

(21)(A1) **2,258,511**
(86) 1997/06/13
(87) 1997/12/18

(72) KUMAR, Rajan, US
 (72) HEANEY, Paul, US
 (71) SARNOFF CORPORATION, US
 (51) Int.Cl.⁶ C12Q 1/68
 (30) 1996/06/14 (08/665,210) US
(54) PROCEDE DE SEQUENCAGE DE POLYNUCLEOTIDES
(54) METHOD FOR POLYNUCLEOTIDE SEQUENCING

(57) La présente invention se rapporte à un procédé de séquençage de polynucléotides mettant en oeuvre un séquençage séquentiel progressif. Ledit procédé consiste tout d'abord à circulariser une matrice de polynucléotides simple brin avec une amorce de façon à former un complexe matrice-amorce. Selon une réalisation, le procédé de l'invention permet d'ajouter un nucléotide marqué à la fois audit complexe matrice-amorce. On peut également mettre en oeuvre ledit procédé avec un nucléotide ou un analogue de nucléotide qui n'arrête pas l'addition après qu'un premier nucléotide ou un premier analogue de nucléotide ait été ajouté à l'amorce. L'invention se rapporte également à des procédés qui consistent à identifier des polynucléotides comportant une séquence à segments contigus non répétitifs ou à segments superposés non répétitifs. La détection d'un segment contigu non répétitif peut servir, par exemple, à identifier un gène particulier. Dans le cas où le segment contigu non répétitif n'est pas associé de manière unique à un gène particulier, le segment peut servir à la création d'une banque génomique qui peut, par la suite, être explorée, par exemple à l'aide d'un second segment. De la même manière, il est possible d'utiliser une séquence de segments superposés non répétitifs pour, par exemple, identifier un gène ou pour explorer une banque génomique. On peut utiliser ces segments pour, par exemple, effectuer des réactions de circularisation et pour explorer informatiquement une base de données comportant un catalogue de polynucléotides séquencés.

(57) The present invention is a method of sequencing polynucleotides by sequential step sequencing. Sequential step sequencing begins with a single-stranded polynucleotide template that is annealed with a primer forming a primer-template complex. In one embodiment of the method, one labeled nucleotide at a time is added to the primer-template complex. Alternatively, this method can be used with a nucleotide or nucleotide analog that does not stop addition after a first nucleotide or nucleotide analog has been added to the primer. Other embodiments of the invention involve identification of polynucleotides having a contiguous non-redundant string or a superimposed non-redundant string pattern. The detection of a non-redundant contiguous string can be used to identify a particular gene. Alternatively, if the non-redundant contiguous string is not unique to a particular gene, the string can be used to form a DNA library that can then be searched with a second string. Similarly, a superimposed non-redundant string pattern can be used to identify a gene or to search a DNA library. Such strings can be used in annealing reactions and in computer searches of a database having a catalog of sequenced polynucleotides.

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12P 19/34, C12Q 1/68		A1	(11) International Publication Number: WO 97/47761 (43) International Publication Date: 18 December 1997 (18.12.97)
(21) International Application Number:	PCT/US97/09664		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).
(22) International Filing Date:	13 June 1997 (13.06.97)		
(30) Priority Data:	08/665,210	14 June 1996 (14.06.96)	US
(71) Applicant:	SARNOFF CORPORATION [US/US]; 201 Washington Road, CN5300, Princeton, NJ 08543 (US).		
(72) Inventors:	KUMAR, Rajan; 18 Buford Road, Robbinsville, NJ 08691 (US). HEANEY, Paul; 62-14 Ravens Crest Drive, Plainsboro, NJ 08536 (US).		
(74) Agent:	SILVERIO, John, V.; Sarnoff Corporation, 201 Washington Road, CN5300, Princeton, NJ 08543 (US).		

(54) Title: METHOD FOR POLYNUCLEOTIDE SEQUENCING

(57) Abstract

The present invention is a method of sequencing polynucleotides by sequential step sequencing. Sequential step sequencing begins with a single-stranded polynucleotide template that is annealed with a primer forming a primer-template complex. In one embodiment of the method, one labeled nucleotide at a time is added to the primer-template complex. Alternatively, this method can be used with a nucleotide or nucleotide analog that does not stop addition after a first nucleotide or nucleotide analog has been added to the primer. Other embodiments of the invention involve identification of polynucleotides having a contiguous non-redundant string or a superimposed non-redundant string pattern. The detection of a non-redundant contiguous string can be used to identify a particular gene. Alternatively, if the non-redundant contiguous string is not unique to a particular gene, the string can be used to form a DNA library that can then be searched with a second string. Similarly, a superimposed non-redundant string pattern can be used to identify a gene or to search a DNA library. Such strings can be used in annealing reactions and in computer searches of a database having a catalog of sequenced polynucleotides.

METHOD FOR POLYNUCLEOTIDE SEQUENCING

Related Co-Pending U.S. Patent Applications

This patent application is being concurrently filed with the following related U.S. patent applications: "Nuclease Protection Assays," R. Kumar, inventor, Attorney Docket No. DSRC 12038; "Microfluidic Method for Nucleic Acid Amplification," Z. Loewy and R. Kumar, inventors, Attorney Docket No. DSRC 12050; "Method for Amplifying a Polynucleotide," Z. Loewy, inventor, Attorney Docket No. DSRC 12081; "Automated Nucleic Acid Preparation" D. Southgate and Z. Loewy, inventors, Attorney Docket No. DSRC 12120; and "Padlock Probe Detection," R. Kumar, inventor, Attorney Docket No. 317913/12162. This patent application is related to the following copending U.S. patent applications: Ser. No. 60/009517, filed November 3, 1995; Ser. No. 60/00602, filed November 3, 1995; and Ser. No. 60/010513, filed January 24, 1996. All of the foregoing patent applications are hereby incorporated by reference herein in their entirety.

