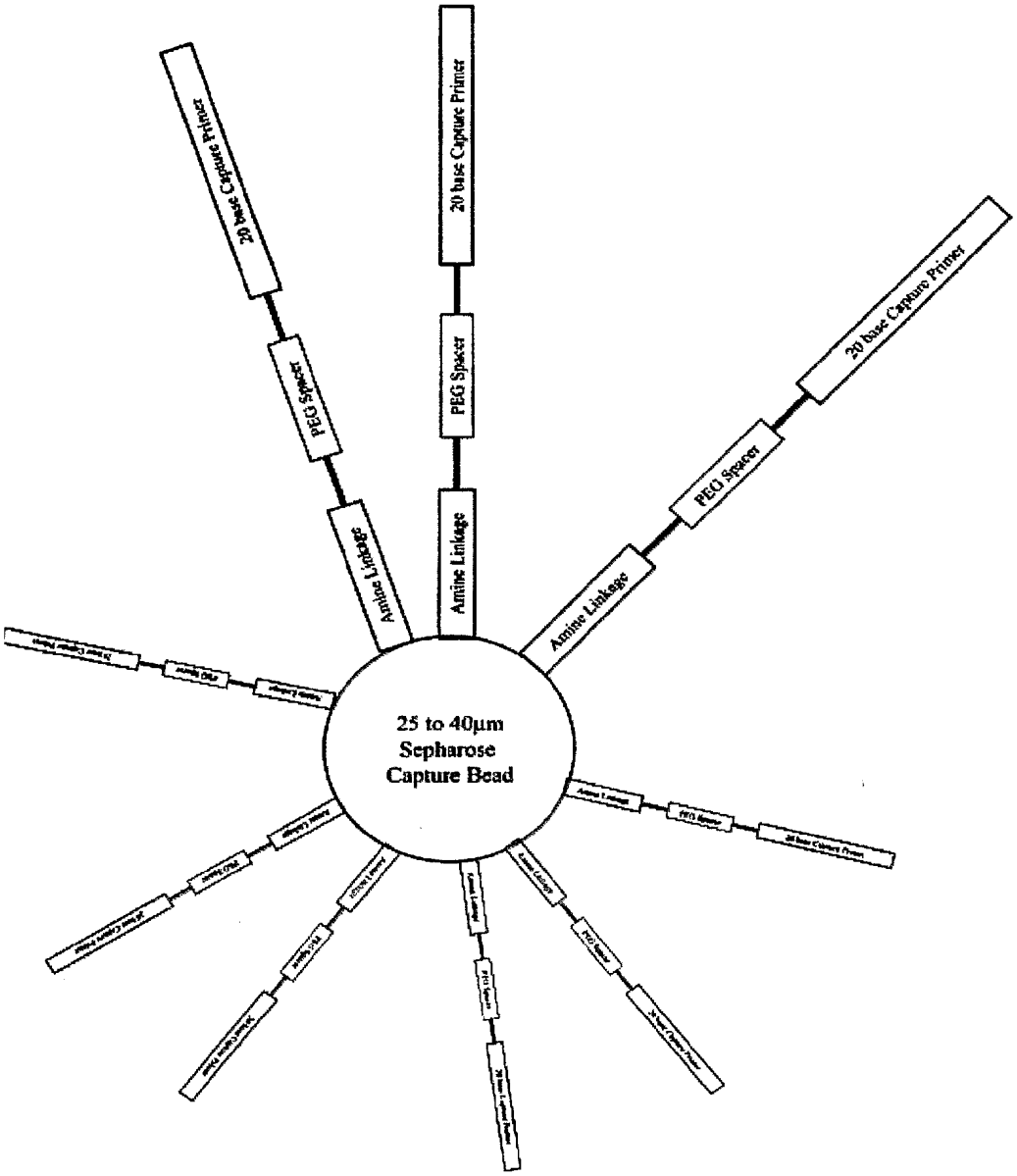

20 base Capture Primer
PEG Spacer
Amine Linkage
20 base Capture Primer
PEG Spacer
Amine Linkage
20 base Capture Primer
PEG Spacer
Amine Linkage
25 to 40µm
Sepharose
Capture Bead

FIGURE 6A

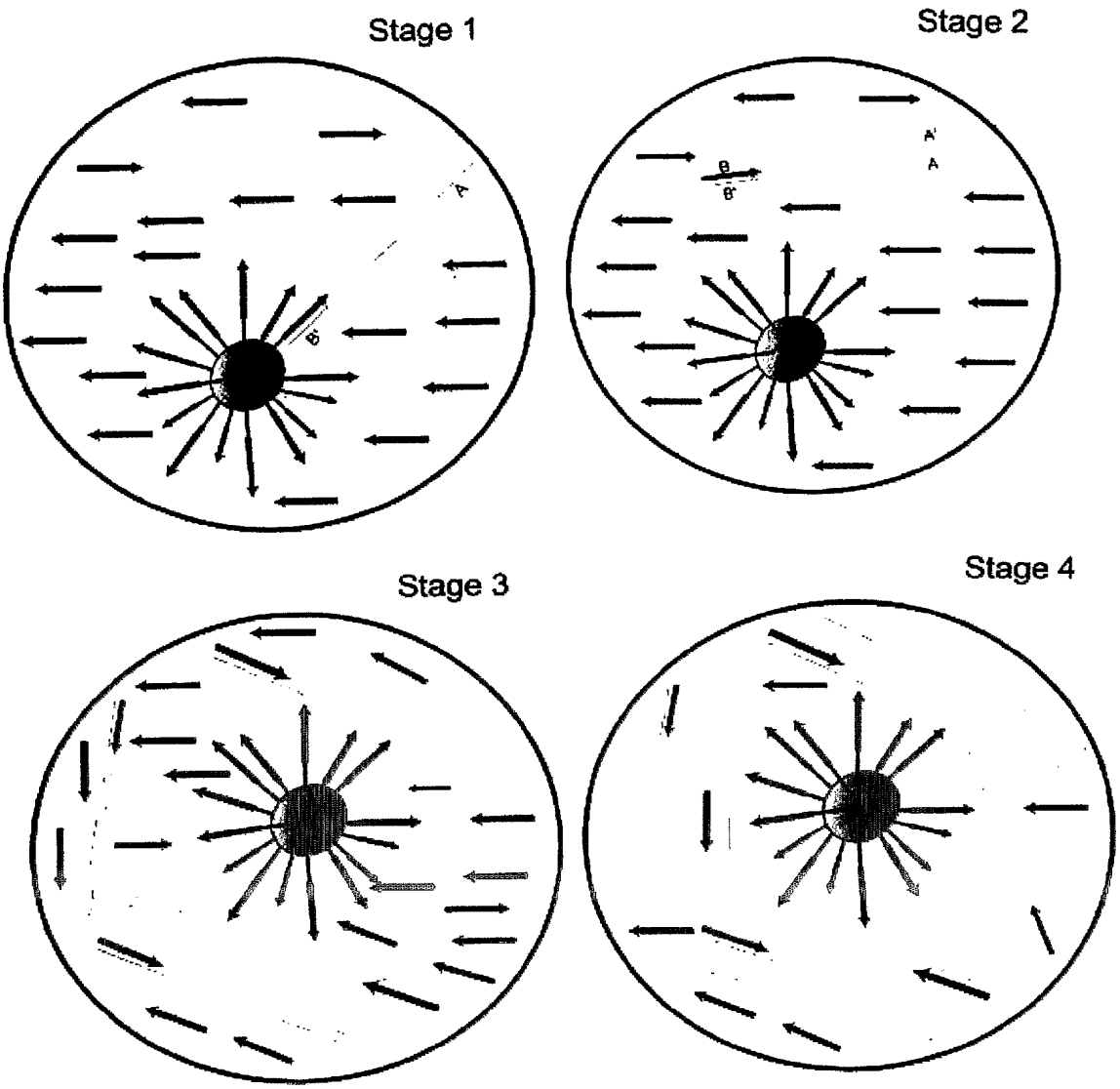


FIGURE 6B

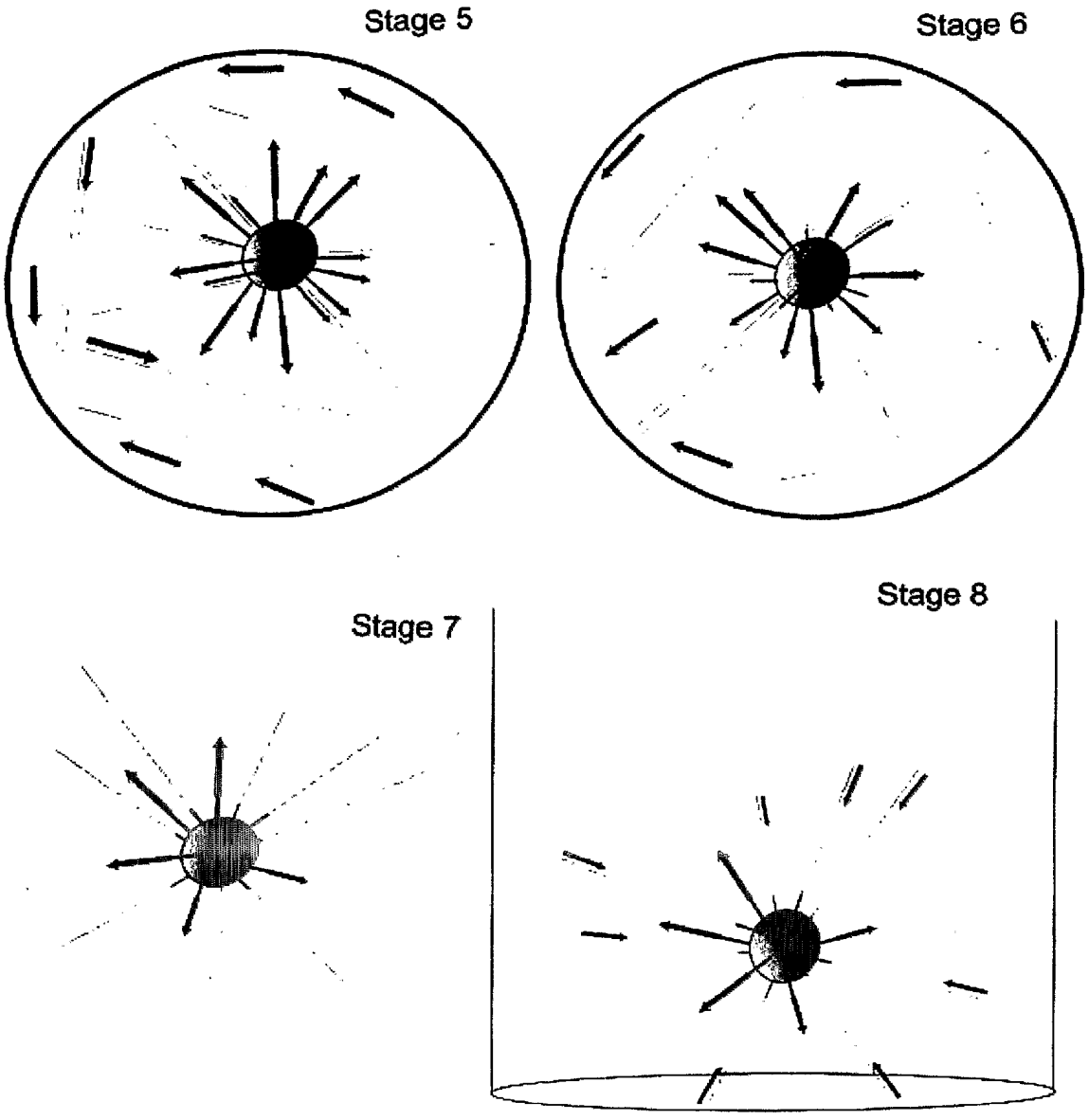


FIGURE 7

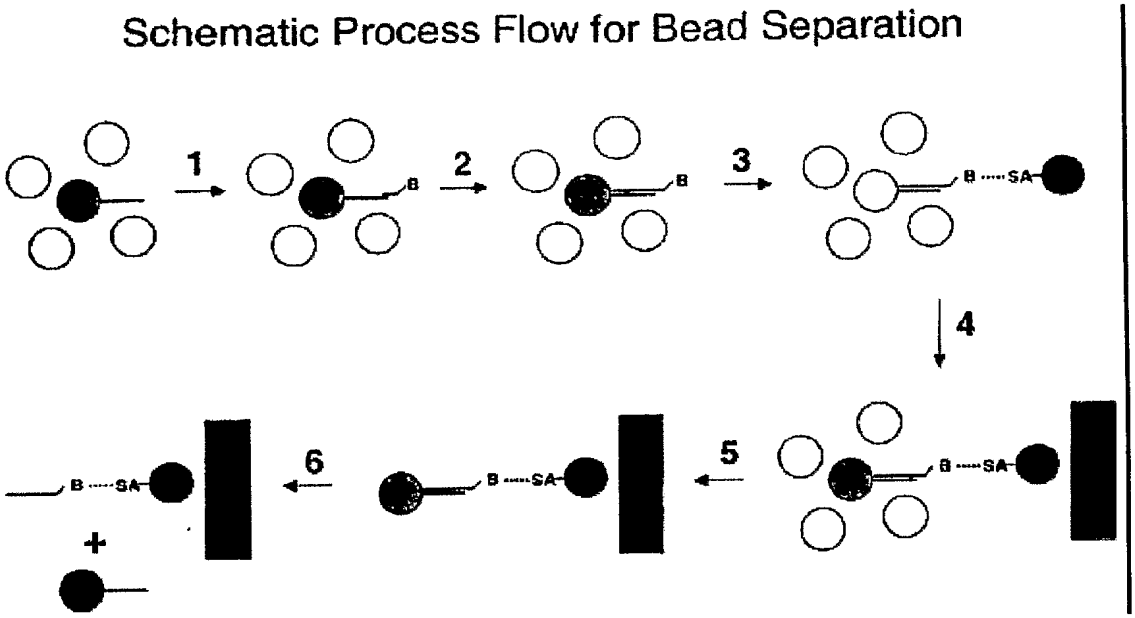


FIGURE 8

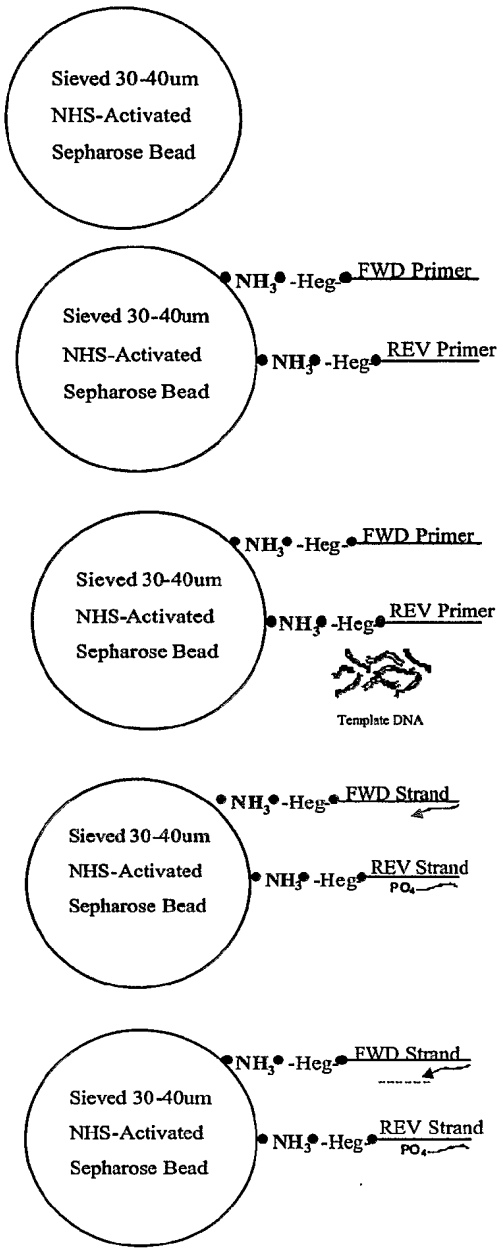

Sieved 30-40um
NHS-Activated
Sepharose Bead
NH3 -Heg FWD Primer
NH3 -Heg REV Primer
Template DNA
FWD Strand
REV Strand
PO4

Figure 8 cont

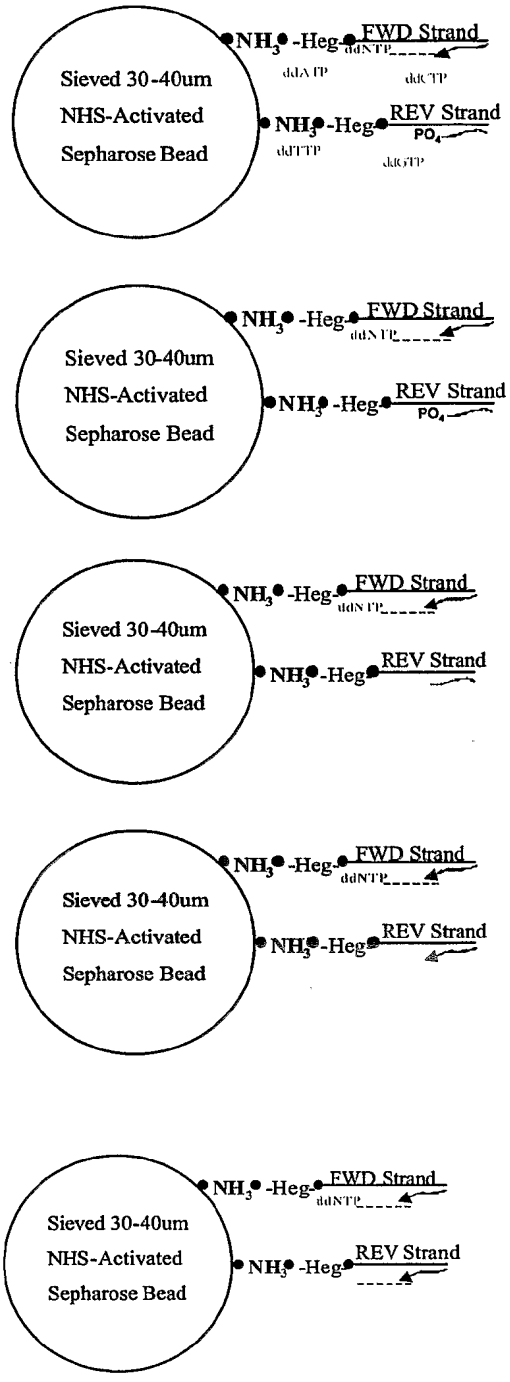


FIGURE 9

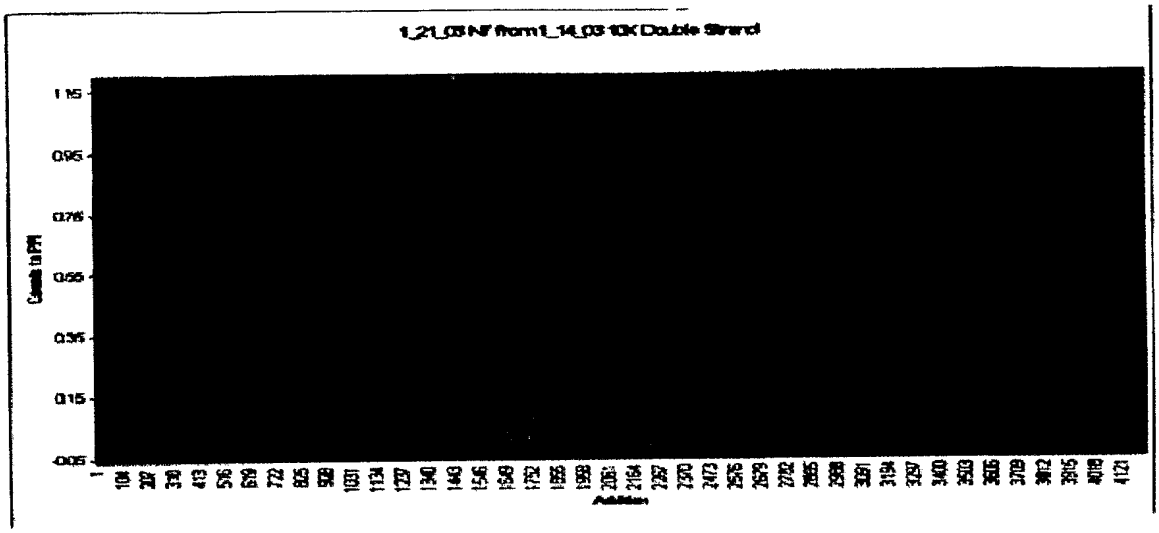
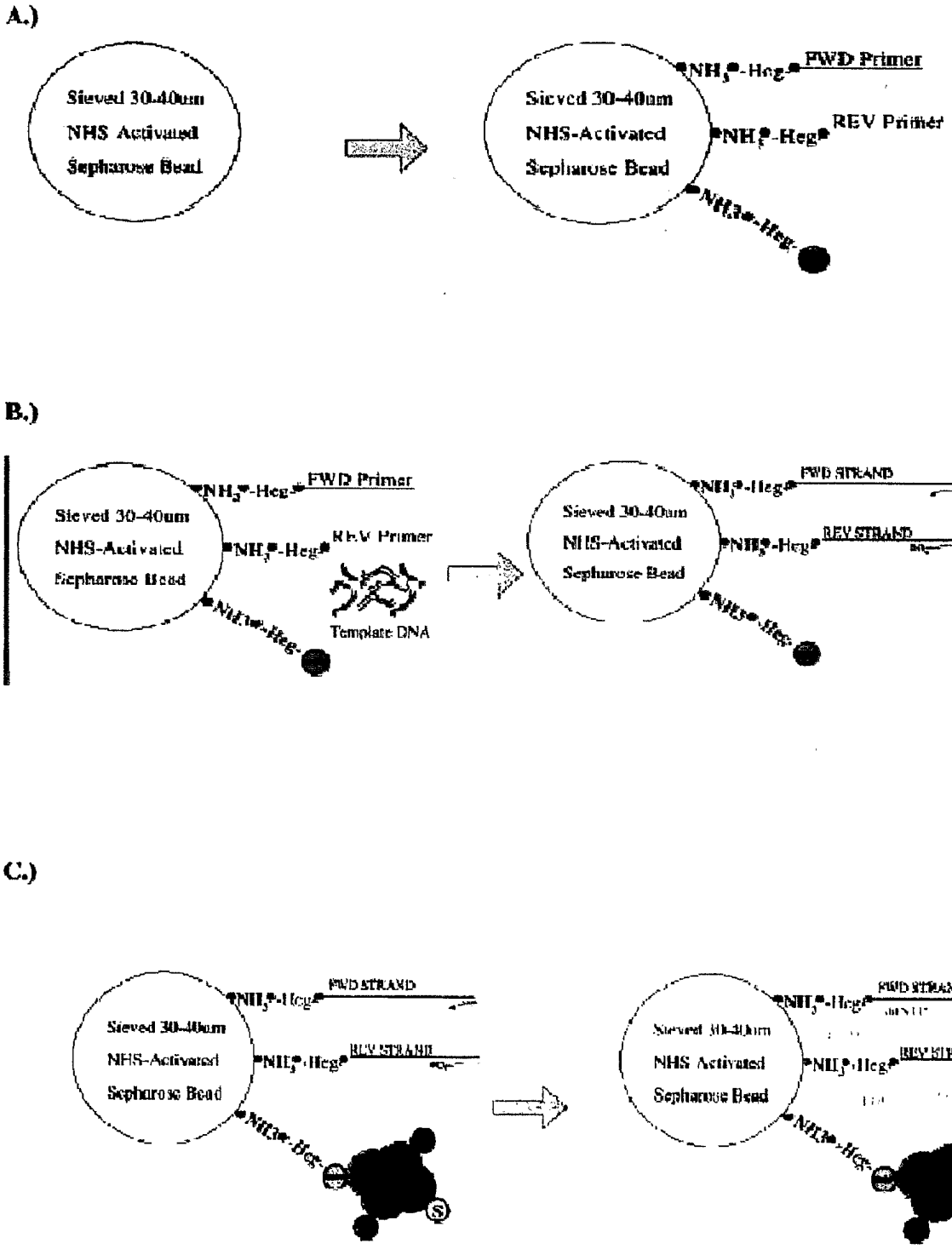
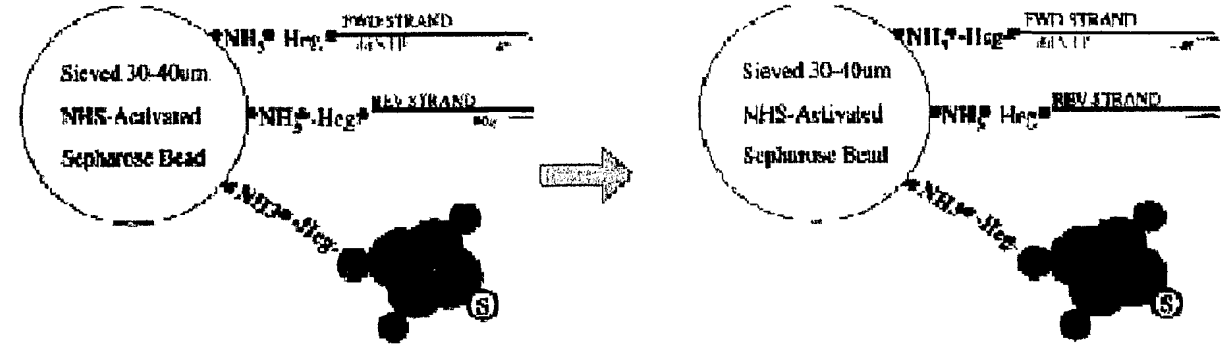


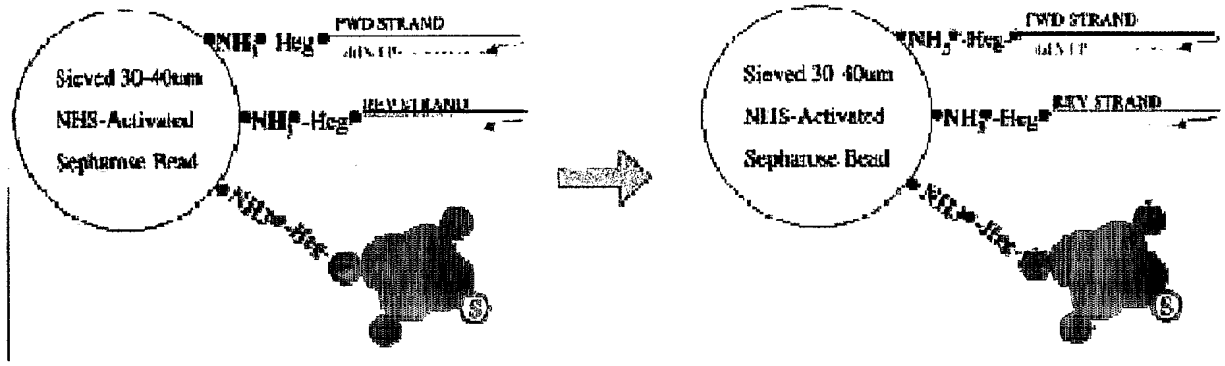
FIGURE 10


A.)
Sieved 30-40um
NHS Activated
Sepharose Bead
Sieved 30-40um
NHS-Activated
Sepharose Bead
FWD Primer
REV Primer
B.)
Sieved 30-40um
NHS-Activated
Sepharose Bead
FWD Primer
REV Primer
Template DNA
Sieved 30-40um
NHS-Activated
Sepharose Bead
FWD STRAND
REV STRAND
C.)
Sieved 30-40um
NHS-Activated
Sepharose Bead
FWD STRAND
REV STRAND
Sieved 30-40um
NHS Activated
Sepharose Bead
FWD STRAND
REV STRAND

D.)


Sieved 30-40um
NHS-Activated
Sepharose Bead
FWD STRAND
REV STRAND
Sieved 30-40um
NHS-Activated
Sepharose Bead
FWD STRAND
REV STRAND

F.)


Sieved 30-40um
NHS-Activated
Sepharose Bead
FWD STRAND
REV STRAND
Sieved 30-40um
NHS-Activated
Sepharose Bead
FWD STRAND
REV STRAND

F.)

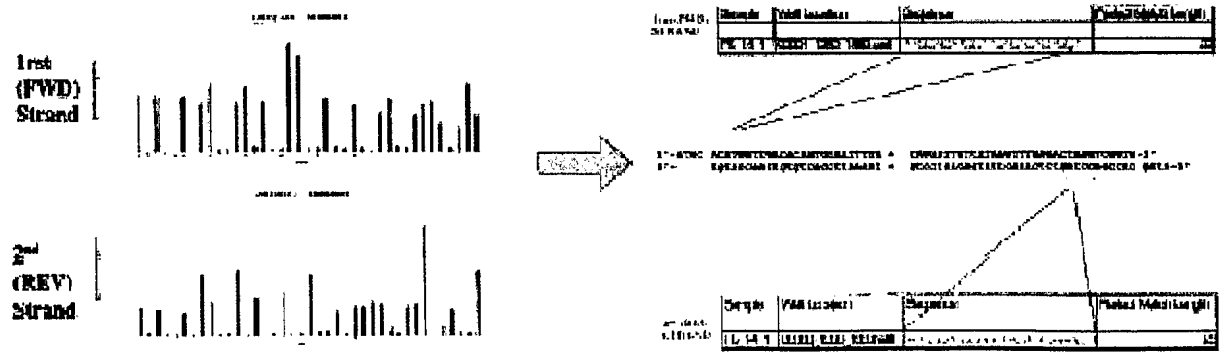

1rst
(FWD)
Strand
2nd
(REV)
Strand

Figure 11A

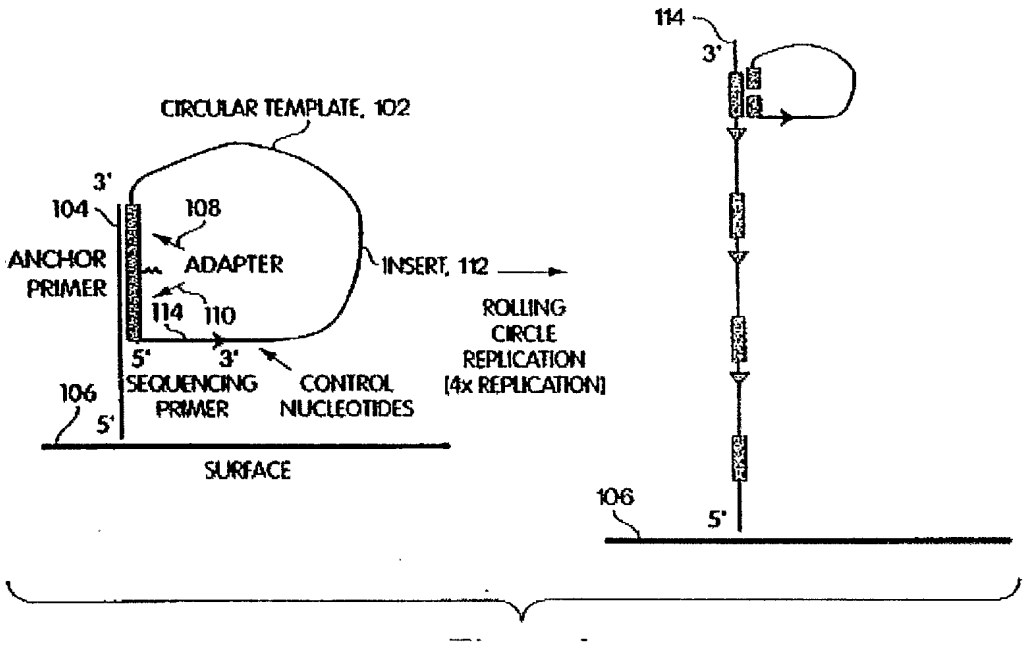


Figure 11B

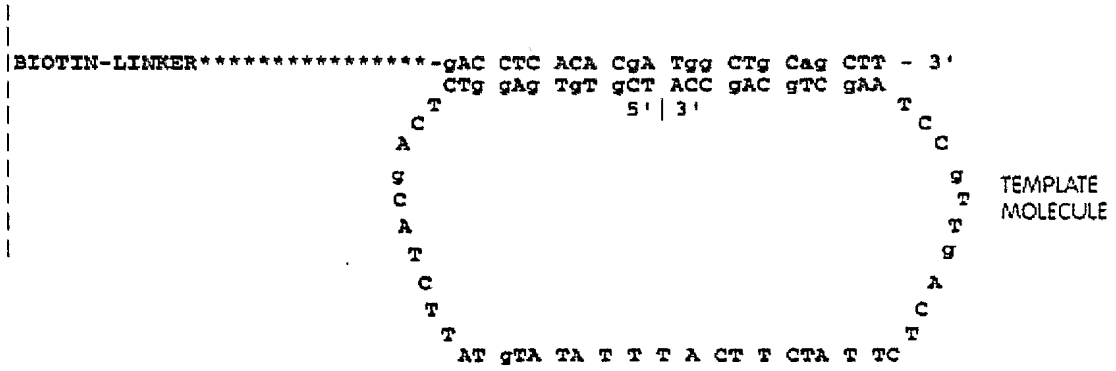


Figure 11C

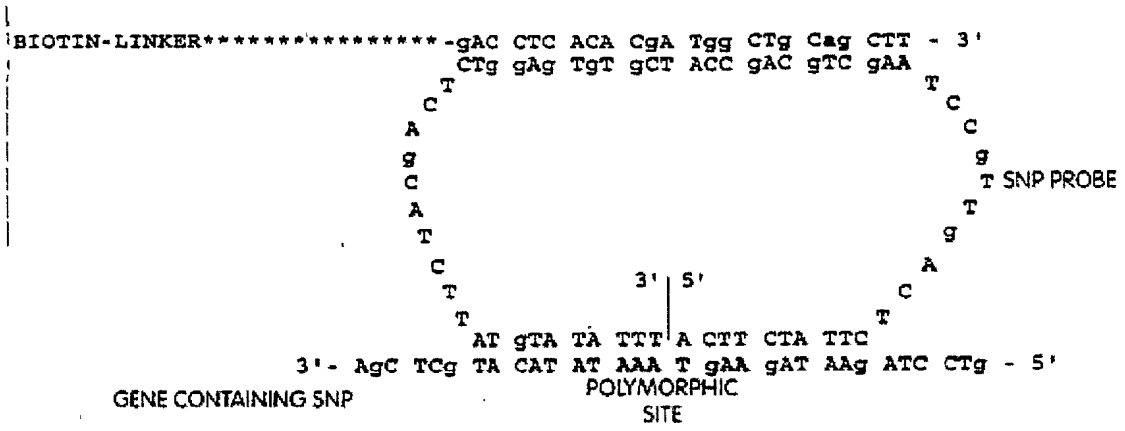


Figure 11D

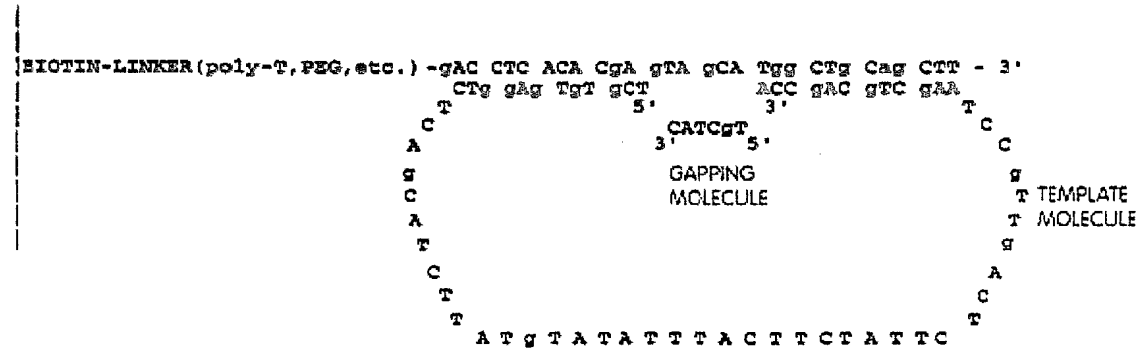


Figure 12

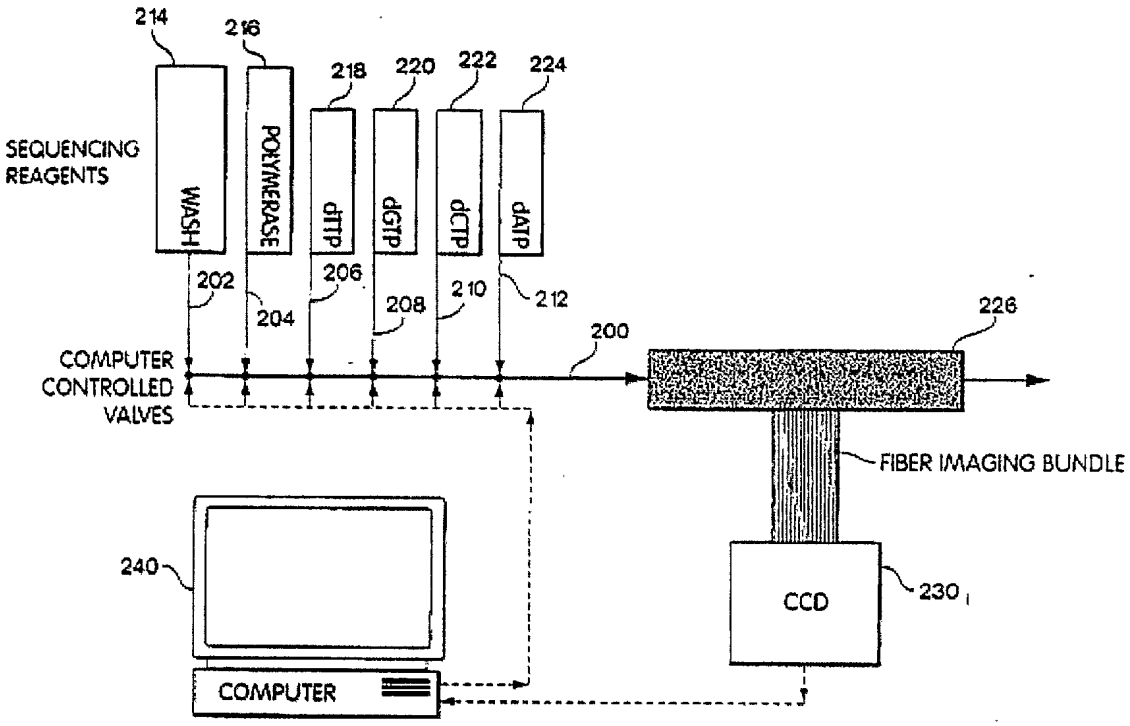

214
216
218
220
222
224
SEQUENCING REAGENTS
WASH
POLYMERASE
dTTP
dGTP
dCTP
dATP
202
204
206
208
210
212
200
226
COMPUTER CONTROLLED VALVES
FIBER IMAGING BUNDLE
240
CCD
230
COMPUTER

