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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 January 2011 (13.01.2011)

PCT

(10) International Publication Number
WO 2011/004994 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01) *C07H 21/00* (2006.01)
C12N 15/11 (2006.01)

(21) International Application Number:
PCT/KR2010/004365

(22) International Filing Date:
6 July 2010 (06.07.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10-2009-0061895 7 July 2009 (07.07.2009) KR

(71) Applicant *(for all designated States except US)*: **KOREA RESEARCH INSTITUTE OF BIOSCIENCE AND BIOTECHNOLOGY** [KR/KR]; 52, Eoeun-dong, Yuseong-gu, Daejeon 305-333 (KR).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **SOHN, Jung-Hoon** [KR/KR]; 614-503, Woorim Fill U 2nd APT., Gakri, Ochang-eup, Cheongwon-gun, Chungcheongbuk-do 363-783 (KR). **BAE, Jung-Hoon** [KR/KR]; No 539-3, Noeun-dong, Yuseong-gu, Daejeon 305-325 (KR).

(74) Agent: **SON, Min**; HANOL Intellectual Property & Law, 6th Floor, City Air Tower, 159-9 Samseong-dong, Gangnam-gu, Seoul 135-973 (KR).

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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

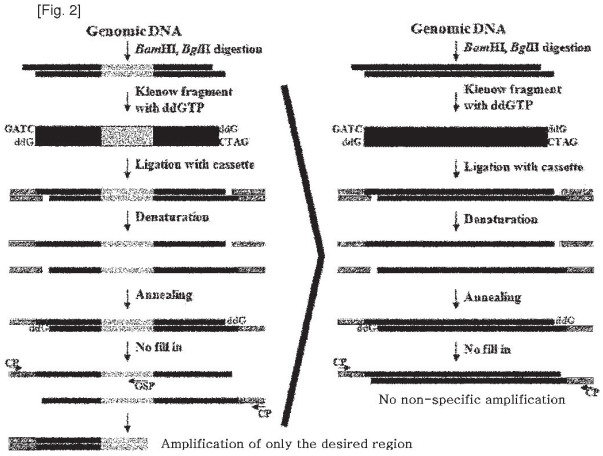
Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

— *with sequence listing part of description (Rule 5.2(a))*

(54) Title: A NOVEL GENOME WALKING METHOD FOR CLONING OF UNKNOWN DNA SEQUENCES ADJACENT TO KNOWN SEQUENCE

[Fig. 2]



(57) Abstract: The present invention relates to a novel genome walking method for cloning unknown genomic regions adjacent to known sequences by PCR. More particularly, the present invention relates to a method for cloning an unknown nucleotide sequence adjacent to a known nucleotide sequence, comprising the steps of 1) cleaving a genomic DNA that contains a known nucleotide sequence and an unknown nucleotide sequence adjacent thereto using a restriction enzyme; 2) modifying the 3'-OH end of the cleaved DNA with a nucleotide analogue; 3) linking the terminal sequence of the cleaved DNA fragment with a linker, adapter or cassette having a terminal sequence complementary thereto; and 4) amplifying the unknown nucleotide sequence using a known nucleotide sequence-specific primer and a cassette sequence-specific primer. Further, the present invention relates to a composition of gene amplification for cloning the unknown nucleotide sequence, and a gene amplification kit comprising the composition.

WO 2011/004994 A2

Description

Title of Invention: A NOVEL GENOME WALKING METHOD FOR CLONING OF UNKNOWN DNA SEQUENCES ADJACENT TO KNOWN SEQUENCE

Technical Field

[1] The present invention relates to a novel genome walking method for cloning unknown genomic regions adjacent to known sequences by PCR. More particularly, the present invention relates to a method for cloning an unknown nucleotide sequence adjacent to a known nucleotide sequence, comprising the steps of 1) cleaving a genomic DNA that contains a known nucleotide sequence and an unknown nucleotide sequence adjacent thereto using a restriction enzyme; 2) modifying the 3'-OH end of the cleaved DNA with a nucleotide analogue; 3) linking the terminal sequence of the cleaved DNA fragment with a linker, adapter or cassette having a terminal sequence complementary thereto; and 4) amplifying the unknown nucleotide sequence using a known nucleotide sequence-specific primer and a cassette sequence-specific primer. Further, the present invention relates to a composition of gene amplification for cloning the unknown nucleotide sequence, and a gene amplification kit comprising the composition.

[2]

Background Art

[3] Genome walking refers to a methodology used to clone unknown genomic regions adjacent to known sequences, and is also called chromosome walking.

[4] Until now, many methods have been developed for genomic walking. The traditional approach consists of the construction and screening of genomic DNA libraries using the known base sequence as a probe. Before the development of Polymerase Chain Reaction (PCR), this method was most widely used, but has disadvantages in that libraries should be constructed and a great deal of manpower and time are required during the screening process.

[5] The most commonly used method for cloning desired regions using PCR is Inverse PCR, where genomic DNA fragments are ligated into circles and the primers designed to recognize the known region in a reverse direction are used to perform amplification (Triglia et al., Nucleic Acid Res. 1988, 16, 8186). One of the drawbacks of Inverse PCR is the requirement for suitable restriction enzyme sites that flank the priming region.

[6] Another method is vectorette PCR, where genomic DNA treated with restriction enzymes is ligated to a cassette containing a mismatch region and then PCR is

performed (Riley et al., Nucleic Acid Res. 1990, 18, 2887), and various types of linker- or adapter-ligated PCR have been developed (Isegawa et al., Mol. Cell. Probe. 1992, 6, 467; Siebert et al., Nucleic Acid Res. 1995, 23, 1087; Kilstrup and Kristialsen, Nucleic Acid Res. 2000, 28, e55; Nthangeni et al., J. Microbiol. Methods. 2005, 61, 225).