This invention was made with U.S. Government support under Contract No. 70NANB5H1037. The U.S. Government has certain rights in this invention.

In one aspect, the present invention provides a new method for determining the base sequence of RNA or DNA, termed sequential step sequencing. In another aspect, the present invention provides new methods of identifying a polynucleotide or polynucleotides using a contiguous string of non-redundant nucleotides or a superimposed non-redundant string pattern.

Prior art methods of sequencing include the Maxam-Gilbert method and the Sanger dideoxy method. In the Maxam-Gilbert method, a substrate DNA is labeled on one strand with ^{32}P at the 5'-hydroxyl terminus. The labeled DNA is then broken preferentially at one of the four nucleotides using one reaction mixture for each base, the reaction conditions causing an average of one break per DNA molecule. In the reaction mixture for each base, each broken chain yields a radiolabeled fragment extending from the ^{32}P 5'-hydroxyl terminus to one of the positions in the DNA in which that base appears. Thus, every time a base appears in a DNA molecule, it generates a fragment of a different size, which are then separated by gel electrophoresis. The autoradiogram of a gel in which all four chemical reactions have been entered into the gel shows a pattern of bands

from which the sequence of the DNA can be read. See, for example, Stryer, *Biochemistry* (3d ed. 1988) at pages 120-121.

In the Sanger dideoxy method, DNA is sequenced by generating fragments through the controlled interruption of enzymatic replication. First, a primer is 5 constructed which is complementary to the DNA sequence. Then, DNA polymerase is used to copy a sequence of a single-stranded DNA using the primer and four labeled deoxyribonucleoside triphosphates and a 2',3'-dideoxy analog of each of the triphosphates. The incorporation of an analog in the new DNA strand being synthesized results in the termination of incorporation of labeled 10 deoxyribonucleoside triphosphates since the dideoxy analogs lack the 3'-hydroxyl terminus needed to form the next phosphodiester bond. Thus, the synthesis results in DNA fragments of various lengths in which the dideoxy analog is at the 3' end. The reaction mixture for each base can then be separately electrophoresed 15 on a gel or electrophoresed together if the deoxyribonucleoside triphosphate corresponding to each base has a separate label. See, for example, Stryer, *Biochemistry* (3d ed. 1988) at pages 121-123.

The sequential step sequencing methods of the invention, unlike the above-described sequencing methods, involve individual reactions for each type of base in every position in which it appears on the DNA molecule, or for every position in 20 which it appears next to a different type of base. The large number of reactions involved would likely have been considered impractical due to being too labor-intensive under the procedures known in the prior art for performing the necessary chemical reactions.

However, the sequential step sequencing methods of the invention can be 25 used, for example, in the context of a microfluidics-based device for automatedly moving fluids in and out of a reaction chamber, which has been disclosed in U.S. Patent Serial Number 60/010513, filed January 24, 1996, the contents of which are incorporated herein by reference. This combination of the microfluidics-based system and the methods of the invention makes sequential step sequencing an 30 attractive alternative to known conventional methods of nucleotide sequencing. Furthermore, the present invention provides an advantage in eliminating the need for electrophoresis, which is one of the most time-consuming steps of the

sequencing reactions of the prior art. Additionally, the present invention provides for an increased rate of sequence read-out since nucleotide addition can occur, for example, at 800 nucleotides per minute.

5 Further, the present invention provides the advantage, for example, of providing a mechanism for more accurate determination of sequences, such as the sequence adjacent to the poly-A tail of a polynucleotide.

SUMMARY OF THE INVENTION

In one aspect, the present invention provides a method of sequential step sequencing of a polynucleotide having x number of nucleotides comprising:

10 (A) providing a single-stranded polynucleotide template and a first complementary primer having n nucleotides, wherein n is an integer greater than three;

(B) causing the template and the primer to anneal, thereby forming a template-primer complex;

15 (C) adding a template-dependent nucleotide polymerase and at least one nucleoside triphosphate or analog thereof having a label attached thereto, wherein the nucleoside triphosphate or analog thereof includes a base selected from the group consisting of adenine, thymine, cytosine, guanine, and uracil; and

(D) determining whether a label is associated with the template-primer 20 complex or which label is associated with the template-primer complex.

DETAILED DESCRIPTION

DEFINITIONS

The following terms shall have the meaning set forth below:

- **sequential step sequencing** - sequencing a polynucleotide by individual reactions, each reaction detecting no more than the sequence of one type of nucleotide at a time. The nucleotide detected can occur, for example, once or more than once in adjacent positions, such as G, GG or GGGGG (SEQ ID NO: 1).
- **actual non-redundant contiguous string** - a base sequence in which the sequence of each base is actually not repeated in the immediately adjacent base. For example, TACATGTACTGCT (SEQ ID NO: 2) is an actual non-redundant contiguous string, whereas TAACATGTACTGCTT (SEQ ID NO: 3) is not,

although the underlined sequence within this sequence, ACATGTACTGCT (SEQ ID NO: 4), is an actual non-redundant contiguous string.

- **superimposed non-redundant string pattern** - pattern derived from an actual sequence. In the pattern, the redundancies in the actual sequence are removed, a redundancy being the duplication of a base in the immediately adjacent base. For example, given an actual sequence CATTAAAGGGAAAAGCCCCAGTCA (SEQ ID NO: 5), the superimposed non-redundant string pattern of the sequence is CATAGAGCAGTCA (SEQ ID NO: 6).