Figure 13

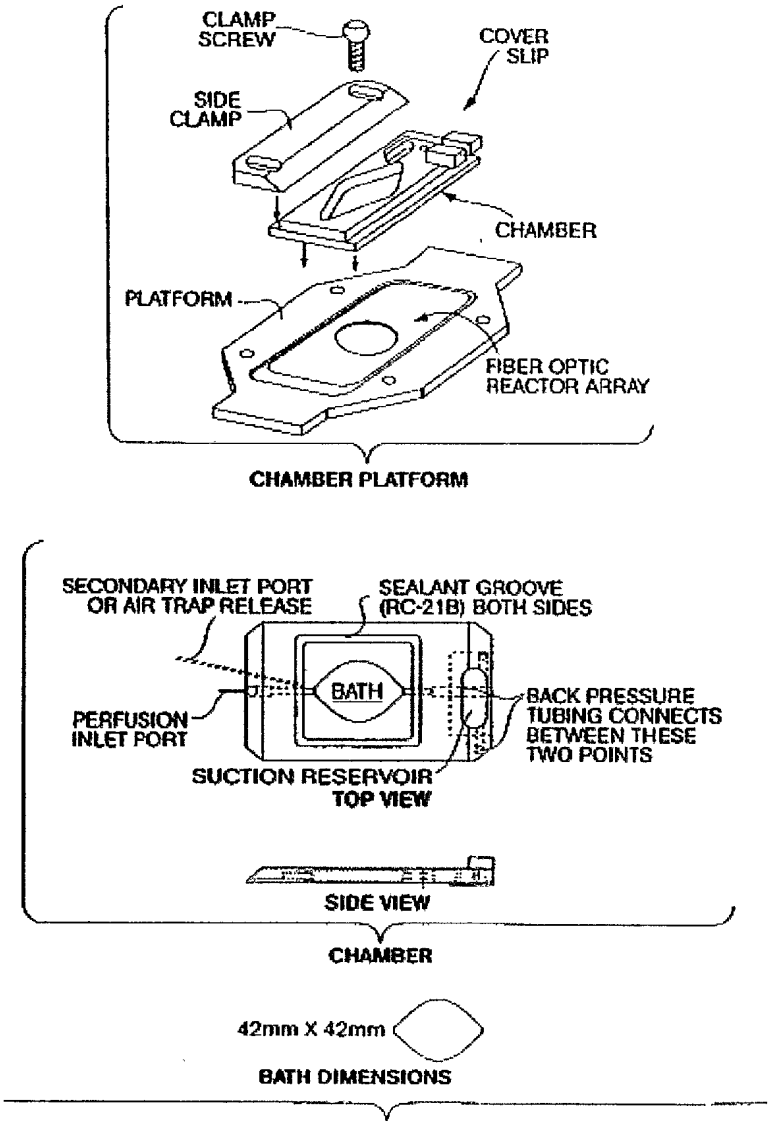
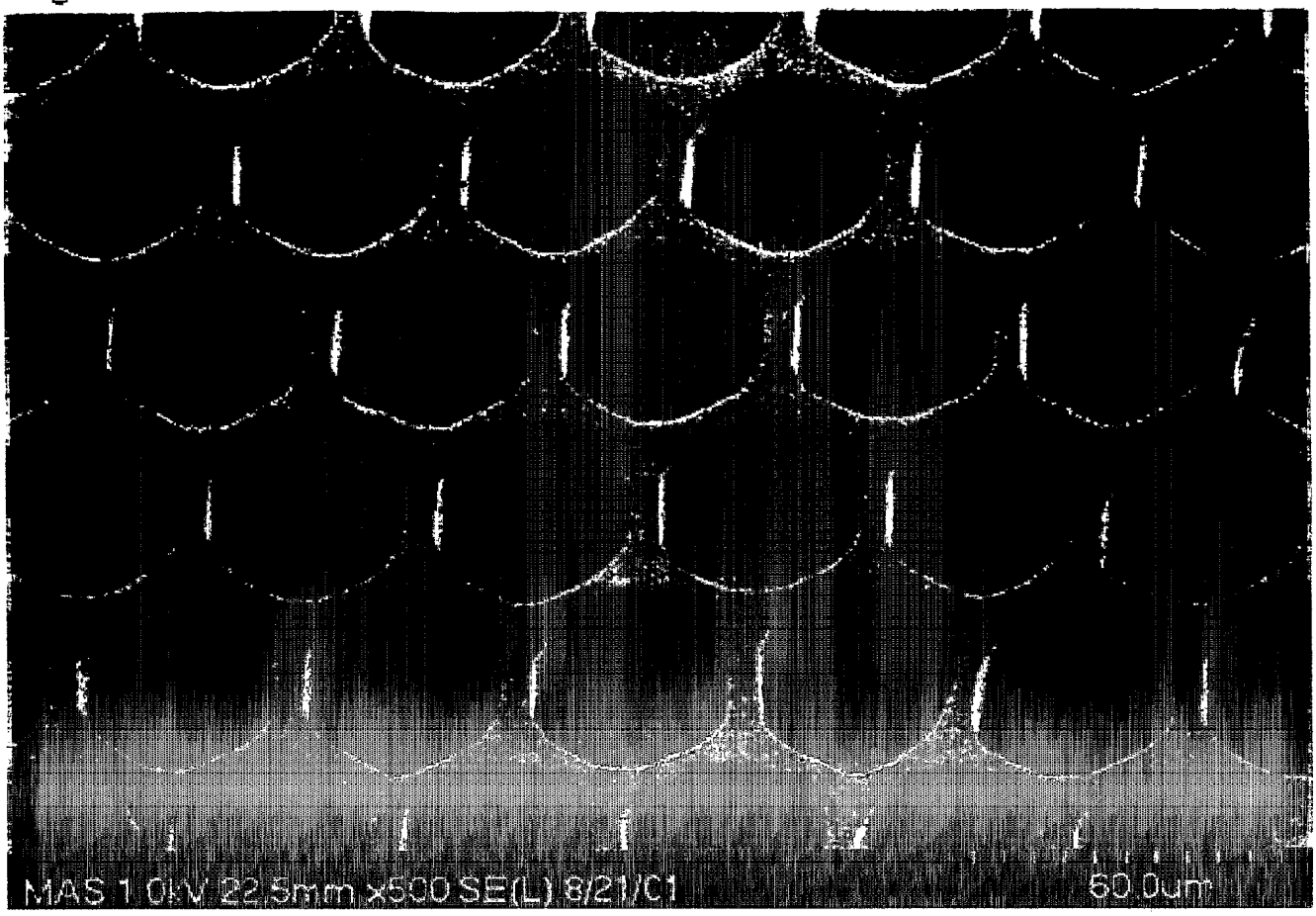

CLAMP SCREW
COVER SLIP
SIDE CLAMP
CHAMBER
PLATFORM
FIBER OPTIC REACTOR ARRAY
CHAMBER PLATFORM
SECONDARY INLET PORT OR AIR TRAP RELEASE
SEALANT GROOVE (RC-21B) BOTH SIDES
BATH
PERFUSION INLET PORT
BACK PRESSURE TUBING CONNECTS BETWEEN THESE TWO POINTS
SUCTION RESERVOIR
TOP VIEW
SIDE VIEW
CHAMBER
42mm X 42mm
BATH DIMENSIONS

Figure 14



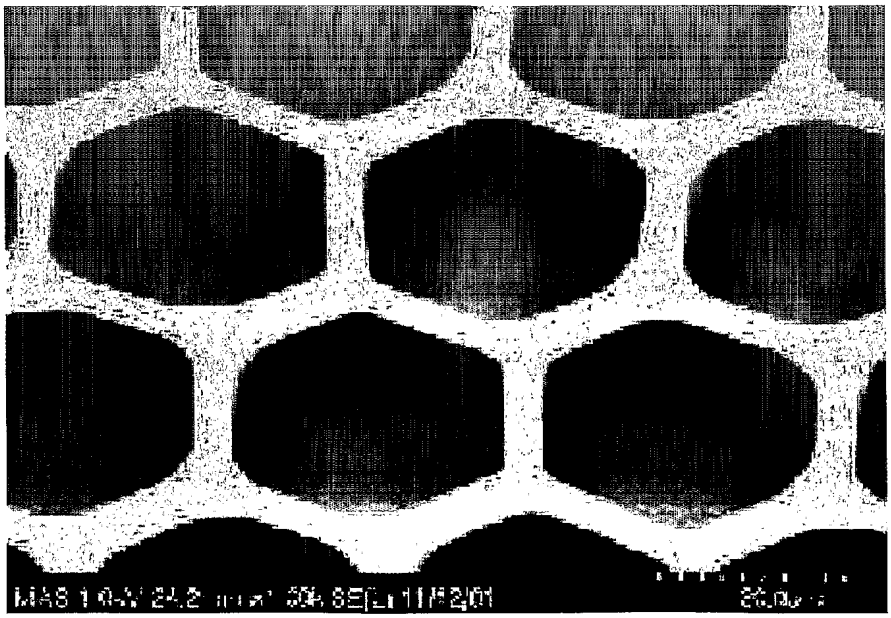
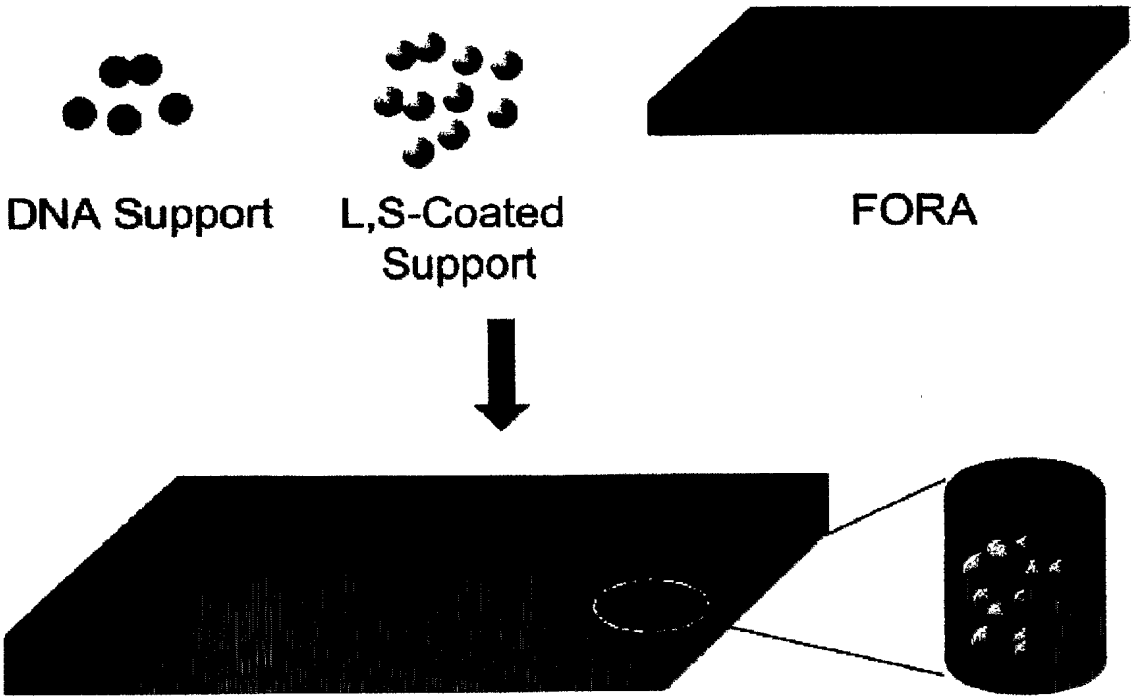


Figure 15


DNA Support
L,S-Coated Support
FORA
"Carpeted" FORA with Randomly Dispersed L,S Support

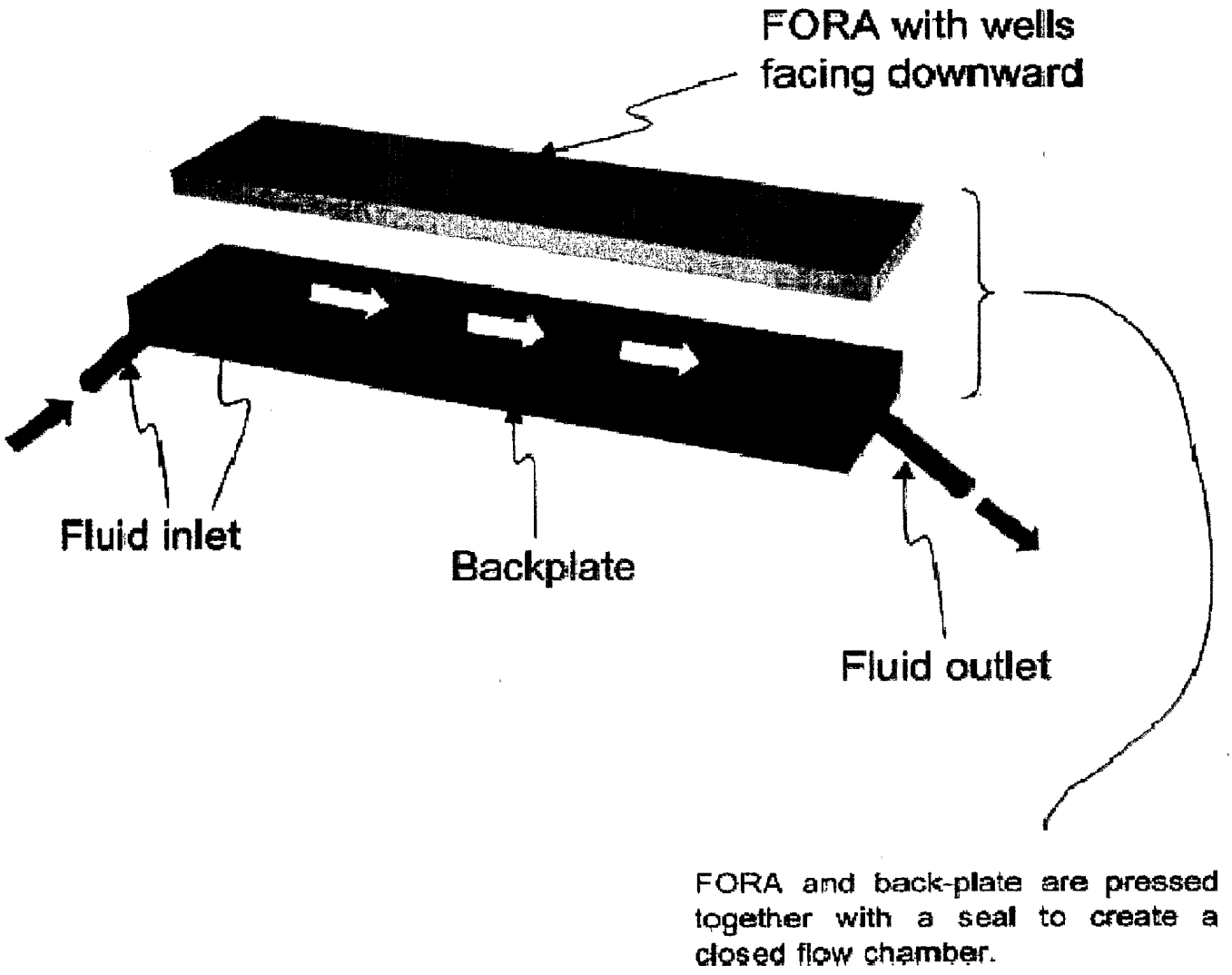

FORA with wells facing downward
Fluid inlet
Backplate
Fluid outlet
FORA and back-plate are pressed together with a seal to create a closed flow chamber.

Figure 16

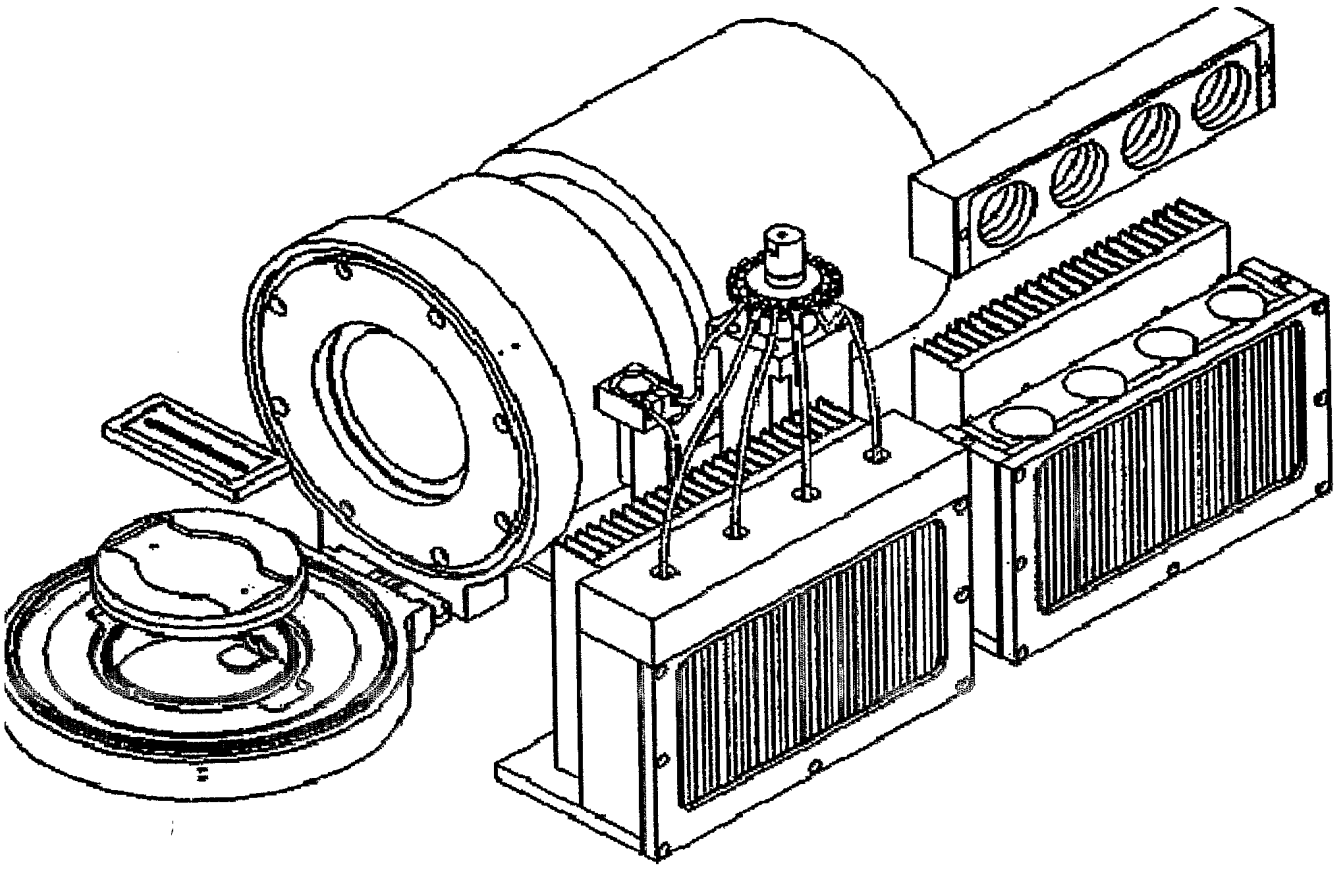


Figure 17

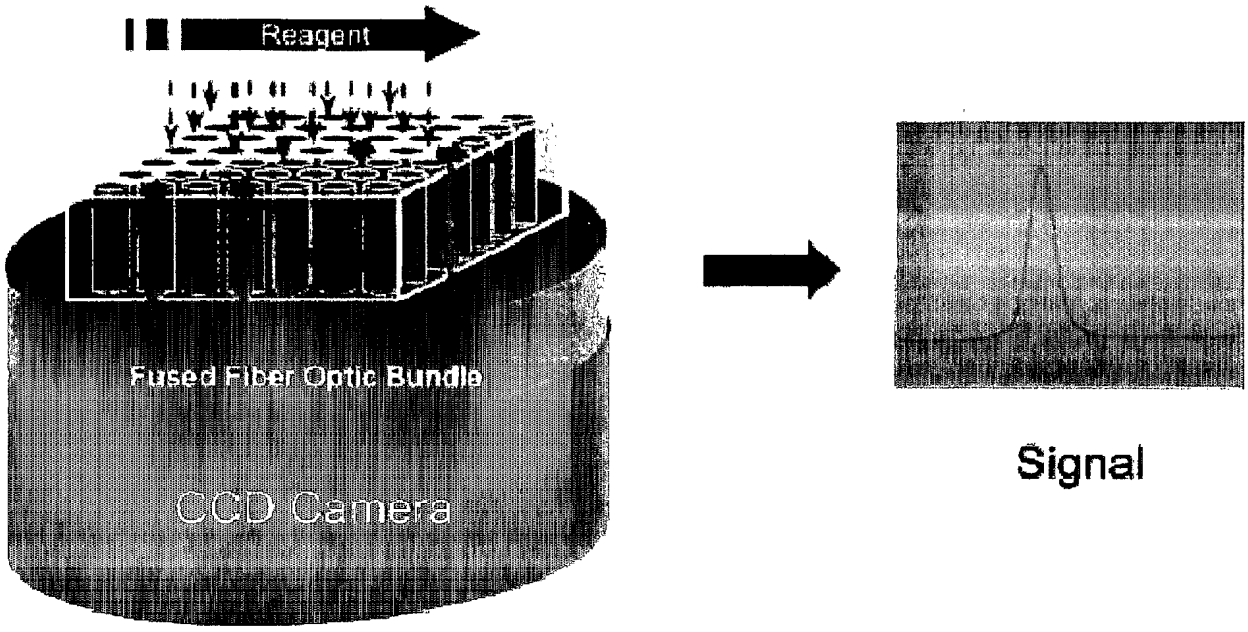
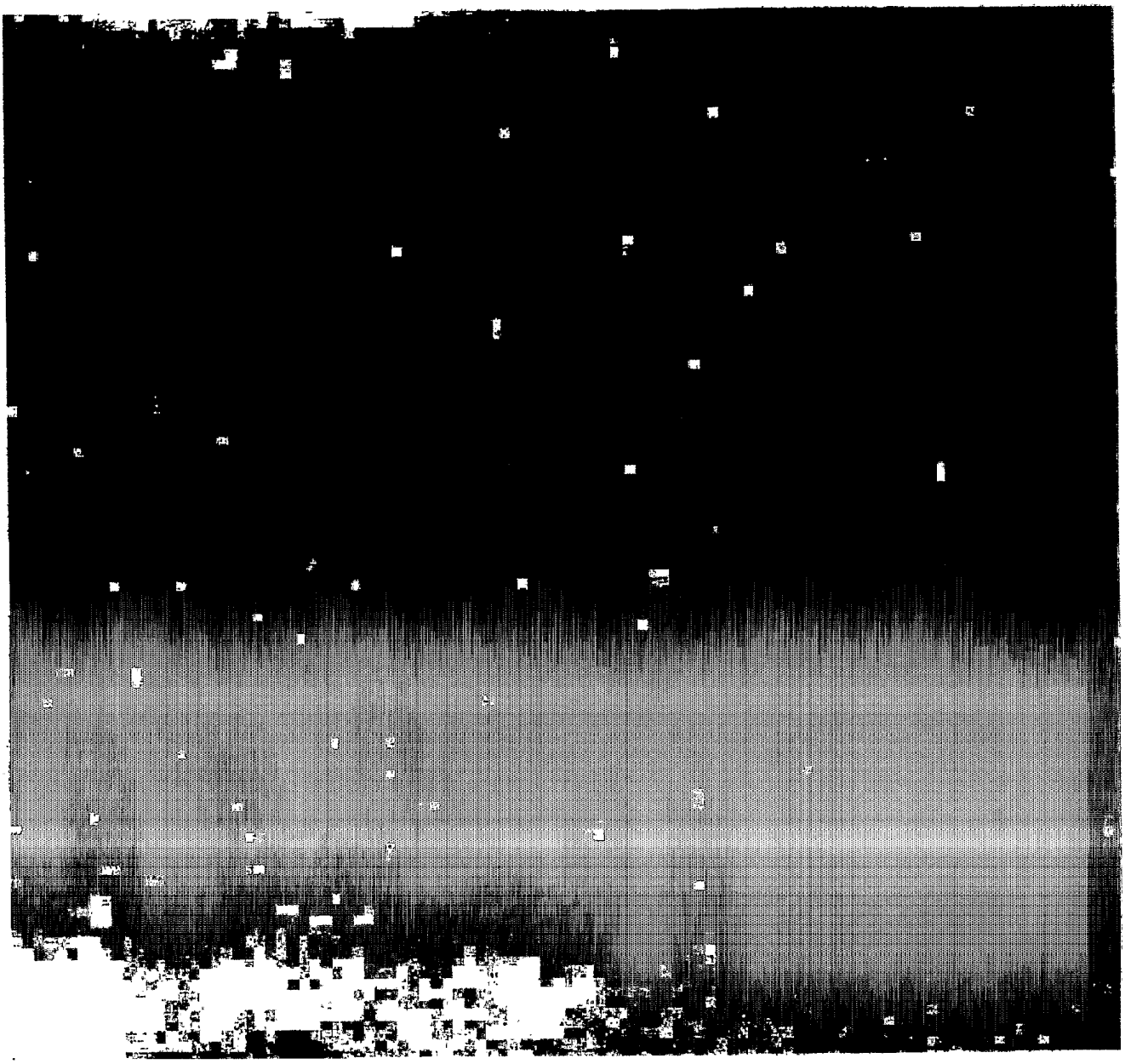


Figure 18



Sulfurylase on mobile support/Luciferase on surface, PPi flows across FORA surface. 1 mobile-support per pixel on camera; 50µm FORA well spanning 4 15µm pixels. Shown is 100 x 100 Pixels (10K pixels), 1.5 mm^2.

Figure 19

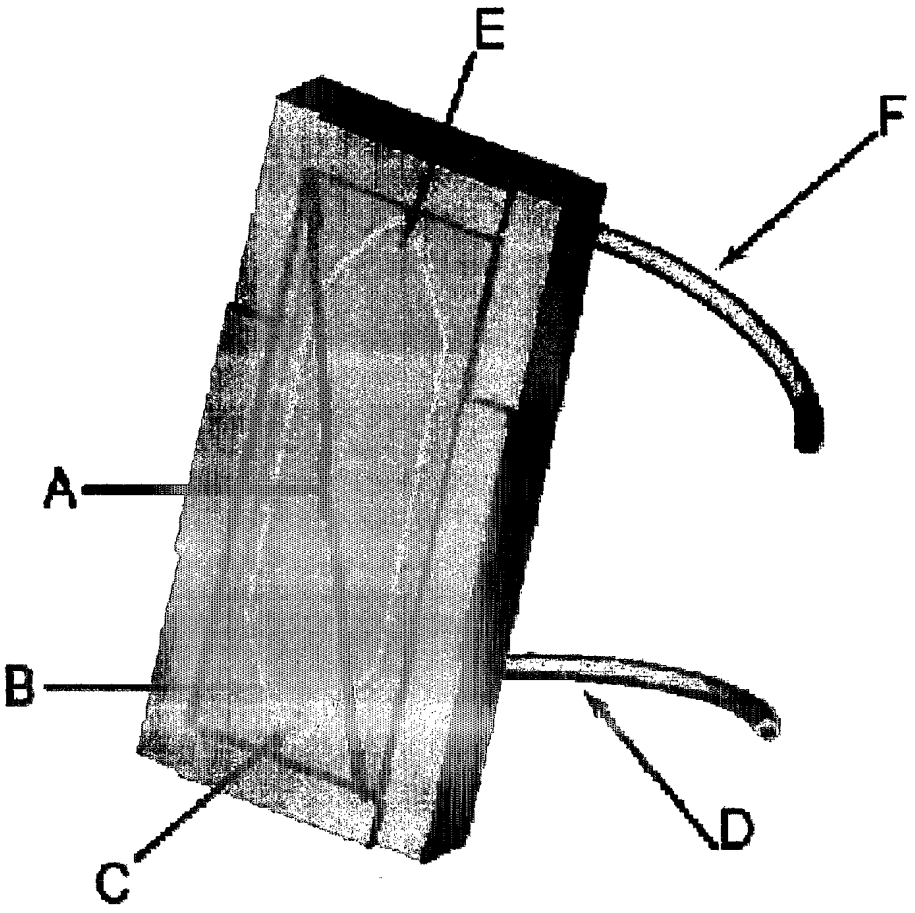


Figure 20

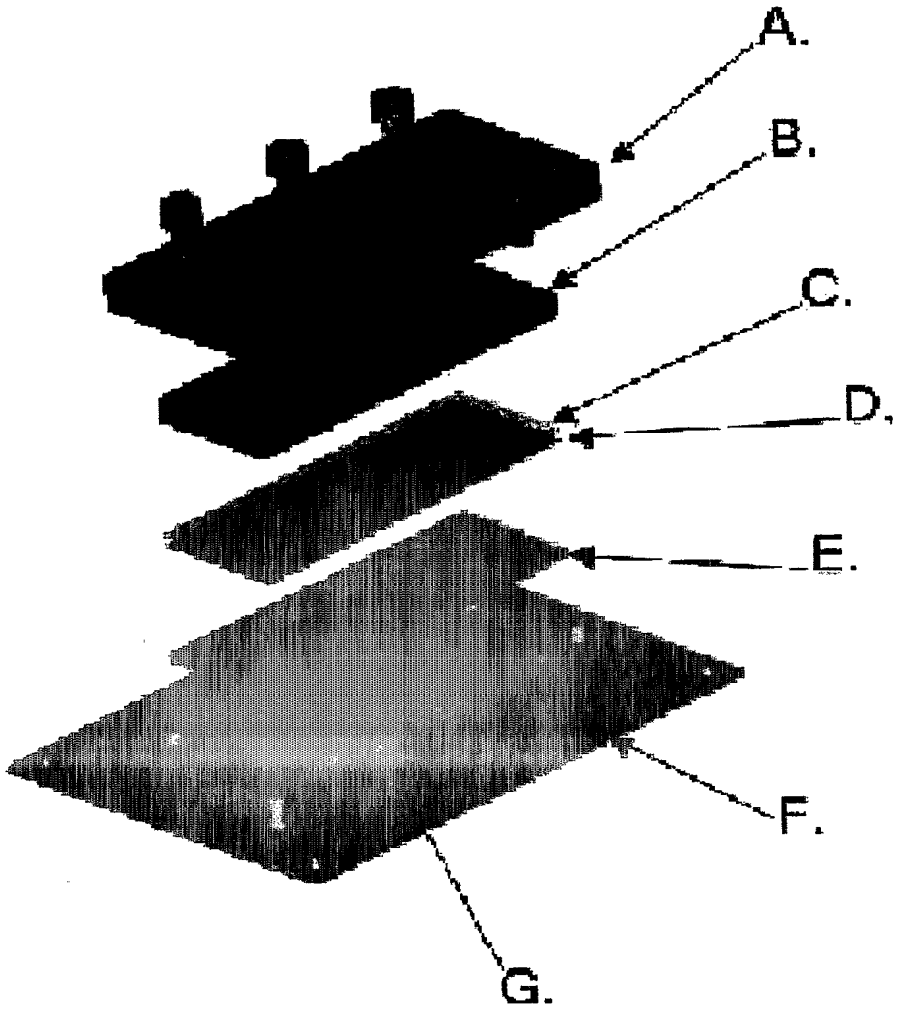


Figure 21

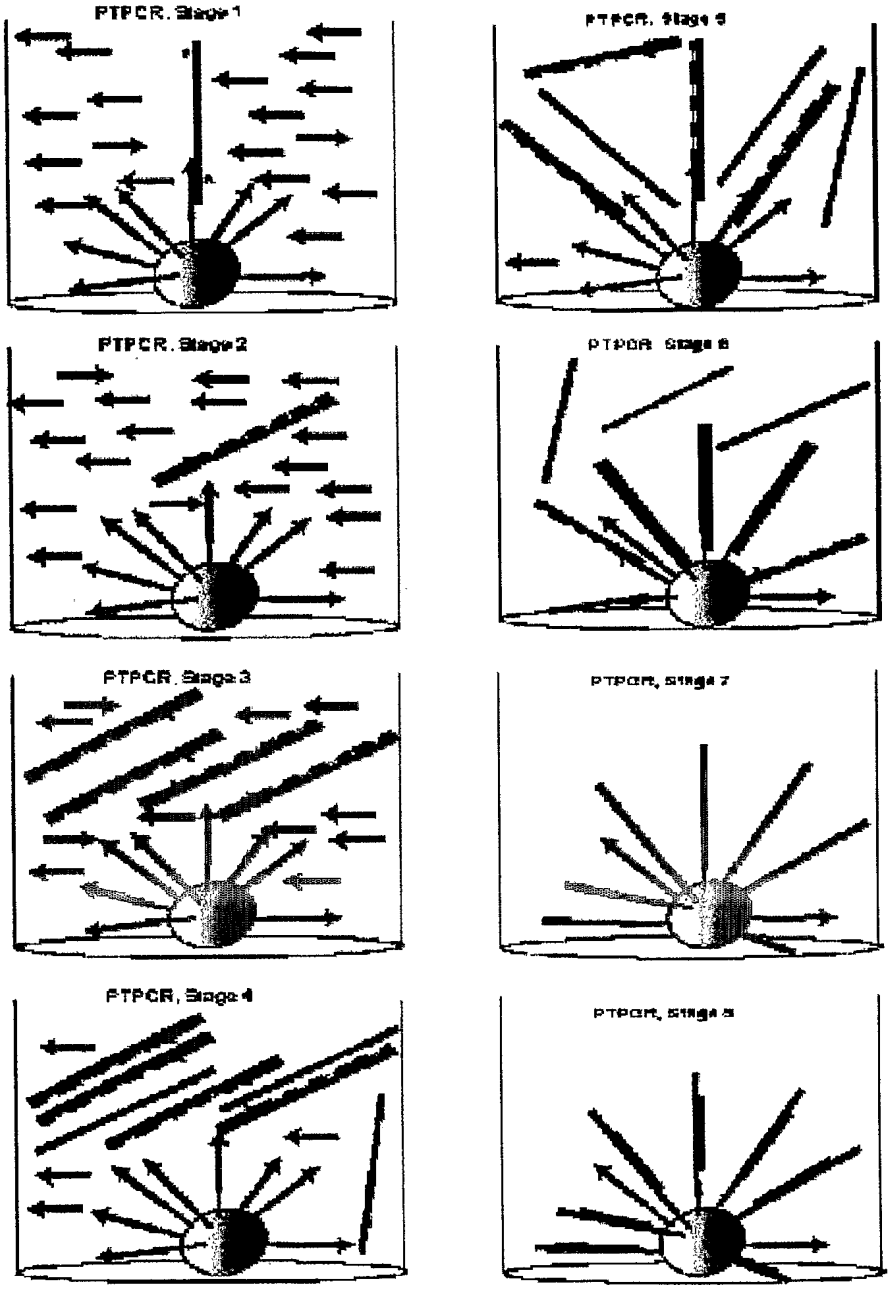

PTPCR, Stage 1
PTPCR, Stage 5
PTPCR, Stage 2
PTPCR, Stage 6
PTPCR, Stage 3
PTPCR, Stage 7
PTPCR, Stage 4
PTPCR, Stage 8