[7] To improve these methods, methods of performing secondary PCR using blocked adapters (Padegimas and Reichert, Anal. Biochem 1998, 260, 149; Reddy et al., Mol. Biotechnol. 2002, 22, 223), biotinylated primers (Rosenthal and Jones, Nucleic Acid Res. 1990, 18, 3095) or nested primers were developed (Rosenthal, Trends. Biotechnol. 1992, 10, 44). In addition, a method of cloning desired regions using random primers without cassette ligation has been developed (Parker et al., Nucleic Acid Res. 1991, 19, 3055; Park et al., Nucleic Acid Res. 1991, 19, 7155; Domingue and Lopez-Larrea, Nucleic Acid Res. 1994, 22, 3247; Liu and Whittier, Genomics, 1995, 25, 674; Trueba and Johnson, Biotechniques, 1996, 21, 20; Tamme et al., Biotechniques, 2000, 28, 895; Karlyshev et al., Biotechniques, 2000, 28, 1078).

[8] As such, many studies have been made to improve the genome walking method, and thus significant progress has been made. However, it is still in need of improvement. The most serious problems of the PCR-based genome walking methods are low amplification efficiency and non-specific amplification. That is, unwanted genes are amplified together, which makes it difficult to distinguish amplification of desired genes. To improve this problem, complicated multi-step methods were developed, including secondary PCR, purification of desired PCR products using biotinylated primers, or improvement of PCR cycle conditions. However, satisfactory performance has not yet been achieved.

[9]

Disclosure of Invention

Technical Problem

[10] Therefore, the present inventors have developed a method capable of notably improving upon the above problems. A low amplification efficiency is caused by insufficient ligation between the restricted genomic DNA and cassette, and non-specific amplification is attributed to PCR reaction by only cassette primers in stead of a combination of the known region-specific primers and cassette primers. Thus, to amplify the desired region only, amplification by only cassette primers should be inhibited, and amplification by a combination of the known region-specific primers and cassette primers should only be performed. By the present inventors, the restricted genomic DNA fragment was treated with ddNTP, and ligated with a complementary cassette to increase the efficiency of ligation between the genome and cassette and to prevent non-specific amplification, thereby completing the present invention.

[11]

Solution to Problem

[12] It is an object of the present invention to provide a method for cloning an unknown nucleotide sequence adjacent to a known nucleotide sequence, comprising the steps of 1) cleaving a genomic DNA that contains a known nucleotide sequence and an unknown nucleotide sequence adjacent thereto using a restriction enzyme; 2) modifying the 3'-OH end of the cleaved DNA with a nucleotide analogue; 3) linking the terminal sequence of the cleaved DNA fragment with a linker, adapter or cassette having a terminal sequence complementary thereto; and 4) amplifying the unknown nucleotide sequence using a known nucleotide sequence-specific primer and a cassette sequence-specific primer.

[13] It is another object of the present invention to provide a composition of gene amplification for cloning an unknown nucleotide sequence, in which the composition of gene amplification comprises 1) a DNA construct containing a DNA fragment that contains a known nucleotide sequence and an unknown nucleotide sequence adjacent thereto and is modified with a nucleotide analogue at the 3'-OH end, and a linker, adapter or cassette having a terminal sequence complementary to that of the modified DNA fragment, 2) a primer specific to the known nucleotide sequence, and 3) a primer specific to the cassette sequence.

[14] It is still another object of the present invention to provide a kit comprising the composition of gene amplification.

Advantageous Effects of Invention

[15] As described above, the genome walking method of the present invention minimizes non-specific amplification and maximizes ligation efficiency of the cassette, and thereby cloning the desired genomic regions can be done by a single PCR reaction without additional experiment. Therefore, the method can be very usefully applied to genome walking technology.

[16]

Brief Description of Drawings

[17] FIG. 1 is a diagram showing a genome walking method using the general cassette-ligation mediated PCR, in which the mark "<" in the drawing represents that the majority of PCR products occurred by non-specific amplification;

[18] FIG. 2 is a diagram showing a genome walking method improved by non-specific amplification-inhibiting PCR, in which the mark ">" in the drawing represents that the majority of PCR products occurred by specific amplification; and

[19] FIG. 3 is the result of testing the genome walking method using the known genes, in which

[20] Left shows cassette-ligated genes (URA3 and pBluescript II) and Right is a photograph showing the result of agarose gel electrophoresis after PCR,

[21] Lane 1: 1 kb DNA ladder

[22] Lane 2: Result of PCR using CP (cassette primer) and GSP (gene-specific primer), after treating 0.5 μg of BamHI-treated pBluescript II and 10 ng of URA3 BamHI fragment with ddGTP, and ligating them with 40 pmol of a cassette

[23] Lane 3: Result of PCR using CP and GSP, after treating 0.5 μg of BamHI-treated pBluescript II and 1 ng of URA3 BamHI fragment with ddGTP, and ligating them with 40 pmol of a cassette

[24] Lane 4: Result of PCR using CP and GSP, after treating 0.5 μg of BamHI-treated pBluescript II and 0.1 ng of URA3 BamHI fragment with ddGTP, and ligating them with 40 pmol of a cassette

[25] Lane 5: Result of PCR using CP and GSP, after treating 0.5 μg of BamHI-treated pBluescript II and 0.01 ng of URA3 BamHI fragment with ddGTP, and ligating them with 40 pmol of a cassette