In one aspect, the present invention relates to methods of sequencing of polynucleotides using sequential step sequencing. The sequential step sequencing methods of the invention, unlike the methods of the prior art, involve individual reactions for each type of base in every position in which it appears on the DNA molecule, or for every position in which it appears next to a different type of base. The methods of the invention are preferably used in the context of a microfluidics-based device for automatedly moving fluids in and out of a reaction chamber, which has been disclosed in U.S. Patent Serial Number 60/010513, filed January 24, 1996, the contents of which are incorporated herein by reference. The present invention provides an advantage over prior art methods, for example, in eliminating the need for electrophoresis, which is one of the most time-consuming steps of sequencing reactions. Additionally, the present invention provides for an increased rate of sequence read-out since nucleotide addition can occur, for example, at 800 nucleotides per minute.

The methods of the invention begin with the provision of a single-stranded polynucleotide template that is annealed with a primer, forming a template-primer complex. In one aspect, the present invention provides for adding one nucleoside triphosphate or analog thereof at a time to the template-primer complex, the nucleoside triphosphate or analog being labelled. If necessary, each of the four nucleoside triphosphates or analogs is added until a label is detected due to the incorporation of a nucleotide into the complex. This method can be used with a nucleoside triphosphate analog that is modified to preclude any subsequent addition of such analog after a first analog has been added to the primer, such as a dideoxynucleoside triphosphate. Alternatively, this method can

be used with a nucleoside triphosphate or analog thereof that does not stop addition after a first nucleotide or nucleotide analog has been added to the primer.

5 In other aspects of the invention, all four nucleoside triphosphates or analogs are added at once to the complex, each nucleoside triphosphate or analog having a different label. Using the latter method, the sequence at this position is determined by identifying the type of label incorporated into the complex. This method is preferably used only with a nucleoside triphosphate analog that is modified to preclude any subsequent addition of such analog after a first analog 10 has been added to the primer.

In one aspect of the invention, when a nucleoside triphosphate or analog is used in the sequential step sequencing methods of the invention, in which the nucleoside triphosphate or analog is *not* modified to preclude any subsequent addition of such nucleoside triphosphate or analog after a first nucleotide has 15 been added to the primer, the sequence obtained may not match the actual sequence of the polynucleotide. Instead, the sequence obtained may be a superimposed non-redundant string pattern. Specifically, if the polynucleotide sequence has a redundancy such that immediately adjacent bases are the same, the sequence obtained using the latter methods of the invention will only detect 20 one of the bases. For example, a polynucleotide with a sequence of CATTAAAGGGAAAGGCCAGTCA (SEQ ID NO:5) will be detected as the corresponding superimposed non-redundant string pattern, CATAGAGCAGTCA (SEQ ID NO: 6). Thus, in one aspect, the methods of the invention provide for the 25 detection of a superimposed non-redundant string pattern in a polynucleotide template.

In another aspect of the invention, when a nucleoside triphosphate or analog is used in the sequential step sequencing methods of the invention, in which the nucleoside triphosphate or analog is *not* modified to preclude any subsequent addition of such nucleoside triphosphate or analog after a first nucleotide has 30 been added to the primer, the sequence obtained will match the actual sequence of the polynucleotide when the amount of label attached to the nucleotide is

quantified, using for example, autoradiography followed by scanning of the autoradiogram to determine signal strength.

More generally, in one aspect, the present invention provides a method of sequential step sequencing of a polynucleotide having x number of nucleotides 5 comprising:

(A) providing a single-stranded polynucleotide template and a first complementary primer having n nucleotides, wherein n is an integer greater than three;

10 (B) causing the template and the primer to anneal, thereby forming a template-primer complex;

(C) adding a template-dependent nucleotide polymerase and at least one nucleoside triphosphate or analog thereof having a label attached thereto, wherein the nucleoside triphosphate or analog thereof includes a base selected from the group consisting of adenine, thymine, cytosine, guanine, and uracil;

15 (D) determining whether a label is associated with the template-primer complex or which label is associated with the template-primer complex.

Preferably, the above method also includes removing unincorporated nucleoside triphosphate or analog from the template-primer complex.

In certain embodiments of the above method, step (C) is limited to using one 20 nucleoside triphosphate or analog thereof, and if no label is associated with the template-primer complex as determined in step (D), then steps (A) to (D) are repeated using another nucleoside triphosphate or analog thereof having a different base than that used previously in step (C), steps (A) to (D) being repeated until it is determined that a label is associated with the template-primer 25 complex.

Preferably, the above methods further comprise step (E), step (E) being, upon having determined which base was added to the first primer by exercise of step (D), a second primer is generated having $n+y$ nucleotides, y being one or the 30 number of identical adjacent nucleotides, wherein the added nucleotide is at its 3' end; and steps (A) to (D) are repeated with the proviso that the second primer is substituted for the first primer. Step (E) is preferably repeated until the primer is

at least x nucleotides long, such that $n=x$, n being the number of nucleotides in the polynucleotide being sequenced.

In certain embodiments, the methods of the invention involve the use of a nucleoside triphosphate analog modified to preclude any subsequent addition of 5 such analog after a first analog has been added to the primer.