Figure 22

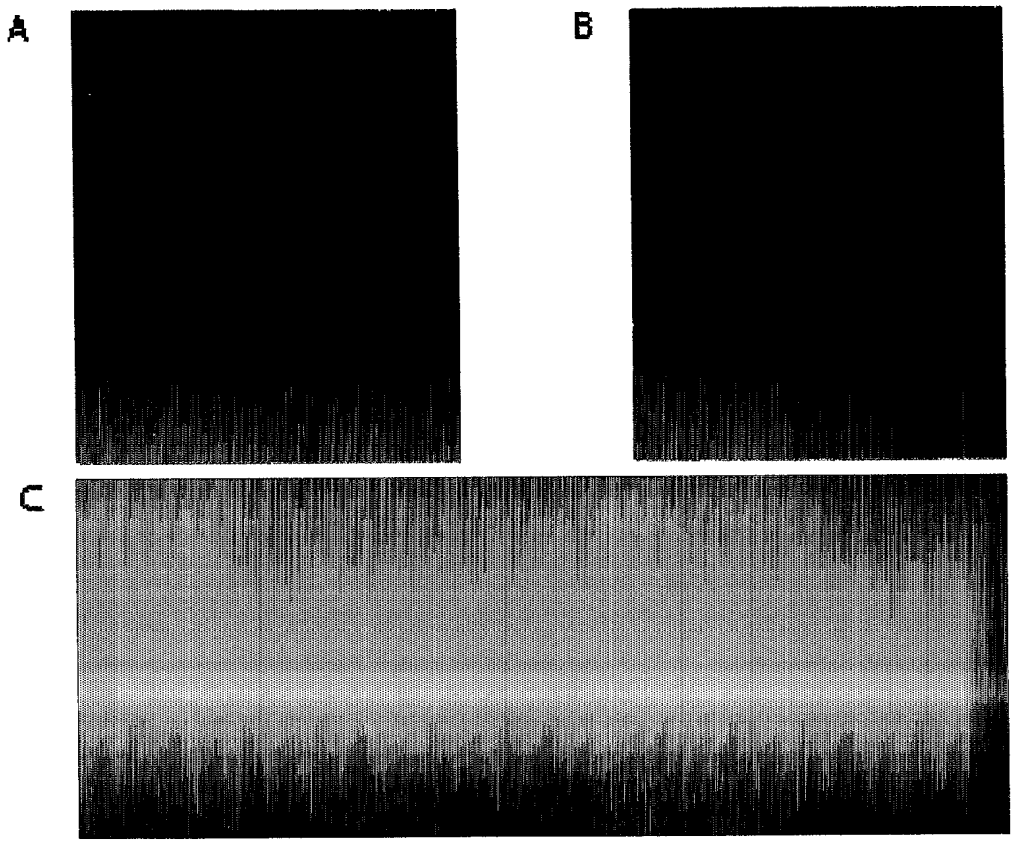


Figure 23

Figure 24

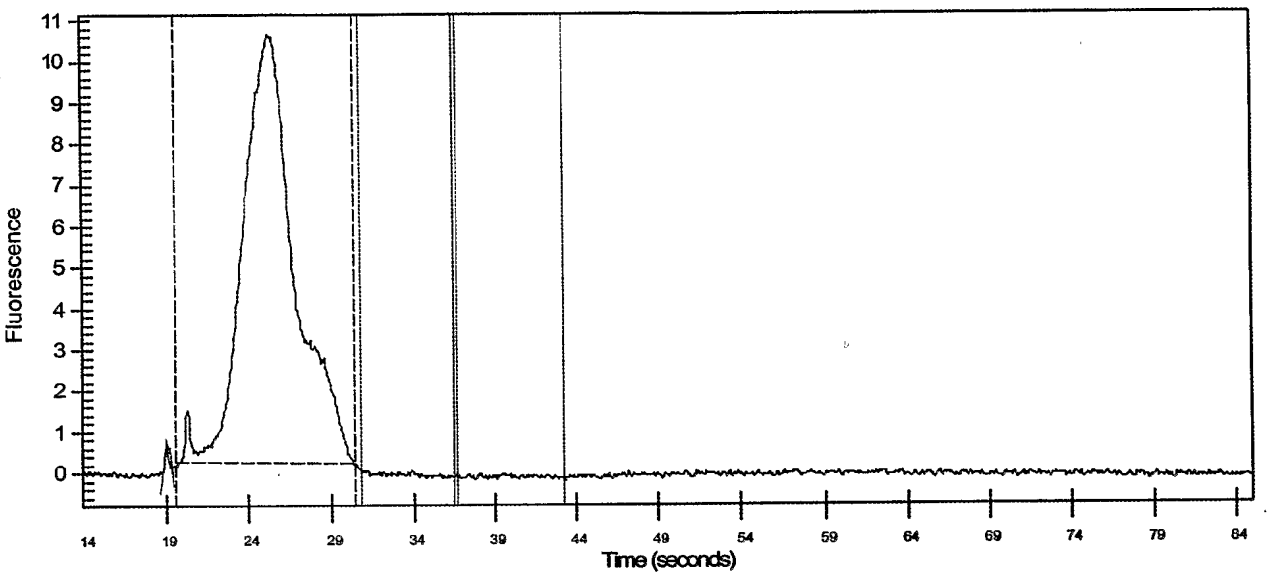

Fluorescence
11
10
9
8
7
6
5
4
3
2
1
0
14
19
24
29
34
39
44
49
54
59
64
69
74
79
84
Time (seconds)

Figure 25

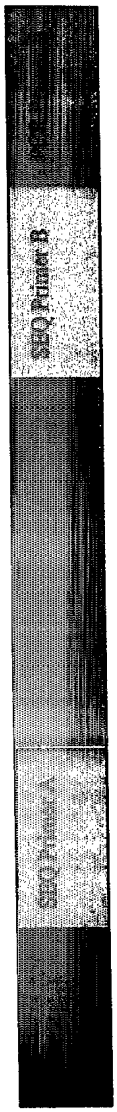


Figure 26

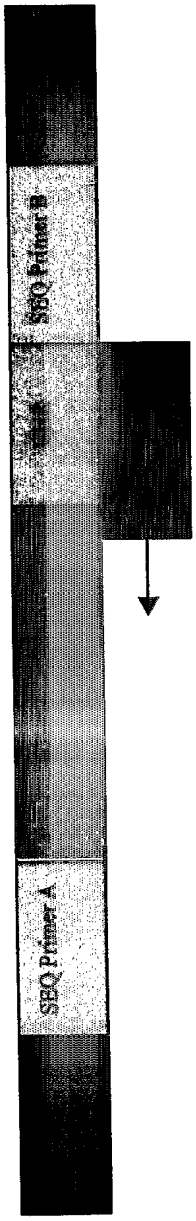


Figure 27

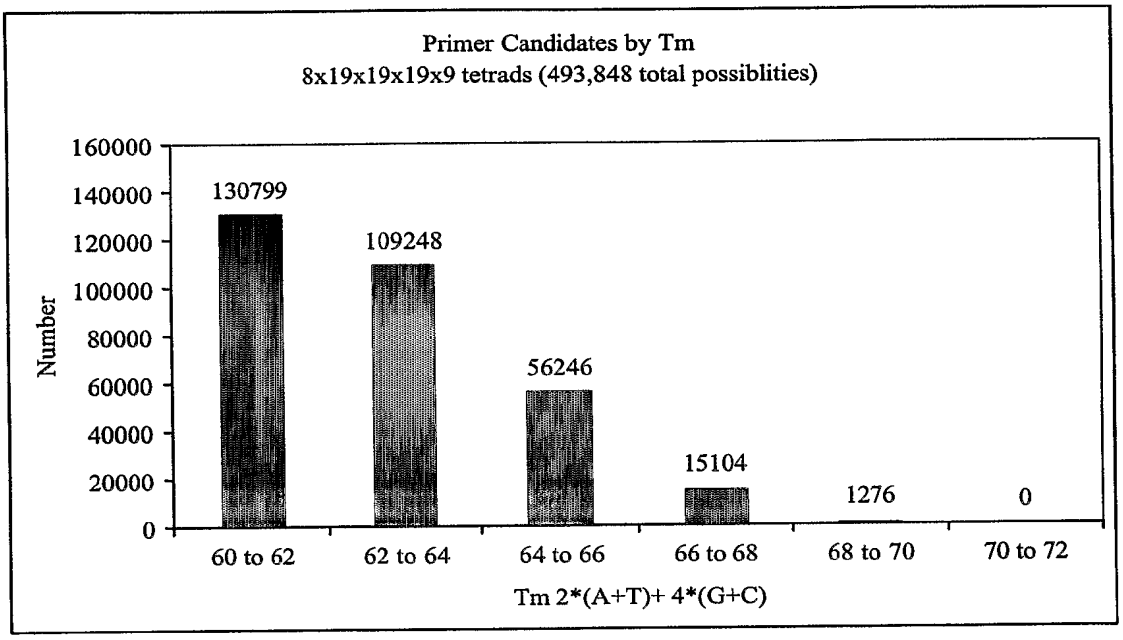

Primer Candidates by Tm
8x19x19x19x9 tetrads (493,848 total possiblities)
Number
160000
140000
120000
100000
80000
60000
40000
20000
0
130799
109248
56246
15104
1276
0
60 to 62
62 to 64
64 to 66
66 to 68
68 to 70
70 to 72
Tm 2*(A+T)+ 4*(G+C)

Figure 28

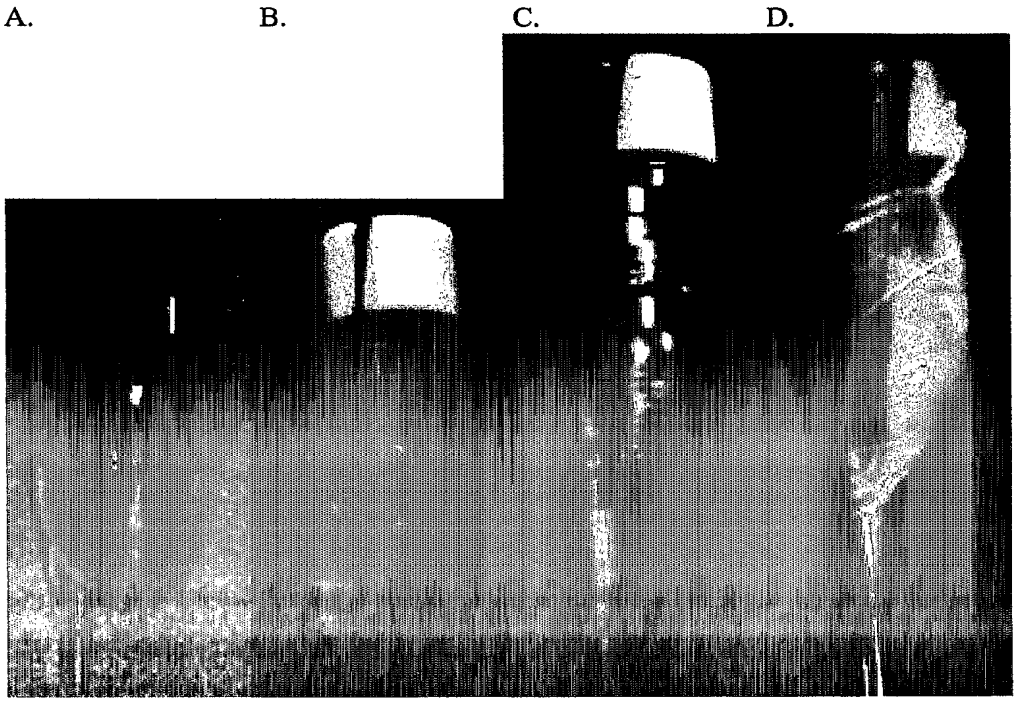


Figure 29A

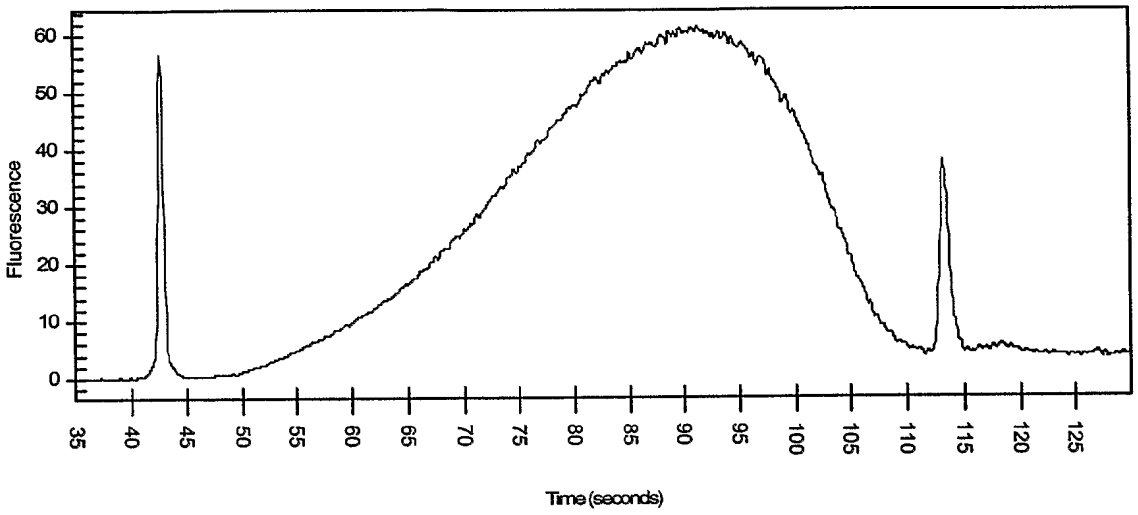
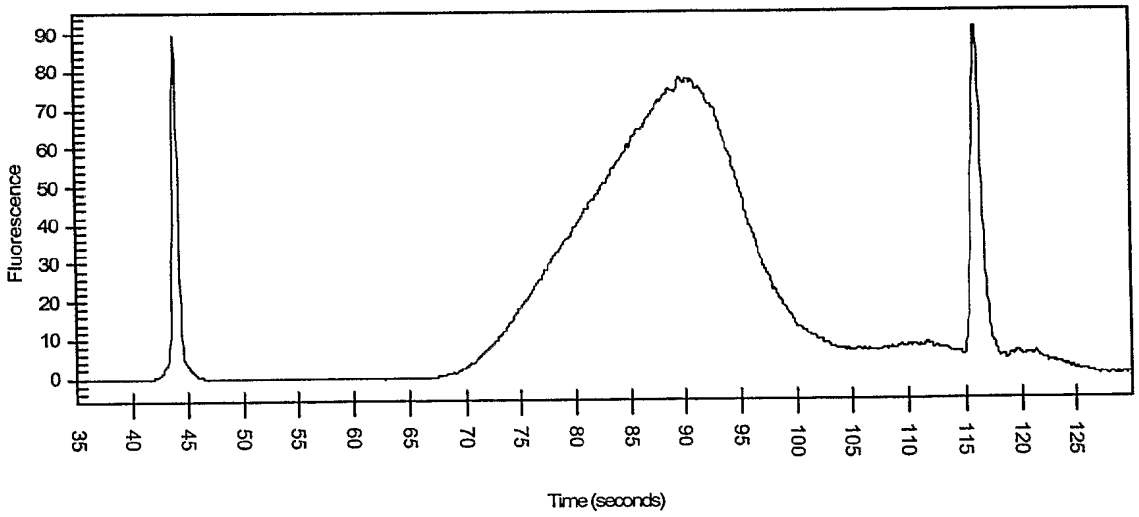


Figure 29B



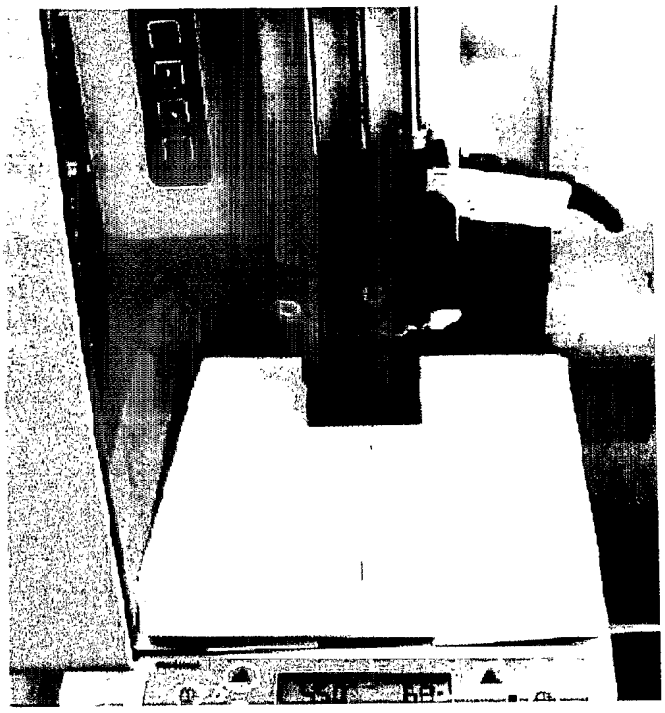


Figure 30

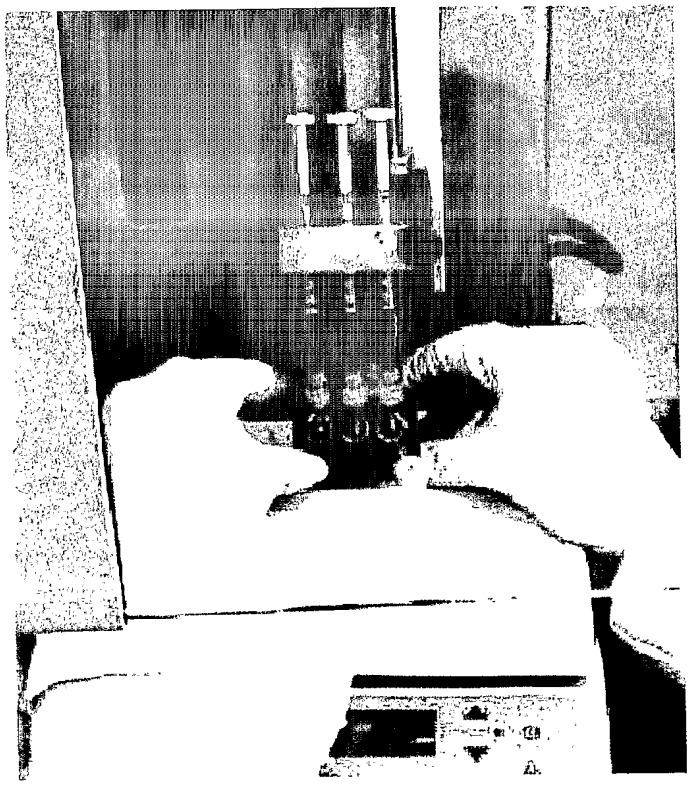


Figure 31

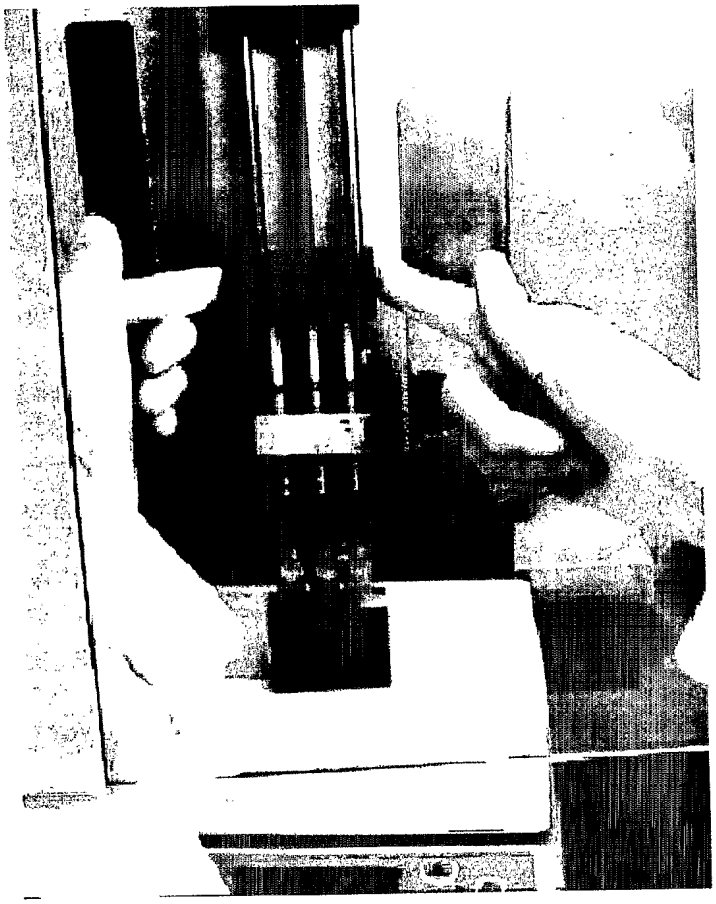


Figure 32

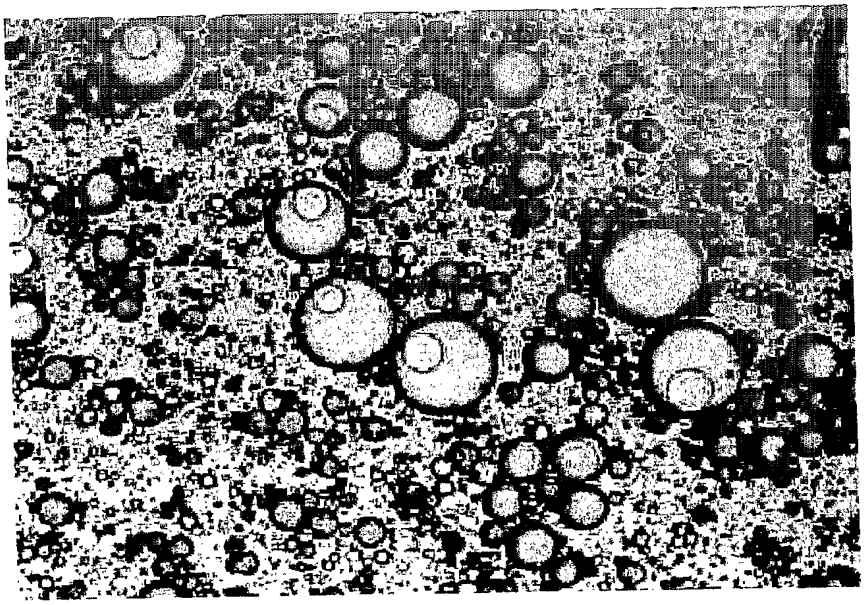
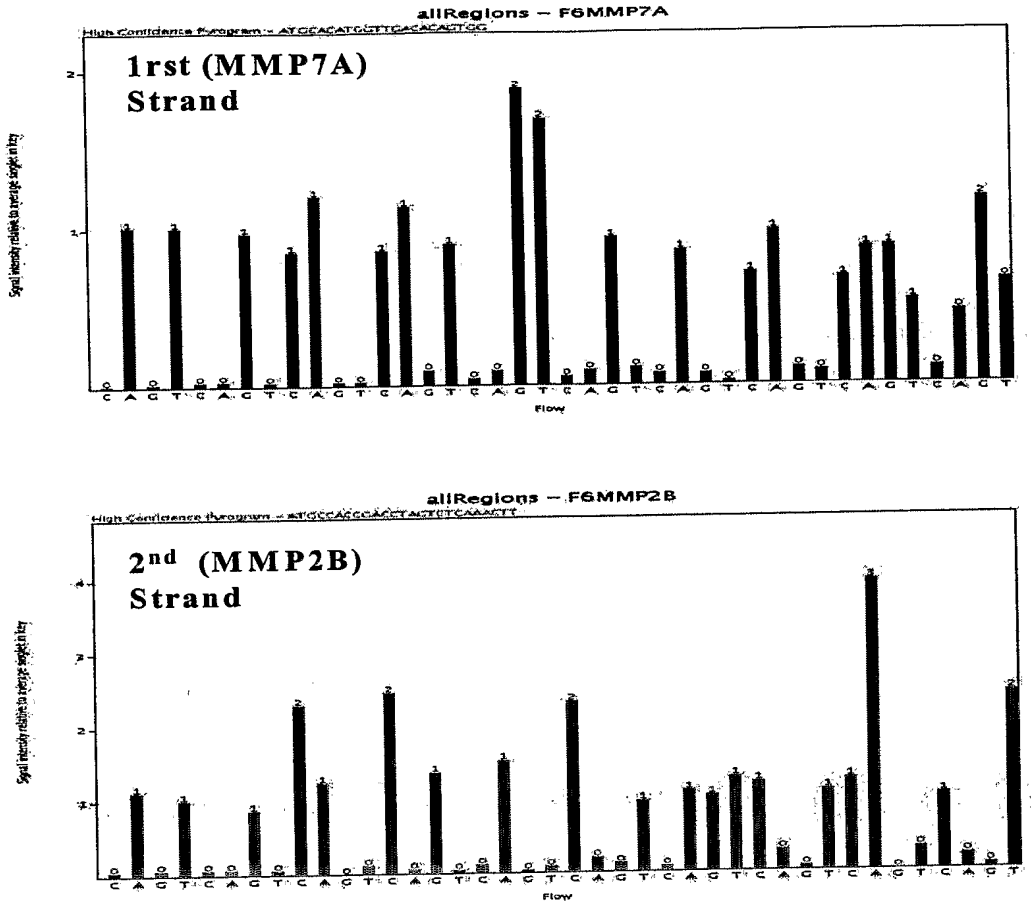


Figure 33

FIGURE 34



1rst Strand (MMP7A)

Sample	Well Location	Sequence	Perfect Match Length
F6_14_1	00001_1362_1660.well	ATGCACATGGTTGACACAGTGGT	22

ATGC ACATGGTTGACACAGTGG

ATGC CACCGACCTAGTCTCAAACTT

2nd Strand (MMP2B)

Sample	Well Location	Sequence	Perfect Match Length
F6_14_1	00003_1363_1660.well	ATGCCACCGACCTAGTCTCAAACTT	25

Fig. 35

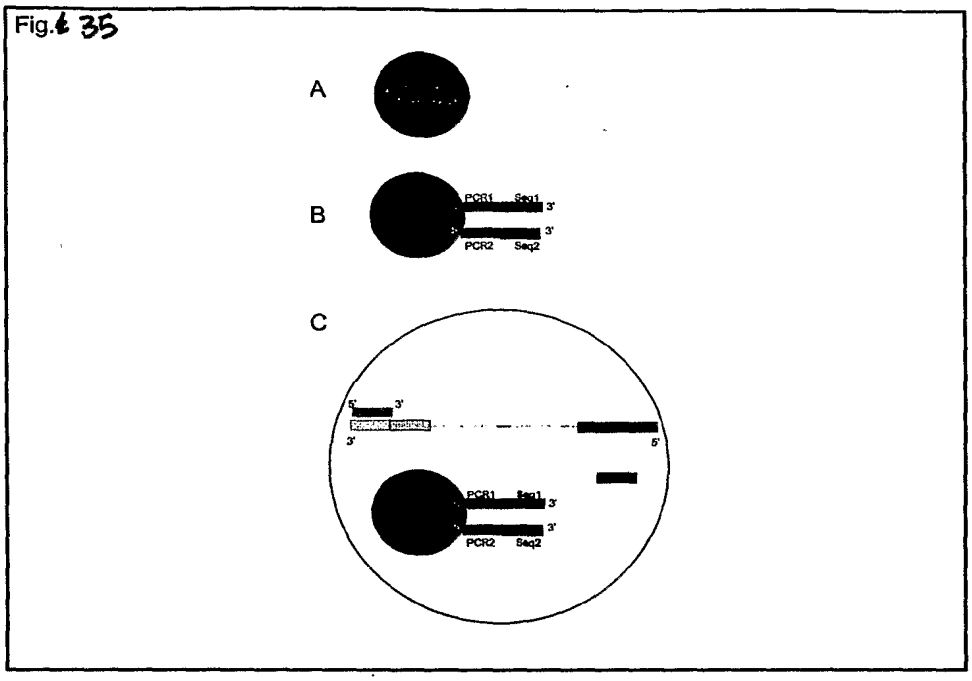

A
B
PCR1 Seq1 3'
PCR2 Seq2 3'
C
5' 3'
3' 5'
PCR1 Seq1 3'
PCR2 Seq2 3'

Fig. 36

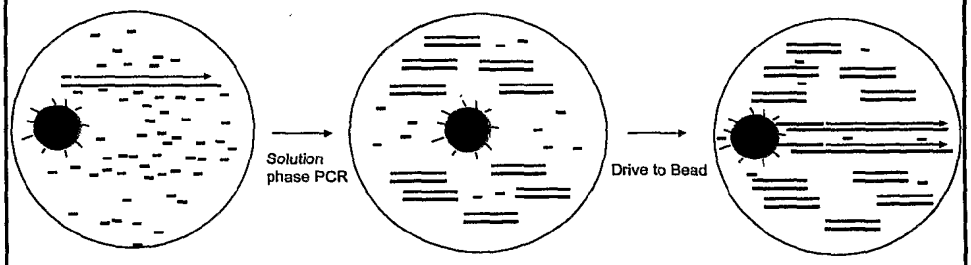

Solution phase PCR
Drive to Bead

Fig.37

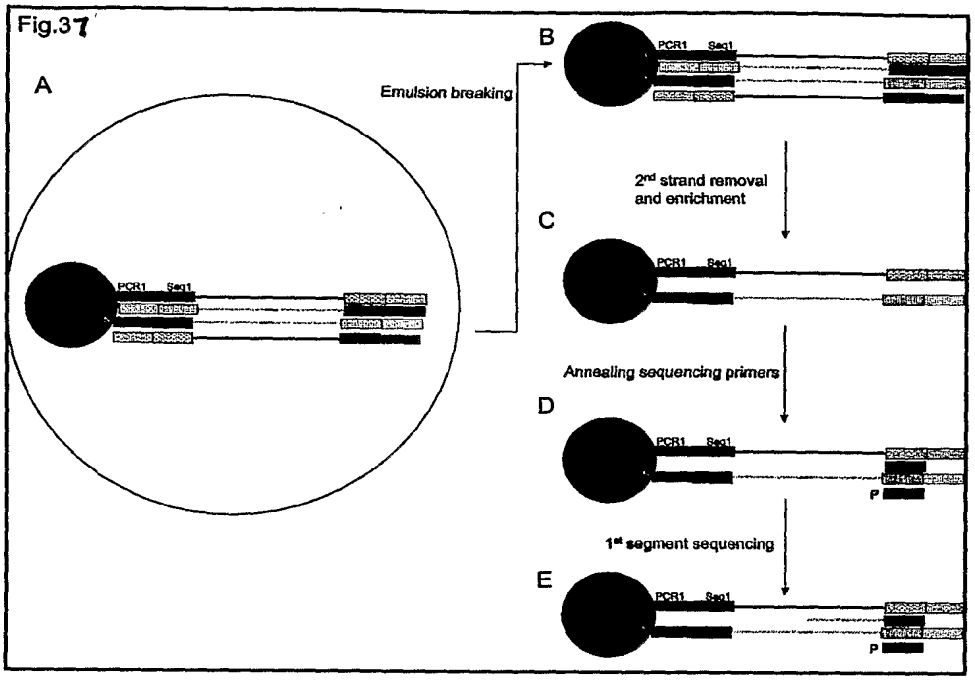

A
Emulsion breaking
B
PCR1
Seq1
2nd strand removal and enrichment
C
Annealing sequencing primers
D
P
1st segment sequencing
E

Fig.38

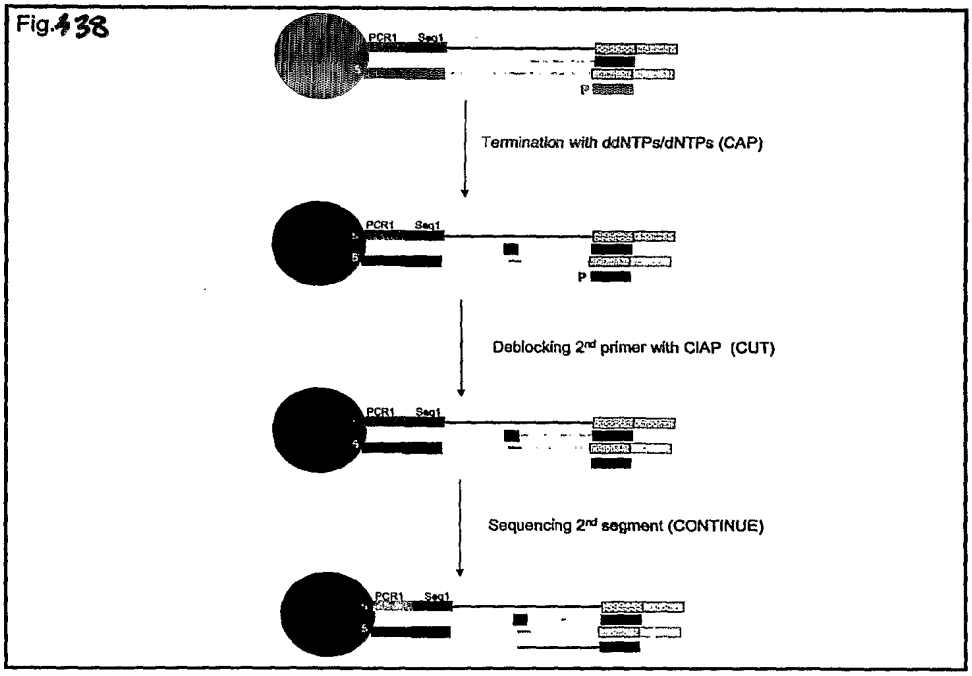
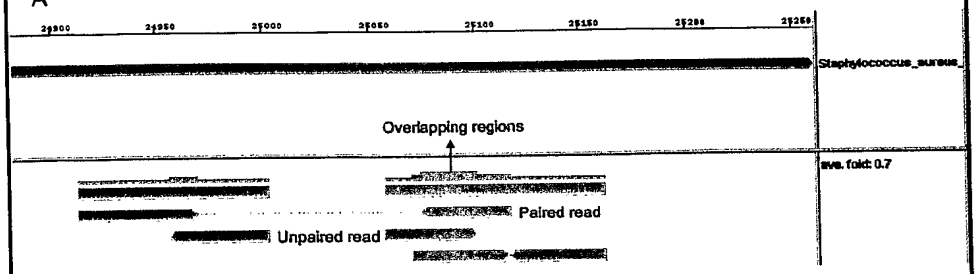

PCR1
Seq1
P
Termination with ddNTPs/dNTPs (CAP)
Deblocking 2nd primer with CIAP (CUT)
Sequencing 2nd segment (CONTINUE)

Fig. 39

A



B

Total Reads	31,785
Total 1st Strand	15,770
Total 2nd strand	16,015
Paired	11,799
Non Paired Reads	8,187
Total Coverage	38%

Fig. 40

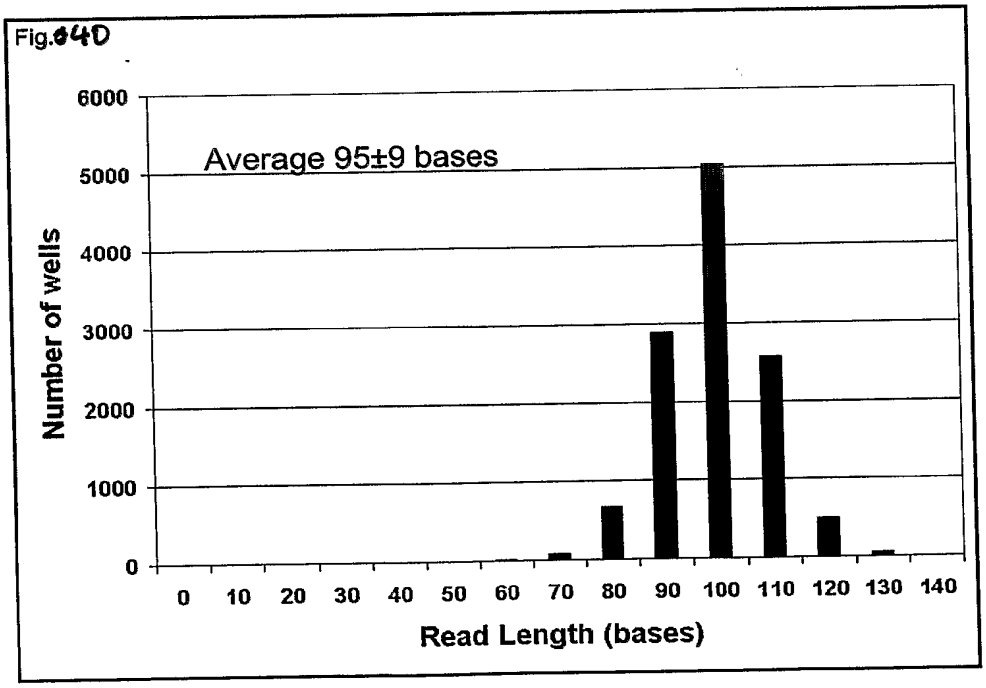


Fig. 41

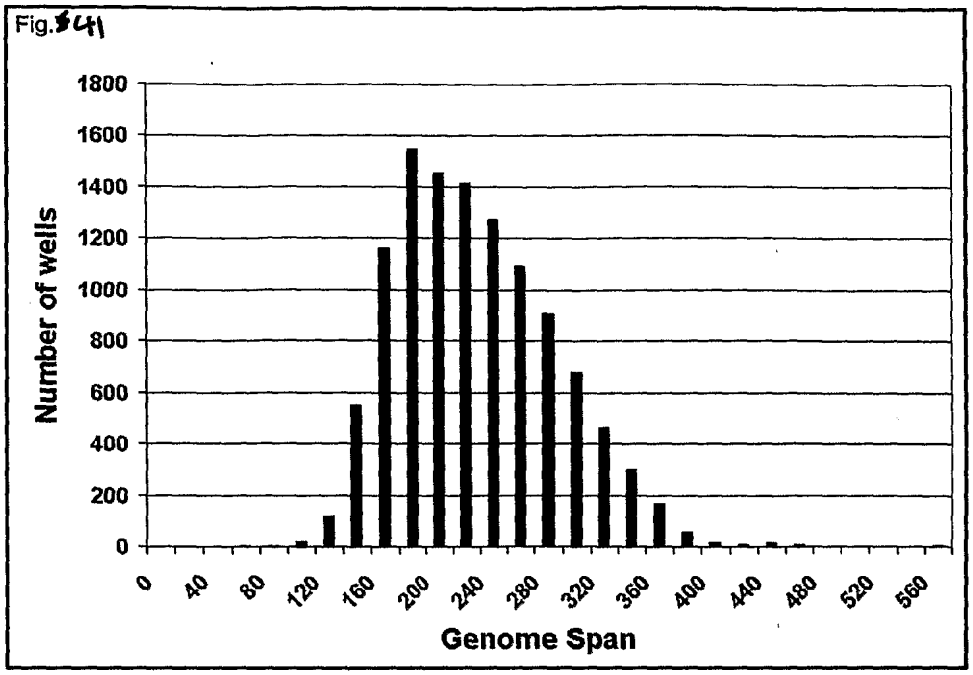


Fig. 42

Well	Genome Position	Orientation	Alignment String
00364_0548_2509	571366	F	TATTGTTGATGCTGTAAAAaGAAGCTACTGGTGTAGtATTTTTATGAAGTT
00364_0548_2509_D2	571512	R	TGCTCAAAGAATTCATTTAAAATATGACCATATTTCATTGTATCTTT
00383_0985_2232	1487890	R	AAGCGAACAGTCAAGTACCACAGTCAGTTGACtTTTACACAAGCGGAT
00383_0985_2232_D2	1487769	F	TACAGGTGTTGGTATGCCATTTGCGATTTGTTGCGCTTGGTTAGCCG
00397_0940_2923	2611093	F	AACATATAAACATCCCCTATCTCAATTTCCGCTTCCATGTAaCAAAAAAAGC
00397_0940_2923_D2	2611164	R	TAGATATCACTTGCGTGTTACTGGTAATGCAGGCATGAG
00417_0611_1933	122001	R	ATTCAACTCTGGAAATGCtTTCTTGATACGCCTCGATGATG
00417_0611_1933_D2	121930	F	GATGAGGAGCTGCAATGGCAATGGGTTAAAGGCATCATCG
00434_0595_0993	2022591	R	TGTATCTCGATTTGGATTAGTTGCtTTTTGCATCTTCATTAGACC
00434_0595_0993_D2	2022473	F	CATTAACATCTGCACCAGAAATAGCTTCTAATACGATTGC
00443_1003_0754	107373	F	GCGACGACGTCCAGCTAATAACGCTGCACCTAAGGCTAATGATAAT
00443_1003_0754_D2	107502	R	AAACCATGCAGATGCTAACAAAGCTCAAGCATTACCAGAAACT
00454_1257_3047	59038	R	TGTTGCTGCATCATAATTTAATACTACATCATTTAAtTCTTTGG
00454_1257_3047_D2	58880	F	GCAGATGGTGTGACTAACCAAGTTGGTCAAAATGCCCTAAATACAAAAGAT

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 2 511 381 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
17.10.2012 Bulletin 2012/42

(51) Int Cl.:
C12Q 1/68 *(2006.01)* ***A01H 1/04*** *(2006.01)*

(21) Application number: **11180635.2**

(22) Date of filing: **09.06.2008**

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MT NL NO PL PT RO SE SI SK TR

(30) Priority: **08.06.2007 US 942707 P**

(62) Document number(s) of the earlier application(s) in accordance with Art. 76 EPC:
08770455.7 / 2 158 336

(27) Previously filed application:
09.06.2008 PCT/US2008/066263

(71) Applicant: **Monsanto Technology LLC**
St. Louis, Missouri 63167 (US)

(72) Inventors:
- **Dotson, Stan**
  St Louis, MO Missouri 63167 (US)
- **Dong, Fenggao**
  St. Louis, MO Missouri 63167 (US)
- **Achard, Fred**
  St. Louis, MO Missouri 63167 (US)
- **Eathington, Sam**
  St. Louis, MO Missouri 63167 (US)
- **Tao, Nengbing**
  St. Louis, MO Missouri 63167 (US)
- **McCuddin, Zoe**
  St. Louis, MO Missouri 63167 (US)

(74) Representative: **Bosch, Henry et al**
Office Hanssens Colens
Square Marie-Louise, 40 bte 19
1000 Bruxelles (BE)

Remarks:
This application was filed on 08-09-2011 as a divisional application to the application mentioned under INID code 62.