[26] Lane 6: Result of PCR using CP and GSP, after treating 0.5 μg of BamHI-treated pBluescript II and 10 ng of URA3 BamHI fragment with dGTP, and ligating them with 40 pmol of a cassette

[27] Lane 7: Result of PCR using CP and GSP, after treating 0.5 μg of BamHI-treated pBluescript II and 1 ng of URA3 BamHI fragment with dGTP, and ligating them with 40 pmol of a cassette

[28] Lane 8: Result of PCR using CP and GSP, after treating 0.5 μg of BamHI-treated pBluescript II and 0.1 ng of URA3 BamHI fragment with dGTP, and ligating them with 40 pmol of a cassette

[29] Lane 9: Result of PCR using CP and GSP, after treating 0.5 μg of BamHI-treated pBluescript II and 0.01 ng of URA3 BamHI fragment with ddGTP, and ligating them with 40 pmol of a cassette

[30] Lane 10: 1 kb DNA ladder; and

[31] FIG. 4 is a diagram showing the cloning of a novel cellulase gene from the fungal genome using the genome walking method, in which A and B show the structure of a cellulase gene and the position of primers linked to the cellulase gene, and C is a photograph showing the result of agarose gel electrophoresis after PCR,.

[32] Lane 1: 1 kb DNA ladder.

[33] Lane 2: PCR result of the BglII cassette library using CP primer and CBH1R primer

[34] Lane 3: PCR result of the BamHI cassette library using CP primer and CBH1F primer

[35] Lane 4: PCR result of the BglII cassette library using CP primer and CBH2R primer

[36] Lane 5: PCR result of the BamHI cassette library using CP primer and CBH2F

primer

[37] Lane 6: 1 kb DNA ladder.

[38]

Best Mode for Carrying out the Invention

[39] In one aspect to achieve the above objects, the present invention relates to a composition of gene amplification, comprising 1) a DNA construct containing a DNA fragment that contains a known nucleotide sequence and an unknown nucleotide sequence adjacent thereto and is modified with a nucleotide analogue at the 3'-OH end, and a linker, adapter or cassette having a terminal sequence complementary to that of the modified DNA fragment, 2) a primer specific to the known nucleotide sequence, and 3) a primer specific to the cassette sequence.

[40]

[41] In the known methods of cloning a desired region by ligation of a cassette to the terminal of genomic DNA, it can be expected that the lack of a phosphate group at the 5'-end of the cassette theoretically prevents non-specific amplification, because one strand of the double-stranded cassette is only ligated with the genomic DNA, and thus cassette primers do not bind to the genomic DNA ligated with the single strand cassette, as shown in FIG. 1. That is, first reaction occurs from the known region-specific primer so as to synthesize a new strand, to which a cassette primer binds, and the newly synthesized strand is used as a template to perform PCR, thereby amplifying a desired region.

[42] In the practical use of the method, however, unexpected non-specific amplification frequently occurs. Thus, even though secondary PCR is performed using a nested primer to clone the desired region, it is difficult to obtain the desired region, and many trials and errors are made to amplify the desired region. As shown in the Right of FIG. 1, the genomic DNA ligated with a single strand cassette is denatured, and then it forms double strands, and a single strand cassette is regenerated to a double-stranded one by DNA polymerase, resulting in non-specific amplification. Since the regenerated double-stranded gene has the double-stranded cassette at both ends thereof, non-specific amplification can be generated by the cassette primer only, irrespective of the known region-specific primer. A target region to be cloned in the genome accounts for only one in several thousands to several ten thousands of the entire genome, and thus non-specific amplification occurs frequently several ten thousand times more than specific amplification. Therefore, it is difficult to detect specific amplification by agarose gel electrophoresis. For successful development of the genome walking method, it is important to prevent non-specific amplification.

[43] Therefore, the present inventors provide a method capable of minimizing non-

specific amplification by modification of the end of a restricted genomic DNA, and a composition of gene amplification.

[44] As used herein, the term 'nucleotide' refers to a deoxyribonucleotide or ribonucleotide in either single or double-stranded form. The term 'known nucleotide' refers to a nucleotide, of which base sequence is known, and is opposite to unknown nucleotide, of which base is not known so as to be a subject of cloning.

[45] Further, the term 'nucleotide analogue' refers to a natural or non-natural occurring analogue being structurally similar to a nucleotide. Herein, the nucleotide analogue is a material that is used not to yield a protruding 3'-OH end by modification at the 3'-end of DNA fragment. Preferably, examples of the nucleotide analogue include, but are not limited to, dideoxynucleotide triphosphate (ddNTP) including ddATP (dideoxyadenosine triphosphate), ddCTP (dideoxycytosine triphosphate), ddGTP (dideoxyguanosine triphosphate), ddTTP (dideoxythymidine triphosphate), ddUTP (dideoxyuridine triphosphate) and ddITP (dideoxyinosine triphosphate). When ddNTP is linked to the end of DNA fragment, further elongation is blocked because ddNTPs lack 3'-OH.

[46] In the known method, as described in FIG. 1, the genomic DNA ligated with a single strand cassette undergoes denaturation/annealing, and then the 3'-OH end thereof is filled by DNA polymerase, thereby producing double-stranded DNA where cassette primers bind. Thus, non-specific amplification occurs frequently. Accordingly, as in the present invention, if the 3'-OH end of genomic DNA is modified with ddNTP for the prevention of further polymerization, non-specific amplification can be remarkably reduced.

[47] As used herein, the term 'cassette' encompasses short double-stranded oligonucleotides that are synthesized to be artificially ligated to the DNA fragment. Because of the lack of a phosphate group at the 5'-end of the cassette, a single strand of double-stranded cassette is only ligated with the genomic DNA.