When such nucleoside triphosphate analog is used with the methods of the invention, in addition to adding one labeled nucleoside triphosphate analog at a time, the methods also include the addition of more than one labeled nucleoside triphosphate analog at a time. Thus, in certain embodiments, the nucleoside 10 triphosphate analog of step (C) is a combination comprising two, three, or four different nucleoside triphosphate analogs having different bases, wherein the nucleoside triphosphate analogs having different bases are differentially labeled. Preferably, the differentially labeled nucleoside triphosphate analogs are labeled 15 with fluorescent dyes, such as fluorescein, rhodamine, 7-amino-4-methylcoumarin, dansyl chloride, Cy3, Hoechst 33258, R-phycoerythrin, Quantum RedTM, Texas Red, suitable analogs and derivatives thereof, and the like, which can be obtained commercially, such as from Sigma.

When a combination of two, three or four different nucleoside triphosphate analogs are used, it is preferred to use a combination of all four different 20 nucleoside triphosphate analogs are used in concert, the four nucleoside triphosphate analogs having the bases adenine, thymine, cytosine and guanine if the polymerase is DNA-dependent, or the bases adenine, uracil, cytosine and guanine if the polymerase is RNA-dependent.

The above-described sequential step sequencing methods of the invention 25 preferably include step (E), step (E) being, upon having determined which base was added to the first primer by exercise of steps (A) to (D), a second primer is generated having $n+1$ nucleotides, wherein the added nucleotide or nucleotides are at its 3' end; and steps (A) to (D) are repeated with the proviso that the second primer is substituted for the first primer. In certain preferred embodiments, step 30 (E) is repeated until n is x nucleotides long, x being the number of nucleotides in the polynucleotide being sequenced, thereby providing for a full sequence.

The methods of the present invention can be used, for example, to sequence the 3' end of an mRNA or a cDNA nucleotide sequence. Using sequential step DNA sequencing and a poly-T or a poly-U primer, nucleoside triphosphates or analogs thereof having the bases thymine or uracil are added until the primer is 5 extended to the beginning of the poly-A tail. The sequence from this point is then determined using the methods of sequential step sequencing described above, and can be used to determine the actual sequence, or a superimposed non-redundant string pattern. This aspect of the invention overcomes the problems associated with the prior art sequencing methods, such as the presence of a smear on the 10 sequencing gel since a poly-T or poly-U primer randomly anneals to different parts of the poly-A tail.

Specifically, the above-described methods of sequential step sequencing can be used to sequence polynucleotides adjacent to a poly-A tail of the template, the methods further comprising the following steps prior to providing the first primer 15 in step (A):

- (a) providing a single-stranded polynucleotide template and an initial complementary primer, the template having a poly-A sequence, and the primer being a poly-T or a poly-U primer;
- (b) causing the template and the primer to anneal, thereby forming a 20 template-primer complex;
- (c) adding a template-dependent nucleotide polymerase and an nucleoside triphosphate or analog thereof including a base, the base being thymine or uracil, thereby forming an elongated initial primer.

In certain embodiments, the sequenced polynucleotides near the poly-A tail of 25 the template are contiguous, and optionally, the sequenced nucleotides form a non-redundant contiguous string. In other embodiments, the sequenced polynucleotides near the poly-A tail of the template are non-contiguous and form a superimposed non-redundant string pattern.

In certain embodiments, the first primer used in step (A) is the elongated 30 initial primer, the primer being complementary to the 5' end of the poly-A sequence in the template polynucleotide. In other embodiments, the elongated initial primer is used to determine at least one nucleotide of the template adjacent

to the poly-A sequence, and the first primer used in step (A) is complementary to this nucleotide and at least a portion of the poly-A sequence.

Preferably, the primers used in the methods of the invention are about 10 to about 50 nucleotides long, and more preferably, about 15 to about 30 nucleotides long.

In preferred embodiments of the sequencing methods of the invention, the label attached to the nucleoside triphosphate or analog is preferably selected from the group consisting of a radioisotope, a fluorescent dye, a signal-generating enzyme, and a first ligand that specifically binds to a second ligand comprising a radioisotope, a fluorescent dye or a signal-generating enzyme, and most preferably, the label is a fluorescent dye. Suitable radioisotopes include, but are not limited to, ^3H , ^{14}C , and ^{32}P . Suitable fluorescent dyes include, but are not limited to, fluorescein, rhodamine, 7-amino-4-methylcoumarin, dansyl chloride, Cy3, Hoechst 33258, R-phycoerythrin, Quantum RedTM, Texas Red, suitable analogs and derivatives thereof, and the like. Suitable signal-generating enzymes include, but are not limited to, alkaline phosphatase, peroxidase, and urease. Any of the aforementioned labels can be obtained commercially, such as from Sigma.

For instance, labeling methods are described in: Sinha and Striepeke, "Oligonucleotides with Reporter Groups Attached to the 5' Terminus" in Oligonucleotides and Analogues: A Practical Approach, Eckstein, Ed., IRL, Oxford, 1991, p. 185 *et seq.*; Sinha and Cook, "The Preparation and Application of Functionalized Synthetic Oligonucleotides: 3. Use of H-Phosphate Derivatives of Protected Amino-Hexanol and Mercapto-Propanol or Mercapto-Hexanol," Nucleic Acids Research, 1988, Vol. 16, p. 2659 *et seq.*; Haugland, Molecular Probes Handbook of Fluorescent Probes and Research Chemicals, Molecular Probes, Inc., Eugene, OR, 1992, p. 20 *et seq.*; Theisen et al., "Fluorescent Dye Phosphoramidite Labelling of Oligonucleotides," Tetrahedron Letters, 1992, Vol. 33, p. 3036 *et seq.*; Rosenthal and Jones, "Genomic Walking and Sequencing by Oligocassette Mediated Polymerase Chain Reaction," Nucleic Acids Research, 1990, Vol. 18, p. 3095 *et seq.*; Smith et al., "The Synthesis of Oligonucleotides containing an Aliphatic Amino Group at the 5' Terminus - Synthesis of Fluorescent DNA