(54) Methods for sequence-directed molecular breeding

(57) The present invention provides breeding methods and compositions to enhance the germplasm of a plant by the use of direct nucleic acid sequence information. The methods describe the identification and accumulation of preferred nucleic acid sequences in the germplasm of a breeding population of plants.

EP 2 511 381 A1

Printed by Jouve, 75001 PARIS (FR)

IDT Exhibit 1020.6725

Description

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Application No. 60/942,707 (filed June 8, 2007), which is incorporated by reference herein in its entirety.

FIELD OF INVENTION

[0002] This invention is in the field of plant breeding. More specifically, this invention relates to the use of high throughput sequencing technology in activities related to germplasm improvement.

BACKGROUND OF INVENTION

[0003] The primary objectives of plant breeding are to select an optimal pair of parents to make a cross and then to select one or more superior progeny resulting from that cross. In hybrid crops, a third objective is to identify a tester to make up high performing hybrid seed. Traditional plant breeding has relied on visual observations and performance data on the plants or lines in order to make selections to meet one of the aforementioned objectives.

[0004] In recent years, molecular breeding has demonstrated promise for improving the breeding process and enhancing the rate of genetic gain. In molecular breeding, molecular markers provide a basis for parental, progeny or tester selections; this process may be used in conjunction with phenotype-based selection as well. Inclusion of genetic markers in breeding programs has accelerated the identification and accumulation of valuable traits into germplasm pools compared to that achieved based only on phenotypic data. Herein, "germplasm" includes breeding germplasm, breeding populations, collection of elite inbred lines, populations of random mating individuals, and biparental crosses.

[0005] For molecular breeding to be effective, the differences in marker genotypes must be heritably associated to one or more phenotypic or performance traits. These associations are established by correlating the marker genotypes to lines or populations segregating for one or more traits. Genetic marker alleles (an "allele" is an alternative sequence at a locus) are used to identify plants that contain a desired genotype at one or more loci, and that are expected to transfer the desired genotype, along with a desired phenotype for one or more traits, to their progeny. Markers that are highly correlated with a phenotype are assumed to be genetically linked to the trait, thus the marker can then be used as a basis for selection decisions in lieu of evaluating the trait per se. Markers that are not correlated will be inherited independently of the trait and are not useful for selections, but can be valuable in comparing similarities and/or measuring genetic distances among varieties and lines. Ideally, the marker will represent the actual genomic variation responsible for a trait and will therefore always segregate with the trait, although the correlations can be masked by phenomena such as environmental interactions or epistatic effects.

[0006] Initial marker platforms for molecular breeding did not require *a priori* knowledge of underlying sequence. These markers were based on restriction fragment length polymorphisms (RFLPs). Random or directed DNA probes were used in Southern hybridization protocols to identify target fragments whose size varied depending on the location and distance between a pair of restriction enzyme recognition sites. These differences in size could be correlated to traits in test populations. The DNA probes were then used as markers that could detect the underlying restriction fragment length polymorphisms and in turn be used to predict a correlated trait. Other types of markers have been used that require *a priori* knowledge of the underlying sequence and include but are not limited to fingerprinting using amplified fragment length polymorphisms (AFLPs) or universal PCR primers (i.e. RICE primers).

[0007] In recent years, markers have been developed based on the knowledge of an underlying sequence. For example, microsatellite or simple sequence repeat (SSR) markers rely on PCR and gel electrophoresis to elucidate variation in the length of DNA repeat sequences. The differences in repeat length, as revealed by the markers, can correlate to associated traits if the target repeat is genetically linked to the trait.

[0008] However, traditional marker platforms are suboptimal because they are not suited for automation or high throughput techniques. In addition, traditional marker platforms are susceptible to false marker-trait associations wherein the identity of a genotype between two lines may not reflect a common parent but a convergent sequence, which is problematic for tracking specific marker alleles across multiple generations.

[0009] Other types of variations useful as traditional markers are single nucleotide polymorphisms (SNPs). These are single base changes which differ between two lines and will segregate with a trait in which they are genetically linked. SNPs can be detected by a variety of commercially available marker technologies. Markers based on SNPs have gained in popularity due to the ease and accuracy of detection, compatibility with information systems and low cost. However, SNP markers are still an indirect tool for querying underlying sequence and a SNP marker is restricted to only detecting two alleles, not the four possible nucleotides that might be found at any given nucleotide position.

[0010] Thus, there is a need in the art for methods to quickly and accurately determine direct sequence information

IDT Exhibit 1020.6726

from at least one plant genome for the purpose of facilitating plant breeding activities such as line development, germplasm diversity analyses, rare allele mining, purity testing, quality assurance, introgression of specific genomic regions, stacking of genomic regions, prediction of line performance, and prediction of hybrid performance.

SUMMARY

[0011] This invention describes novel methods that utilize high throughput sequencing and molecular breeding methodologies to enable the use of direct sequencing information in molecular plant breeding. The invention also includes means to selectively target specific loci and to DNA tag samples prior to sequence determination. Taken together, the methods of the invention enable plant breeders better tools for parent selection, progeny selection, choosing tester combinations, developing pedigrees, fingerprinting samples, screening for haplotype diversity, ensuring quality, assessing germplasm diversity, measure breeding progress, providing variety or line descriptions and for building databases of sequence associations to trait and performance data. Such databases provide the basis for calculating nucleic acid effect estimates for one or more traits, wherein associations can be made *de novo* or by leveraging historical nucleic acid sequence-trait association data.

[0012] The present invention provides methods for Sequence Directed Selection (SDS), Sequence Directed Breeding (SDB) and Sequence Directed Fingerprinting (SDF) and its novel application for making parent selections, progeny selections, tester combinations, introgression of allelic variants and directed selection of at least one variant between at least two germplasm entries, fingerprints, pedigrees and for building databases of haplotype and phenotype information which can be used to calculate nucleic acid sequence effect estimates and, ultimately, breeding values. This a priori information facilitates decision making for Sequence Directed Predictive Breeding (SDPB).

[0013] In the present invention, breeding selections are conducted directly on a sequence, rather than indirectly on a marker, basis, wherein a first plant is crossed with a second plant that contains at least one sequence that is different from the first plant sequence or sequences; and at least one progeny plant is selected by detecting the sequence or set of sequences of the first plant, wherein the progeny plant comprises in its genome one or more sequences of interest of the first plant and at least one sequence of interest of the second plant; and the progeny plant is used in activities related to germplasm improvement, herein defined as including using the plant for line and variety development, hybrid development, transgenic event selection, making breeding crosses, testing and advancing a plant through self fertilization, purification of lines or sublines, using plant or parts thereof for transformation, using plants or parts thereof for candidates for expression constructs, and using plant or parts thereof for mutagenesis.

[0014] The present invention includes a method for breeding of a plant, such as maize (*Zea mays*), soybean (*Glycine max*), cotton (*Gossypium hirsutum*), peanut (*Arachis hypogaea*), barley (*Hordeum vulgare*); oats (*Avena sativa*); orchard grass (*Dactylis glomerata);* rice *(Oryza sativa,* including indica and japonica varieties); sorghum *(Sorghum bicolor);* sugar cane *(Saccharum sp);* tall fescue *(Festuca arundinacea);* turfgrass species (e.g. species: *Agrostis stolonifera, Poa pratensis, Stenotaphrum secundatum);* wheat (*Triticum aestivum*), and alfalfa (*Medicago sativa*), members of the genus *Brassica,* broccoli, cabbage, carrot, cauliflower, Chinese cabbage, cucumber, dry bean, eggplant, fennel, garden beans, gourd, leek, lettuce, melon, okra, onion, pea, pepper, pumpkin, radish, spinach, squash, sweet corn, tomato, watermelon, ornamental plants, and other fruit, vegetable, tuber, oilseed, and root crops, wherein oilseed crops include soybean, canola, oil seed rape, oil palm, sunflower, olive, corn, cottonseed, peanut, flaxseed, safflower, and coconut, with enhanced traits comprising at least one sequence of interest, further defined as conferring a preferred property selected from the group consisting of herbicide tolerance, disease resistance, insect or pest resistance, altered fatty acid, protein or carbohydrate metabolism, increased grain yield, increased oil, increased nutritional content, increased growth rates, enhanced stress tolerance, preferred maturity, enhanced organoleptic properties, altered morphological characteristics, other agronomic traits, traits for industrial uses, or traits for improved consumer appeal, wherein said traits may be nontransgenic or transgenic.

[0015] In one embodiment, the invention is directed to a method of plant breeding. The method comprises determining the sequence of a plurality of nucleic acids within the genome of at least one or more plants in a breeding population; associating each of the nucleic acid sequences with a numerical value wherein the numerical value is related to one or more phenotypic traits: and making a plant breeding decision for the one or more plants based on the association.

[0016] In another embodiment, the invention is directed to a method of plant breeding. The method comprises providing a breeding population comprising one or more plants wherein at least one nucleic acid is sequenced for at least one locus for each plant in the population; utilizing historical marker-phenotypic trait associations to determine a nucleic acid sequence effect estimate for a nucleic acid sequence at a locus; and ranking nucleic acid sequences based on the determined nucleic acid sequence effect estimate for any given phenotypic trait. The ranking is then used to make plant breeding decisions.

[0017] In another embodiment, the invention is directed to a method of plant breeding. The method comprises establishing a fingerprint map defining a plurality of loci within the genome of a breeding population; associating a QTL allele with known map location with a phenotypic trait in a mapping population; and assaying for presence of the QTL allele

IDT Exhibit 1020.6727

and at least one nucleic acid sequence within the plurality of loci to predict expression of the phenotypic trait in a population other than the mapping population.

[0018] In another embodiment, the invention is directed to a method of marker assisted breeding. The method comprises providing a breeding population comprising at least two plants and associating at least one phenotypic trait with a locus of the genome of the plants, provided that the locus is defined by at least one nucleic acid sequence. The population is then assayed for the presence of at least one nucleic acid sequence of the locus to predict the expression of at least one phenotypic trait in a progeny plant of the breeding population.

[0019] In another embodiment, the invention is directed to a method of selecting a breeding population for use in a breeding program. The method comprises providing at least two distinct breeding populations; establishing a database of breeding values for at least two loci of up to 10 centimorgans for each breeding population; ranking the breeding values of the alleles for each breeding population; and selecting a breeding population with a higher composite breeding value.

[0020] Further areas of applicability will be more particularly described below in relation to the detailed description. It should be understood that the description and specific examples are intended for purposes of illustration only and are not intended to limit the scope of the present disclosure.

DESCRIPTION OF DRAWINGS

[0021] Figure 1 is a generic flow diagram illustrating the molecular process of high throughput nucleic acid sequencing.

[0022] Figure 2 illustrates a method for reducing complexity of template nucleic acids from selective digestion.

[0023] Figure 3 illustrates a method for targeted complexity reduction from the transcriptome.

[0024] Figure 4 illustrates a method for targeted complexity reduction by amplification of at least one genomic region of interest.

[0025] Figure 5 illustrates a method for targeted complexity reduction, including sample tagging, by allele specific extension/ligation.

[0026] Figure 6 illustrates a method for the multiplexing of samples using DNA tags attached to the template nucleic acids through ligation.

[0027] Figure 7 illustrates a method for the multiplexing of samples using DNA tags attached to the template nucleic acids through PCR.

[0028] Figure 8 illustrates a workflow for high throughput nucleic acid sequencing.

[0029] Figure 9 illustrates a method for preparing samples for sequence directed selection for a SNP and an indel.

[0030] Figure 10 is a scatter plot for results of genotyping for the purpose of sequence-directed selection using high throughput sequencing for the Fad3b SNP as described in Example 1.

[0031] Figure 11 is a scatter plot for results of genotyping for the purpose of sequence-directed selection using high throughput sequencing for the Fad3c indel as described in Example 1.

[0032] Figure 12 illustrates a strategy for adding sample DNA tags with allele-specific extension/ligation as described in Example 4.

[0033] Figure 13 illustrates the success rate of fingerprinting using high throughput sequencing technology for 1536 SNPs in 96 soybean varieties as described in Example 4.

DETAILED DESCRIPTION

[0034] The definitions and methods provided define the present invention and guide those of ordinary skill in the art in the practice of the present invention. Unless otherwise noted, terms are to be understood according to conventional usage by those of ordinary skill in the relevant art. Definitions of common terms in molecular biology may also be found in Albers et al., Molecular Biology of The Cell, 5th Edition, Garland Science Publishing, Inc.: New York, 2007; Rieger et al., Glossary of Genetics: Classical and Molecular, 5th edition, Springer-Verlag: New York, 1991; King et al, A Dictionary of Genetics, 6th ed, Oxford University Press: New York, 2002; and Lewin, Genes IX, Oxford University Press: New York, 2007. The nomenclature for DNA bases as set forth at 37 CFR § 1.822 is used.

[0035] An "allele" refers to an alternative sequence at a particular locus; the length of an allele can be as small as 1 nucleotide base. Allelic sequence can be denoted as nucleic acid sequence or as amino acid sequence that is encoded by the nucleic acid sequence.

[0036] A "locus" is a position on a genomic sequence that is usually found by a point of reference; e.g., a short DNA sequence that is a gene, or part of a gene or intergenic region. A locus may refer to a nucleotide position at a reference point on a chromosome, such as a position from the end of the chromosome. The ordered list of loci known for a particular genome is called a genetic map. A variant of the DNA sequence at a given locus is called an allele and variation at a locus, i.e., two or more alleles, constitutes a polymorphism. The polymorphic sites of any nucleic acid sequence can be determined by comparing the nucleic acid sequences at one or more loci in two or more germplasm entries.

IDT Exhibit 1020.6728

[0037] As used herein, a "nucleic acid sequence" comprises a contiguous region of nucleotides at a locus within the genome. Further, a nucleic acid sequence, as used herein, may comprise one or more haplotypes, portions of one or more haplotypes, one or more genes, portions of one or more genes, one or more QTL, and portions of one or more QTL. In addition, a plurality of nucleic acid sequences can comprise one or more haplotypes, portions of one or more haplotypes, one or more genes, portions of one or more genes, one or more QTL, and portions of one or more QTL. The sequence may originate from a DNA or RNA template, either directly or indirectly (i.e., cDNA obtained from reverse transcription of mRNA).

[0038] As used herein, "polymorphism" means the presence of one or more variations of a nucleic acid sequence at one or more loci in a population of one or more individuals. The variation may comprise but is not limited to one or more base changes, the insertion of one or more nucleotides or the deletion of one or more nucleotides. A polymorphism may arise from random processes in nucleic acid replication, through mutagenesis, as a result of mobile genomic elements, from copy number variation and during the process of meiosis, such as unequal crossing over, genome duplication and chromosome breaks and fusions. The variation can be commonly found, or may exist at low frequency within a population, the former having greater utility in general plant breeding and the latter may be associated with rare but important phenotypic variation. Useful polymorphisms may include single nucleotide polymorphisms (SNPs), insertions or deletions in DNA sequence (Indels), simple sequence repeats of DNA sequence (SSRs) a restriction fragment length polymorphism, and a tag SNP. A genetic marker, a gene, a DNA-derived sequence, a haplotype, a RNA-derived sequence, a promoter, a 5' untranslated region of a gene, a 3' untranslated region of a gene, microRNA, siRNA, a QTL, a satellite marker, a transgene, mRNA, ds mRNA, a transcriptional profile, and a methylation pattern may comprise polymorphisms. In addition, the presence, absence, or variation in copy number of the preceding may comprise a polymorphism.

[0039] As used herein, "nucleic acid effect estimate" means a predicted effect estimate for a nucleic acid sequence reflecting association with one or more phenotypic traits, wherein said associations can be made *de novo* or by leveraging historical nucleic acid sequence-trait association data.

[0040] As used herein, "breeding value" means a calculation based on nucleic acid sequence effect estimates and nucleic acid sequence frequency values, the breeding value of a specific nucleic acid sequence relative to other nucleic acid sequences at the same locus (i.e., haplotype window), or across loci (i.e., haplotype windows), can also be determined. In other words, the change in population mean by fixing said nucleic acid sequence is determined. In addition, in the context of evaluating the effect of substituting a specific region in the genome, either by introgression or a transgenic event, breeding values provide the basis for comparing specific nucleic acid sequences for substitution effects. Also, in hybrid crops, the breeding value of nucleic acid sequences can be calculated in the context of the nucleic acid sequence in the tester used to produce the hybrid.

[0041] As used herein, "genotype" is the actual nucleic acid sequence at a locus in an individual plant. As opposed to a genetic marker such as a SNP, where the genotype comprises a single nucleotide, the genotype identified with the present invention is a plurality of nucleotides, where the length of the genotype is contingent on the length of the nucleic acid sequence. Notably, a genetic marker assay as known in the art (e.g., SNP detection via TaqMan) detects only two alleles. An advantage of the present invention is the ability to directly query all four nucleotides (adenine, A; thymine, T; cytosine, C; and guanine, G) simultaneously at any one nucleotide position. That is, for any one base pair position, there will be twice the information when using direct nucleic acid sequencing versus genetic marker assays. This can be very important in determining whether two lines share DNA that is identical by descent. With a SNP genotype, one can only assess whether a pair of alternative nucleic acid bases exist at a single nucleotide locus. For example, one might query whether two lines have a C or a T at a single nucleotide locus and find that one line has a C but the other has neither. However, unlike directly assessing the sequence at the single nucleotide locus, the genetic marker assay will not distinguish a failed reaction or whether an alternative base, such as an adenine or guanidine, is present at that locus. Therefore, the present invention provides greater certainty whether a given region is identical by descent by observing the nucleic acid sequence for that region.

[0042] As used herein, a nucleic acid sequence can comprise 1 or more nucleotides (for example, 2 or more nucleotides, 25 or more nucleotides, 250 or more nucleotides, 1,000 or more nucleotides, even 20,000 or more nucleotides). In certain embodiments, adjacent nucleic acid sequence fragments can be ligated *in vitro* or aligned *in silico* for the purpose of obtaining a longer nucleic acid sequence. As used herein, a nucleic acid sequence from each of two or more individual plants from the same genomic region, that may or may not be associated with one or more phenotypic trait values, provides the basis for decisions related to germplasm improvement activities, wherein one or more loci can be evaluated. Knowing whether two sequences at a locus are completely identical or if they contain combinations of identical and non-identical loci can aid in determining whether the loci have the same trait value, are linked to the same traits or are identical by descent. Therefore in another aspect, one or more nucleic acid sequences from one or more individual plants that are associated with a phenotypic trait value can provide the basis for decisions related to germplasm improvement activities.

[0043] As used herein, the term "haplotype" means a chromosomal region within a haplotype window. Typically, the unique marker fingerprint combinations in each haplotype window define and differentiate individual haplotypes for that

IDT Exhibit 1020.6729

window. As used herein, a haplotype is defined and differentiated by one or more nucleic acid sequences at one or more loci within a "haplotype window."

[0044] As used herein, the term "haplotype window" means a chromosomal region that is established by statistical analyses known to those of skill in the art and is in linkage disequilibrium. In the art, identity by state between two inbred individuals (or two gametes) at one or more molecular marker loci located within this region is taken as evidence of identity-by-descent of the entire region, wherein each haplotype window includes at least one polymorphic molecular marker. As used herein, haplotype windows are defined by two or more nucleic acid sequence genotypes. Haplotype windows can be mapped along each chromosome in the genome and do not necessarily need to be contiguous. Haplotype windows are not fixed per se and, given the ever-increasing amount of nucleic acid sequence information, this invention anticipates the number and size of haplotype windows to evolve, with the number of windows increasing and their respective sizes decreasing, thus resulting in an ever-increasing degree confidence in ascertaining identity by descent based on the identity by state of genotypes. Haplotype windows are useful in delineating nucleic acid sequences of interest because these genomic regions tend to be inherited as linkage blocks and thus are informative for association mapping and for tracking across multiple generations.

[0045] As used herein, "phenotype" means the detectable characteristics of a cell or organism which can be influenced by genotype.

[0046] As used herein, "marker" means a detectable characteristic that can be used to discriminate between organisms. Examples of such characteristics may include genetic markers, protein composition, protein levels, oil composition, oil levels, carbohydrate composition, carbohydrate levels, fatty acid composition, fatty acid levels, amino acid composition, amino acid levels, biopolymers, pharmaceuticals, starch composition, starch levels, fermentable starch, fermentation yield, fermentation efficiency, energy yield, secondary compounds, metabolites, morphological characteristics, and agronomic characteristics. As used herein, "genetic marker" means polymorphic nucleic acid sequence or nucleic acid feature.

[0047] As used herein, "marker assay" means a method for detecting a polymorphism at a particular locus using a particular method, e.g. measurement of at least one phenotype (such as seed color, flower color, or other visually detectable trait), restriction fragment length polymorphism (RFLP), single base extension, electrophoresis, sequence alignment, allelic specific oligonucleotide hybridization (ASO), random amplified polymorphic DNA (RAPD), microarray-based technologies, and nucleic acid sequencing technologies, etc.

[0048] As used herein, "consensus sequence" means a constructed DNA sequence which identifies single nucleotide and Indel polymorphisms in alleles at a locus. Consensus sequence can be based on either strand of DNA at the locus and states the nucleotide base of either one of each SNP in the locus and the nucleotide bases of all Indels in the locus. Thus, although a consensus sequence may not be a copy of an actual DNA sequence, a consensus sequence is useful for precisely designing primers and probes for actual polymorphisms in the locus.

[0049] As used herein, "linkage" refers to relative frequency at which types of gametes are produced in a cross. For example, if locus A has genes "A" or "a" and locus B has genes "B" or "b" and a cross between parent I with AABB and parent B with aabb will produce four possible gametes where the genes are segregated into AB, Ab, aB and ab. The null expectation is that there will be independent equal segregation into each of the four possible genotypes, i.e. with no linkage ¼ of the gametes will of each genotype. Segregation of gametes into a genotypes differing from ¼ are attributed to linkage.

[0050] As used herein, "linkage disequilibrium" is defined in the context of the relative frequency of gamete types in a population of many individuals in a single generation. If the frequency of allele A is p, a is p', B is q and b is q', then the expected frequency (with no linkage disequilibrium) of genotype AB is pq, Ab is pq', aB is p'q and ab is p'q'. Any deviation from the expected frequency is called linkage disequilibrium. Two loci are said to be "genetically linked" when they are in linkage disequilibrium.

[0051] As used herein, "quantitative trait locus (QTL)" means a locus that controls to some degree numerically representable traits that are usually continuously distributed.

[0052] As used herein, "complexity reduction" refers to methods to reduce the complexity of a nucleic acid sample, such as by restriction enzyme digestion, reverse transcription, targeted amplification by PCR methods, or random amplification by PCR methods. Complexity reduction can be performed on total genomic nucleic acids or a subset thereof. In a preferred aspect, a method with reproducible results will be used. Methods for complexity reduction are included in WO 06/137734, WO 06/137733, and EP 0 534 858 which are specifically incorporated herein by reference in their entirety.

[0053] As used herein, "DNA tag" means a short segment of DNA used as an identifier for a nucleic acid sample. A DNA tag, also known as a molecular barcode, can range from about 2 to about 20 base pairs in length and can be added during complexity reduction of the template nucleic acid sample(s). For examples, sets of DNA tags are available in U.S. Patent No. 7,157,564. The tag can be identified via sequencing or microarray methods as described in EP 1 724 348. In other embodiments, such as in the case of oligonucleotides mass tags, mass spectrometry methods have been used to differentiate tags (Zhang et al. PNAS 2007 104:3061-3066). Further, molecular barcodes have been developed for detection by other imaging platforms, including surface plasmon resonance, fluorescent, or Raman spectroscopy, as

IDT Exhibit 1020.6730

described in U.S. Patent Application 2007/0054288. In another embodiment, spike-in tags of RNA or protein have been used which are distinct from molecules of the target sample and are co-analyzed with a plurality of samples for the purpose of sample discrimination, methods of which are included in WO 03/052101. In a preferred embodiment of this invention, the identity of the tag is assessed by sequencing either directly before or directly after the sequencing of a trait locus. In this way, the sequence of the tag conjugated to the sequence of the locus and can be used to maintain a linkage between the locus sequence and the sample origin. In another embodiment, the tag may be combinatorial or hierarchical. For example one portion of the tag may indicate multiple nucleic acids are from the same sample and another portion of the tag may indicate the nucleic acids were derived from different subsamples. The number of hierarchical levels or combinations of tags is only limited by the amount of sequencing which can be dedicated to the DNA tag vs. the trait locus.

[0054] As used herein, a "tagged sample" means a sample of nucleic acid to which the same tag has been attached to each individual nucleic acid in the sample. As used herein, a tagged sample includes a samples tagged with a hierarchical or combinatorial tag, wherein at least a portion of the tag is identical and attached to each nucleic acid sequence in the sample.

[0055] As used herein, an "allele-specific tag" is a DNA tag that corresponds to a particular allele in the target sequence. In a preferred embodiment, only the allele-specific tag, rather than the polymorphism plus any linked DNA tags, needs to be sequenced to be able to genotype the corresponding polymorphism.

[0056] As used herein, "nucleic acid sequencing" means the determination of the order of nucleotides in a sample of nucleic acids, wherein nucleic acids include DNA and RNA molecules. "High throughput nucleic acid sequencing" means an automated and massively parallel approach for the determination of nucleotides in a sample of nucleic acids wherein examples of high throughput nucleic acid sequencing technology include, but are not limited to, platforms provided by 454 Life Sciences, Agencourt Bioscience, Applied Biosystems, LI-COR Biosciences, Microchip Biotechnologies, Network Biosystems, NimbleGen Systems, Illumina, and VisiGen Biotechnologies, comprising but not limited to formats such as parallel bead arrays, sequencing by synthesis, sequencing by ligation, capillary electrophoresis, electronic microchips, "biochips," microarrays, parallel microchips, and single-molecule arrays, as reviewed by Service (Science 2006 311: 1544-1546).

[0057] As used herein, "aligning" or "alignment" of two or more nucleic acid sequences is the comparison of the nucleic acid sequences found at the same locus. Several methods of alignment are known in the art and are included in most of the popular bioinformatics packages.

[0058] As used herein, the term "primer" means a single strand of synthetic oligonucleotide, preferably from about 10 to about 120 nucleotides, which can be synthesized chemically or assembled from several chemically synthesized oligonucleotides. As used herein, primers may be used to initiate sequencing reactions and polymerase reactions, such as in gap fill reactions and PCR. As used herein, a primer will hybridize under the assay conditions specifically to a desired target sequence. As used herein, primers may be used to introduce a DNA tag, to introduce chemically modified bases, such as biotin labeled bases, or to introduce a hybridization sequence that can subsequently be used for capture, such as capture to a sequencing matrix or to an avidin-containing surface.

[0059] As used herein, the term "adapters" means a double stranded nucleic acid molecule of a known composition, typically about 10 to 120 base pairs in length, which are designed such that they can be ligated, for example through the use of a DNA ligase, to one or both ends of a second nucleic acid molecule(s). Adapters can be designed to be ligated to the blunt end of a nucleic acid (blunt end adapters) or by first annealing to a specific overhang sequence and then ligated. In this embodiment, adapters may be used to provide primer sites, to tag a nucleic acid with a DNA tag, to provide sequences that enable hybridization for the purposes of capture and to add chemically modified nucleic acid sequences such as biotin containing adapters.

[0060] As used herein, the term "ligation" means the biochemical reaction catalyzed by the enzyme ligase wherein two DNA molecules are covalently joined.

[0061] As used herein, "DNA amplification" means the in vitro synthesis of double stranded DNA through the use of a DNA polymerase. Typically, this is accomplished in a polymerase chain reaction (PCR) assay but may also include other methods such as a gap-fill reaction, mis-match repair, Klenow reaction, etc. DNA amplification is used to provide detectable or excess amounts of a specific DNA. It can also be used to incorporate into a target nucleic acid, hybridized probes, annealed adaptors and primers which may include specific functionality or information.

[0062] As used herein, the term "transgene" means nucleic acid molecules in form of DNA, such as cDNA or genomic DNA, and RNA, such as mRNA or microRNA, which may be single or double stranded.

[0063] As used herein, the term "inbred" means a line that has been bred for genetic homogeneity.

[0064] As used herein, the term "hybrid" means a progeny of mating between at least two genetically dissimilar parents. Without limitation, examples of mating schemes include single crosses, modified single cross, double modified single cross, three-way cross, modified three-way cross, and double cross wherein at least one parent in a modified cross is the progeny of a cross between sister lines.

[0065] As used herein, the term "tester" means a line used in a testcross with another line wherein the tester and the

IDT Exhibit 1020.6731

lines tested are from different germplasm pools. A tester may be isogenic or nonisogenic.

[0066] As used herein, the term "corn" means *Zea mays* or maize and includes all plant varieties that can be bred with corn, including wild maize species. More specifically, corn plants from the species *Zea mays and* the subspecies *Zea mays* L. ssp. *Mays* can be genotyped using the compositions and methods of the present invention. In an additional aspect, the corn plant is from the group *Zea mays* L. subsp. *mays* Indentata, otherwise known as dent corn. In another aspect, the corn plant is from the group *Zea mays* L. subsp. *mays* Indurata, otherwise known as flint corn. In another aspect, the corn plant is from the group *Zea mays* L. subsp. *mays* Saccharata, otherwise known as sweet corn. In another aspect, the corn plant is from the group *Zea mays* L. subsp. *mays* Amylacea, otherwise known as flour corn. In a further aspect, the corn plant is from the group *Zea mays* L. subsp. *mays* Everta, otherwise known as pop corn. *Zea* or corn plants that can be genotyped with the compositions and methods described herein include hybrids, inbreds, partial inbreds, or members of defined or undefined populations.

[0067] As used herein, the term "soybean" means *Glycine max* and includes all plant varieties that can be bred with soybean, including wild soybean species. More specifically, soybean plants from the species *Glycine max* and the subspecies *Glycine max* L. ssp. *max* or *Glycine max ssp. formosana* can be genotyped using the compositions and methods of the present invention. In an additional aspect, the soybean plant is from the species *Glycine soja,* otherwise known as wild soybean, can be genotyped using these compositions and methods. Alternatively, soybean germplasm derived from any of *Glycine max*, *Glycine max* L. ssp. *max, Glycine max* ssp. *Formosana,* and/or *Glycine soja* can be genotyped using compositions and methods provided herein.

[0068] As used herein, the term "comprising" means "including but not limited to".

[0069] As used herein, the term "elite line" means any line that has resulted from breeding and selection for superior agronomic performance. An elite plant is any plant from an elite line.

[0070] In accordance with the present invention, Applicants have discovered methods for making breeding decisions genotypically on nucleic acid sequences per se. For example, the methods of the present invention provide for direct, sequence-based analysis instead of using genetic markers as indirect tools for selecting a locus of interest. Further, the methods of the present invention allow for improved flexibility in using nucleic acid information in a breeding program, wherein the entire genome of a plant or animal can be queried without reliance on pre-determined genetic markers and the development of genetic marker detection assays. In addition, any length of sequence from any locus can be leveraged to 1) determine genotype-trait associations, 2) discriminate between two or more lines, 3) predict line performance or hybrid performance and, ultimately, 4) provide the basis for decisions in activities related to germplasm improvement.

[0071] Molecular breeding is often referred to as marker-assisted selection (MAS) and marker-assisted breeding (MAB), wherein MAS refers to making breeding decisions on the basis of molecular marker genotypes for at least one locus and MAB is a general term representing the use of molecular markers in plant breeding. In these types of molecular breeding programs, genetic marker alleles can be used to identify plants that contain the desired genotype at one marker locus, several loci, or a haplotype, and that would therefore be expected to transfer the desired genotype, along with an associated desired phenotype, to their progeny. Markers are highly useful in plant breeding because, once established, they are not subject to environmental or epistatic interactions. Furthermore, certain types of markers are suited for high throughput detection, enabling rapid identification in a cost effective manner.

[0072] Marker discovery and development in crops provides the initial framework for applications to MAB (U.S. Patent No. 5,437,697; U.S. Patent Applications 2005000204780, 2005000216545, 2005000218305, and 2006000504538). The resulting "genetic map" is the representation of the relative position of characterized loci (DNA markers or any other locus for which alleles can be identified) along the chromosomes. The measure of distance on this map is relative to the frequency of crossover events between sister chromatids at meiosis. As a set, polyallelic markers have served as a useful tool for fingerprinting plants to inform the degree of identity of lines or varieties (U.S. Patent No. 6,207,367). These markers form the basis for determining associations with phenotype and can be used to drive genetic gain. The implementation of MAS, wherein selection decisions are based on marker genotypes, is dependent on the ability to detect underlying genetic differences between individuals.

[0073] Because of allelic differences in these molecular markers, QTL can be identified by statistical evaluation of the genotypes and phenotypes of segregating populations. Processes to map QTL are well-described (WO 90/04651; U.S. Patent Nos. 5,492,547, 5,981,832, 6,455,758; reviewed in Flint-Garcia et al. 2003 Ann. Rev. Plant Biol. 54:357-374). Using markers to infer phenotype in these cases results in the economization of a breeding program by substitution of costly, time-intensive phenotyping with genotyping. Marker approaches allow selection to occur before the plant reaches maturity, thus saving time and leading to more efficient use of plots. In fact, selection can even occur at the seed level so only preferred seeds are planted (U.S. Patent Applications 2005000213435 and 2007000680611). Further, breeding programs can be designed to explicitly drive the frequency of specific, favorable phenotypes by targeting particular genotypes (U.S. Patent No. 6,399,855). Fidelity of these associations may be monitored continuously to ensure maintained predictive ability and, thus, informed breeding decisions (U.S. Patent Application 2005/0015827).

[0074] This process has evolved to the application of markers as a tool for the selection of "new and superior plants" via introgression of preferred loci as determined by statistical analyses (U.S. Patent No. 6,219,964). Marker-assisted

IDT Exhibit 1020.6732

introgression involves the transfer of a chromosomal region, defined by one or more markers, from one germplasm to a second germplasm. The initial step in that process is the localization of the genomic region or transgene by gene mapping, which is the process of determining the position of a gene or genomic region relative to other genes and genetic markers through linkage analysis. The basic principle for linkage mapping is that the closer together two genes are on a chromosome, the more likely they are to be inherited together. Briefly, a cross is generally made between two genetically compatible but divergent parents relative to the traits of interest. Genetic markers can then be used to follow the segregation of these traits in the progeny from the cross, often a backcross (BC1), F_2, or recombinant inbred population.

[0075] Historically, genetic markers were not appropriate for distinguishing identity by state or by descent. It has long been recognized that genes and genomic sequences may be identical by state (i.e., identical by independent origins; IBS) or identical by descent (i.e., through historical inheritance from a common progenitor; IBD) which has tremendous bearing on studies of linkage disequilibrium and, ultimately, mapping studies (Nordborg et al. 2002 Trends Gen. 18: 83-90). Notably, newer classes of markers such as SNPs (single nucleotide polymorphisms), are more diagnostic of origin. The likelihood that a particular SNP allele is derived from independent origins in the extant populations of a particular species is very low. Polymorphisms occurring in linked genes are randomly assorted at a slow, but predictable rate, described by the decay of linkage disequilibrium or, alternatively, the approach of linkage equilibrium. Consequences of this well-established scientific discovery are that long stretches of coding DNA, defined by a specific combination of polymorphisms, are very unique and extremely improbable of existing in duplication except through linkage disequilibrium, which is indicative of recent co-ancestry from a common progenitor. The probability that a particular genomic region, as defined by some combination of alleles, indicates absolute identity of the entire intervening genetic sequence is dependent on the number of linked polymorphisms in this genomic region, barring the occurrence of recent mutations in the interval. Such loci are also referred to as haplotype windows. Each haplotype within that window is defined by specific combinations of alleles; the greater the number of alleles, the greater the number of potential haplotypes, and the greater the certainty that identity by state is a result of identity by descent at that region. The present invention permits the direct determination of IBD by using direct nucleic acid sequence information, rather than inferred by marker information.