[48] Similarly, the term 'linker' refers to a short double-stranded DNA synthesized in vitro. Although the linker has blunt ends, but also a restriction enzyme recognition site, both ends thereof can have cohesive ends by treatment of the restriction enzyme even after ligation.

[49] In addition, the term adapter refers to a short double-stranded oligonucleotide, like the linker. However, since the adapter has one blunt end and one cohesive end unlike the linker, it binds to the cohesive end of restricted DNA fragment to form a new DNA molecule having blunt ends.

[50] As used herein, the term 'DNA construct' refers to a naturally occurring or synthesized nucleic acid construct that contains nucleic acid elements necessary for the expression of structural genes in the host cell. In the present invention, DNA construct

refers to a recombinant DNA fragment comprising a DNA fragment that contains a known nucleotide and an unknown nucleotide adjacent thereto and that is modified with a nucleotide analogue at the 3'-OH end, and a linker, adapter or cassette having a terminal sequence complementary to that of the modified DNA fragment.

[51] As used herein, the term 'primer' refers to a short single-stranded nucleic acid sequence having a free 3' hydroxyl group, which is able to undergo base-pairing interaction with a complementary template and serves as a starting point for replicating the template strand. A primer is able to initiate DNA synthesis in the presence of a reagent for polymerization (e.g., DNA polymerase or reverse transcriptase) and four different dNTPs in suitable buffers and at a suitable temperature. A primer specific to the known nucleotide sequence and a primer specific to the cassette sequence are included in the composition of gene amplification of the present invention.

[52]

[53] Using the composition of gene amplification, the present inventors had designed a new genome walking model, where non-specific amplification is inhibited, as shown in FIG. 2.

[54] In still another aspect, the present invention relates to a method for cloning an unknown nucleotide sequence adjacent to a known nucleotide sequence, comprising the steps of:

[55] 1) cleaving a genomic DNA that contains a known nucleotide sequence and an unknown nucleotide sequence adjacent thereto using a restriction enzyme;

[56] 2) modifying the 3'-OH end of the cleaved DNA with a nucleotide analogue;

[57] 3) linking the terminal sequence of the cleaved DNA fragment with a linker, adapter or cassette having a terminal sequence complementary thereto; and

[58] 4) amplifying the unknown nucleotide sequence using a known nucleotide sequence-specific primer and a cassette sequence-specific primer.

[59]

[60] Step 1) is a step of cleaving a genomic DNA using a restriction enzyme to prepare a DNA fragment containing a known nucleotide and an unknown nucleotide adjacent thereto.

[61] As used herein, the term 'restriction enzymeis' an enzyme that recognizes and cleaves a specific base sequence of DNA, and there are restriction enzyme generating blunt or flush ends and restriction enzyme generating sticky or cohesive ends. In the present invention, it is preferable to use restriction enzyme generating cohesive ends, in order to maximize the efficiency of ligation between the cleaved DNA fragment and cassette.

[62] For example, genomic DNA cleaved by the restriction enzyme, BamHI, Bgl II or Sau3A has protruding GATC sequences at its 5' end. In addition, the genomic DNA

cleaved by Hind has protruding AGTC sequences at its 5' end. The sequences of cohesive ends generated by restriction enzymes are widely known in the art, and thus, this information is used to prepare a DNA fragment having a desired end sequence.

[63] Step 2) is a step of modifying the 3'-OH end of the restricted DNA with a nucleotide analogue. In the present invention, the 3'-OH end of genomic DNA is modified with ddNTP for the prevention of further polymerization, non-specific amplification can be remarkably reduced. For example, when the genomic DNA fragment having protruding GATC sequences at its 5' end by treatment of the restriction enzyme BamHI or Bgl II is reacted with a DNA polymerase Klenow fragment together with dideoxy GTP (ddGTP), ddGTP is inserted into each 3' end of the double-stranded DNA fragment. Further elongation is blocked at each 3' end, because ddGTP lacks 3'-OH needed for DNA polymerization. Consequently, a cassette primer cannot bind thereto, thereby minimizing non-specific amplification.

[64] Step 3) is a step of linking the DNA fragment, which is modified with the nucleotide analogue, with a linker, adapter or cassette having a terminal sequence complementary thereto.

[65] In the case where a DNA fragment is treated with a restriction enzyme so as to have cohesive ends, the cohesive ends of cassette are designed to have a sequence complementary to those of the DNA fragment, and they are ligated with each other to give a new DNA molecule having cassettes at its both ends. This ligation between the genomic DNA and complementary cassette can improve the efficiency of ligation between the genomic DNA and cassette. For example, when the genomic DNA is cleaved with the restriction enzyme BamHI, Bgl II or Sau3A, and modified with ddGTP so as to have protruding GAT sequences at each 5' end of double strand, a cassette is designed to have a protruding ATC sequences at its 5' end and one blunt end at the other end, so that it can be ligated with both ends of the double-stranded genomic DNA fragment. Finally, the cassette is able to complementarily bind to the end of genomic DNA. It is also applicable to a linker or adapter.

[66] Further, the efficiency of ligation between the genomic DNA and cassette is more improved by modification of the 3' end of genomic DNA with ddNTP. That is, if the ends of restricted DNA are partially filled with 1 to 3 bases by DNA polymerase, it can be prevented that the cleaved genes are reconnected to form a circle or multiple binding, because the ends are not complementary to each other. In the same way, the cassette, which is complementary to the ends of a gene partially filled with bases, cannot be connected to each other. Therefore, the cassette ligation method provided in the present invention minimizes unwanted reaction, and is able to ligate only between the cleaved DNA fragment and cassette at 1:1, thereby maximizing the ligation efficiency.