Primers for Use in DNA-Sequence Analysis," Nucleic Acids Research, 1985, Vol. 13, 2399 *et seq.*

The detection used in conjunction with the invention will depend on the nature of the label. Where a colorimetric or fluorescent label is used visual 5 inspection or an optical instrument such as the fluorescence microscope from Olympus (Lake Success, NY), the Plate Reader device from BioTek Instruments (Winooski, VT) and the CCD (charge-coupled device) camera from Princeton Instruments (Princeton, NJ). Where radioisotopes are used, detection can comprise such spatially sensitive detection devices as the Phosphor Imager device 10 (Molecular Dynamics, Sunnyvale, CA), or can comprise separately detecting individual solid surfaces in a detection apparatus such as a gamma-counter or a liquid scintillation counter.

Further, the template-primer complex is preferably attached to a solid surface, such as a microparticle, which is preferably paramagnetic. A 15 microparticle can have any shape, and preferably it is spherical. Preferably, it has a diameter of less than 1 mm, and more preferably, less than 500 microns. In certain prefererred embodiments, the microparticles have a diameter from about 0.5 micron to about 25 microns, and more preferably about 1 micron to about 5 microns, and even more preferably, about 2 microns to about 4 microns. 20 Microparticles are comprised of any suitable material, the choice of material being guided by its characteristics, which preferably include minimal non-specific absorptive characteristics, such as that of polystyrene. In other embodiments, the microparticles are comprised of, for example, plastic, glass, cellulose, a cellulose derivative, nylon, polytetrafluoroethylene ("TEFLON"), ceramic and the like. A 25 paramagnetic bead can be comprised of, for example, iron dispersed in a polystyrene matrix. A paramagnetic bead can be comprised of, for example, iron dispersed in a polystyrene matrix, and can be obtained with an associated biomolecule, for example, from Dynal (Oslo, Norway), or without an associated biomolecule, for example, from Bang Laboratories (Carmel, Indiana).

30 Additionally, in preferred embodiments, the template-dependent nucleotide polymerase is a DNA polymerase or an RNA polymerase or a fragment thereof having polymerase activity. Most preferably, the DNA polymerase or a fragment

thereof having polymerase activity is T7 DNA polymerase, the Klenow fragment of *E.coli* DNA polymerase I or *Taq* polymerase and the RNA polymerase or a fragment thereof having polymerase activity is derived from *E.coli* or *S.cerevisiae*.

Furthermore, the modified nucleoside triphosphate is preferably a dideoxynucleoside triphosphate. Reaction conditions for the methods of the invention can be found, for example, in EP 0 223 618 and Maniatis et al., *Molecular Cloning* (Cold Spring Harbor 1982) which are hereby incorporated by reference herein in their entirety. Additionally, where methodologies are referred to herein without specific enumeration of well-known methods steps, generally, the following text can be referenced for further details: Ausubel et al., *Short Protocols in Molecular Biology*; Sambrook et al., *DNA Cloning, A Laboratory Manual*; and *Molecular Biology Protocols*, web-site: listeria.nwfsc.noaa.gov/protocols.html.

In preferred embodiments, the methods of the invention are used in the context of a microfluidics-based device for automatedly moving fluids in and out of a reaction chamber, which has been disclosed in U.S. Patent Serial Number 60/010513, filed January 24, 1996, the contents of which are incorporated herein by reference. The microfluidics device is designed specifically for moving small volumes of fluids through fluid exchange channels that connect various sorts of fluid chambers. In particular, such a device comprises a fluid chamber, which is a generic term that describes chambers designed for storage of fluid reagents or reactants, i.e., a supply chamber, for locating reactants undergoing a reaction, i.e., a reaction chamber, for measuring a volume of a fluid, i.e., a metering chamber, and more. More particularly, the device includes a reaction chamber. The reaction chamber is comprised of any suitable material, as are all fluid chambers, such as, for example, glass, plastic, ceramic, or combinations thereof, and is connected to at least two fluid exchange channels for passaging material in and out of the reaction chamber. The reaction chamber preferably remains at a constant temperature of within about two degrees centigrade, wherein the temperature is between about 20°C and 65°C, and alternatively can have adjustable temperatures as in accordance with the requisites of the reactions to take place therein.

The liquid distribution system can conduct synthesis in a great number of separate reaction wells, such as 10,000 reaction wells. The synthesis in each reaction well can occur on a bead or microparticle or can occur on the surfaces of the wells, where these surfaces have been appropriately treated. The wells are 5 formed on a plate that is separable from the portions of the liquid distribution system used to shuttle reagents to a multitude of reaction wells. Another way of forming an array is to apply the photolithographic synthesis procedures described in a number of patents and patent applications owned by Affymax, Inc. These include Fodor et al., "Very Large Scale Immobilized Polymer Synthesis," 10 WO92/10092; Dovor et al., "Method of Synthesizing Diverse Collections of Oligomers," WO93/06121; Campbell et al., "Methods for Synthesis of Phosphonate Esters," U.S. Pat. 5,359,115; Campbell, "Methods for Synthesis of Phosphonate Esters," U.S. Pat. 5,420,328; Fodor et al., "Very Large Scale Immobilized Polymer Synthesis," U.S. Pat. 5,424,186; and Pirrung et al., "Large 15 Scale Photolithographic Solid Phase Synthesis of Polypeptides and Receptor Binding Screening Thereof," U.S. Pat. 5,143,854.