[0076] During the development of new lines, ancestral haplotypes are maintained through the process and are typically thought of as 'linkage blocks' that are inherited as a unit through a pedigree. Further, if a specific haplotype has a known effect, or phenotype, it is possible to extrapolate its effect in other lines with the same haplotype. Currently, haplotypes are identified and tracked in germplasm using one or more diagnostic markers for that haplotype window. The present invention provides a method to directly identify haplotypes by using nucleic acid sequence information. Further, by using direct sequence information, more polymorphisms within any genomic region may be identified versus using only genetic markers, thus resulting in the identification of additional haplotypes. One can also better assess haplotypes that may share identity by descent. By discriminating haplotypes on a deeper level, greater fidelity in haplotype-phenotype associations can be gained. In another aspect, exotic germplasm can be queried for novel haplotypes by using direct sequence information, thus enabling the identification and subsequent leveraging of unique haplotypes.

[0077] In another approach, regions of IBD can be queried across at least one germplasm pool in order to assess genetic diversity. For example, allelic variants have been queried in order to infer genetic bottlenecks in the domestication of crop plants (reviewed in Doebley et al. 2006 Cell 127:1309-1321). However, using a marker platform to query diversity may be limiting since a single marker queries only a single position in the sequence.

[0078] Further, one theory of heterosis predicts that regions of IBD between the male and female lines used to produce a hybrid will reduce hybrid performance. Identity by descent has historically been inferred from patterns of marker alleles in different lines, wherein an identical string of markers at a series of adjacent loci may be considered identical by descent if it is unlikely to occur independently by chance. Analysis of marker fingerprints in male and female lines can identify regions of IBD. In the present invention, the genome can be directly queried for at least one locus within the genome to evaluate IBD between lines. Knowledge of these regions can inform the choice of hybrid parents, since avoiding IBD in hybrids is likely to improve performance. This knowledge may also inform breeding programs in that crosses could be designed to produce pairs of inbred lines (one male and one female) that show little or no IBD.

[0079] In one aspect of the present invention, heterosis is evaluated for at least one genomic region, wherein heterozygosity between parents in a cross as determined on an allele basis can be presumed to confer a phenotypic advantage. In another aspect of the present invention, methods are provided to evaluate heterosis in terms of genomic synteny, wherein non-colinearity for at least one locus can result in a heterotic advantage and improved performance in the hybrid.

[0080] Markers have traditionally been used to fingerprint lines and thus provide estimates of genetic purity, facilitate QA/QC operations, and assess genetic diversity. The present invention improves upon traditional marker protocols by providing methods to directly assess base pair sequences, instead of estimating underlying sequence identity from a single base position as with traditional marker protocols. For example, a typical biallelic SNP marker provides information on only one base pair position and it can only distinguish between 2, rather than 4, nucleotides.

[0081] The methods of the present invention take advantage of recent breakthroughs in high throughput sequencing to provide novel methods for molecular breeding. High throughput (HT) sequencing methodologies have recently been developed whereby information can be generated for 100MB or more of sequence in a single sequencing machine run.

IDT Exhibit 1020.6733

It is contemplated that any commercially available HT sequencing technology, or any other commercially available nucleic acid sequencing platform that may be developed in the future, can be employed as long as the platform is capable of determining the sequence of a single nucleic acid molecule. Non-limiting examples of commercially available HT sequencing technologies are provided by 454 Life Sciences (Branford, CT), Agencourt Bioscience (Beverly, MA), Applied Biosystems (Foster City, CA), LI-COR Biosciences (Lincoln, NE), NimbleGen Systems (Madison, WI), Illumina (San Diego, CA), and VisiGen Biotechnologies (Houston, TX) (see also, www.solexa.com, www.454.com or www.abi.com). Commercially available HT sequencing technologies are also reviewed in Service (Science 2006 311:1544-1546), which is incorporated herein by reference in its entirety. In essence the Illumina Genome Analyzer, 454 Flex and the ABI Solid technology are able to determine the sequence of a single DNA molecule although that molecule may be amplified in the process. Some of these examples employ sequencing by synthesis although this is not a pre-requisite. Preferred HT sequencing platforms will generate 100 megabases, 1 gigabase or even more sequence information per run. Highly preferred HT sequencing platforms will simultaneously determine the sequence on the maximum number of individual DNA molecules. Such systems are said to be highly parallel. For this reason, the Illumina Genome Analyzer platform is generally preferred because it can sequence many more DNA molecules by generating only a small read per molecule. Platforms which generate longer reads on fewer sequences will work but may present additional challenges for time and cost efficiency.

[0082] Direct determination of the polymorphic nucleotides has key advantages over marker technologies. Although marker technologies are generally robust, they can still incorrectly report an underlying sequence, be subject to noise, and be subject to failure. Further, a marker may not span the actual genomic region of interest and, depending on the degree of linkage to the genomic region of interest, lose value in breeding populations due to recombination and loss of the linkage. Direct determination of the nucleic acid sequences overcomes the inherent limitations of a marker based system by sequencing through not just the nucleotide(s) of interest but the surrounding sequences as well. In addition, the present invention provides methods for "indirect" polymorphism detection wherein allele-specific tags are used that are immediately adjacent to the SNP (Figure 5) so the sequencing reaction only needs to be completed as far as the tag, which is especially useful for technologies generating short reads. Indirect sequencing still overcomes the shortcomings of typical markers' tendency to be linked, vs. comprising, causal polymorphisms since the tag is essentially physically linked to the SNP. Use of nucleic acid sequencing also provides more sequence information about the loci that correlate to traits of importance, which will help breeders better understand and utilize the loci or traits. Furthermore, direct determination of nucleic acid sequences may eliminate the need for extensive up-front sequencing for marker development.

[0083] In one embodiment, the method of the present invention comprises sequencing the whole genome of one plant, comparing the sequenced genome to the genotype of a second plant and then making a decision to cross them, select either one or both to advance, or test the combination of the two. Alternatively, the whole genome information can be used to develop pedigrees by grouping lines that share similarities and separating lines on the basis of genetic differences in order to leverage heterosis. The whole genome sequence provides the complete listing of polymorphic nucleotides and the complete listing of haplotypes.

[0084] The HT sequencing technology as described in the public domain is enabling yet still inherently limited in its application to plant genotyping, even with the ability to sequence 100 megabases or even 1 gigabases of sequence per sample. The limitation arises from the need to sequence 10,000s of thousands of individuals or lines needed to support a modern breeding program. The large number of individuals or lines are needed to identify rare recombinants between two loci or the sub-population with the highest frequency of favorable alleles at multiple loci. The ability to sequence the whole genomes of such a large number of individuals is still impractical. A means to reduce the genome to a smaller number of informative polymorphic regions is needed as well as a means to combine samples from multiple individuals into a smaller number of sequencing runs or reactions. One aspect of this invention is the use of a reproducible method to reduce the complexity of a whole genome to a representative subset of sequences which can be analyzed, compared and used for plant breeding decisions. An additional aspect of this invention is the ability to apply DNA tagging so that multiple samples can be combined in a single sequencing run. The sequences from the combined samples that are determined in parallel in a single run can then be de-convoluted and tracked back to the individual plant or plant pool which they originated.

[0085] In one aspect, the present invention provides subsets of total genomic DNA or RNA for nucleic acid sequencing such that a reduced representation sample is obtained to narrow the target for sequencing, i.e., to coding regions or regions including at least one polymorphism of interest. These subsets may sometimes be referred to as reduced complexity samples or libraries.

[0086] In another aspect of this invention, the reduced representation sample is targeted to or limited to one or more selected regions, or loci, in the genome. The selected loci can be selected based on one or more associations with one or more traits or performance characteristics or they can be a representative subset of the all loci within a genome, such as a subset evenly spaced along the chromosomes and which are segregating in the target breeding population. A preferred subset of the loci are polymorphic loci. A polymorphic locus is defined by one or more nucleotides that vary

IDT Exhibit 1020.6734

between a pair of or multiple individuals or lines. Any type of polymorphic locus may be used with this technology including but not limited to sequence length polymorphisms, repetitive sequence length polymorphisms, restriction site polymorphisms and single nucleotide polymorphisms. Single nucleotide polymorphisms are detected in a preferred embodiment of this invention. The sequence of a targeted locus can be determined by priming the locus to synthesize a complementary oligonucleotide and then directly sequencing the complementary oligonucleotide. The targeted regions can be synthesized through a gap fill reaction, primer extension reaction, a polymerase chain reaction or a combination of these reactions. Alternatively, in the case of polymorphic targeted loci, mis-match repair enzymes or ribozymes or other such nucleotide specific enzymes can be used to specifically repair a complementary oligonucleotide that is mis-matched at the polymorphic nucleotide. Once the complementary nucleotide has been extended, amplified, repaired or gap-filled, the sequence of the in vitro generated oligonucleotide can be determined and represents the sequence of the polymorphic locus. Any of these methods can be employed to directly determine the nucleotide sequence of one or both strands of one or many nucleotide regions. Since the high throughput sequencing methodologies can generate greater than 100 MB of sequence information in a single run, oligonucleotides from large number of loci can be combined and sequenced simultaneously such that the sequences of large numbers of loci can be determined in parallel in one sequencing reaction. In such an embodiment, the invention provides high-throughput and cost effective methods for the direct determination of polymorphic, or non-polymorphic, nucleotides.

[0087] In another aspect, a reduced representation sample can be prepared that consists of a specific class of genome fragments. In one preferred embodiment, a sample is prepared using restriction enzymes. For the purpose of comparing at least two plants of a species, each sample is prepared by digesting with one or more restriction endonucleases, fractionating the digested DNA fragments based on size of nucleotide sequence and comparing the sequence of a fragments in a fraction. More particularly, the method of identifying at least one locus in genomic DNA comprises digesting total genomic DNA from at least two variants of a eukaryotic species with a methylation sensitive endonuclease to provide a pool of digested DNA fragments. The average nucleotide length of fragments is smaller for DNA regions characterized by a lower percent of 5-methylated cytosine. Such fragments are separable, e.g. by gel electrophoresis, based on nucleotide length. A fraction of DNA with less than average nucleotide length is separated from the pool of digested DNA. As compared to coding sequence, repetitive sequence is more likely to comprise 5-methylated cytosine, e.g. in -CG- and -CNG-sequence segments. In a preferred aspect of the method, genomic DNA from at least two different inbred varieties of a crop plant is digested with a with a methylation sensitive endonuclease selected from the group consisting of Aci I, Apa I, Age I, Bsr F I, BssH II, Eag I, Eae I, Hha I, HinP1 I, Hpa II, Msp I, MspM II, Nar I, Not I, Pst I, Pvu I, Sac II, Sma I, Stu I and Xho I to provide a pool of digested DNA which can be physically separated, e.g. by gel electrophoresis. Comparable size fractions of DNA are obtained from digested DNA of each of said varieties and then sequenced.

[0088] In another embodiment, RNA can be used as a reduced representation of the genome, i.e. the subset of the genome which is expressed. The RNA may be polyA RNA, small RNA or other RNA fractions which may be used directly after extraction or experimentally manipulated to further reduce complexity or improve reproducibility. Prior to sequencing, the RNA is converted by reverse transcription methods to cDNA which can be directly sequenced or experimentally manipulated to further reduce complexity or improve reproducibility.

[0089] In a preferred embodiment of this invention, multiple nucleic acid samples can be combined into a sample multiplex, i.e. pool, and sequenced in parallel in the same run to maximize sample throughput per sequencing run. To achieve this, a DNA tag, comprising one or more nucleotides unique for that sample, is added to the nucleic acid prepared from an individual sample. Typical DNA tags comprise 1 to 10 nucleotides but can extend to any length as long as the tag does not interfere with the ability to determine the sample sequence. For example, a DNA tag of 2 nucleotides can be use to separate a mixture of 16 samples. DNA tags of 3, 4, 5 or 6 nucleotides can be used to separate mixtures of 64, 256, 1024 or 4096 samples, and so on. Shorter DNA tags place less constraints on sequence read length but limit the number of samples which can be mixed. In one embodiment of the invention, the DNA tags are simply synthesized as part of one or both PCR primers and then incorporated in a PCR reaction. In another aspect, the DNA tag can be ligated onto the sample nucleic acids using a DNA ligase. After fully incorporating a DNA tag into the nucleic acid sample, multiple DNA preparations, each with a unique tag, can be multiplexed, i.e. pooled or combined. The multiplexed mixtures are then subjected to a single HT sequencing reaction. The number of samples that are multiplexed is based on optimally using the full sequencing capacity of a single sequencing run. Parameters that influence the complexity of a sample mixture include the number of loci being assessed, the size of the loci, the information content per run of the HT platform, the length of the DNA tag, the presence, if any, of an adapter or primer sequence and the read length of a given sequence. The level of multiplexing can be balanced to achieve optimum cost per sample, redundancy per sequence read. The minimum length of a single sequence read needs to be sufficient to read a sample DNA tag (for example, 2-5 nucleotides, depending on the number of samples which are pooled), a sequence specific tag (6-20 nucleotides) and one or more adjacent nucleotides. After the HT sequencing reaction, sequences with the same DNA tag are first separated logically into separate pools which represent the individual or line or pool which the DNA was extracted. The sequences with identical DNA tags can then be read to determine the nucleotide identity within the loci which were selected to be queried.

IDT Exhibit 1020.6735

[0090] In this invention, the sequence of nucleic acids can be associated to traits of interest or to plant performance and then used to make selections of parents, progeny or testers. Sequences will be useful if they are genetically linked to the trait or performance characteristic. Typically, they are genetically linked if they are causative for the trait or performance characteristic or are closely physically linked to the trait or performance loci. In the case of physically linked sequences, no knowledge of the gene(s) and/or causative variation for the trait or performance information is required. One only needs to determine the sequence of the physically linked nucleotides. Once a sequence has been genetically linked to a trait or performance character, the sequence of the nucleic acids can be directly used to select parents, progeny or testers which will exemplify that trait or performance without the need to first measure the trait or performance characteristic. The knowledge of the nucleotide sequences can also be used to fingerprint a plant or line and be used to measure genetic similarity/distance among plants or lines and to build pedigrees. The pedigrees can then be used to make selections of parents or to manage the diversity in a germplasm pool.

[0091] In another embodiment, plants can be screened for one or more markers, such as nucleic acid sequences, using high throughput, non-destructive seed sampling. In a preferred aspect, seed is sampled in this manner and only seed with at least one genotype of interest is advanced. Apparatus and methods for the high-throughput, non-destructive sampling of seeds have been described which would overcome the obstacles of statistical samples by allowing for individual seed analysis. For example, published U.S. Patent Applications US 2006/0042527, US 2006/0046244, US 2006/0046264, US 2006/0048247, US 2006/0048248, US 2007/0204366, and US 2007/0207485, which are incorporated herein by reference in their entirety, disclose apparatus and systems for the automated sampling of seeds as well as methods of sampling, testing and bulking seeds.

[0092] As provided by the present invention, the knowledge of nucleic acid sequences can be applied to make decisions at multiple stages of the breeding program:

a) Among segregating progeny, as a pre-selection method, to increase the selection index and drive the frequency of favorable nucleic acid sequences among breeding populations, wherein pre-selection is defined as selection among offspring of a breeding cross based on the genotype of these progenies at a selected set of two or more nucleic acid sequences at one or more loci as determined by HT sequencing, and leveraging of nucleic acid sequence-trait associations identified in previous breeding crosses.

b) Among segregating progeny from a breeding population, to increase the frequency of the favorable nucleic acid sequences for the purpose of line or variety development.

c) Among segregating progeny from a breeding population, to increase the frequency of the favorable nucleic acid sequences prior to QTL mapping within this breeding population.

d) For hybrid crops, among parental lines from different heterotic groups to predict the performance potential of different hybrids.

[0093] In another embodiment, the present invention provides a method for improving plant germplasm by accumulation of nucleic acid sequences of interest in a germplasm comprising determining nucleic acid sequences for at least two loci in the genome of a species of plant, and associating the nucleic acid sequences with at least one trait, and using this nucleic acid sequence effect estimates to direct breeding decisions. These nucleic acid sequence effect estimates can be derived using historical nucleic acid sequence -trait associations or *de novo* from mapping populations. The nucleic acid sequence effect estimates for one or more traits provide the basis for making decisions in a breeding program. This invention also provides an alternative basis for decision-making using breeding value calculations based on the estimated effect and frequency of nucleic acid sequences in the germplasm. Nucleic acid sequence breeding values can be used to rank a specified set of nucleic acid sequences. In the context of the specified set of nucleic acid sequences, these breeding values form the basis for calculating an index to rank the alleles both within and between loci.

[0094] For example, any given chromosome segment can be represented in a given population by a number of nucleic acid sequences that can vary from 1 (region is fixed), to the size of the population times the ploidy level of that species (2 in a diploid species), in a population in which every chromosome has a different nucleic acid sequence. Identity-by-descent among nucleic acid sequences carried by multiple individuals in a non-fixed population will result in an intermediate number of different nucleic acid sequences and possibly a differing frequency among the different nucleic acid sequences. New nucleic acid sequences may arise, through recombination at meiosis between existing nucleic acid sequences in heterozygous progenitors. The frequency of each nucleic acid sequence may be estimated by several means known to one versed in the art (e.g. by direct counting, or by using an EM algorithm). Let us assume that "k" different nucleic acid sequences, wherein a nucleic acid sequence represents at least one nucleotide and may constitute an allele or haplotype, identified as "n_i" (i=1,...,k), are known, that their frequency in the population is "f_i" (i=1,...,k), and for each of these nucleic acid sequences we have an effect estimate "Est_i" (i=1,...,k). If we call the "breeding value" (BV_i) the effect on that population of fixing that nucleic acid sequence, then this breeding value corresponds to the change in mean for the trait(s) of interest of that population between its original state of haplotypic distribution at the window and a final state at which nucleic acid sequence "n_i" encounters itself at a frequency of 100%.

IDT Exhibit 1020.6736

The breeding value of n_i in this population can be calculated as:

$$BV_i = Est_i - \sum_{i=1}^{k} Est_i f_i$$

[0095] One skilled in the art will recognize that nucleic acid sequences that are rare in the population in which effects are estimated tend to be less precisely estimated, this difference of confidence may lead to adjustment in the calculation. For example one can ignore the effects of rare nucleic acid sequences, by calculating breeding value of better known nucleic acid sequence after adjusting the frequency of these (by dividing it by the sum of frequency of the better known nucleic acid sequences). One could also provide confidence intervals for the breeding value of each nucleic acid sequences.

[0096] This breeding value will change according to the population for which it is calculated, as a function of difference of nucleic acid sequence frequencies. The term population can then assume different meanings, below are two examples of special cases. First, it can be a single inbred line in which one intend to replace its current nucleic acid sequence n_j by a new nucleic acid sequence n_i, in this case BV_i=Est_i-Est_j. Second, it can be a F2 population in which the two parental nucleic acid sequence n_i and n_j are originally present in equal frequency (50%), in which case BV_i= V_2 (Est_i-Est_j).

[0097] These statistical approaches enable nucleic acid sequence effect estimates to inform breeding decisions in multiple contexts. Other statistical approaches to calculate breeding values are known to those skilled in the art and can be used in substitution without departing from the spirit and scope of this invention.

[0098] Further, methods for determining the statistical significance of a correlation between a phenotype and a genotype, in this case a nucleic acid sequence, may be determined by any statistical test known in the art and with any accepted threshold of statistical significance being required. The application of particular methods and thresholds of significance are well with in the skill of the ordinary practitioner of the art.

[0099] Nucleic acid sequence effect estimates and/or breeding values for one or more traits of interest provide the basis for determining one or more nucleic acid sequences of interest in comparisons of two or more nucleic acid sequences. With this *a priori* information, breeding selections are conducted on a nucleic acid sequence, rather than marker, basis, wherein a first plant is crossed with a second plant that contains at least one locus where the nucleic acid sequence of the second plant is different from the first plant nucleic acid sequence; and at least one progeny plant is selected by detecting the nucleic acid sequence or set of nucleic acid sequences of the first plant, wherein the progeny plant comprises in its genome one or more nucleic acid sequences of interest of the first plant and at least one nucleic acid sequence of interest of the second plant; and the progeny plant is used in activities related to germplasm improvement, herein defined as including using the plant for line and variety development, hybrid development, transgenic event selection, making breeding crosses, testing and advancing a plant through self fertilization, purification of lines or sublines, using plant or parts thereof for transformation, using plants or parts thereof for candidates for expression constructs, and using plant or parts thereof for mutagenesis.

[0100] In one aspect, this invention provides high throughput sequencing to identify large segments of nucleic acids, in one or more regions of a plant genome, that provide a basis to compare two or more germplasm entries. These regions of contiguous nucleic acid sequence are indicative of the conservation of genetic identity of all intervening genes from a common progenitor. In cases where conserved sequence segments are coincident with segments in which QTL have been identified it is possible to deduce with high probability that QTL inferences can be extrapolated to other germplasm having an identical sequence in that locus. This *a priori* information provides the basis to select for favorable QTLs prior to QTL mapping within a given population.

For example, plant breeding decisions could comprise:

a) Selection among new breeding populations to determine which populations have the highest frequency of favorable nucleic acid sequences, wherein sequences are designated as favorable based on coincidence with previous QTL mapping; or

b) Selection of progeny containing said favorable nucleic acid sequences in breeding populations prior to, or in substitution for, QTL mapping within that population, wherein selection could be done at any stage of breeding and could also be used to drive multiple generations of recurrent selection; or

c) Prediction of progeny performance for specific breeding crosses; or

d) Selection of lines for germplasm improvement activities based on said favorable haplotypes, including line development, hybrid development, selection among transgenic events based on the breeding value of the haplotype that the transgene was inserted into, making breeding crosses, testing and advancing a plant through self fertilization, using plant or parts thereof for transformation, using plants or parts thereof for candidates for expression constructs, and using plant or parts thereof for mutagenesis.

IDT Exhibit 1020.6737

[0101] An additional unique aspect of this invention is the ability to select for specific genes or gene alleles, as they are targeted by high throughput sequencing. For example, in cases where the nucleic acid sequence is coincident with segments in which genes have been identified it is possible to deduce with high probability that gene inferences can be extrapolated to other germplasm having an identical genotype in that locus. This *a priori* information provides the basis to select for favorable genes or gene alleles on the basis of nucleic acid sequencing within a given population. For example, plant breeding decisions could comprise:

a) Selection among new breeding populations to determine which populations have the highest frequency of favorable nucleic acid sequences, wherein sequences are designated as favorable based on coincidence with previous gene mapping; or

b) Selection of progeny containing said favorable nucleic acid sequences in breeding populations, wherein selection is effectively enabled at the gene level, wherein selection could be done at any stage of inbreeding and could also be used to drive multiple generations of recurrent selection; or

c) Prediction of progeny performance for specific breeding crosses; or

d) Selection of lines for germplasm improvement activities based on said favorable haplotypes, including line development, hybrid development, selection among transgenic events based on the breeding value of the haplotype that the transgene was inserted into, making breeding crosses, testing and advancing a plant through self fertilization, using plant or parts thereof for transformation, using plants or parts thereof for candidates for expression constructs, and using plant or parts thereof for mutagenesis.

[0102] Further, in another preferred embodiment of this invention, the *a priori* information on the frequency of favorable nucleic acid sequences in breeding populations enables pre-selection. That is, parental lines are selected based on the historical genotype-phenotype association information for the purpose of driving favorable nucleic acid frequency for multiple traits simultaneously. In pre-selection, breeders can predict the phenotypic contribution for multiple traits of any line based on that line's fingerprint information, which corresponds to a composition of pre-defined sequences. This multi-trait sequence selection approach economizes a breeding program by initiating selection at the initial stage of choosing parental crosses and it also reduces the need for costly, time-consuming phenotyping of progeny.

[0103] A preferred sequence provides a preferred property to a parent plant and to the progeny of the parent when selected by a marker means or phenotypic means. The method of the present invention provides for selection of preferred sequences, or sequences of interest, and the accumulation of these sequences in a breeding population.

[0104] In another embodiment, this invention enables indirect selection through selection decisions for at least one nucleic acid sequence based on at least one nucleic acid sequence effect estimate such that additional phenotypes are indirectly selected upon due to the additional nucleic acid sequence effect estimates for other phenotypic traits.

[0105] Another preferred embodiment of the present invention is to build additional value by selecting a composition of nucleic acid sequences wherein each sequence has an estimated associated phenotype that is not negative with respect to yield, or is not positive with respect to maturity, or is null with respect to maturity, or amongst the best 50 percent with respect an agronomic trait, transgene, and/or a multiple trait index when compared to any other nucleic acid sequence at the same locus in a set of germplasm, or amongst the best 50 percent with respect to an agronomic trait, transgene, and/or a multiple trait index when compared to any other loci across the entire genome in a set of germplasm, or the nucleic acid sequence being present with a frequency of 75 percent or more in a breeding population or a set of germplasm can be taken as evidence of its high value, or any combination of these.

[0106] This invention anticipates a stacking of nucleic acid sequences from at least two loci into plants or lines by crossing parent plants or lines containing different nucleic acid sequences, that is, different genotypes. The value of the plant or line comprising in its genome stacked nucleic acid sequences from two or more loci can be estimated by a composite breeding value, which depends on a combination of the value of the traits and the value of the nucleic acid sequence(s) to which the traits are linked. The present invention further anticipates that the composite breeding value of a plant or line can be improved by modifying the components of one or each of the nucleic acid sequences. Additionally, the present invention anticipates that additional value can be built into the composite breeding value of a plant or line by selection of at least one recipient nucleic acid sequence with a preferred nucleic acid sequence effect estimate or, in conjunction with the frequency of said nucleic acid sequence in the germplasm pool, breeding value to which one or any of the other nucleic acid sequences are linked, or by selection of plants or lines for stacking two or more nucleic acid sequences from two or more loci by breeding.

[0107] Another embodiment of this invention is a method for enhancing breeding populations by accumulation of one or more nucleic acid sequences in one or more loci, in a germplasm. Loci include genetic information and provide phenotypic traits to the plant. Variations in the genetic information can result in variation of the phenotypic trait and the value of the phenotype can be measured. The genetic mapping of the nucleic acid sequences allows for a determination of linkage across sequences. The nucleic acid sequence of interest is novel in the genome of the progeny plant and can in itself serve as a genetic marker of a locus of interest. Notably, this nucleic acid sequence can also be used as an

IDT Exhibit 1020.6738

identifier for a gene or QTL. For example, in the event of multiple traits or trait effects associated with the nucleic acid sequence, only one marker would be necessary for selection purposes. Additionally, the locus of interest may provide a means to select for plants that have the linked locus.

[0108] In another embodiment, at least one preferred nucleic acid of the present invention is stacked with at least one transgene. In another aspect, at least one transgenic event is advanced based on linkage with or insertion in a preferred nucleic acid, as disclosed in published U.S. Patent Application US 2006/0282911, which is incorporated herein by reference in its entirety.

[0109] In still another embodiment, the present invention acknowledges that preferred nucleic acids identified by the methods presented herein may be advanced as candidate genes for inclusion in expression constructs, i.e., transgenes. Nucleic acids of interest may be expressed in plant cells by operably linking them to a promoter functional in plants. In another aspect, nucleic acids of interest may have their expression modified by double-stranded RNA-mediated gene suppression, also known as RNA interference s("RNAi"), which includes suppression mediated by small interfering RNAs ("siRNA"), trans-acting small interfering RNAs ("ta-siRNA"), or microRNAs ("miRNA"). Examples of RNAi methodology suitable for use in plants are described in detail in U. S. patent application publications 2006/0200878 and 2007/0011775.

[0110] Methods are known in the art for assembling and introducing constructs into a cell in such a manner that the nucleic acid molecule for a trait is transcribed into a functional mRNA molecule that is translated and expressed as a protein product. For the practice of the present invention, conventional compositions and methods for preparing and using constructs and host cells are well known to one skilled in the art, see for example, Molecular Cloning: A Laboratory Manual, 3rd edition Volumes 1, 2, and 3 (2000) J.F. Sambrook, D.W. Russell, and N. Irwin, Cold Spring Harbor Laboratory Press. Methods for making transformation constructs particularly suited to plant transformation include, without limitation, those described in U.S. Patent Nos. 4,971,908, 4,940,835, 4,769,061 and 4,757,011, all of which are herein incorporated by reference in their entirety. Transformation methods for the introduction of expression units into plants are known in the art and include electroporation as illustrated in U.S. Patent No. 5,384,253; microprojectile bombardment as illustrated in U.S. Patent Nos. 5,015,580; 5,550,318; 5,538,880; 6,160,208; 6,399,861; and 6,403,865; protoplast transformation as illustrated in U.S. Patent No. 5,508,184; and *Agrobacterium*-mediated transformation as illustrated in U.S. Patent Nos. 5,635,055; 5,824,877; 5,591,616; 5,981,840; and 6,384,301.

[0111] The present invention also provides for the screening of progeny plants' loci of interest and using the nucleic acid effect estimate as the basis for selection for use in a breeding program to enhance the accumulation of preferred nucleic acid sequences.

[0112] Using this method, the present invention contemplates that nucleic acid sequences of interest are selected from a large population of plants. Additionally, these nucleic acid sequences can be used in the described breeding methods to accumulate other beneficial and preferred loci and maintain these in a breeding population to enhance the overall germplasm of the plant. Plants considered for use in the method include but are not limited to, corn, soybean, cotton, wheat, rice, canola, oilseed rape, sugar beet, sorghum, millet, alfalfa, forage crops, oilseed crops, grain crops, fruit crops, ornamental plants, vegetable crops, fiber crops, spice crops, nut crops, turf crops, sugar crops, beverage crops, tuber crops, root crops, and forest crops.

[0113] In summary, this invention describes the novel combination of high throughput sequencing and molecular breeding methodologies to enable the use of direct nucleic acid sequence information to carry out molecular plant breeding. The invention also includes means to selectively target polymorphic nucleotide sites and to DNA tag samples prior to sequence determination. Taken together, this invention enables the plant breeder to use sequence information in parent selection, progeny selection, choosing tester combinations, developing pedigrees, fingerprinting samples, screening for haplotype diversity, and for building databases of sequence associations to trait and performance data.

Examples

Example 1. Sequence Directed Selection

[0114] An important aim of any breeding program is to incorporate economically or otherwise important traits into a breeding line or population. The ability to directly determine the sequence of region linked to the trait or to directly determine the sequence(s) of the loci which are causative to the trait will allow the breeder to determine which individuals or lines in a population likely exhibit the trait of interest and thus inform advancement decisions. A sample workflow for high throughput sequencing is depicted in Figure 1. The present example demonstrates a method of the invention for making sequence-directed selection. The method is differentiated from traditional marker-assisted selection in that it uses direct nucleic acid sequence information for selection instead of a marker.

[0115] Low linolenic acid soybean oil is of commercial interest because it does not result in trans fats during processing and use and therefore is healthier for human consumption. A gene that is essential for linolenic acid biosynthesis is the fad3 gene. In soybeans, there are at least three fad3 genes and mutations in two of the genes, fad3b and fad3c, can result in low linolenic acid. Exemplary primers and probes for the detection of mutations in these genes are set forth in

IDT Exhibit 1020.6739

published U.S. Patent Application 20060107348, which is incorporated herein by reference in its entirety.

[0116] In one aspect, a first step for sequence-directed selection can be genome complexity reduction, wherein different strategies are exemplified in Figures 2-5. That is, a reduced representation library can be obtained by selective digestion and purification, using enzymes known in the art (Figure 2). In other aspects, the library can be targeted from the transcriptome (Figure 3). In still other aspects, SNP-containing regions are isolated using allele-specific extension/ligation (Figure 5).

[0117] In still other aspects, the sequence-targeted genomic regions are selectively amplified (Figure 4). In the present example, the Fad3c indel region was amplified using specific primers for insertion and deletion. This method is useful when the region of interest comprises an indel and is especially useful in screening for transgenes. Alternatively, the region spanning the nucleic acid of interest is amplified. In the present example, a second complexity reduction strategy was employed, wherein the SNP assay for the Fad3b region was used to amplify the region containing the SNP for the purpose of sequencing. In general, this approach is especially useful for leveraging existing SNP PCR-based assay libraries and using the known primer sets as a tool in complexity reduction. The present invention anticipates using SNPs provided by published U.S. Patent Applications US 2005/0204780, US 2005/0216545, US 2005/0218305, and US 2006/0504538, as both targets for sequencing as well as for use in genome complexity reduction as described herein.

[0118] A second step that can be useful for sequence-directed selection is use of DNA tags in order to enable sample multiplexing. In the present example, each sample in a multiplex set was assigned a unique DNA tag, i.e., a sequence tag differing by at least one base pair from the other barcodes in the set. In a preferred aspect, the percentage of G and C bases is balanced to minimize bias in the sequencing process. The DNA tag can range in length from about 2 to about 20 bp. In the present examples, with 384 PCR samples, representing 192 germplasm entries assayed for both the Fad3b SNP and Fad3c indel, 6 bp sequences were used and each sample was sequenced for both the SNP and the indel regions.

[0119] In one aspect, the DNA tags are added to the PCR primers as shown in Figure 7. Alternatively, they can be incorporated into allele-specific extension/ligation as shown in Figure 5, with the barcode ligated to the allele-specific extension/ligation products or added to the products using PCR. In the present example, the DNA tags were included in the PCR primers. Figure 9 illustrates a schematic of the resulting template that will be used for sequencing, showing both the Fad3b SNP and the Fad3c indel. Specifically, a pair of oligonucleotides were synthesized to aid in the sequence determination for the fad3b locus. A forward oligonucleotide primer is synthesized to include a 6-nucleotide DNA tag (Table 1) and a sequence that matches the nucleotide sequence that is 5' to the fad3b mutation which is known to affect gene function. For the purposes of this invention, a mutation is the same as a polymorphic nucleotide and represents a polymorphic locus. A reverse oligonucleotide primer is synthesized to a sequence complementary to the region 3' of the fad3a mutation. A second pair of forward and reverse PCR primers are generated in the similar fashion to match a mutation which deletes the fad3c gene which is also known to reduce linolenic acid in soybean oil. Since the deletion extends beyond the boundaries of the fad3 gene, one pair of primers is designed within the genes coding region to determine the sequence of the fad3c gene if the gene is present and a second set of primers is designed to span.the deletion of the fad3c locus, if the gene is absent. The distance between the nucleotide pairs is designed to be between 10 and 200 nucleotides and the mutation adjacent to the end of the forward primer, therefore in close proximity to the DNA tag. The more similar the distance between the primers, the more likely the PCR amplification of the template will be unbiased across the multiple loci, however, longer distances can be required in some examples to find stretches of nucleotides appropriate for robust primer design, e.g. devoid of repetitive sequences, non-existent sequence structure and balanced GC content. The same DNA tag can be used for the forward primer in the three pairs of primers. The three pairs represented a genotyping or fingerprinting set which can be used for one sample. Specifically, the following primer pairs were utilized in the present example: Fad3B (SNP NS0193115), 192 forward primers ACACTCTTTCCCTACACGACGCTCTTCCGATCT plus 192 DNA tags plus CATTGGCACCCATGTTATCC; Single Fad3B reverse primer CAAGCAGAAGACGGCATACGAGCTCTTCCGATCT plus GACTTAGATCACATAGGCAGACATAC; Fad3C insertion, 192 forward primers ACACTCTTTCCCTACACGACGCTCTTCCGATCT plus 192 DNA tags plus TAAGTGACACTGGAGATGTGG; Fad3C deletion, 192 forward primers ACACTCTTTCCCTACACGACGCTCTTCCGATCT plus 192 DNA tags plus CAGAAAGTATTGGTAAAGTACTGGTA; Single Fad3C reverse primer CAAGCAGAAGACGGCATACGAGCTCTTCCGATCT plus TAAATATTCCATTGAGGCCCACTA, wherein equal molar amount of primers were mixed.

IDT Exhibit 1020.6740

Table 1. Exemplary 6 nucleotide DNA tags for the 192 soybean varieties genotyped in the present example.