[67] Step 4) is a step of amplifying the unknown nucleotide sequence using a known nucleotide sequence-specific primer and a cassette sequence-specific primer to clone the unknown gene.

[68] When the known region-specific primer binds to the known region, as in the Left in FIG. 2, a new DNA strand, to which the cassette primer binds, is generated from the known region-specific primer, and therefore, a desired region can be amplified without influence of ddGTP treatment. In addition, without complicated multistep PCR, simultaneous treatment of the known region-specific primer and cassette-specific primer is performed to readily amplify the desired region by a single step PCR.

[69] In the present invention, the amplification of nucleotide sequence may be performed by Polymerase Chain Reaction (PCR). PCR is a technique in molecular biology to selectively amplify only a specific DNA fragment in a trace amount of DNA solution, and calls for 30-40 cycles that have three steps of Denaturation, Annealing, and Extension. The type of PCR includes Multiplex-PCR, Nested PCR, Quantitative PCR, real-time PCR, RT-PCR, Touchdown PCR or the like, and any suitable type can be used in the present invention without limitation.

[70]

[71] In still another aspect, the present invention relates to a gene amplification kit comprising the composition of gene amplification.

[72] Preferably, the kit comprises a DNA construct containing a DNA fragment that contains a known nucleotide sequence and an unknown nucleotide sequence adjacent thereto and is modified with a nucleotide analogue at the 3'-OH end, and a linker, adapter or cassette having a terminal sequence complementary to that of the modified DNA fragment, a primer specific to the known nucleotide sequence, and a primer specific to the cassette sequence, and further comprises a test tube or other proper container, a reaction buffer solution, deoxynucleotides (dNTP), enzymes such as Taq polymerase, sterilized water or the like for the amplification of genomic DNA fragment.

[73]

[74] In the specific embodiment of the present invention, prior to application of the method of the present invention to the experiment of cloning a specific region from genomic DNA, a plasmid and a gene, of which base sequence has been known, are used to examine its effect. Specifically, to use yeast URA3 gene as a target for specific amplification, 2.5 kb of BamHI-restriction fragment containing the yeast URA3 gene was used, and to confirm the degree of non-specific amplification, 3 kb of BamHI-restricted pBluescript II plasmid containing no URA3 gene was used. They were mixed with each other at a proper concentration ratio, and used as PCR template. Under the same conditions, the 3' end of each of them was filled-in with one com-

plementary base using ddGTP or using dGTP as a control group. A cassette having a terminal sequence complementary to that of each fragment is ligated at both ends of the fragment, and then the cassette primer and URA3-specific primer were used to perform PCR.

[75] As shown in FIG. 3, in the dGTP-treated sample, it was found that 3kb of pBluescript II was amplified (non-specific amplification), irrespective of the concentration ratio of PCR template, and URA3-specific amplification (2.2 kb) occurred at certain concentration ratios. In contrast, in the ddGTP-treated sample, it was found that non-specific amplification did not occur at all concentration ratios, and URA3-specific amplification only occurred.

[76] To practically apply the method of the present invention to the experiment of cloning a specific region from genomic DNA, the fungus Penicillium sp. cf4 strain having cellulase activity was identified, and a novel cellulase gene was cloned therefrom. That is, based on the cloned cellulase gene sequence, 11 pairs of degenerate primers were designed to perform PCR of the genomic DNA of the strain. 0.4-kb and 0.45-kb PCR products were obtained, and used as the known regions. Finally, two cellulase genes (PeCBH1 and PeCBH2) containing an unknown sequence adjacent thereto were cloned according to the present invention.

[77] As such, the method of the present invention is advantageous in that it is used to increase the efficiency of ligation between genomic DNA and cassette and inhibit non-specific amplification, thereby readily cloning the desired genomic regions by a single PCR reaction without complicated multistep PCR, unlike the known genome walking methods.

[78]

Mode for the Invention

[79] Hereinafter, the present invention will be described in detail with reference to Examples. However, these Examples are for illustrative purposes only, and the invention is not intended to be limited thereby.

[80]

[81] Example 1: Cassette construction

[82] To examine and confirm the method of the present invention, a double-stranded cassette was first constructed. In order to ligate with genomic DNA which was digested with BamHI, BglII or Sau3A and had one base filled-in with ddGTP, a CSF primer of SEQ ID NO. 1 (GACGCGTAATACGACTCACTATAGGGA) and a CSR primer of SEQ ID NO. 2 (ATCTCCCTATAGTGAGTCGTATTACGCGTC) were synthesized. 40 ㎕ of 100 μM CSF primer and 100 μM CSR primer were mixed with each other, and then 250 mM Tris.Cl (pH 8.0), 10 ㎕ of 100 mM MgCl2 buffer, and 10

μℓ of distilled water were added, followed by heating in a water bath at 95 ℃ for 5 min. The water bath was left to cool to room temperature. The cassette constructed by this method had one blunt end and a protruding ATC sequences at its 5'-end.

[83]

[84] Example 2: Test on genome walking method using the known genes

[85] Prior to application of the method of the present invention to the experiment of cloning a specific region from genomic DNA, the method was tested using a plasmid and a gene, of which base sequence has been known. As shown in FIG. 3, 2.5 kb of BamHI-restriction fragment containing the URA3 gene and 3 kb of BamHI-restricted pBluescript II plasmid were used to examine the effect of terminal modification with ddGTP.