In another aspect, the methods of the invention involve the identification of a polynucleotide or polynucleotides having a contiguous non-redundant string or a superimposed non-redundant string pattern. The detection of the presence of a 20 non-redundant contiguous string can be used, for example, to identify a particular gene. Alternatively, for example, if the non-redundant contiguous string is not unique to a particular gene, the string can be used to form a DNA library that can then be searched, for example, with a second string. Similarly, a superimposed non-redundant string pattern can be used, for example, to identify a gene or to 25 search a DNA library. Preferably the string is at least about 10 nucleotides long, and more preferably, the string is at least about 12 nucleotides long.

Specifically, one method of identifying a polynucleotide or a group of nucleotides, comprises:

- (A) providing a primer complementary to a contiguous string of non-redundant nucleotides, said primer having a label attached thereto;
- 30 (B) providing a single-stranded polynucleotide template;

(C) causing the template and the primer to anneal, thereby forming a template-primer complex;

(D) determining whether a label is associated with the template-primer complex or which label is associated with the template-primer complex.

5 Another method of identifying a polynucleotide or a group of polynucleotides comprises:

(A) providing a base sequence of a string, the string being a superimposed non-redundant string pattern or a contiguous non-redundant string; and

10 (B) searching a computer database of polynucleotide base sequences using the base sequence of the string. In embodiments wherein the string is a superimposed non-redundant string pattern, the above method preferably further comprises providing a computer program for searching for the superimposed string pattern in the polynucleotide sequences, the computer program being capable of identifying a superimposed string pattern despite the presence of a 15 redundancy or redundancies within a sequence that includes the string pattern located in the base sequence of a polynucleotide or polynucleotides in the database.

20 While this invention has been described with an emphasis upon a preferred embodiment, it will be obvious to those of ordinary skill in the art that variations in the preferred composition and method may be used and that it is intended that the invention may be practiced otherwise than as specifically described herein. Accordingly, this invention includes all modifications encompassed within the spirit and scope of the invention as defined by the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

5 (i) APPLICANT: Kumar, Rajan and Heaney, Paul

(ii) TITLE OF INVENTION: METHOD FOR POLYNUCLEOTIDE
SEQUENCING

10 (iii) NUMBER OF SEQUENCES: 6

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: SARNOFF Corporation

(B) STREET: CN 5300

15 (C) CITY: Princeton

(D) STATE: NJ

(E) COUNTRY: USA

(F) ZIP: 08543-5300

20 (v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk

(B) COMPUTER: IBM PC compatible Pentium Pro

(C) OPERATING SYSTEM: WINDOWS NT

(D) SOFTWARE: Microsoft WORD 97

25

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE:

(C) CLASSIFICATION:

30

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Silverio, John V.

WO 97/47761

PCT/US97/09664

(B) REGISTRATION NUMBER: 34,014
(C) REFERENCE/DOCKET NUMBER: SAR 12024PCT

5 (i) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 609-734-2454
(B) TELEFAX: 609-734-2673

10 (2) INFORMATION FOR SEQ ID NO:1:

15 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

15

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

GGGGG

5

20 (3) INFORMATION FOR SEQ ID NO:2:

25 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 13 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

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

TACATGTACTGCT

13

(4) INFORMATION FOR SEQ ID NO:3:

5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

10

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

TAACATGTACTGCTT

15

15

(5) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

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

25

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

ACATGTACTGCT

12

30

(6) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- 5 (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

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

CATTAAAGGGAAAAGCCCAGTCA

23

15 (7) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 base pairs
- (B) TYPE: nucleic acid
- 20 (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

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

CATAGAGCAGTCA

13

WHAT IS CLAIMED:

1. A method of sequential step sequencing of a polynucleotide having x number of nucleotides comprising:

5 (A) providing a single-stranded polynucleotide template and a first complementary primer having n nucleotides, wherein n is an integer greater than three;

(B) causing the template and the primer to anneal, thereby forming a template-primer complex;

10 (C) adding a template-dependent nucleotide polymerase and at least one nucleoside triphosphate or analog thereof having a label attached thereto, wherein the nucleoside triphosphate or analog includes a base selected from the group consisting of adenine, thymine, cytosine, guanine, and uracil;

(D) determining whether a label is associated with the template-primer 15 complex or which label is associated with the template-primer complex.

2. The method of claim 1, further comprising removing unincorporated nucleoside triphosphate from the template-primer complex.

3. The method of claim 1, wherein step (C) is limited to using one nucleoside triphosphate or analog thereof, further comprising:

20 if no label is associated with the template-primer complex as determined in step (D), then steps (A) to (D) are repeated using another nucleoside triphosphate or analog thereof having a different base than that used previously in step (C), steps (A) to (D) being repeated until it is determined that a label is associated with the template-primer complex.

25 4. The method of claim 1, further comprising

(E) upon having determined which base was added to the first primer by exercise of step (E), a second primer is generated having n+y nucleotides, y being one if each nucleotide is added one at a time, or, if each nucleotide is added more than one at a time, y being the number of identical adjacent nucleotides, wherein 30 the added nucleotide or nucleotides are at the primer's 3' end; and steps (A) to (D) are repeated with the proviso that the second primer is substituted for the first primer.

5. The method of claim 1, wherein, in step (C), the nucleoside triphosphate analog has a modification that precludes any subsequent addition of such analog after a first analog has been added to the primer.