W	Code	W	Code	W	Code	W	Code	W	Code	W	Code	W	Code	W	Code
1	AACGCT	33	AGGAAG	65	CCTCTA	97	GAATGC	129	GTACCA	161	TCGTGA				
2	AACGTC	34	AGGAGT	66	CCTGAT	98	GACACA	130	GTAGAC	162	TCGTTG				
3	AACTCG	35	AGGTCA	67	CGAAAG	99	GACATC	131	GTAGGA	163	TCTGGA				
4	AAGACC	36	AGTCGT	68	CGAAGA	100	GACGTA	132	GTCAAG	164	TCTTCC				
5	AAGGGT	37	ATCACG	69	CGAATC	101	GACTAG	133	GTCCAA	165	TGAAGG				
6	AAGTGC	38	ATCCAC	70	CGACAT	102	GACTCT	134	GTCGAT	166	TGAGCA				
7	ACACGA	39	ATCCCA	71	CGCATT	103	GACTGA	135	GTCTAC	167	TGAGTC				
8	ACACTG	40	ATCGGT	72	CGGAAT	104	GAGACT	136	GTCTCA	168	TGCAAG				
9	ACAGGT	41	ATCTGC	73	CGTATG	105	GAGGTT	137	GTCTGT	169	TGCATC				
10	ACATCG	42	ATGAGG	74	CGTCTT	106	GATAGC	138	GTGAGA	170	TGCGAA				
11	ACCAAC	43	ATGCCT	75	CGTTCT	107	GATCGT	139	GTGCTA	171	TGCGTT				
12	ACCACT	44	ATGCTG	76	CGTTTC	108	GATGAG	140	GTGTTG	172	TGCTAC				
13	ACCATG	45	ATGGAG	77	CTAAGG	109	GATGGA	141	GTTACC	173	TGCTGA				
14	ACCTA	46	ATGGCA	78	CTACTG	110	GATTGG	142	GTTCAG	174	TGGAAC				
15	ACCGAA	47	ATGTCG	79	CTAGCA	111	GCAGAT	143	GTTCCT	175	TGGATG				
16	ACCGTT	48	ATTGCG	80	CTAGGT	112	GCATGT	144	GTTTCG	176	TGGCAT				
17	ACCTTC	49	CAATGG	81	CTAGTC	113	GCCATA	145	TACAGC	177	TGGTCT				
18	ACGAGA	50	CAGAGT	82	CTATCG	114	GCCTAA	146	TACCTC	178	TGTACC				
19	ACGCTT	51	CAGATG	83	CTCAGT	115	GCTAAG	147	TACGAG	179	TGTCAG				
20	ACGGAT	52	CAGCTT	84	CTCATG	116	GCTAGA	148	TAGCTG	180	TGTCGA				
21	ACGTGT	53	CAGTAG	85	CTCGTA	117	GCTCTT	149	TAGTCG	181	TGTCTC				
22	ACTACC	54	CAGTTC	86	CTCTCT	118	GCTGTA	150	TCAAGC	182	TGTGAC				
23	ACTCCA	55	CATAGG	87	CTCTGA	119	GCTTCT	151	TCACCA	183	TGTGCT				
24	ACTCTC	56	CATTCG	88	CTGACA	120	GCTTTG	152	TCACTC	184	TGTGTG				
25	ACTGAC	57	CCAAGT	89	CTGGAA	121	GGAACA	153	TCAGTG	185	TGTTCG				

IDT Exhibit 1020.6741

(continued)

W	Code	W	Code	W	Code	W	Code	W	Code	W	Code
26	ACTGCT	58	CCACTT	90	CTTACG	122	GGAATG	154	TCATGG	186	TGTTGC
27	AGAACG	59	CCATAG	91	CTTCCA	123	GGAGTA	155	TCCAGT	187	TTCAGG
28	AGACAC	60	CCATGA	92	CTTGAC	124	GGCTAT	156	TCCCTT	188	TTCGCT
29	AGAGAG	61	CCATTC	93	CTTGCT	125	GGCTTA	157	TCGACA	189	TTCGTG
30	AGCGTA	62	CCGATA	94	CTTTCC	126	GGTCAA	158	TCGATC	190	TTCTCC
31	AGCTAG	63	CCGTAA	95	GAAGCT	127	GGTTAG	159	TCGCAA	191	TTGACG
32	AGCTGT	64	CCTATC	96	GAAGTC	128	GGTTCA	160	TCGTAC	192	TTGCAC

IDT Exhibit 1020.6742

[0120] An additional 192 genotyping sets were then generated where each set is identical except that the DNA tag in the forward primer of the three pairs of oligonucleotides is exchanged for a unique tag from the list of 4096 possible tags. The sequences of the fad3b and fad3c mutations were then determined for a population of 192 soybean varieties in the following manner. A single seed from each of the 192 lines was sampled to remove a portion of seed tissue while maintaining seed viability as described, for example, in US 2006/0046264 and US 2007/0204366, each of which is incorporated herein by reference.

[0121] To prepare template for sequencing, DNA was prepared for each of the tissue samples and then 10 ng was dispensed into 2, 96 well microtitre plates. To each well, a PCR master mix was added along with Taq polymerase, according to manufacturer's recommendations (Roche, ABI). Finally, 100 μM of a selected genotyping primer set, including matching DNA tags, was added to each well. The plate was heated to 95°C for nine minutes to denature the DNA. Twenty cycles of PCR were then completed using the following conditions: 94°C for 30 sec, 55°C for 30 sec, 72°C for 2 min, followed by a final 10 minute extension at 72 °C.

[0122] After PCR, all 192 lines were combined into a single well which was then used for a HT sequencing reaction, according to the manufacturer's directions (Illumina Genome Analyzer). Briefly, equal amounts of the 384 PCR products were mixed and subsequently purified using PCR purification methods known in the art. Approximately 5-10 ng of the purified template was amplified with enrich PCR per Illumina Genome Analyzer specifications. The enrich PCR also adds the adapter required for the downstream bridge PCR reaction if the adapters were not already incorporated in the primers. The enrich PCR product is purified, again using PCR purification methods known in the art, and the resulting template is sequencing per Illumina Genome Analyzer specifications.

[0123] The sequences obtained from the sequencing reaction were binned according to the DNA tag sequence. Within each bin, the sequences were analyzed by alignment to the SNP and indel forward primers in order to determine if the known mutation, any other variation or wild-type nucleotides were present next to the 3' complementary oligonucleotide. SNP genotypes were called based on the sequences at the SNP position (see Figure 10 for resulting scatter plot). Indel genotypes were determined by the matches to the sequences of the two forward primers (see Figure 11 for resulting scatter plot). Matching sequence counts can be plotted to handle backgrounds. Advanced scoring tools can be used/developed for normalization/ calibration and more reliable genotype calls. If both mutant and wild type sequences were identified, the sample was predicted to be heterozygous. If only nucleotide sequences that corresponded to wild type sequences were present, then the sample was classified as normal linolenic acid. If known mutant sequences were identified, then the samples were classified as low linolenic acid. Identifying and classifying the sequences at the fad3b locus, the fad3c locus and at the fad3c deletion locus allows a breeder to screen plants to characterize the low linolenic acid-associated genotype and then decide which low linolenic acid varieties to advance to yield testing.

Example 2. Sequence-directed introgression

[0124] A powerful tool in plant breeding is back-crossing. Back-crossing allows a breeder to extract one or more of the best characteristics in a donor line and systematically introgress them into a recurrent parent line. In essence, the genomic region(s) at one or more selected donor DNA loci are systematically introgressed into a recurrent parent genome, replacing the nucleic acids at the corresponding loci in the recurrent parent genome. The types of characteristics that are typically introgressed between lines include, but are not limited too, transgenes, disease resistance, pest resistance, quality traits, agronomic traits, etc. Traditionally, this process can take five or more generations to obtain the traits of interest in a progeny that also shows equivalency to the recurrent parent and has the recurrent parent's agronomic performance. If the performance of the converted line does not equal the predicted performance of the recurrent parent plus the new trait, it can often be very difficult to understand the issue and how to correct.

[0125] Sequence directed back-crossing (SDBC) can greatly accelerate the process and result in a more quantifiable outcome. Using sequences, the progeny from each back-cross generation are examined for both the nucleic acid sequences of the donor parent that encode or are linked to the characteristics of interest and nucleic sequences in the recurrent parent genome. The examination takes into account both differences (polymorphisms) and identity between the sequences. Back-cross progeny are selected and advanced based on their nucleic acid sequence composition, which includes both the nucleic acid sequences encoding or linked to the target trait and the highest percent of nucleic acid sequences matching the recurrent parent sequence. By directing the process with sequence rather than marker information, the process can be completed in fewer generations, with a higher recovery of the recurrent parent.

[0126] A particular example of SDBC is the directed introgression of a transgene from a donor line to a recurrent parent line and an example of a transgene encodes resistance to the herbicide, also known as the bacterial CP4 gene, which is a critical part of the sequence required for the Roundup Ready® trait. In this example, a donor line is fixed or homozygous for the CP4 gene and it is desirable to introgress the CP4 into a recurrent parent line. The breeder planted 15 seeds of recurrent parent in a row next to a row of 15 seeds of the CP4 donor parent. Four crosses are made by pollinating the donor ears with pollen from the recurrent parent. The resulting seed is the F1 seed. A triplet is planted with a recurrent parent row planted between two rows of F1 seed obtained from the one or two best looking F1 ears. At

IDT Exhibit 1020.6743

the time of pollen shed, the recurrent parent is used to pollinate 4 F1's in each of the flanking rows (8 total crosses). The best two BC1 ears are harvested from each row and the BC1 seed is bulked. On average, it is expected that the BC1 seed would contain 25% of the donor genome and 75% of the recurrent parent genome, however, the exact content of any one individual plant would vary within a normal distribution. Subsequent back crossing efforts would be enhanced by selecting the subset of seeds with the highest recurrent parent genome and which contain the transgene. The BC1 seed would also be segregating for the CP4 transgene. Sequencing is used to identify which of 93 BC1 plants had the highest amount of recurrent parent nucleic acid sequences and contained the transgene. The desirable subset can be identified by inspecting the sequence at a number of loci, for example 96, where one of the loci is the CP4 locus.

[0127] Seed from each of the parents, the F1 bulk and from each of 93 BC1 is planted in rows and plants cultivated. At the V4 stage (4th leaf stage), a leaf tear is taken from each plant and placed in a single well of a 96 well block. The DNA is prepped according to the method described in Dellaporta et al., 1983 Plant Mol Biol Rep 1: 19-21, which is incorporated by reference herein in its entirety. The DNA from each of the 96 loci are further prepared using an initial amplification. In this example, amplification is used to incorporate the DNA tag and adapters but other methods are known and applicable. A locus specific forward primer is designed which contained 18 nucleotides at the 3' end which would hybridize to the 5' of the target locus. The 5' end of the forward primer also contained 15 nucleotides which matched the 15 nucleotides 3' to a universal forward PCR primer. In the similar way, a reverse PCR primer is designed where 18 bases at the 3' end complemented the nucleotides 3' of the target locus. The reverse primer also contained 15 base pairs on the 5' end which matched the 3' end of a universal reverse primer. In this example, the target loci are 6-10 nucleotides, however they could range from just 2 nucleotides to several hundred or more. This process is repeated for each of the 96 loci where one of the loci is the CP4 locus. Ninety-five of the loci are selected to cover each arm of every chromosome and included a few extra markers flanking the CP4 locus.

[0128] In addition to the gene specific primers, universal primers are also designed. The forward universal primer is synthesized to contain the 15 nucleotides at the 5' end of the forward gene specific primer. The reverse universal primer is synthesized to hybridize to the universal PCR nucleotides on the reverse gene specific primer and in addition, contained a 5 nucleotide tag at the 5' end. Ninety-six (96) different universal reverse primers are synthesized with each primer containing a unique tag sequence chosen from the 1024 possible combinations provided by one of 4 bases at each of the 5 nucleotide positions. The samples are subject to PCR using standard conditions. The initial rounds of PCR have the objective of incorporating the universal primers and DNA tag in a limited number of copies of each locus. The 96 gene specific forward and reverse primer pairs are diluted and then combined to make a multiplexed, equimolar stock solution at a final concentration of 10 μmol total oligonucleotide per litre solution. PCR assays contained 1X PCR buffer, 2.5 mM $MgCl_2$, 0.2mM dNTP mix, 1U Taq DNA polymerase, 1 μM of the forward universal primer, 100 nM of the multiplexed primers and 1 μl of DNA extract. In addition, to each unique sample, a uniquely tagged reverse universal primer is added for a final concentration of 1 μM. Cycling is performed in an ABI 7900 with the following cycling program: Initial denaturation at 94°C for 90 seconds; followed by 4 cycles of 94°C for 30 seconds, 55°C for 30 seconds and 72°C for 30 seconds; followed by 22 cycles of 94°C for 30 seconds and 72°C for 60 seconds. The incorporation of the DNA tag through PCR or ligation is essential to the method, however, the subsequent amplification is not always necessary but can facilitate the downstream sample handling steps in preparation for sequencing. After PCR, 2 μl of amplification product is examined by agarose gel electrophoresis on 2% agarose gels subsequently stained with ethidium bromide to confirm the presence of a single product. Assays with a positive PCR reaction are combined into a single pool and purified using a Qiagen kit (Qiagen, USA). The purified products are then subjected to high throughput sequencing according to the manufacturer's protocol (Illumina Genome Analyzer 1G Analyzer, Illumina, Inc.). Two reads are obtained from each sequenced molecule. The first read is obtained by using a primer that corresponded to the forward universal PCR primer sequence. This sequencing primer resulted in a short read of the sequence at the locus for which the primer is designed and within a given sample, as identified by the tag. The tag is read using a short run from a sequencing primer designed to hybridize to the reverse universal primer sequence. This second sequence read is reinitiated after the read of the locus sequence is completed.

[0129] The sequences obtained from the sequencing reaction are binned according to the DNA tag sequence. This is done by trimming the second sequence read down to the DNA tag and then blasting the tags within one run to each other. Within each sample bin, the sequences are clustered to combine multiple reads of the same locus. The sequences at a given locus are then compared (using BLAST) to the expected sequence for the recurrent parent and the donor parent and the CP4 gene. If all the sequence reads matched the recurrent parent, the locus is designated as fixed for the recurrent parent. If the all the sequences matched the donor parent, the locus is fixed for the donor parent and one or more additional backcrosses would be needed to re-introduce the recurrent parent nucleic acids for that locus into the population. If both recurrent parent and donor parent sequences are observed, the locus is called heterozygous and the line could be selfed or further back-crossed to fix for the recurrent parent. This logic is followed for all 95 loci and for the CP4 locus. The progeny with the largest number of recurrent parent loci and which contained the CP4 locus are advanced to further back-crossing so as to further introgress and fix the recurrent parent nucleic acids at all the loci except for the donor nucleic acids at the CP4 locus.

IDT Exhibit 1020.6744

[0130] The present invention further anticipates use of the methods described herein for introgression of 2 or more genomic regions, which may be transgenic or conventional (i.e, QTL).

Example 3. Molecular Fingerprinting Using HT sequencing (Sequence Directed Fingerprinting)

[0131] Nucleotide sequences are ultimate assessment and measurement of genetic makeup of individual plants and genetic similarities among plant varieties/lines. Molecular fingerprints based on nucleotide profiles may provide general information across the genome that can be used, among other applications, to assess germplasm diversity, to help the selection of high performing parents and testers, to query new germplasm pools for potential introgression targets, to query new or existing germplasm pools for genomic regions associated with at least one phenotype of interest, as well as to protect germplasm intellectual properties. If two lines are sufficiently diverse, they are likely in different heterotic groups. That is, they can complement each other, and, when hybridized, have a high probability of generating a productive breeding cross or a hybrid combination. On the other hand, similarity among lines may suggest a potential suboptimal cross. Further, fingerprint similarity provides a basis for evaluation of intellectual property infringement.

[0132] Molecular fingerprints may focus on selected regions of the genome and reveal sequence information at specific loci including, but not limited to, those that are causative or linked to traits of economical importance. The presence or absence of particular nucleotide sequences or particular nucleotide sequence variants at one or more loci can be associated with the traits of interests, and used to predict the performance of these traits, and to select high performing lines in lieu of direct phenotyping.

Molecular fingerprints can be generated based on whole genome sequences, which is costly and time consuming, and often times not practical. The genome complexity could be reduced using various methods before sequencing to produce fingerprints that are based on a small representation (selected regions or loci) of the genome. The present invention provides a more efficient and cost effective approach than the current art, which involves PCR-based detection of a plurality of genetic polymorphisms. Herein, selected polymorphic regions/loci are PCR amplified and then directly genotyped using HT sequencing. Multiplex PCR can be used to amplify as many as hundreds of thousands of such regions/loci simultaneously. Multiplexing samples by using DNA tags can further take advantage of the massive sequence information generated per run by HT sequencing methodologies.

[0133] For molecular fingerprinting, the first step is to select the polymorphic regions or loci to be used to generate the nucleotide sequence-based molecular fingerprints. SNPs are one source of candidate loci although they are not the only source. The number of loci used is determined by many factors including, but not limited to, the objectives and budgets of the projects as well as the structure of the genomes under investigation.

[0134] For example, we select 384 corn SNPs to demonstrate the molecular fingerprinting process although the capacity of a single HT sequencing run allows for the use of a much larger set of SNPs. A single channel of the Illumina Genome Analyzer flow cell can generate around 6 million sequence reads per sequencing run. Therefore approximately 300,000 loci can be genotyped simultaneously with about 20X sequence redundancy. If a smaller number of loci are needed, ~3,000 loci from 96 different samples can be sequenced at the same time by multiplexing samples (see below). These 384 SNPs are chosen from a larger pool of SNPs on the basis of features including even distribution in the corn genome and polymorphism information content (PIC) values of more than 3.0 in an attempt to maximize the information content. A portion of the SNPs are linked to important performance-related characteristics in corn.

[0135] The second step is to amplify the selected loci using multiplexing PCR. A pair of oligonucleotides is synthesized for each SNP, with one of them matching the nucleotide sequence that is 5' to the polymorphic nucleotide in the SNP and the other complementary to the region 3' of polymorphic nucleotide. For optimal sequencing results, although not necessary, the two oligonucleotides are separated by a length that matches the fragment size suggested by the HT sequencing methodologies (50 to 150 nucleotides for Illumina Genome Analyzer), with one of them adjacent to but not overlapping with the polymorphic nucleotide. To increase the efficiency of multiplexing PCR, the oligonucleotides for the 384 loci are designed so that they interfere with one another the least and that the resulting 384 PCR products have similar length and GC content. Two-stage PCR with bipartite oligonucleotides that containing a genome-specific sequence and a universal PCR primer can also help increase multiplexing PCR efficiency. When two-stage PCR is used, the employed HT sequencing methodology needs to be able to sequence through the universal PCR primer and the genome-specific oligonucleotides to reach the polymorphic nucleotide(s) of interest. Otherwise, the PCR products need to be processed to ensure that sequencing read into the polymorphic nucleotide(s). Another option would be to use the sequencing primer as part of universal PCR primer (see example 2) to cut down the number of nucleotides between the sequencing primer and the nucleotide(s) to be sequenced.

[0136] Although it is possible to pool loci "as you go" based on the objective of the experiment and/or the informativeness of individual locus in a given sample population, for molecular fingerprinting the selected loci are usually used as a fixed set. The 384 pairs of oligonucleotides (one for each chosen locus) are diluted in water and pooled together to a final concentration of 5 nM for each oligonucleotide.

[0137] DNA is prepared from each corn line to be fingerprinted using standard extraction protocols. About 100 ng of

IDT Exhibit 1020.6745

each DNA (varying depending on the number of loci used and the size of the genome) is dispensed into 96- or 384-well microtitre plates depending on the number of lines in an experiment and sample multiplexing format. In this example, we fingerprint 96 corn inbred lines. To each well, a PCR master mix is added along with high fidelity DNA polymerase according to standard PCR protocols. Finally, the mixture of the 384 pairs of oligonucleotides is added to each well to a final concentration of 0.5 nM per oligonucleotide and a final volume of 10 μL. An example PCR profile would be 94C for 1 min, 55C for 2 min, and ramping from 55C to 72C within 7 min for 25 cycles, followed by 72C for 7 min. Any PCR protocol can be used as long as enough specific products from all the selected loci are generated for HT sequencing. To minimize PCR amplification errors and uneven amplification among loci, amplification is controlled by reducing the number of cycles and/or amount of oligonucleotides. The goal is to generate the amount of PCR products that are equivalent of the starting DNA suggested by the HT sequencing methodologies.

[0138] The PCR products are then purified according to the HT sequencing requirements before being ligated to sequencing adapters. The template genomic DNA used in PCR will not compete with the PCR products significantly in the downstream sequencing reactions due to the large size of genomic DNA. For optimal results, the template DNA can be removed from the PCR products using methods that are known in the art. In fact, if Qiagen purification columns are used to purify the PCR products for ligation, the majority of genomic DNA will be removed. In this example, Qiagen PCR purification kits (96 well format, according to manufacturer's instructions) are used to purify the PCR products and to remove the template genomic DNA (genomic DNA binds to the columns very tight due to its size and is difficult to elute).

[0139] Finally, the PCR products are ligated to the sequencing adapters for Illumina Genome Analyzer HT sequencing. Other methodologies are known in the art and are within the spirit and scope of this invention. In fact, if universal primers are used in a two-stage PCR scheme and the adapter sequences are used as universal primers, ligation of PCR products to adapters is not necessary since they are introduced through PCR already.

[0140] To take advantage of the massive sequence information generated by the Illumina Genome Analyzer sequencing technology, multiple samples are pooled in sequencing reactions and then deconvoluted using DNA tag sequences. DNA tags are usually 2-6 nucleotides (16 to 4096 unique tags for multiplexing) although longer sequences are desired so that samples are distinguished by more than one nucleotide difference to reduce error. The level of sample multiplexing is determined by the number of sequencing reads generated per run, the number of loci used and the desired level of redundancy, among other factors. The DNA tags can be introduced into the sequencing templates (PCR products in this case) using various methods including the one in example 2, i.e. including the DNA tag sequences in PCR primers. Or different versions of the adapters can be synthesized, with each version having one of the unique DNA tag sequences added at the 3' end; then each version is used for one of the samples in a multiplexing set. In this example, we use the set of 96 adapters provided by Illumina Genome Analyzer, and each adapter, according to the manufacturer's instructions, is ligated to the PCR product in one of the 96 wells in the PCR plate that corresponds to one of the 96 samples in a sample multiplexing format. The ligated products in the 96 wells are then combined into a single well, and used for HT sequencing reaction according to Illumina Genome Analyzer's sequencing protocols. The same oligonucleotide mixture of 384 SNPs can be used to amplify more samples, and PCR products from each plate of 96 samples can be ligated to the 96 versions of the adapters and pooled into one well for HT sequencing. Each Illumina Genome Analyzer flow cell can process up to 8 such pools per sequencing run.

[0141] The sequences obtained from the HT sequencing reactions are first binned according to the DNA tag sequences, assigning sequences to the 96 samples in a pool. Within each bin, the sequences are further grouped based on the sequences of the oligonucleotides that are adjacent to the polymorphic nucleotide(s) and used to amplify the PCR products. There should be 384 groups of sequences in each bin, with each one corresponding to each of the 384 SNP loci. The sequences are then analyzed to determine which allele is present at each of the 384 loci in each of the 96 samples.

[0142] The sequence information is used to determine the presence or absence of a particular nucleotide sequence or a particular variant of the nucleotide sequence at a locus that can be used to correlate with the performance of economically important traits. Once the association is established, with a particular sequence or sequence variant being the cause of or being tightly linked to the trait(s) of interest, the sequence can be used to predict the performance of these trait(s) and to select high performing parents, testers or progenies in lieu of direct phenotyping. The sequences or sequence variants can also be used to estimate, and for the purpose of increasing, the frequency of favorable sequences or sequence variants.

[0143] Sometimes, combinations of several nucleotide sequences or variants of nucleotide sequences at multiple loci are more predictive of certain traits. Using the sequence or variant combinations at closely linked loci, that is, defining haplotypes within pre-determined haplotype windows, is more informative and predictive than treating the loci individually. The other advantage of using combinations of sequences at linked loci is that only a subset of loci is needed to have information about the whole genome because chromosomes are inherited in linkage disequilibrium blocks (haplotype windows) and sequence information at selected loci (tagging loci) from one block can give information for all the loci on the block.

IDT Exhibit 1020.6746

Example 4. Molecular Fingerprinting Soybeans Using HT sequencing (Sequence Directed Fingerprinting)

[0144] The present invention provides a more efficient and cost effective approach than the current art, which involves PCR-based detection of a plurality of genetic polymorphisms. Herein, selected soybean polymorphic regions/loci were amplified and then directly genotyped using HT sequencing. In the present example 1536 loci were evaluated using Illumina Genome Analyzer HT sequencing technology. The present example also provides methods for indirect sequencing, wherein allele-specific tags were incorporated into corresponding template so that only the tag needs to be sequenced to infer the polymorphism.

[0145] As depicted in figures 2-5, there are multiple strategies for genome complexity reduction. For the purpose of fingerprinting, one may wish to employ one or more of the complexity reduction methods known in the art. In the present example, existing PCR-based SNP assays were leveraged to target known polymorphisms using PCR primers corresponding to the SNPs as shown in Figure 4 (direct fingerprinting) or allele-specific extension/ligation as illustrated in Figure 5 (indirect fingerprinting). Leveraging an existing SNP library is particularly advantageous for referencing one or more databases with historic genotype information with a core set of SNPs.

[0146] Next, incorporation of DNA tags is used in order to enable sample multiplexing. In the present example, each sample in a multiplex set was assigned a unique DNA tag, i.e., a sequence tag differing by at least one base pair from the other barcodes in the set. In a preferred aspect, the percentage of G and C bases is balanced to minimize bias in the sequencing process. The DNA tag can range in length from about 2 to about 20 bp. In the present example, with 96 samples (germplasm samples), 5 bp sequences were used for the DNA tags with each DNA tag differing by 2 or more nucleotides (Table 2). These sample DNA tags were incorporated into the allele-specific tags and these allele-specific oligonucleotides were added to the allele-specific extension/ligation projects using PCR. In other aspects the allele-specific tags could be added to the extension/ligation products using a ligation reaction.

Table 2. Exemplary 5 nucleotide DNA tags for 96 samples.

Well	Code	Well	Code	Well	Code	Well	Code
A1	AAGCT	A4	CAAGA	A7	CTGTC	A10	GTTCA
B1	ACACA	B4	CACAG	B7	CTTAG	B10	GTTTC
C1	ACAGC	C4	CACGT	C7	CTTCT	C10	TACCT
D1	ACCAG	D4	CACTC	D7	GAACA	D10	TACGC
E1	ACCTA	E4	CAGAC	E7	GAATC	E10	TACTG
F1	ACGAA	F4	CAGCA	F7	GACAT	F10	TAGGT
G1	ACGGT	G4	CAGTG	G7	GACGA	G10	TAGTC
H1	ACGTC	H4	CATCC	H7	GAGAG	H10	TATCG
A2	ACTAC	A5	CCAAG	A8	GATAC	A11	TCATC
B2	ACTCT	B5	CCCAA	B8	GATGT	B11	TCCAC
C2	AGACT	C5	CCTAT	C8	GATTG	C11	TCCCA
D2	AGCAC	D5	CCTGA	D8	GCAAC	D11	TCGAT
E2	AGCGA	E5	CCTTC	E8	GCACT	E11	TCGGA
F2	AGGAT	F5	CGATC	F8	GCAGA	F11	TCGTG
G2	AGGCA	G5	CGCAT	G8	GCATG	G11	TCTAG
H2	AGGTG	H5	CGGTA	H8	GCGTA	H11	TCTCC
A3	AGTAG	A6	CGTAC	A9	GGTCT	A12	TCTGT
B3	AGTTC	B6	CGTGT	B9	GGTGA	B12	TGACC
C3	ATCCA	C6	CGTTG	C9	GTACG	C12	TGAGA
D3	ATCTC	D6	CTACC	D9	GTAGC	D12	TGATG
E3	ATGAG	E6	CTCAC	E9	GTCCT	E12	TGCAA
F3	ATGCC	F6	CTCGA	F9	GTCTA	F12	TGCTC

IDT Exhibit 1020.6747

(continued)

Well	Code	Well	Code	Well	Code	Well	Code
G3	ATGGA	G6	CTCTG	G9	GTGGT	G12	TGGAG
H3	CAACT	H6	CTGAT	H9	GTGTG	H12	TGTCA

[0147] This fingerprinting example included 1536 soybean SNPs, wherein each SNP was treated as bi-allelic and thus had two allele-specific oligonucleotides (allele-specific tag plus sample DNA tag) and one locus-specific oligonucleotide (Figure 7). The locus-specific oligonucleotide comprised a universal adapter sequence at the 3' end, herein GTCTGCCTATAGTGAG, though the universal adapter sequence could also be part of the primer needed for downstream sequencing (i.e., the Illumina PCR 2.1 primer). The allele-specific oligonucleotides were about 15 nucleotides in length, with balanced melting temperatures.

[0148] To prepare template for sequencing, DNA was prepared for each of the tissue samples as described above. To generate the allele-specific extension/ligation products, the allele-specific tags and locus-specific oligonucleotides were mixed with template, with an initial heating at 70°C , then cool down gradually, followed by 15 minutes at 45°C for DNA polymerase and ligase reactions, as depicted in Figure 5.

[0149] After extension/ligation, products were purified using magnetic beads as known in the art. A subsequent PCR was conducted to add the sample DNA tag, which was added next to the allele-specific tag as illustrated in Figure 12. 96 (x2) forward primers were used, corresponding to 96 germplasm samples. In addition, the Illumina Genome Analyzer genomic sequencing primer was added to the 5' end, wherein the 5' end of the sequence reads: ACACTCTTTCCCTACACGACGCTCTTCCGATCT plus 5-nt sample barcodes (96 versions) plus 15/16-nt allele codes (2 versions). A single reverse primer was used, which corresponds to the universal adapter sequence, and the Illumina Genome Analyzer PCR primer 2.1 was added to the 5' end of this reverse primer, wherein the 3' end of the sequence reads: CAAGCAGAAGACGGCATACGAGCTCTTCCGATCT plus CTCACTATAGGCAGAC. PCR master mix was added to 5 μL of extension/ligation products, along with 0.3 U high fidelity DNA polymerase according to standard PCR protocol, with a final reaction concentration of 0.16 μM primers, 0.1 mM dNTPs in a final volume of 25 μL. The plate was heated to 95°C for nine minutes to denature the DNA. Fifteen cycles of PCR were then completed using the following conditions: 94°C for 30 sec, 50°C for 30 sec, 72°C for 2 min, followed by a final 10 minute extension at 72 °C.

[0150] Approximately 5-10 ng of the purified template was amplified with enrich PCR per Illumina Genome Analyzer specifications. The enrich PCR also adds the adapter required for the downstream bridge PCR reaction if the adapters were not already incorporated in the primers. The enrich PCR product is purified, again using PCR purification methods known in the art, and the resulting template is sequencing per Illumina Genome Analyzer specifications.

[0151] The sequences obtained from the sequencing reaction were binned according to the DNA tag sequence and allele-specific tag sequence. Figure 13 shows the success rate for the markers and the soybean samples, with nearly 90% of the markers and the germplasm entries having a call rate between 90 and 100%. The present example used allele-specific tags which offers an advantage in sequence deconvolution such that the genotype of a sample could be assigned based on the first 20 base pairs since the first 5 base pairs identified the germplasm sample and the next 15 base pairs represented the allele. In other embodiments, the DNA tag could be as short as 2 base pairs and the allele-specific tag could be as short as two base pairs to further reduce the sequence read needed to genotype. In a preferred embodiment, the methods of the present invention anticipate inferring genotype based on just a 2 base pair tag, depending on the degree of multiplexing. In still another aspect, the methods of the present invention anticipate inferring genotype based on a single base pair.

[0152] The ability to simultaneously generate large amounts of fingerprint data, coupled with the flexibility to either saturate specific regions with contiguous sequence or leverage known polymorphic sites for fingerprint data across a haplotype, chromosome, or even genome provides a valuable tool for germplasm improvement activities, experimental work to identify genomic regions of interest, quality assurance and control, and monitoring IP protection.

Example 5. Complexity reduction sequencing using DNA-tagged random primers

[0153] One aspect of this invention is the ability to simultaneously sequence multiple nucleic acid templates which may comprise samples from different individuals or pooled individuals as well as multiple loci.

[0154] In this example, we utilize random primers (hexamer to decamers depending on the project) labeled with a coding system. The coding system will consist of a series of non-native nucleotide sequences ranging from two nucleotides to half the length of the random primer. Mixtures of random primers labeled with at least two DNA tags will be created to amplify and identify any number of genomes or portions of genomes. The amplified sequences are then determined by any number of sequencing methods including, but not limited to, Sanger sequencing using the ABI 3730 or similar platform, pyrosequencing using a 454 or similar platform, and sequencing by synthesis using a Illumina Genome Analyzer

IDT Exhibit 1020.6748

sequencing instrument or similar platform. It is anticipated that this method will be used on new sequencing technologies as they arise.

[0155] This aspect of the present invention will allow researchers to pool DNA samples saving valuable monetary and time resources on sequencing. To evaluate multiple genomes or genomic regions simultaneously, each template will be amplified independently with a different set of DNA-tagged random primers. The length of the random primer is be dictated by level of complexity of the genome; the more repeat sequences, the longer the primer will need to be in order to selectively exclude these regions. Once the genomes are amplified they can be purified by standard methods specific to a given sequencing technology. To make later steps easier, the random primers could also be labeled with a capture molecule such as biotin.

[0156] After amplification, the purified DNA can be sequenced by any number of nucleic acid sequencing methods and compared to identify genome diversity and which specific genomes contribute to the diversity. The present invention could be used without the DNA tags but then once pooled for sequencing there is no way to "de-pool" the sequences and further evaluation either through sequencing or specific genotyping reactions are required.

[0157] This method provides a highly novel method of applying sequence tags to multiplex genome sequencing and genotyping.

Example 6. Mining rare alleles.

[0158] The use of direct nucleic acid sequence data enables detection of rare allleles or haplotypes in the genome of a plant. This is particularly important for leveraging rare but important genomic regions in a breeding program, such as a disease resistance locus from exotic or unadapted germplasm, wherein rare alleles are defined as occurring in low frequency within the germplasm pool and potentially being previously undetected within the germplasm pool. The present example provides methods for rare allele detection, experimental design (i.e., selecting exotic germplasm, germplasm with known phenotype of interest, screening non-elite gp), and utility (i.e., introgression programs for beneficial rare variants for specific traits and/or to expand germplasm diversity in one or more specific germplasm pools such as per maturity zone).

[0159] A set of germplasm comprising at least 2 germplasm entries is provided. Non-limiting factors influencing inclusion in a sequencing project for at least one locus include germplasm origin or geography, at least one genotype of interest, at least one phenotype of interest, performance in hybrid crosses, performance of a transgene, and other observations of the germplasm or predictions relating the germplasm and its performace.

[0160] Using the methods and approaches presented herein, at least one base pair is sequenced for at least 2 germplasm entries. Using methods known in the art for sequence alignment and *in silico* evaluation, differences and similarities are identified and linked to the source germplasm entry. Following identification of alleles of interest, selection decisions can be made.

[0161] In the case of rare allele mining, the rare allele may be associated with a known phenotype. In addition, the identification of the rare allele can provide the basis for additional phenotyping, association studies, and other assays to evaluate the effect of the rare allele on plant phenotype and breeding performance. Further, the direct nucleic acid sequence of the rare allele can be immediately leveraged for use as a marker via methods known in the art and described herein to detect this rare allele in additional germplasm entries, to be used as a basis for selection, and to facilitate introgression of the rare allele in germplasm entries lacking the rare allele. In other aspects, the rare allele is isolated and the isolated nucleic acid is transformed into a plant using methods known in the art in order to confer a preferred phenotype to the recipient plant. The recipient plant can subsequently be used as a donor for conversion programs to cross with elite germplasm for trait integration purposes.

[0162] The identification of rare alleles is useful for leveraging the full genetic potential of any germplasm pool, i.e., set of 2 or more germplasm entries. This is useful for determining breeding cross strategy, increasing the diversity between 2 or more germplasm pools, evaluating heterotic pools, and informing breeding decisions. High throughput sequencing both accelerates the identification of the alleles and allows simultaneous detection of rare alleles and identification of associated markers.

Claims

1. A method of identifying a phenotypic trait of interest in a plant, the method comprising
providing a population of plants comprising at least two plants;
assaying the population of plants for at least one nucleic acid sequencegenetically linked to at least one phenotypic trait of interest; and selectingone or more plantsbased on the presence of the at least one genetically linked nucleic acid sequence.

IDT Exhibit 1020.6749

2. The method as set forth in claim1, wherein the plant is selected from the group consisting of a forage crop, oilseed crop, grain crop, fruit crop, ornamental plants, vegetable crop, fiber crop, spice crop, nut crop, turf crop, sugar crop, beverage crop, tuber crop, root crop, and forest crop.

3. The method as set forth in claim 1, wherein the phenotypic trait is selected from the group consisting of herbicide tolerance, disease resistance, insect or pest resistance, altered fatty acid, protein or carbohydrate metabolism, increased grain yield, increased oil, enhanced nutritional content, increased growth rates, enhanced stress tolerance, preferred maturity, enhanced organoleptic properties, altered morphological characteristics, sterility, other agronomic traits, traits for industrial uses, or traits for improved consumer appeal.

4. The method of claim 1, wherein the assayed nucleic acid sequence is an allele-specific tag.

IDT Exhibit 1020.6750

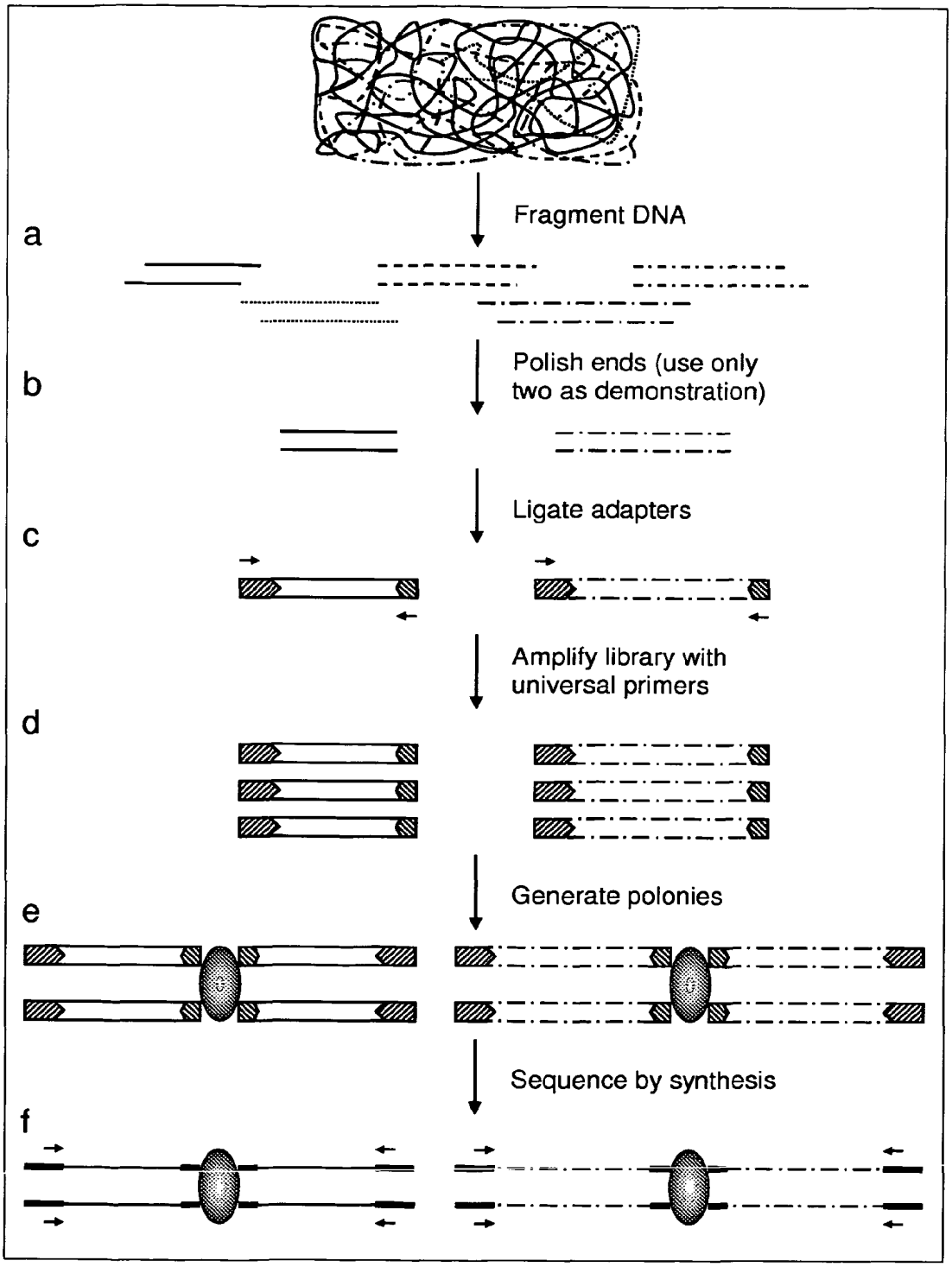


Figure 1.