[86] 0.5 μg of BamHI-restricted pBluescript II plasmid used as a non-specific DNA and 10 ng, 1 ng, 0.1 ng, and 0.01 ng of URA3 BamHI fragment used as a target gene were mixed, respectively. 1 μℓ of 2 mM ddGTP and 2 μℓ of 10X Klenow fragment buffer solution [100 mM Tris.Cl(pH7.5), 70 mM MgCl2, 1 mM DTT] were added, and then distilled water was added to a final volume of 20 μℓ. Subsequently, 5 units of Klenow fragment (Takara) were added, and reacted at 37 ℃ for 30 min. For comparison, dGTP was used instead of ddGTP to prepare a control group under the same conditions. DNAs were purified using a PCR purification kit (Axygen, USA) and ligated with 40 pmol of the cassette using a DNA Ligation kit (Takara). The ligation was performed at 16 ℃ for 30 min, and each DNA was purified using the PCR purification kit (Axygen), and then eluted using 10 μℓ of distilled water. 1 μℓ of each eluted sample was used as a template to perform PCR using a cassette primer (CP) of SEQ ID NO. 3 (ACGCGTAATACGACTCACTATAGGGAGATC) and a URA3 gene-specific primer (GSP). The reaction composition was as follows: [1 μℓ of cassette-ligated DNA, 1 pmol of CP, 1 pmol of GSP, 5 μℓ of EX taq buffer (Takara), 5 μℓ of 2.5 mM dNTP, 37 μℓ of distilled water, EX taq DNA polymerase(Takara) 1]. PCR was performed under the conditions of 95 ℃ 5 min, 30 cycles of 95 ℃ 30 sec, 60 ℃ 30 sec, and 72 ℃ 3min, and 72 ℃ 10 min. After reaction was terminated, 10 μℓ of the reaction solution was analyzed by agarose gel electrophoresis.

[87] As shown in FIG. 3, 3kb of pBluescript II was amplified (non-specific amplification) in the dGTP-treated samples (Lanes 6-9), and in contrast, non-specific amplification did not occur in the ddGTP-treated samples (Lanes 2-5). In addition, a 2.2 kb band corresponding to target-specific amplification was detected from 0.01 ng of target gene (1/5000 of pBluescript II) in the ddGTP-treated samples (Lane 4). Therefore, it can be seen that the method of the present invention minimizes non-specific amplification and maximizes specific amplification, thereby being very useful for genome walking.

[88]

[89] Example 3: Cellulase gene cloning using genome walking method

[90] In order to clone a novel cellulase gene, the present inventors selected the fungus having cellulase activity, and identified by sequence analysis a 26 S ribosomal RNA gene, resulting in the identification of a Penicillium sp. cf4 strain belonging to the genus Penicillium. Based on the cloned cellulase gene sequence, 11 pairs of degenerate primers were designed in order to clone a cellulase gene from the strain. The genomic DNA of the strain was subjected to PCR, and 0.4-kb and 0.45-kb PCR products containing a part of cellulase gene were obtained. Each PCR product was cloned into a pGEM-T Easy vector (Invitrogen), followed by sequence analysis. As a result, both of the PCR products were a part of the exo-glucanase gene, and identical sequences have not been discovered in any genes reported to date. Thus, since there was a high probability that the cellulase gene is a novel gene, a genome cassette library was constructed using other regions of the gene by the method of the present invention, and cloned by the genome walking method.

[91] Penicillium sp. cf4 genomic DNA 2 was reacted with 10 units of BamHI and BglII in the presence of 1X BamHI buffer solution (Takara) and 1X BglII buffer solution (Takara) at 37 ℃ for 2 hrs, and agarose gel electrophoresis was performed to confirm complete cleavage. The reactant was treated with in a 70 ℃ water bath for 20 min to inactivate the restriction enzymes, and then ddGTP was added to be 0.2 mM, and reacted with 8 units of klenow fragment (Takara) at 37 ℃ for 30 min. DNA was purified using a PCR purification kit (Axygen), and ligated with 40 pmol of a cassette using a DNA ligation kit (Takara). The ligation was performed at 16 ℃ for 10 hrs, and then DNA was further purified using the PCR purification kit (Axygen), and eluted using 20 ㎕ of distilled water to construct a BamHI cassette library and a BglII cassette library, respectively.

[92] To clone the first cellulase gene, as shown in FIG. 4A, 25 bases of cellulase-specific primers (CBH1F, CBH1R) were designed, based on the base sequence of 0.4 kb PCR product (known region I), and PCR was performed using the cassette library 1 ㎕ as a template and the cassette primer CP (SEQ ID NO. 3). The reaction composition for cloning the 5 ㎕ region of cellulase gene was as follows: [1 ㎕ of genome cassette library, 1 pmol of CP, 1 pmol of CBH1R, 5 ㎕ of EX taq buffer (Takara), 5 ㎕ of 2.5 mM dNTP, 37 ㎕ of distilled water, 1 ㎕ of EX taq DNA polymerase(Takara)]. To clone the 3' region of cellulase gene, the CBH1F primer was used instead of the CBH1R primer. PCR was performed under the conditions of 95℃ 5 min, 30 cycles of 95℃ 30 sec, 60℃ 30 sec, and 72℃ 3min, and 72℃ 10 min. After reaction was terminated, 10 ㎕ of the reaction solution was analyzed by agarose gel electrophoresis. As a result, a 2.6 kb fragment was amplified from the BglII cassette library in the cloning of 5' region, and a 3.7 kb fragment was amplified from the BamHI cassette

library in the cloning of 3' region (FIG. 4C, lanes 2 and 3). Each PCR product was cloned into a pGEM-T Easy vector (Invitrogen), and sequence analysis was performed. As a result, it was confirmed that the desired region was amplified, and the entirety of the sequences of the cellulase gene were identified. The cellulase gene (PeCBH1) derived from Penicillium sp. cf4, cloned by the above method, was found to be exo-cellobiohydrolase consisting of 547 amino acids including no intron, 26 secretion signals, and cellulose binding module (CBM) and showed 74.3% similarity to exo-cellobiohydrolase derived from Penicillium oxalicum.