6. The method of claim 1, wherein the method is used to sequence 5 polynucleotides adjacent to a poly-A tail of the template, further comprising the following steps prior to providing the first primer in step (A):

(a) providing a single-stranded polynucleotide template and an initial complementary primer, the template having a poly-A sequence, and the primer being a poly-T or a poly-U primer;

10 (b) causing the template and the primer to anneal, thereby forming a template-primer complex;

(c) adding a template-dependent nucleotide polymerase and an nucleoside triphosphate or analog thereof including a base, the base being thymine or uracil, thereby forming an elongated initial primer.

15 7. The method of claim 1, wherein the label is selected from the group consisting of a radioisotope, a fluorescent dye, a signal-generating enzyme, and a first ligand that specifically binds to a second ligand comprising a radioisotope, a fluorescent dye or a signal-generating enzyme.

8. The method of claim 1, wherein the template-primer complex is attached to 20 a solid surface.

9. The method of claim 8, wherein the solid surface is a microparticle.

10. The method of claim 1, wherein the template-dependent nucleotide polymerase is a DNA polymerase or an RNA polymerase or a fragment thereof having polymerase activity.

25 11. A method of identifying a polynucleotide or polynucleotides, comprising:

(A) providing a primer complementary to a contiguous string of non-redundant nucleotides, said primer having a label attached thereto;

(B) providing a single-stranded polynucleotide template;

30 (C) causing the template and the primer to anneal, thereby forming a template-primer complex;

(D) determining whether a label is associated with the template-primer complex or which label is associated with the template-primer complex.

12. A method of identifying a polynucleotide or polynucleotides, comprising:

(A) providing a base sequence of a string, the string being a superimposed non-redundant string pattern or a contiguous non-redundant string; and

5 (B) searching a computer database of polynucleotide base sequences using the base sequence of the string.

13. The method of claim 12, wherein the string is a superimposed non-redundant string, further comprising providing a computer program for searching for the superimposed string pattern in the polynucleotide sequences, the computer

10 program being capable of identifying a superimposed string pattern despite the presence of a redundancy or redundancies within a sequence that includes the string pattern located in the polynucleotide or polynucleotides in the database.

CHECKLIST

1. International Application No. : PCT/US97/09664.
2. Date of Entry into National Phase : (D/M/Y) 14 Dec 1998.
3. a) Priority Date : 14 Jan 96
International Filing Date : 13 June 97.
- b) Demand-before 19 months : Yes
- c) Entry at : 20 32 mo. (Chapter I)
30 or 42 mo. (Chapter II)
4. Filing Fee (KC 29101) \$ 150 Small Entity \$ 300 Large Entity
5. Entry Form submitted & checked : Yes
6. Maintenance Fees paid (24 months from the International
Filing Date) (KC 51501) : Yes No
7. Drawings submitted : Yes No
8. Late Payment/Reinstatement (12 months from N.E. due date) : Yes No
9. Chapter I or Chapter II - Amendment under Article 19 or Article 34
Pre. Exam. Rep.. Missing Yes
Publication as filed Amendment entered
Abstract Specification 1-17 Claims 1-13 Drawings: (Pages)
Sequence Listing: 4 pages (Figures)
10. Agent's Code : 0 1803.
11. Pub. No: WO9 147761
12. Formalities:

Submission of Information:

Cdn. Rep. & Agent: Yes No

Evidence : Yes No Change of name : Yes No

Name & address of inventor : Yes No

13. PCT. Language of (Pamphlet) : English French Foreign

14. Agent's Copy : English French
As Filed: _____ Amended: _____

MM

PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF ELECTION
(PCT Rule 61.2)

To:

Canadian Patent Office
The Commissioner of Patents
50 Victoria Street
Ontario K1A 0C9
CANADA

Date of mailing (day/month/year)

09 February 1998 (09.02.98)

in its capacity as elected Office

International application No.

PCT/US97/09664

Applicant's or agent's file reference

SAR 12024PCT

International filing date (day/month/year)

13 June 1997 (13.06.97)

Priority date (day/month/year)

14 June 1996 (14.06.96)

Applicant

SARNOFF CORPORATION

1. The designated Office is hereby notified of its election made:

 in the demand filed with the International Preliminary Examining Authority on:

12 January 1998 (12.01.98)

 in a notice effecting later election filed with the International Bureau on:2. The election was was not

made before the expiration of 19 months from the priority date or, where Rule 32 applies, within the time limit under Rule 32.2(b).

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No.: (41-22) 740.14.35

Authorized officer

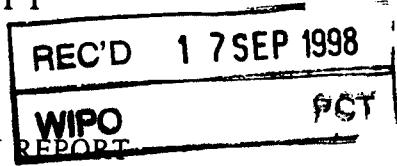
Ting Zhao

Telephone No.: (41-22) 338.83.38

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY EXAMINATION



(PCT Article 36 and Rule 70)

Applicant's or agent's file reference SAR 12024PCT	FOR FURTHER ACTION	See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)
International application No. PCT/US97/09664	International filing date (day/month/year) 13 JUNE 1997	Priority date (day/month/year) 14 JUNE 1996
International Patent Classification (IPC) or national classification and IPC IPC(6): C12P 19/34; C12Q 1/68 and US Cl.: 435/91.1, 6		
Applicant SARNOFF CORPORATION		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of 4 sheets. AR FILER / TEL. QUE DEPOSE

This report is also accompanied by ANNEXES, i.e., sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority. (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

These annexes consist of a total of 1 sheets.