IDT Exhibit 1020.6751

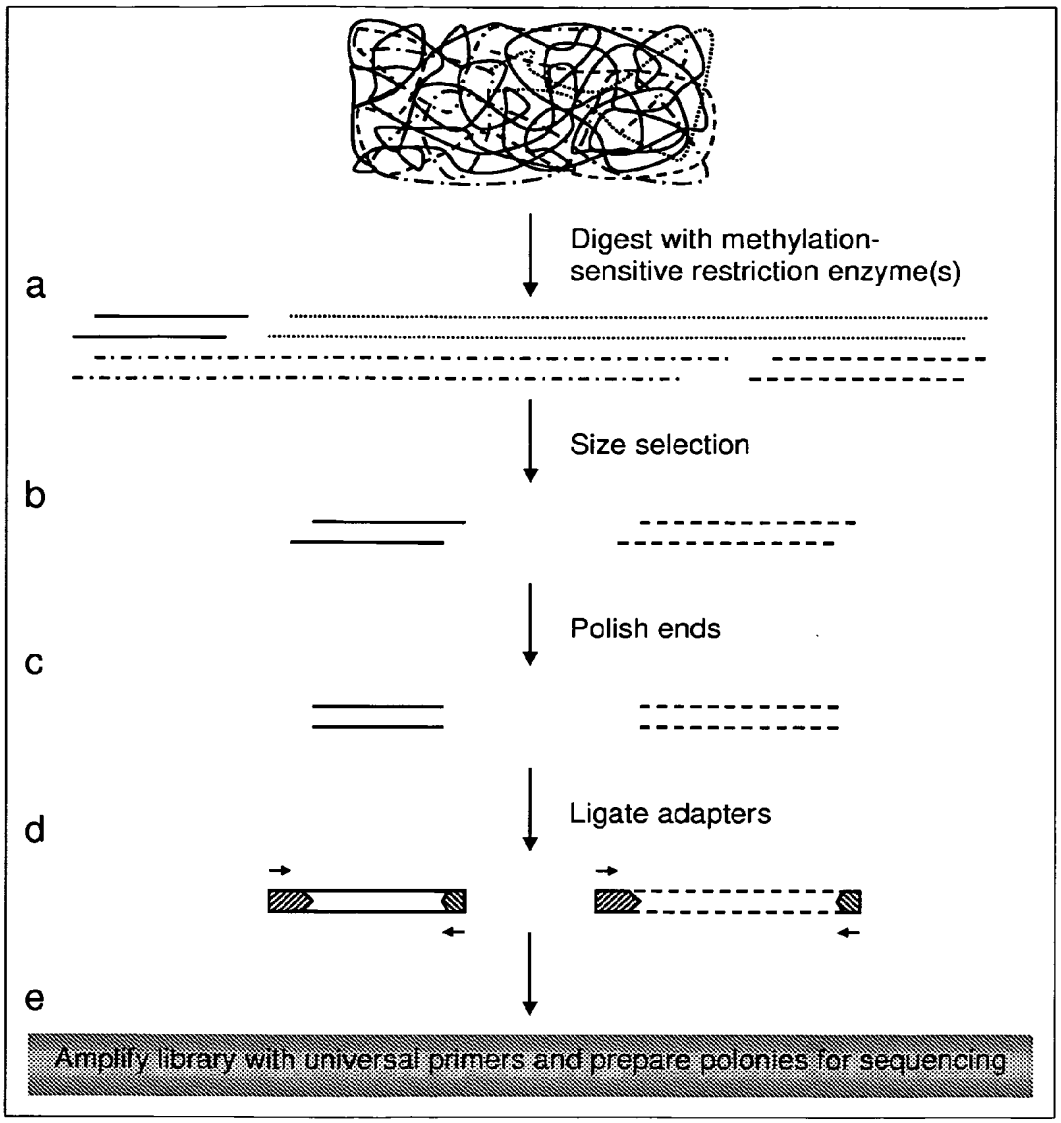


Figure 2.

IDT Exhibit 1020.6752

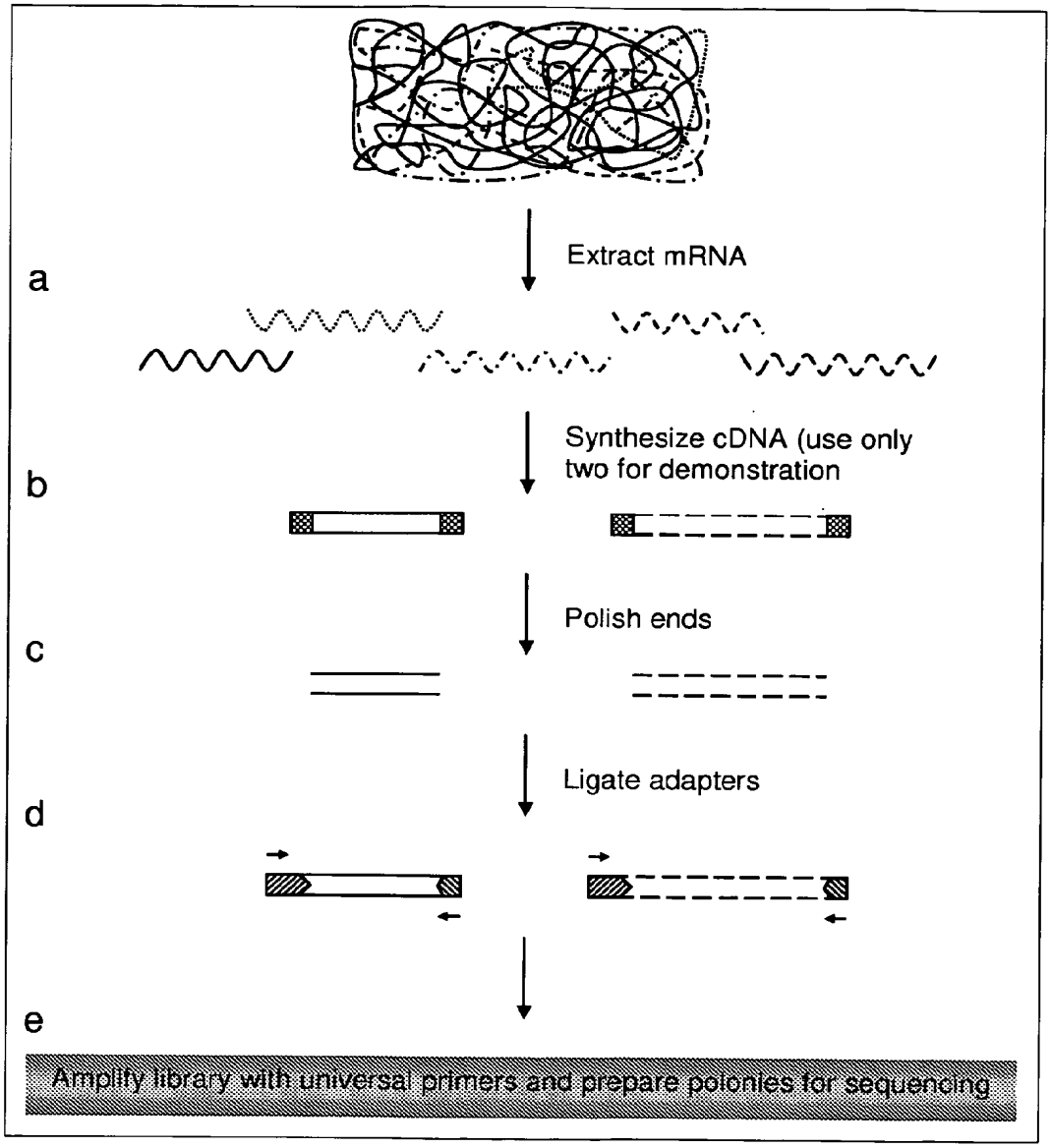


Figure 3.

IDT Exhibit 1020.6753

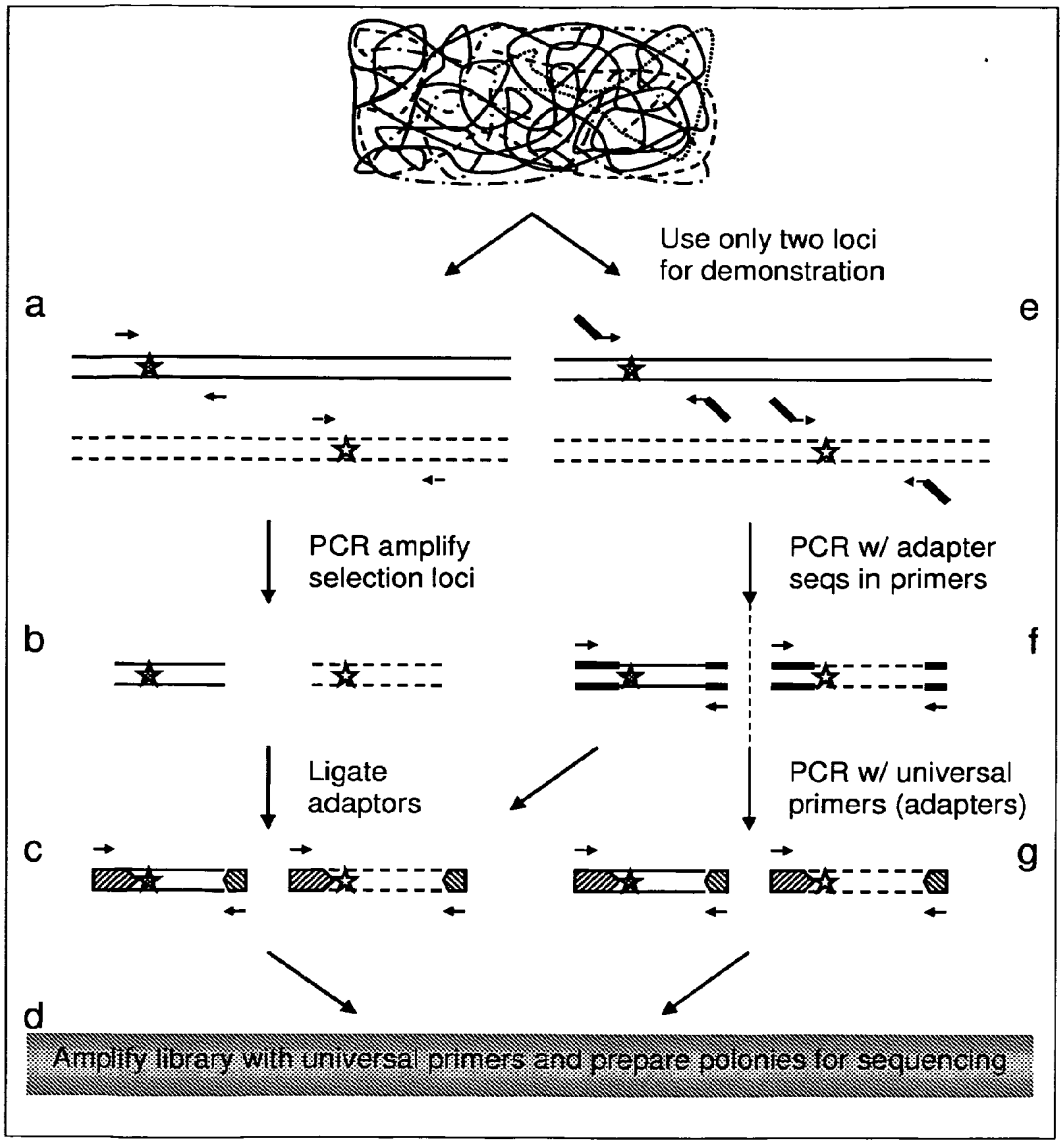


Figure 4.

IDT Exhibit 1020.6754

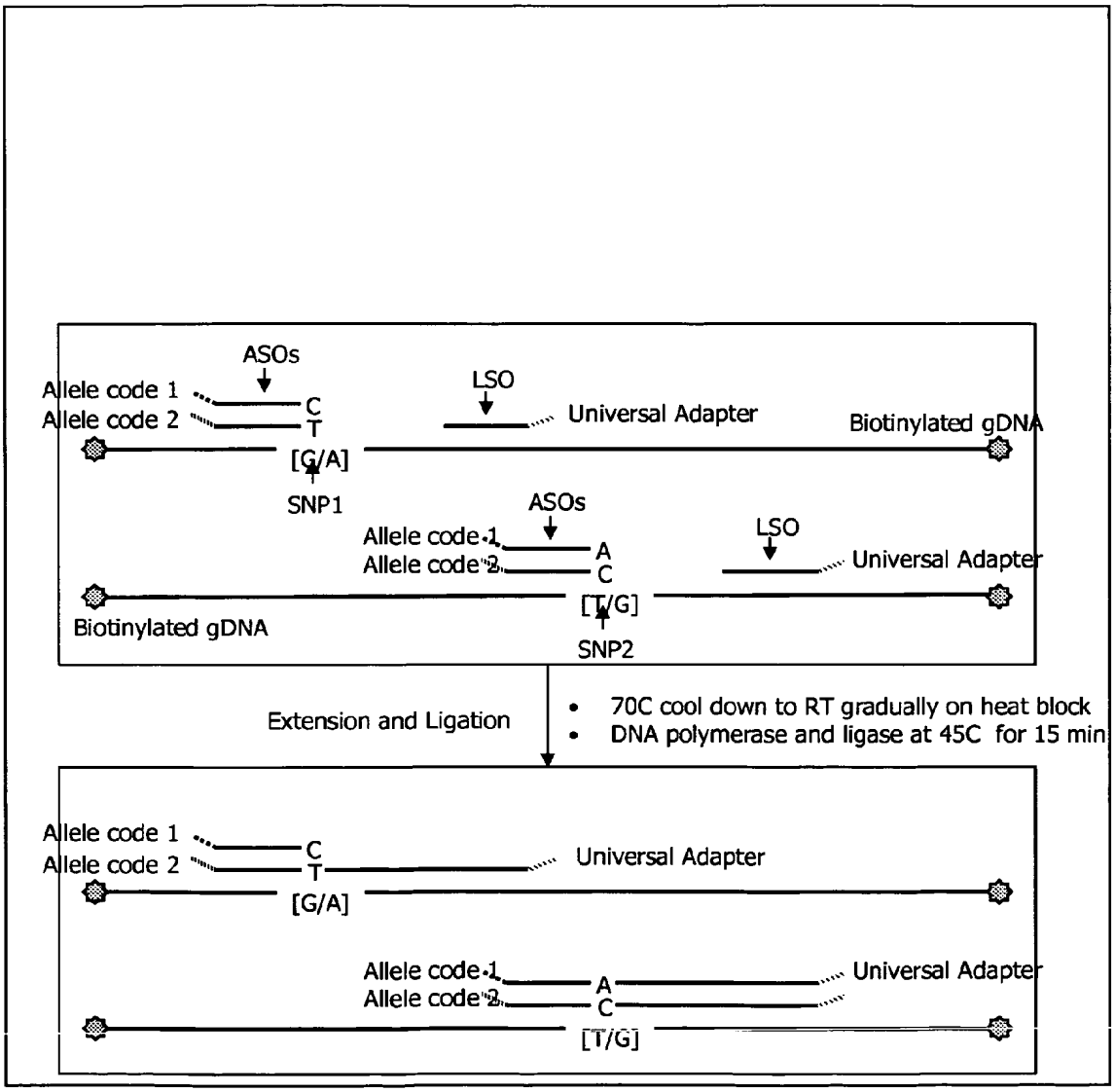


Figure 5.

IDT Exhibit 1020.6755

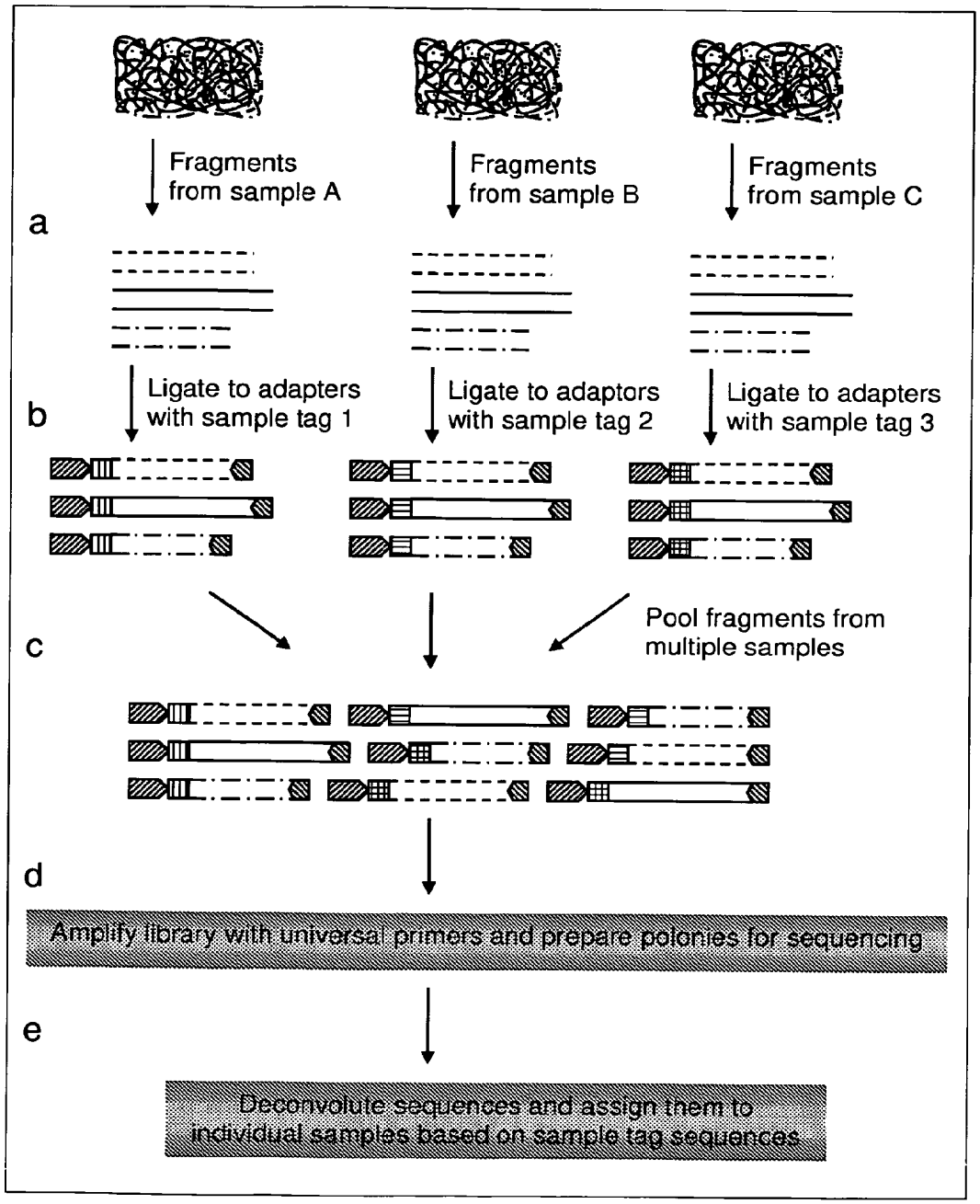


Figure 6.

IDT Exhibit 1020.6756

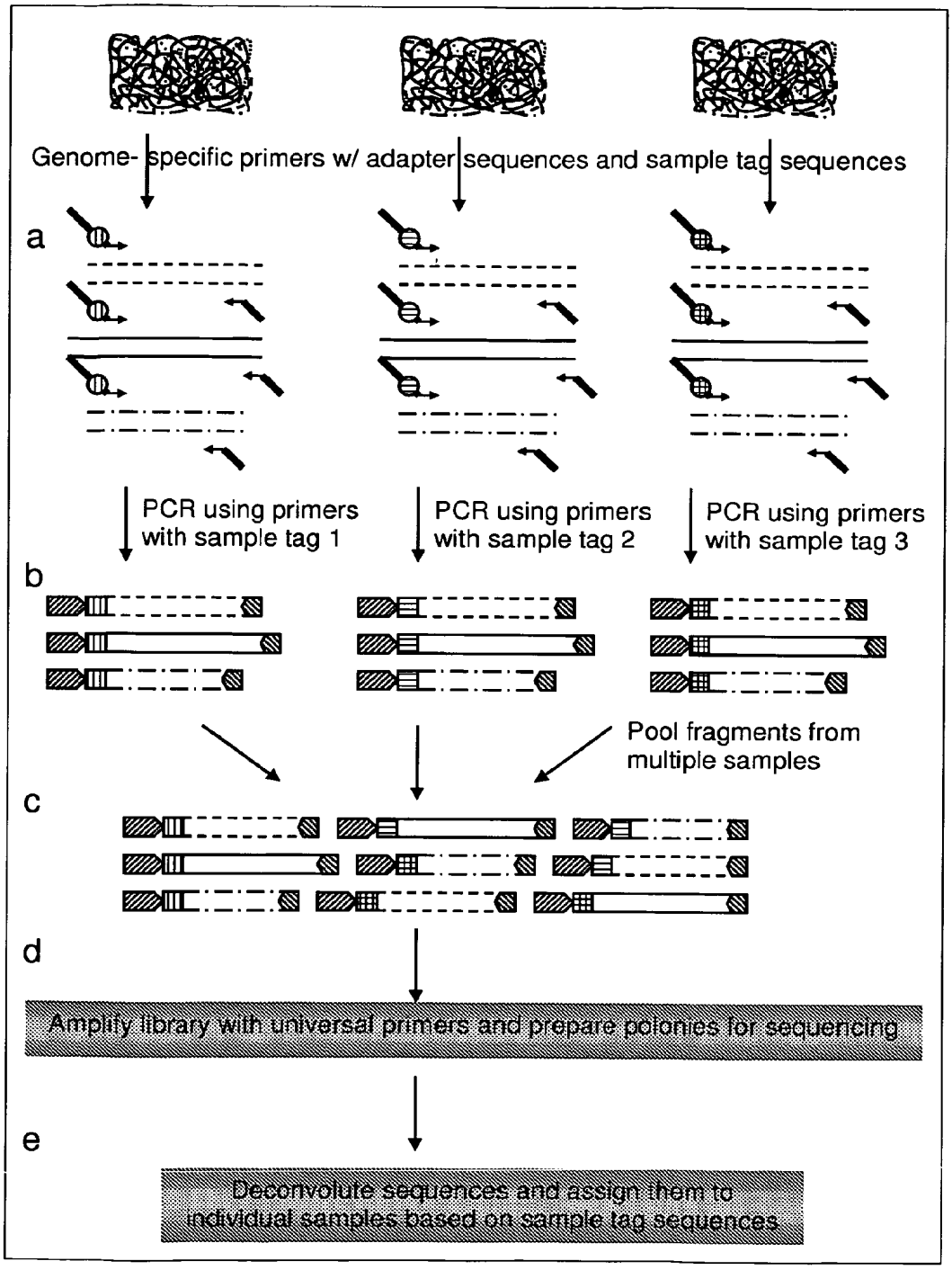


Figure 7.

IDT Exhibit 1020.6757

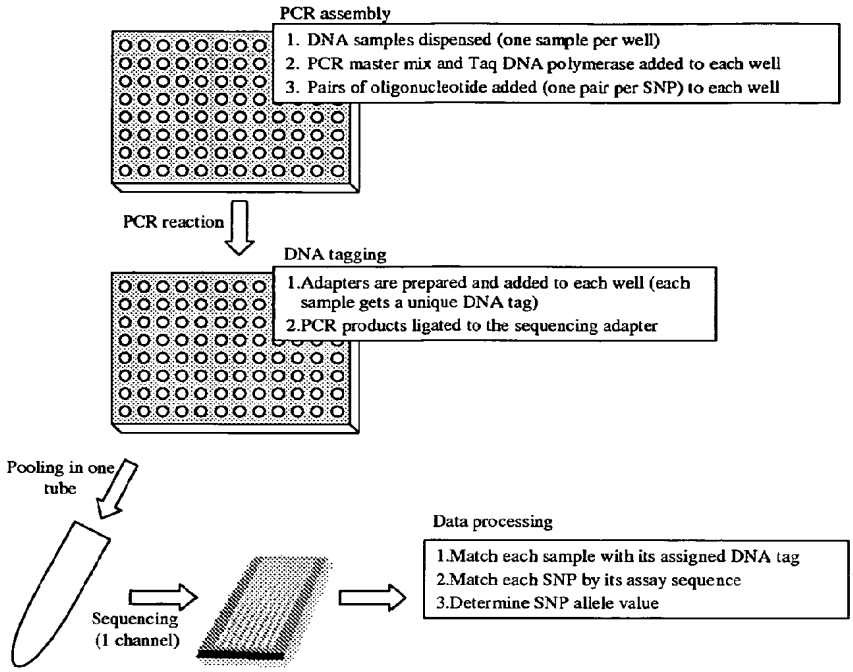


Figure 8.

IDT Exhibit 1020.6758

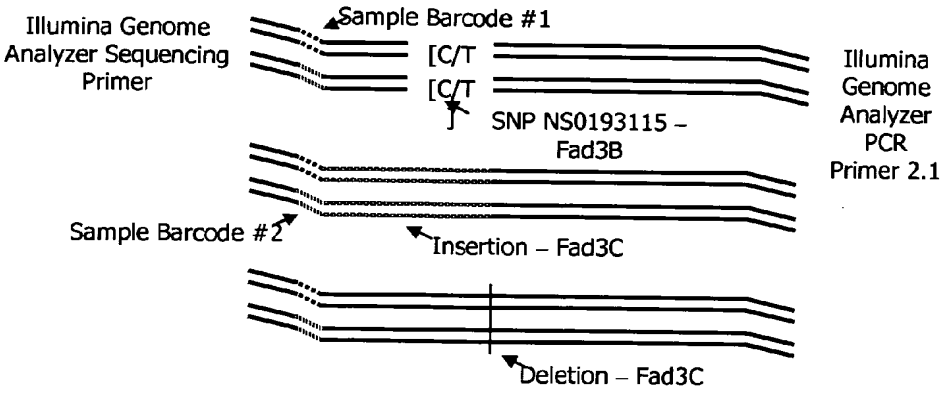


Figure 9.

IDT Exhibit 1020.6759

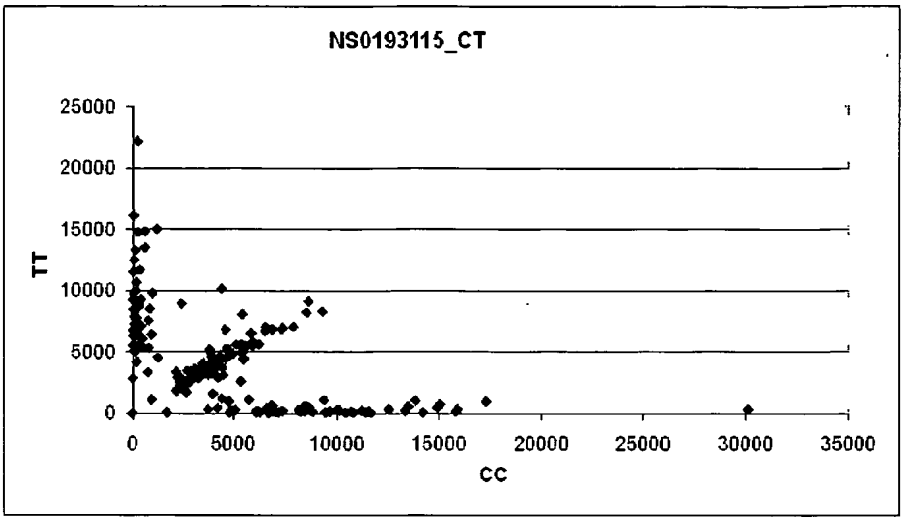


Figure 10.

IDT Exhibit 1020.6760

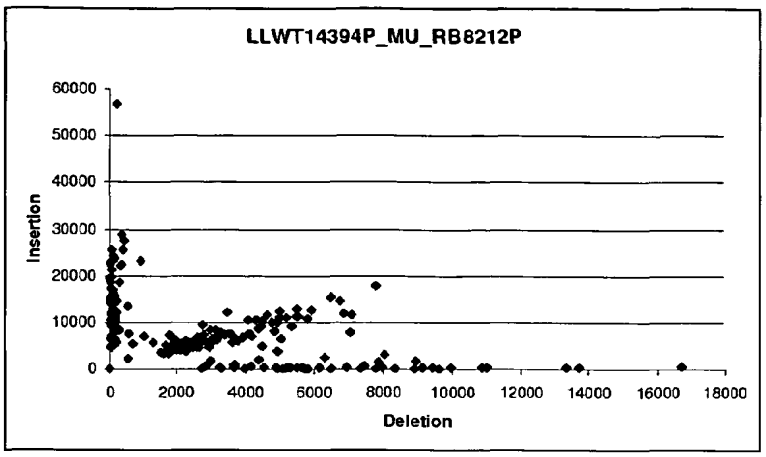


Figure 11.

IDT Exhibit 1020.6761

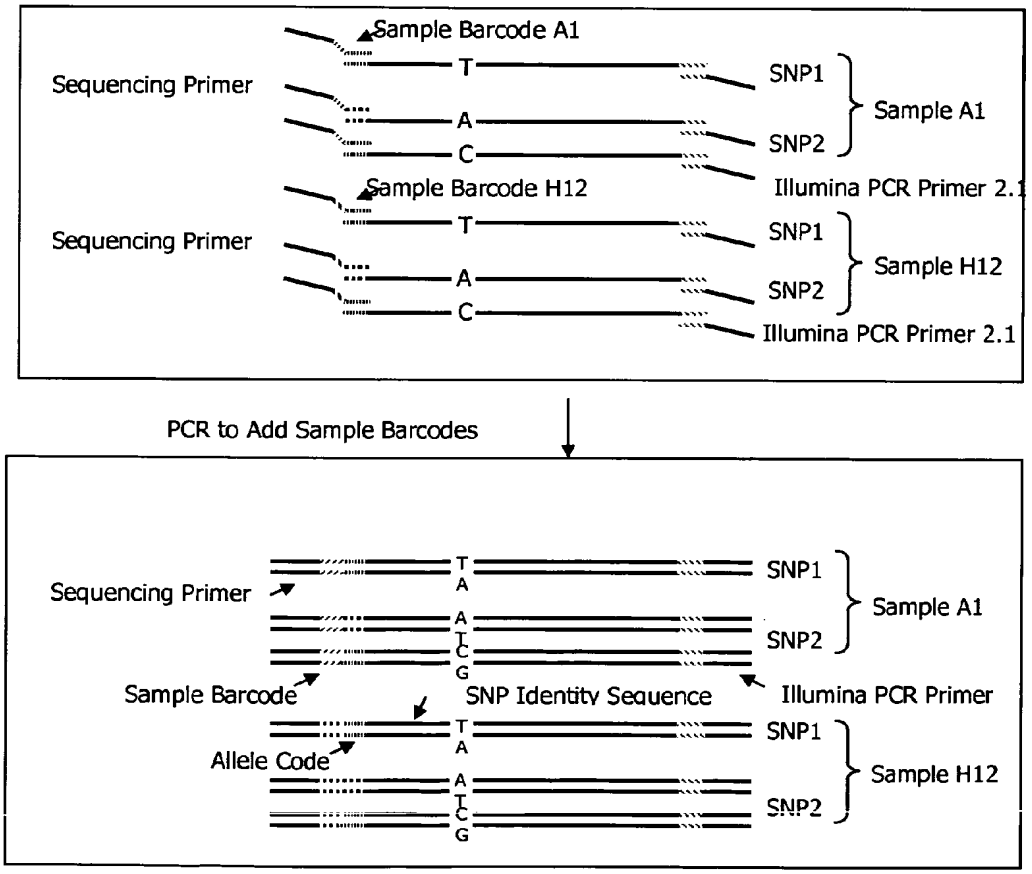


Figure 12.

IDT Exhibit 1020.6762

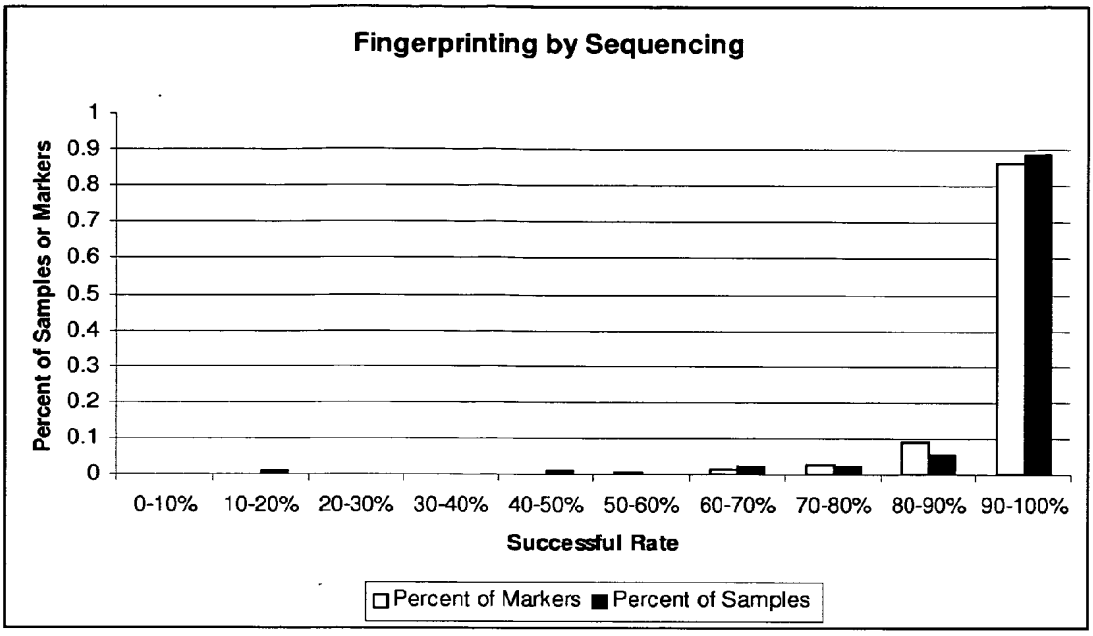


Figure 13.