[93] To clone the second cellulase gene, as shown in FIG. 4B, 25 bases of cellulase-specific primers (CBH2F, CBH2R) were designed, based on the base sequence of 0.45 kb PCR product (known region II), and PCR was performed in the same manner as above. A 1.5 kb fragment was amplified from the Bgl II cassette library in the cloning of 5' region, and 4.0 kb fragment was amplified from the BamHI cassette library in the cloning of 3' region (FIG. 4C, lanes 4 and 5). Each PCR product was cloned into a pGEM-T Easy vector (Invitrogen), and sequence analysis was performed. As a result, it was confirmed that the desired region was amplified, and the entire sequences of the second cellulase gene were identified. The second cellulase gene (PeCBH2) derived from Penicillium sp. cf4 was found to be exo-cellobiohydrolase consisting of 452 amino acids including two introns and 17 secretion signals, and showed 78.4% similarity to exo-cellobiohydrolase derived from Penicillium decumbens.

Claims

[Claim 1] A novel genome walking method for cloning an unknown nucleotide sequence adjacent to a known nucleotide sequence, comprising the steps of:
1) cleaving a genomic DNA that contains a known nucleotide sequence and an unknown nucleotide sequence adjacent thereto using a restriction enzyme;
2) modifying the 3'-OH end of the cleaved DNA with a nucleotide analogue;
3) linking the terminal sequence of the cleaved DNA fragment with a linker, adapter or cassette having a terminal sequence complementary thereto; and
4) amplifying the unknown nucleotide sequence using a known nucleotide sequence-specific primer and a cassette sequence-specific primer.

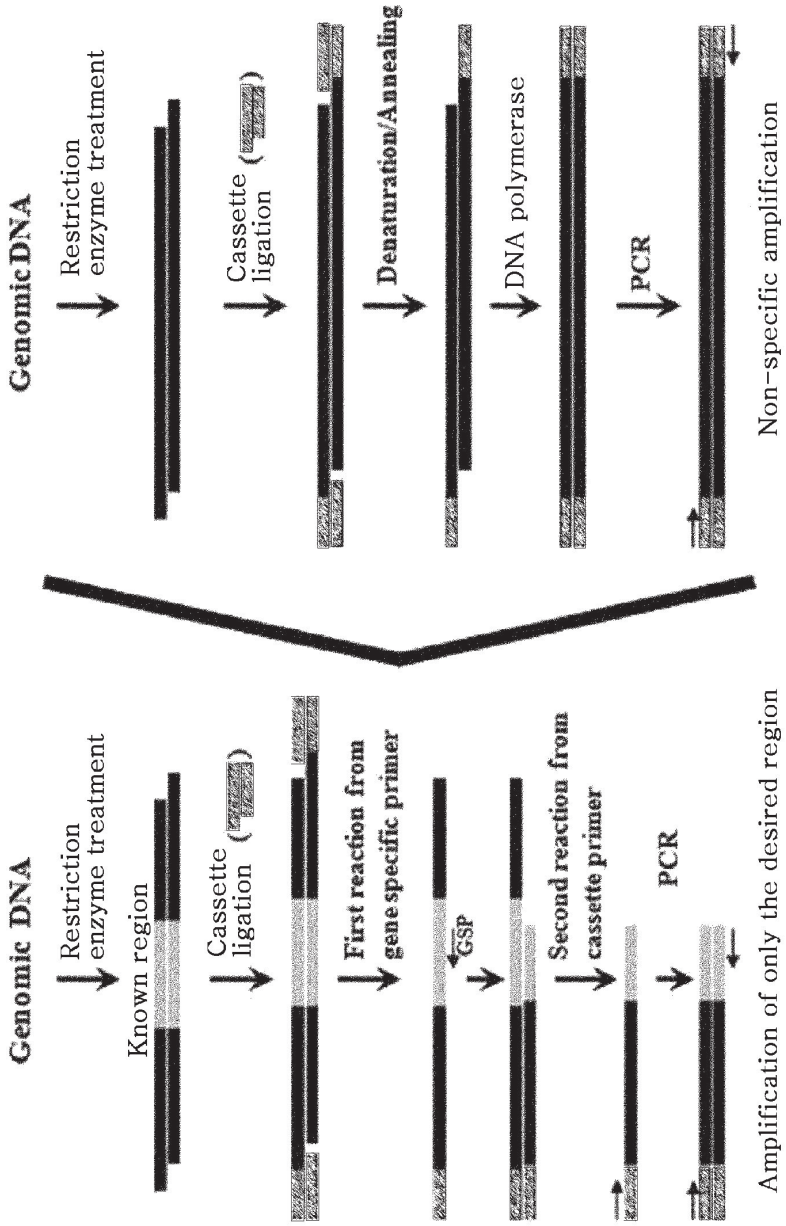
[Claim 2] The method according to claim 1, wherein the nucleotide analogue is ddNTP selected from the group consisting of ddATP, ddTTP, ddGTP, ddCTP, ddUTP, and ddITP.