3. This report contains indications relating to the following items:

- I Basis of the report
- II Priority
- III Non-establishment of report with regard to novelty, inventive step or industrial applicability
- IV Lack of unity of invention
- V Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
- VI Certain documents cited
- VII Certain defects in the international application
- VIII Certain observations on the international application

Date of submission of the demand 12 JANUARY 1998	Date of completion of this report 17 AUGUST 1998
Name and mailing address of the IPEA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	Authorized officer Kenneth B. Neffick Telephone No. (703) 308-0196
Facsimile No. (703) 305-3230	

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.

PCT/US97/09664

L Basis of the report

1. This report has been drawn on the basis of (Substitute sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to the report since they do not contain amendments):

the international application as originally filed.

the description, pages 1-17, as originally filed.
pages NONE, filed with the demand.
pages NONE, filed with the letter of _____.
pages , filed with the letter of _____.

the claims, Nos. 1-13, as originally filed.
Nos. NONE, as amended under Article 19.
Nos. NONE, filed with the demand.
Nos. NONE, filed with the letter of _____.
Nos. , filed with the letter of _____.

the drawings, sheets/fig none, as originally filed.
sheets/fig NONE, filed with the demand.
sheets/fig NONE, filed with the letter of _____.
sheets/fig , filed with the letter of _____.

2. The amendments have resulted in the cancellation of:

the description, pages NONE.

the claims, Nos. NONE.

the drawings, sheets/fig NONE.

3. This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed, as indicated in the ~~Supplemental Box~~ Additional observations below (Rule 70.2(c)).

4. Additional observations, if necessary:

NONE

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.

PCT/US97/09664

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement**1. STATEMENT**

Novelty (N)	Claims <u>12 and 13</u>	YES
	Claims <u>1-11</u>	NO
Inventive Step (IS)	Claims <u>12 and 13</u>	YES
	Claims <u>1-11</u>	NO
Industrial Applicability (IA)	Claims <u>1-13</u>	YES
	Claims <u>none</u>	NO

2. CITATIONS AND EXPLANATIONS

Claim 11 lacks novelty under PCT Article 33(2) as being anticipated by US 5,403,708.

This claim is drawn to a method comprising: providing a primer having a label; providing a template; annealing the primer and template to form a complex; and determining whether a label is present in the complex.

US 5,403,708 teaches such a method in columns 2-9. The method of this patent applies to the use of any primers, including those which contain either redundant or non-redundant sequences.

Claims 1-10 lack novelty under PCT Article 33(2) as being anticipated by WO 90/13666.

These claims are drawn to a sequencing method comprising: providing a template and a primer; annealing the template and primer to form a complex; adding a polymerase and at least one nucleoside triphosphate having a label, and determining whether the label is incorporated into the complex.

WO 90/13666 discloses such a method; see especially pages 3-5, 7, and 14, as well as Fig. 1.

Claims 1-10 lack novelty under PCT Article 33(2) as being anticipated by US 4,863,849.

These claims are drawn to a sequencing method comprising: providing a template and a primer; annealing the template and primer to form a complex; adding a polymerase and at least one nucleoside triphosphate having a label, and determining whether the label is incorporated into the complex.

US 4,863,849 discloses such a method; see especially columns 6-10 and 13, as well as the claims.

Claims 12 and 13 satisfy the criteria set out in PCT Article 33(2)-(4), because the prior art does not teach or fairly suggest: providing a modified nucleotide sequence corresponding to an unknown sequence of interest, said modified sequence being a non-redundant string pattern in which there are no identical consecutive nucleotides present; and searching a database using the non-redundant string pattern sequence. The sequencing (Continued on Supplemental Sheet.)

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.

PCT/US97/09664

Supplemental Box

(To be used when the space in any of the preceding boxes is not sufficient)

Continuation of: Boxes I - VIII

Sheet 10

V. 2. REASONED STATEMENTS - CITATIONS AND EXPLANATIONS (Continued):

methods of WO 90/13666 and US 4,863,849 intentionally avoid generating redundant sequence information by quantitating the amount of incorporated label.

— NEW CITATIONS —

NONE

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US97/09664

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C12P 19/34; C12Q 1/68

US CL :435/91.1, 6

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/91.1, 6

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

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

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5,403, 708 A (BRENNAN et al.) 04 April 1995, see columns 2-9.	1-3, 5-9, 11
X	WO 90/13666 A1 (AMERSHAM INTERNATIONAL PLC) 15 November 1990, see pages 3-5, 7, 14, figure 1.	1-11
X	KHURANA et al. (CA) repeat polymorphism in the chromosome 18 encoded dystrophin-like protein. Human Molecular Genetics. 1994. Vol. 3, No. 5. page 841, see column 2.	12, 13
X	BORODOVSKY et al. Intrinsic and extrinsic approaches for detecting genes in a bacterial genome. Nucleic Acids Research. 1994. Vol. 22. No. 22. pages 4756-4767, especially pages 4756 and 4758, column 2.	12, 13



Further documents are listed in the continuation of Box C.



See patent family annex.

•	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reasons (as specified)	"Z"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

23 SEPTEMBER 1997

Date of mailing of the international search report

31 OCT 1997

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

AMY ATZEL, P.M.D.

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US97/09664

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 4,863,849 A (MELAMEDE, R. J.) 05 September 1989, see especially 6-10, 13 and claims	1-11

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/09664

B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

APS, INPADOC, CAPLUS, MEDLINE, BIOSIS, EMBASE, SCISEARCH, LIFESCI

search terms:

sequence, sequencing, DNA, nucleotide, nucleotides, nucleic acid, polynucleotide, algorithm, blast, search, Melamede, Garland