IDT Exhibit 1020.6763

Europäisches Patentamt
European Patent Office
Office européen des brevets

EUROPEAN SEARCH REPORT

Application Number
EP 11 18 0635

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X	WO 01/49104 A2 (PIONEER HI BRED INT [US]; JANSEN RITSERT C [NL]; BEAVIS WILLIAM D [US]) 12 July 2001 (2001-07-12) * claims *	1-4	INV. C12Q1/68 A01H1/04
X	WO 2007/057402 A1 (BIOGEMMA FR [FR]; PEREZ PASCUAL [FR]; CONSONNI GABRIELLA [IT]; PAUL WY) 24 May 2007 (2007-05-24) * paragraph [0402]; claims *	1-4	
X	US 2005/015827 A1 (PODLICH DEAN [US] ET AL) 20 January 2005 (2005-01-20) * claims *	1-4	
X	WO 2005/000006 A2 (PIONEER HI BRED INT [US]; SMITH OSCAR S [US]; COOPER MARK [US]; TINGEY) 6 January 2005 (2005-01-06) * claim 20 *	1-4	
X,P	WO 2007/103786 A2 (MONSANTO TECHNOLOGY LLC [US]; FORBES HEATHER [US]; DEPPERMANN KEVIN L) 13 September 2007 (2007-09-13) * the whole document *	1-4	TECHNICAL FIELDS SEARCHED (IPC) C12Q C12N A01H
X,P	WO 2008/021413 A1 (MONSANTO TECHNOLOGY LLC [US]; BULL JASON [US]; BUTRUILLE DAVID [US]; E) 21 February 2008 (2008-02-21) * the whole document *	1-4	
A	SOMERS D J ET AL: "Molecular breeding for multiple pest resistance in wheat", DEVELOPMENTS IN PLANT BREEDING (7TH INTERNATIONAL WHEAT CONFERENCE),, 1 January 2007 (2007-01-01), pages 667-676, XP009111292, * the whole document *	1-4	

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
Munich	30 August 2012	Chakravarty, Ashok

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

IDT Exhibit 1020.6764

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 11 18 0635

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

30-08-2012

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0149104 A2	12-07-2001	AU 2591501 A	16-07-2001
		EP 1265476 A2	18-12-2002
		WO 0149104 A2	12-07-2001
WO 2007057402 A1	24-05-2007	AU 2006314535 A1	24-05-2007
		CA 2629363 A1	24-05-2007
		EP 1948683 A1	30-07-2008
		US 2008313778 A1	18-12-2008
		WO 2007057402 A1	24-05-2007
US 2005015827 A1	20-01-2005	AU 2004263840 A1	17-02-2005
		BR PI0412377 A	19-09-2006
		CA 2531119 A1	17-02-2005
		CN 1849064 A	18-10-2006
		EP 1643828 A2	12-04-2006
		US 2005015827 A1	20-01-2005
		WO 2005014858 A2	17-02-2005
WO 2005000006 A2	06-01-2005	AU 2004251624 A1	06-01-2005
		BR PI0410656 A	18-07-2006
		CA 2525956 A1	06-01-2005
		CN 101410008 A	15-04-2009
		EP 1626621 A2	22-02-2006
		US 2005144664 A1	30-06-2005
		WO 2005000006 A2	06-01-2005
WO 2007103786 A2	13-09-2007	AR 059718 A1	23-04-2008
		BR PI0708486 A2	31-05-2011
		CA 2644700 A1	13-09-2007
		EP 1993348 A2	26-11-2008
		RU 2008139096 A	10-04-2010
		US 2007204366 A1	30-08-2007
		US 2010263087 A1	14-10-2010
		US 2011217700 A1	08-09-2011
		WO 2007103786 A2	13-09-2007
WO 2008021413 A1	21-02-2008	AR 063688 A1	11-02-2009
		CA 2660526 A1	21-02-2008
		CN 101528029 A	09-09-2009
		EP 2056667 A1	13-05-2009
		US 2010293673 A1	18-11-2010
		WO 2008021413 A1	21-02-2008

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82

IDT Exhibit 1020.6765

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US 94270707 P **[0001]**
- WO 06137734 A **[0052]**
- WO 06137733 A **[0052]**
- EP 0534858 A **[0052]**
- US 7157564 B **[0053]**
- EP 1724348 A **[0053]**
- US 20070054288 A **[0053]**
- WO 03052101 A **[0053]**
- US 5437697 A **[0072]**
- US 2005000204780 A **[0072]**
- US 2005000216545 A **[0072]**
- US 2005000218305 A **[0072]**
- US 2006000504538 A **[0072]**
- US 6207367 B **[0072]**
- WO 9004651 A **[0073]**
- US 5492547 A **[0073]**
- US 5981832 A **[0073]**
- US 6455758 B **[0073]**
- US 2005000213435 A **[0073]**
- US 2007000680611 A **[0073]**
- US 6399855 B **[0073]**
- US 20050015827 A **[0073]**
- US 6219964 B **[0074]**
- US 20060042527 A **[0091]**
- US 20060046244 A **[0091]**
- US 20060046264 A **[0091] [0120]**
- US 20060048247 A **[0091]**
- US 20060048248 A **[0091]**
- US 20070204366 A **[0091] [0120]**
- US 20070207485 A **[0091]**
- US 20060282911 A **[0108]**
- US 20060200878 A **[0109]**
- US 20070011775 A **[0109]**
- US 4971908 A **[0110]**
- US 4940835 A **[0110]**
- US 4769061 A **[0110]**
- US 4757011 A **[0110]**
- US 5384253 A **[0110]**
- US 5015580 A **[0110]**
- US 5550318 A **[0110]**
- US 5538880 A **[0110]**
- US 6160208 A **[0110]**
- US 6399861 B **[0110]**
- US 6403865 B **[0110]**
- US 5508184 A **[0110]**
- US 5635055 A **[0110]**
- US 5824877 A **[0110]**
- US 5591616 A **[0110]**
- US 5981840 A **[0110]**
- US 6384301 B **[0110]**
- US 20060107348 A **[0115]**
- US 20050204780 A **[0117]**
- US 20050216545 A **[0117]**
- US 20050218305 A **[0117]**
- US 20060504538 A **[0117]**

Non-patent literature cited in the description

- **ALBERS et al.** Molecular Biology of The Cell. Garland Science Publishing, Inc, 2007 **[0034]**
- **RIEGER et al.** Glossary of Genetics: Classical and Molecular. Springer-Verlag, 1991 **[0034]**
- **KING et al.** A Dictionary of Genetics. Oxford University Press, 2002 **[0034]**
- **LEWIN.** Genes IX. Oxford University Press, 2007 **[0034]**
- **ZHANG et al.** *PNAS,* 2007, vol. 104, 3061-3066 **[0053]**
- *Science,* 2006, vol. 311, 1544-1546 **[0056] [0081]**
- **FLINT-GARCIA et al.** *Ann. Rev. Plant Biol.,* 2003, vol. 54, 357-374 **[0073]**
- **NORDBORG et al.** *Trends Gen.,* 2002, vol. 18, 83-90 **[0075]**
- **DOEBLEY et al.** *Cell,* 2006, vol. 127, 1309-1321 **[0077]**
- **J.F. SAMBROOK ; D.W. RUSSELL ; N. IRWIN.** Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, 2000, vol. 1, 2, 3 **[0110]**
- **DELLAPORTA et al.** *Plant Mol Biol Rep,* 1983, vol. 1, 19-21 **[0127]**

IDT Exhibit 1020.6766

Electronic Acknowledgement Receipt

EFS ID:	46340448
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:	21:44:54
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	Information Disclosure Statement (IDS) Form (SB08)	34079_550US_-_Information_Disclosure_Statement_Fillable_PDF_1_of_3.pdf	1275371 ed55bd917f8cf4528bc51af1d767d459fa599dde	no	40

Warnings:

Information:					
2	Information Disclosure Statement (IDS) Form (SB08)	34079_550US_-_Information_Disclosure_Statement_Fillable_PDF_2_of_3.pdf	1267330 c5abd295aa244c24eaf2e04bb46b6a33c0027640	no	17
Warnings:					
Information:					
3	Information Disclosure Statement (IDS) Form (SB08)	34079_550US_-_Information_Disclosure_Statement_Fillable_PDF_3_of_3.pdf	1264080 16141a71b3d9389b59b31ee08660059f7277107b	no	6
Warnings:					
Information:					
A U.S. Patent Number Citation or a U.S. Publication Number Citation is required in the Information Disclosure Statement (IDS) form for autoloading of data into USPTO systems. You may remove the form to add the required data in order to correct the Informational Message if you are citing U.S. References. If you chose not to include U.S. References, the image of the form will be processed and be made available within the Image File Wrapper (IFW) system. However, no data will be extracted from this form. Any additional data such as Foreign Patent Documents or Non Patent Literature will be manually reviewed and keyed into USPTO systems.					
4	Foreign Reference	CA2444926A1.pdf	6747050 27c9bb6a51b324a04cfb08c8ae70bc5a8617ae51	no	164
Warnings:					
Information:					
5	Foreign Reference	34079_135US_-_Translation_of_CA2444926A1.pdf	727406 febc611d2a460e927c16967d979bc56f8b1d218b	no	63
Warnings:					
Information:					
6	Foreign Reference	CN101565746A.pdf	1214408 b2c8031711bc336fee166469cbc4fc32fdcdbb7c	no	20
Warnings:					
Information:					
7	Foreign Reference	translate.pdf	640011 62e4b771cdc7d813483808102e37a8617d5f034a	no	18
Warnings:					
Information:					

8	Foreign Reference	CN105890722A.pdf	207251 9cd71363f151ff6b06ad73a78ae0289dd5691ead	no	5
Warnings:					
Information:					
9	Foreign Reference	34079_516US_-_Translation_of_CN105890722A.pdf	257759 ec5cbc1104a2e8af5ae817d5ef621c1e43789e5f	no	5
Warnings:					
Information:					
10	Foreign Reference	CN1661102A.pdf	626505 a0750ccab47fe58d02d7e12aff1b3e0400905e92	no	12
Warnings:					
Information:					
11	Foreign Reference	34079_516US_-_Translation_of_CN1661102A.pdf	278036 1bd6f02a1ea8e7736c835ceef563ffc4d4c97813	no	7
Warnings:					
Information:					
12	Foreign Reference	EP0329822A2.pdf	2154028 212028dde07bd4d6ebe4ae0501990614f3be90fb	no	23
Warnings:					
Information:					
13	Foreign Reference	EP0365627B1.pdf	2455678 3830068c8b19806d9e9b52bbe9a5adb53ae29525	no	31
Warnings:					
Information:					
14	Foreign Reference	EP0667393A2.pdf	2818540 5c6ae502b4b4de3f0db766d04ab4ac68a09e1339	no	49
Warnings:					
Information:					

15	Foreign Reference	EP0843735B1.pdf	1892862 51f0bbb00d2cf000fcf2874a1a91c9b11f37ab81	no	29
Warnings:					
Information:					
16	Foreign Reference	EP1071811B1.pdf	2049393 29b986c92632848829de0729fdf87a1c211b51eb	no	41
Warnings:					
Information:					
17	Foreign Reference	EP1929039B2.pdf	1880964 9041a7b041de4e3f3a85bb933301bb27570553e5	no	46
Warnings:					
Information:					
18	Foreign Reference	EP2272976A1.pdf	2959032 1c204e1811f9937dee83b1efecc4604d8e728465	no	82
Warnings:					
Information:					
19	Foreign Reference	EP2322612A1.pdf	1135001 0bf24e2ed6b5e1bd245371bb4fb631b4f160a489	no	22
Warnings:					
Information:					
20	Foreign Reference	EP2451973A1.pdf	3993995 efdfc2b6b8963267ec41bf245ddf27ecc2e4a545	no	84
Warnings:					
Information:					
21	Foreign Reference	EP2511381A1.pdf	2699113 17852e1c18d7458a88909ff59f90ba4017162359	no	42
Warnings:					
Information:					

22	Foreign Reference	WO2000018957.pdf	4448929 fe4d7f96ab629f57703290579247da823191da76	no	104
Warnings:					
Information:					
23	Foreign Reference	WO0008208A2.pdf	2223750 aba67a0b011315cb2d8528ae9c0addad46f56e21	no	52
Warnings:					
Information:					
24	Foreign Reference	WO0039345A1.pdf	3890847 b39b6bfb566301d2920c62897a41942b91a5e80b	no	77
Warnings:					
Information:					
25	Foreign Reference	WO0052191A1.pdf	2818551 eb535658b29396371be28afca411d2f452ab49c9	no	58
Warnings:					
Information:					
26	Foreign Reference	WO0070039A1.pdf	1578728 78c5f86d9f1f1e38b8ecf32ff79ee9ddcee20593	no	43
Warnings:					
Information:					
27	Foreign Reference	WO0120035A2.pdf	5347236 99780db10b4b09aea9fab8461ac61d244f55633b	no	116
Warnings:					
Information:					
28	Foreign Reference	WO0123613A1.pdf	7063831 09fee3b47f2863f660d1dd2479872ed0139bc57e	no	179
Warnings:					
Information:					

29	Foreign Reference	WO0146464A1.pdf	1799893 cf0ff1e311314751b0527d982538b5b85d4ea970	no	47
Warnings:					
Information:					
30	Foreign Reference	WO0157248A2.pdf	724064 6ed3fba04385cd131f2cda2ad35e7f6d2fa8bbe8	no	16
Warnings:					
Information:					
31	Foreign Reference	WO0164952A2.pdf	1770930 e616306da9e52a1929d4a90d2e40a934b223da2e	no	41
Warnings:					
Information:					
32	Foreign Reference	WO0200938A2.pdf	7055331 7f671923b0874f87e6a519fe9d377818a9cba98f	no	122
Warnings:					
Information:					
33	Foreign Reference	WO02060318A2.pdf	7014084 25916829176629f527bcb2b4bc7e170a112e4223	no	134
Warnings:					
Information:					
34	Foreign Reference	WO02072772A2.pdf	8573535 8d99101b38bda6faec4434475cad4a7afbf9c68d	no	149
Warnings:					
Information:					
35	Foreign Reference	WO02072773A2.pdf	5814977 619f1495374c5136bbadf336204153e0ab4dabc7	no	111
Warnings:					
Information:					

36	Foreign Reference	WO02081753A1.pdf	2044172 820d314a94b0eeaa818db0c550c64e26d79458f0	no	34
Warnings:					
Information:					
37	Foreign Reference	WO02090584A2.pdf	6540634 3ed6dd1f9897dde2f2ec34c000b5003f5ea410ce	no	163
Warnings:					
Information:					
38	Foreign Reference	WO0228876A2.pdf	3715353 075baa9d4cdd776015485466cc8bda471b020b18	no	97
Warnings:					
Information:					
39	Foreign Reference	WO0229117A2.pdf	4384957 8e56646a154dc1d1a6ea06d66f2880d4db6b17da	no	82
Warnings:					
Information:					
40	Foreign Reference	WO0236821A2.pdf	3487137 ae7186cd2a6c0450ef479e110304e83985ccf61c	no	77
Warnings:					
Information:					
41	Foreign Reference	WO0248402A2.pdf	9001783 d9c8789db6c81800ff7d046354d4478cdf3fbf99	no	172
Warnings:					
Information:					
42	Foreign Reference	WO2003004690.pdf	3805408 7f463b11c2eb019f5bf13070eba115477fad784c	no	77
Warnings:					
Information:					

43	Foreign Reference	WO03027259A2.pdf	2369559 ccb279f2ae5475b825d3d4d34bb59712ee8aef80	no	49
Warnings:					
Information:					
44	Foreign Reference	WO03078645A2.pdf	4814705 3d8d07c52fcfc4632f154d06afd08412673849ac	no	81
Warnings:					
Information:					
45	Foreign Reference	WO03083435A2.pdf	5893654 484aef2bb82312038400f7f16e3f8d3618a1281b	no	114
Warnings:					
Information:					
46	Foreign Reference	WO03106642A2.pdf	3399953 34e45d418404966676 31a6c5a584beb00fb5ce30	no	71
Warnings:					
Information:					
47	Foreign Reference	WO2004011665A2.pdf	6663836 75b94175ffb20d07f334afb6d42e14800282f8d4	no	114
Warnings:					
Information:					
48	Foreign Reference	WO0009756A1.pdf	2439950 ec8919b45ac3b2f50992ba6582373f387a63fda0	no	49
Warnings:					
Information:					
49	Foreign Reference	WO0055364A2.pdf	5236658 929d06077e8f8a90845c21e330b1d6e3af7275bb	no	168
Warnings:					
Information:					

50	Foreign Reference	WO03002736A2.pdf	3507273 41087e94b669a728739e892267589ebd8a02c6b8	no	66
Warnings:					
Information:					
51	Foreign Reference	WO03012118A1.pdf	1899370 415d962452db1c23e5a5ec1bb11a57fce89bbc28	no	39
Warnings:					
Information:					
52	Foreign Reference	WO2004070007.pdf	3695746 bab2c11456c933cbd4f57d293e2d2269641c5b0b	no	73
Warnings:					
Information:					
53	Foreign Reference	WO2004092418A2.pdf	6086132 3be8358ba52774042d57482ac3912dc9c060d0fe	no	96
Warnings:					
Information:					
54	Foreign Reference	WO2005003375A2.PDF	11674420 6844686253543437 23a0fe8ed736e98dd5106def	no	200
Warnings:					
Information:					
55	Foreign Reference	WO2005038427.pdf	1599606 8eeff33bc4061b945bd1cfeb94db39c25544ea77	no	51
Warnings:					
Information:					
56	Foreign Reference	WO2005065321A2.pdf	5854379 53ba149f09d55c3aeab73560f962d05462130ac6	no	95
Warnings:					
Information:					

57	Foreign Reference	WO2006081222A2.pdf	2587188 d535e68c6ec955c8c6b8adc5cf662afc2438fe2c	no	40
Warnings:					
Information:					
58	Foreign Reference	WO2006086668A2.pdf	6034034 377d7437715e4f7a131446b84c65235b424b9b43	no	98
Warnings:					
Information:					
59	Foreign Reference	WO2006137733.pdf	3359055 ef603385fe8e2c637ff28cceb77c28711606bfee	no	75
Warnings:					
Information:					
60	Foreign Reference	WO2007018601A1.pdf	19862831 89cbd0a94dc6b7893bc60a3c1136805fb5e027fe	no	152
Warnings:					
Information:					
			Total Files Size (in bytes):	228626292	

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.

[19] 中华人民共和国国家知识产权局

[51] Int. Cl[7]
C12Q 1/68

[12] 发明专利申请公开说明书

[21] 申请号 200410093095.2

[43] 公开日 2005 年 8 月 31 日

[11] 公开号 CN 1661102A

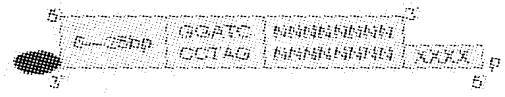
[22] 申请日 2004.12.16
[21] 申请号 200410093095.2
[71] 申请人 上海交通大学
地址 200240 上海市闵行区东川路 800 号
[72] 发明人 张治洲 胡 钧 贺 林
[74] 专利代理机构 上海交达专利事务所
代理人 王锡麟 王桂忠

权利要求书 2 页 说明书 7 页 附图 2 页

[54] 发明名称 基于 DNA 自动机过程的双链 DNA 序列测定的方法

[57] 摘要

一种基于 DNA 自动机过程的双链 DNA 序列测定的方法。本发明中双链 DNA 序列作为输入分子，由限制性内切酶对 DNA 进行切割，露出粘性末端，与状态转移分子通过碱基配对相结合并发生连接。由于状态转移分子仍然含有限制性内切酶的识别位点，连接后的 DNA 分子继续新一轮的切割和连接循环并使作为输入分子的 DNA 序列逐渐变短，作为输入分子的 DNA 序列逐渐变短，输入分子的长度变化反映了 DNA 序列的信息，该长度变化是可阅读的，通过设计状态转移分子的结构可以实现输入分子的长度递减的速度为每次 1 个碱基。本发明有效地避免单链 DNA 分子高级结构引起的测序困难，克服现有技术中单链模板二级结构影响测序的不足，并在此基础上降低成本，具有很大的应用前景。



知识产权出版社出版

IDT Exhibit 1020.6777

1．一种基于 DNA 自动机过程的双链 DNA 序列测定的方法，其特征在于，双链 DNA 序列作为输入分子，由限制性内切酶对 DNA 进行切割，露出粘性末端，与状态转移分子通过碱基配对相结合，在连接酶作用下发生连接；由于状态转移分子仍然含有限制性内切酶的识别位点，连接后的 DNA 分子继续新一轮的切割和连接循环并使作为输入分子的 DNA 序列逐渐变短，作为输入分子的 DNA 序列逐渐变短，输入分子的长度变化反映了 DNA 序列的信息，通过设计状态转移分子的结构可以实现输入分子的长度递减的速度为每次 1 个碱基。

2．根据权利要求 1 所述的基于 DNA 自动机过程的双链 DNA 序列测定的方法其特征是，所述的输入分子为双链 DNA，或由单链核酸分子转换而成的双链。

3.根据权利要求 1 所述的基于 DNA 自动机过程的双链 DNA 序列测定的方法，其特征是，所述的限制性内切酶为 II 型限制性内切酶，切割出来的 DNA 分子末端为粘性末端，长度为 2-5 个碱基，粘性末端既可以是 5’序列，也可以是 3’序列，两个或两个以上的限制性内切酶可同时使用。

4.根据权利要求 2 所述的基于 DNA 自动机过程的双链 DNA 序列测定的方法，其特征是，基于双链为模板进行单链连接的 DNA 连接酶或者是能够直接进行单链 DNA 连接的连接系统。

5.根据权利要求 1 所述的基于 DNA 自动机过程的双链 DNA 序列测定的方法，其特征是，所述的状态转移分子，使用小分子荧光标记物对状态转移分子进行四色荧光标记，状态转移分子在连接酶的作用下与逐渐变短的输入分子发生共价连接，并通过荧光基团来标识正在变短的输入分子。

6．根据权利要求 1 或者 5 所述的基于 DNA 自动机过程的双链 DNA 序列测定的方法，其特征是，所述的状态转移分子，用于 DNA 计算反应的状态转移分子的基本结构为：由长度不同的两条 DNA 单链组成双链 DNA 分子，上链长度在 19—38 碱基，下链长度在 21--43 碱基。上链与下链的 3’端序列完全配对，这样一来下链的 5’端有 2-5 个碱基为粘性末端用于状态转换；粘性末端的长度（n）决定了需要合成的状态转移分子的总个数 4^n，完全配对的区域含有上述 3 中酶的识别位点，识别位点距离下链的 3’末端有 6--25 个碱基，这一段对应的双链为保

IDT Exhibit 1020.6778

护序列；识别位点与粘性末端之间的双链序列为间隔序列，间隔序列的长度设计使得 DNA 计算过程中一次状态转换只移动一个碱基，下链的 3’末端为荧光标记位点；根据下链的 3’末端是 A、G、C 或 T 来选择四种互不相同的荧光标记或任意一种可以用于检测的物理化学标记，下链的 5’末端要进行磷酸化标记。

7．根据权利要求 1 或者 6 所述的基于 DNA 自动机过程的双链 DNA 序列测定的方法，其特征是，所述的序列长度变化可以通过毛细管电泳技术来阅读。

8．根据权利要求 1 所述的基于 DNA 自动机过程的双链 DNA 序列测定的方法，其特征是，DNA 计算反应过程还包括：

1）无蛋白质性质的酶所组成的反应系统，但是使用了核酶或脱氧核酶，在该系统中核酶或脱氧核酶可以代替限制性内切酶驱动 DNA 分子进行的计算反应，但是系统还含有 DNA 连接酶；

2）由生物分子构成某一组合经过物理化学过程而组成的自动机反应系统，以双链 DNA 分子为输入分子进行类似图灵机运算并反映 DNA 序列信息；

在上述反应体系中可以直接观察双链 DNA 输入分子与状态转移分子连接产物的长度变化来检测 DNA 序列信息。

9．根据权利要求 1 所述的基于 DNA 自动机过程的双链 DNA 序列测定的方法，其特征是， DNA 计算反应的反应温度为 16-25° C，各自组分的浓度为：每种状态转换分子为 0.374 um/L，FokI：0.2 umit/uL， Ligase 18.7 units/uL, ATP 0.8 mmol/L, Input DNA 0.15 umol/L. Stock solutions: FokI: 4 units/uL, ligase 400 units/uL, ATP 100 mmol L。

IDT Exhibit 1020.6779

基于 DNA 自动机过程的双链 DNA 序列测定的方法

技术领域

本发明涉及的是一种用于生物技术领域的 DNA 序列测定的方法，特别是一种基于 DNA 自动机过程的双链 DNA 序列测定的方法。

背景技术

自从人类基因组序列公布以来，全世界进入了后基因组时代。成千上万的生物物种需要进行基因组序列测定。在目前的 DNA 测序方案中，最流行的是 Sanger 的双脱氧法，目前 DNA 测序已经有一批国际专利技术[如美国专利，欧洲专利和日本专利]，其中除了少量实用性不是很强的方法外，大部分是围绕双脱氧法进行的一系列技术问题实施的专利。人类基因组序列就是利用双脱氧法测定的。近年来 DNA 计算及 DNA 计算机引起了人们的广泛兴趣。DNA 计算是一个新的领域，其中，DNA 分子作为计算的介质用于解决数学问题。

经对现有技术文献的检索发现，2001 年以色列魏兹曼研究所的 Shapiro 等人在“自然”刊物（Nature, 2001, 414, 430-434)发表了类似图灵机的 DNA 有限自动机。我们发现，在该 DNA 自动机中，作为输入分子的 DNA 双链在限制性内切酶和 DNA 连接酶的循环作用下通过状态转移实现了连续运算，在该过程中双链 DNA 分子的序列长度由长变短。由于在 DNA 自动机反应过程中输入的双链 DNA 分子序列逐渐变短，其长度变化可以即时跟踪，而且长度变化的方式完全可以编程设计，可以设计成每次变短 1 个碱基，所以构成了一个完备的 DNA 测序方案。而且该方案与双脱氧法所描写的技术是完全不一样的技术。

发明内容

本发明的目的在于克服现有技术中的不足，提供一种基于 DNA 自动机过程的双链 DNA 序列测定的方法。使其有效地避免单链 DNA 分子高级结构引起的测序困难，克服现有双脱氧测序技术中单链模板二级结构影响测序的问题，并在此基础上降低成本，本发明在技术路线上不仅能够替代现有技术，而且克服了现有技术的缺陷，具有很大的应用前景。

IDT Exhibit 1020.6780

本发明是通过以下技术方案实现的，本发明的方法具体如下：双链 DNA 序列作为输入分子，由限制性内切酶对 DNA 进行切割，露出粘性末端，与状态转移分子通过碱基配对相结合，在连接酶作用下发生连接；由于状态转移分子仍然含有限制性内切酶的识别位点，连接后的 DNA 分子继续新一轮的切割和连接循环并使作为输入分子的 DNA 序列逐渐变短，作为输入分子的 DNA 序列逐渐变短，输入分子的长度变化反映了 DNA 序列的信息，通过设计状态转移分子的结构可以实现输入分子的长度递减的速度为每次 1 个碱基。由于是基于双链分子，可以有效地避免单链 DNA 分子高级结构引起的测序困难。

所述的输入分子为双链 DNA，或由单链核酸分子转换而成的双链。

基于双链为模板进行单链连接的 DNA 连接酶或者是能够直接进行单链 DNA 连接的连接系统。

所述的限制性内切酶为 II 型限制性内切酶，其识别位点离切割位点有一定的距离，但是又有高度的切割专一性；切割出来的 DNA 分子末端为粘性末端，长度为 2-5 个碱基，而且粘性末端既可以是 5’序列，也可以是 3’序列。两个或两个以上不同类型的限制性内切酶也可以同时使用。

使用小分子荧光标记物对状态转移分子进行四色荧光标记。状态转移分子在连接酶的作用下与逐渐变短的输入分子发生共价连接，并通过荧光基团来标识正在变短的输入分子。

用于 DNA 计算反应的状态转移分子的基本结构为：由长度不同的两条 DNA 单链组成双链 DNA 分子，上链长度在 19—38 碱基，下链长度在 21--43 碱基。上链与下链的 3’端序列完全配对，这样一来下链的 5’端有 2-5 个碱基为粘性末端用于状态转换；粘性末端的长度（n）决定了需要合成的状态转移分子的总个数 4^n。完全配对的区域含有上述 3 中酶的识别位点，识别位点距离下链的 3’末端有 6--25 个碱基，这一段对应的双链为保护序列；识别位点与粘性末端之间的双链序列为间隔序列，间隔序列的长度设计使得 DNA 计算过程中一次状态转换只移动一个碱基。下链的 3’末端为荧光标记位点；根据下链的 3’末端是 A、G、C 或 T 来选择四种互不相同的荧光标记或任意一种可以用于检测的物理化学标记。下链的 5’末端要进行磷酸化标记。

DNA 计算反应体系组分包括输入分子，状态转换分子，连接酶，内切酶，ATP，缓冲液，该反应体系同时包含了内切酶和连接酶。 DNA 计算反应过程还包括：1）

IDT Exhibit 1020.6781

无蛋白质性质的酶所组成的反应系统，但是使用了核酶或脱氧核酶，在该系统中核酶或脱氧核酶可以代替限制性内切酶驱动 DNA 分子进行的计算反应，但是系统还含有 DNA 连接酶；2）由生物分子构成某一组合经过物理化学过程而组成的自动机反应系统，以双链 DNA 分子为输入分子进行类似图灵机运算并反映 DNA 序列信息。在上述反应体系中可以直接观察双链 DNA 输入分子与状态转移分子连接产物的长度变化来检测 DNA 序列信息。

序列长度变化可以通过毛细管电泳技术来阅读。

与目前流行的 DNA 序列双脱氧测定方法相比，本发明描述的方法使用的荧光标记是完全一样的。但是设计的一套状态转移分子对任意 DNA 序列是通用的，而双脱氧法经常需要针对具体的 DNA 序列设计相应的引物。加上本发明是以双链 DNA 为模板，克服了双脱氧法以 DNA 单链为模板带来的二级结构问题，所以本发明明显具备了高于现有技术的技术优势。

以下对本发明的方法进一步具体描述：

（1） 未知 DNA 序列，将其转换成双链并与前导分子连接起来，该前导分子含有 DNA 自动机硬件内切酶所特定的识别位点。前导分子长度在 10 到 20bp（碱基对）左右，可以是一般亚克隆载体上的多克隆位点。PCR（聚合酶链式反应）产物可以与前导分子连接起来，也可以在引物上设计硬件内切酶所特定的识别位点。未知 DNA 序列也可以是由单链核酸（DNA 或 RNA）分子转换而成的 DNA 双链。

（2） 硬件内切酶准备两套， FokI 是其中一种酶，也可以使用 BbsI, HgaI 或 BsMFI。未知序列在测定之前使用这些酶先消化，如果没有某酶的识别位点则可以使用该酶，否则使用上述酶中的另外一个。

（3） 状态转移分子的设计。FokI 的限制性内切酶切割后露出的粘性末端为 4 个碱基，这样需要的状态转移分子个数是 4^4 即 256 种。如果使用的内切酶切割双链 DNA 后粘性末端为 3 个碱基或其它个数，相应的状态转移分子个数需要相应地调整。状态转移分子的结构见图 2。间隔序列使得 DNA 计算过程中一次状态转换只移动一个碱基。全部 256 种状态转移分子中 8 碱基对序列可以全部相同，也可以有所不同。露出的粘性末端需要 5’ 磷酸化以利于连接（在无连接酶系统中 5’ 磷酸化对反应速度没有太大影响），全部状态转移分子可以一次合成。 状态转移分子的荧光分子标记在下链的 3’ 最末端。为了防止状态转移分子的自身连接并保证状态转移分子可以高效地被限制性内切酶识别和切割，左侧保护序

IDT Exhibit 1020.6782

列可以进行类似图 1 显示的特殊设计，即保护序列所在的双链末端不一定是平头末端，也可以让下链的 3’末端比上链多出几个碱基。系统描述用于 DNA 测序的 DNA 计算反应系统的状态转移分子为：由长度不同但是完全配对的两条 DNA 单链组成双链 DNA 分子，上链长度在 19—38 碱基，下链长度在 21--43 碱基。上链与下链的 3’端序列完全配对，这样一来下链的 5’端有 2-5 个碱基为粘性末端用于状态转换；完全配对的区域含有上述 3 中酶的识别位点，识别位点距离下链的 3’末端有 6--25 个碱基，这一段对应的双链为保护序列；识别位点与粘性末端之间的双链序列为间隔序列，间隔序列的长度设计使得 DNA 计算过程中一次状态转换只移动一个碱基。下链的 3’末端为荧光标记位点；根据下链的 3’末端是 A、G、C 或 T 来选择四种互不相同的荧光标记或任意一种可以用于检测的物理化学标记。下链的 5’末端要进行磷酸化标记。

（4） 荧光标记的选择。小分子荧光标记物。可以参考 Applied BioSystems 公司的荧光标记。根据状态转移分子中的碱基粘性末端最后一个碱基来规定荧光标记基团的不同，比如 A 标记 FAM, T 标记 HEX, C 标记 CY3, G 标记 CY5 等，根据序列阅读仪器的要求选择标记方法. 序列阅读仪器将产生反映 DNA 双链分子长度变化的四色荧光图。

（5） 连接酶的选择。一般基于双链为模板进行单链连接的 DNA 连接酶或者能够直接进行单链连接的 DNA 连接酶或连接系统都是可以的。

（6）DNA 计算反应的一般体系:组分包括双链 DNA 输入分子（Input），状态转换分子，DNA 连接酶（Ligase），II 型限制性内切酶，ATP，缓冲液，DNA 计算反应的反应温度为 16-25° C，各自浓度为：每种状态转换分子为 0.374 um/L，FokI：0.2 umit/uL， Ligase 18.7 units/uL, ATP 0.8 mmol/L, Input DNA 0.15 umol/L. Stock solutions: FokI: 4 units/uL, ligase 400 units/uL, ATP 100 mmol L。在上述条件下可以继续进行反应体系的优化。

（7）序列长度变化的阅读方式。本发明的核心在于一个全新的 DNA 测序反应系统即 DNA 计算反应系统；至于观察序列长度的变化可以使用任何现有的技术，如毛细管电泳技术(如 ABI3100/MegaBase 测序仪)。在 DNA 计算反应的一般体系中不同长度的 DNA 片段可以利用毛细管电泳技术和荧光阅读系统来进行 DNA 序列识别。需要知道被测序列的前 3 个碱基的序列以及反映 DNA 双链分子长度变化的四色图，则全部后续序列即可以推算出来，比如如果已经知道前 3 个序列碱基是

IDT Exhibit 1020.6783

AGT,随后的第一个碱基完全由荧光基团判定,比如 A 标记 FAM, T 标记 HEX, C 标记 CY3,G 标记 CY5,依据荧光颜色就可以得到碱基种类。在 DNA 计算其它反应体系中，可以采取直接检测双链 DNA 输入分子长度的变化来识别 DNA 序列信息。

附图说明

图 1 类似图灵机的 DNA 自动机工作原理示意图。

图中：方框中的数字(7,21,22,15,311)表示方框内的双链 DNA 序列长度，T1、T4 和 T6 是状态转移(Transition)分子的 3 个例子，Input SO 表示双链 DNA 分子开始进行 DNA 计算的初始状态，aba 代表 DNA 自动机计算的状态转换程序例子，terminator 代表自动机反应的终止符号，计算到达这里后会给出最终状态。通过设计状态转移(Transition)分子中间隔序列的长度，可以控制自动机每次前进的碱基个数，当自动机每次前进的碱基个数为 1 时，就能利用 DNA 自动机直接测序。在 DNA 计算反应中，上述所有序列和符号都是必需的；当 DNA 计算用于 DNA 序列测定时，初始状态和自动机反应的终止符号都可以不必特意规定。

图 2 用于 FokI 反应系统的状态转移分子示意图。

图中：左侧黑色椭圆球表示荧光标记，标记在下链（3’），GGATC 为 FokI 识别位点，6—25bp 为保护序列，NNNNNNNN 为 8 碱基间隔序列，XXXX 为所有可能的碱基组合，p 为下链 5’磷酸标记，共需要 256 种状态转移分子。

具体实施方式

实施例 1

测定已经插入质粒多克隆位点的未知序列的 DNA 序列信息。利用 DNA 重组技术把把待测序列插入质粒，多克隆位点要包含一个用于 DNA 计算反应的限制性内切酶的识别位点如 FokI；多克隆位点的已知序列为测定的起始序列。设计状态转移分子使每次 DNA 自动机移动切割距离为 1 个碱基，则可以直接阅读 DNA 输入分子的序列。如果利用 FokI 来进行自动机的运行，则露出的粘性末端为 4 个碱基，这样需要考虑的全部状态转移分子个数为 4^4=256 个。只需要知道被测定序列的前 3 个碱基（多克隆位点的已知序列）就可以依据状态转移分子的荧光颜色来推断后面的全部序列。反应之前先用 FokI 切割含有未知序列的质粒，判断未知序列中是否含有另外的 FokI 识别位点。如果没有另外的 FokI 识别位点，则直接进行测序反应，即整个质粒作为 DNA 计算的输入分子加入 DNA 计算反应体系；如果有另外的 FokI 识别位点，则按最佳实施例 3 中的方案进行。

IDT Exhibit 1020.6784

实验研究中所使用的输入分子举例（其中 NNNN 表示长度不限的任意序列）：

FokI

gtcactatggcgtgctgctaggatgtgtgatcctggctcgcagcctggcttgtcgcNNNN

cagtgataccgcacgacgatcctacacactaggaccgagcgtcggaccgaacagcgNNNN

状态转移分子举例 T1：

acgttgcatacggactgctgatggatgtgtgatcc

FAM-tgcaacgtatgcctgacgactacctacacactaggccga-p

在反应体系中加入输入分子 ,T1, FokI 和 T4 DNA ligase. 输入分子 经 FokI 酶切得到：

ggctcgcagcctggcttgtcgcNNNN

gcgtcggaccgaacagcgNNNN

带有 ccga 粘性末端的 T1 在 T4 DNA Ligase 的作用下连接，产物记为分子 M:

acgttgcatacggactgctgatggatgtgtgatccggctcgcagcctggcttgtcgcNNNN

FAM-tgcaacgtatgcctgacgactacctacacactaggccgagcgtcggaccgaacagcgNNNN

下链带有荧光标记，可以用毛细管电泳直接观察。由于 FoKI 的存在，分子 M 还会继续被酶切为两个片断：

gctcgcagcctggcttgtcgcNNNN

cgtcggaccgaacagcgNNNN

和

acgttgcatacggactgctgatggatgtgtgatccg

FAM-tgcaacgtatgcctgacgactacctacacactaggccgag

状态转移分子举例 T2：

acgttgcatacggactgctgatggatgtgtgatcc

FAM-tgcaacgtatgcctgacgactacctacacactaggcgag-p

T2 有可以与输入分子继续连接，然后再被切割。从上述反应中能看出，输入分子一次向右移动一个碱基。上述含有 FoKI 识别位点的输入分子的部分序列可以

IDT Exhibit 1020.6785

设计成多克隆位点序列的一部分。根据已有的实验结果，利用 DNA 计算反应进行 DNA 序列测定可以阅读的序列程度在未经优化的反应条件下将可以达到 500 个碱基左右，已经接近目前双脱氧法的较好水平。

实施例 2

测定 PCR 产物双链 DNA 的序列。需要把 PCR 引物中设计有一个 FokI 位点，这样通过 PCR 扩增（PCR 反应的具体方式不限）出来的双链 DNA 分子的一端含有一个 FokI 位点，该位点离双链 DNA 分子的最末端有 6-10 个碱基的距离，保证该位点可以有效地被 FokI 切割。这样就可以利用 FokI 对应的 DNA 计算反应系统实现 PCR 产物的序列测定。反应之前先用 FokI 切割 PCR 扩增的双链 DNA 序列，判断序列中是否含有另外的 FokI 识别位点。如果没有另外的 FokI 识别位点，则直接进行测序反应，即整个 PCR 产物双链 DNA 作为 DNA 计算的输入分子加入 DNA 计算反应体系；如果有另外的 FokI 识别位点，则按最佳实施例 3 中的方案进行。

实施例 3

对于已经含有 FokI 识别位点的未知序列可以使用如下策略。利用 FokI 切割该序列，切割后的片段进行回收并直接进行测序反应。片段的回收方式有多种，可以将 FokI 切割反应的产物用琼脂糖胶按片段大小进行分离，每个片段分别进行割胶纯化。这些回收纯化过程利用的都是标准的 DNA 重组技术，可以选择市场上任意一种试剂盒来进行。纯化后的 DNA 片段可以直接进行 DNA 计算反应来测定序列，即纯化后的 DNA 双链片段作为 DNA 计算的输入分子加入 DNA 计算反应体系。最后把这些片段的序列拼接起来。

IDT Exhibit 1020.6786