[Claim 3] A composition of gene amplification, comprising 1) a DNA construct containing a DNA fragment that contains a known nucleotide sequence and an unknown nucleotide sequence adjacent thereto and is modified with a nucleotide analogue at the 3'-OH end, and a linker, adapter or cassette having a terminal sequence complementary to that of the modified DNA fragment, 2) a primer specific to the known nucleotide sequence, and 3) a primer specific to the cassette sequence.

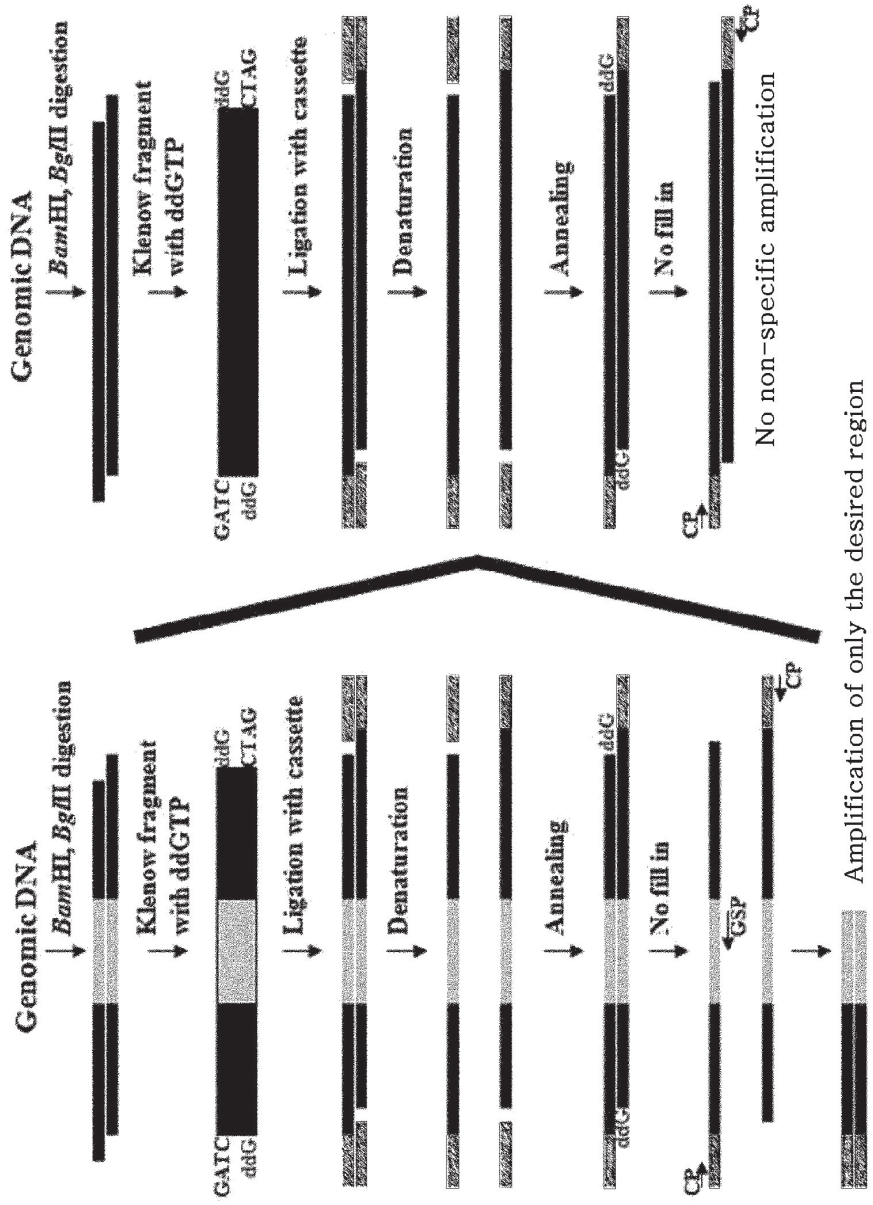
[Claim 4] The composition according to claim 3, wherein the nucleotide analogue is ddNTP selected from the group consisting of ddATP, ddTTP, ddGTP, ddCTP, ddUTP, and ddITP.

[Claim 5] A gene amplification kit comprising the composition of claim 3.

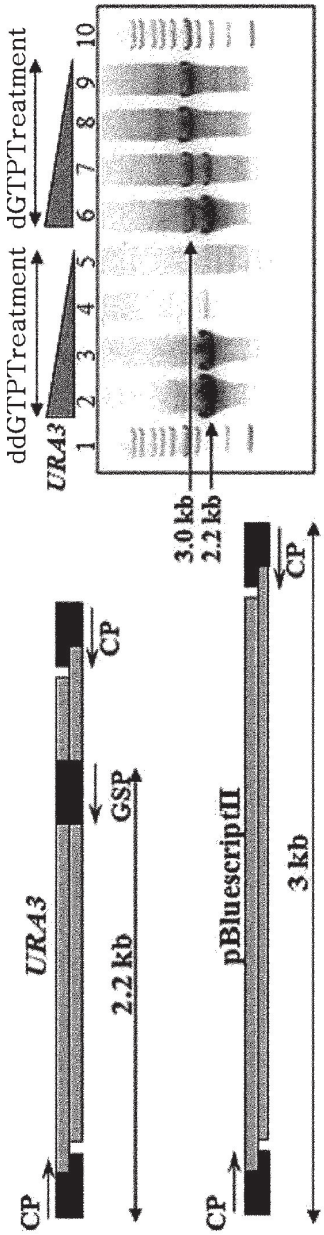
[Fig. 1]


Genomic DNA
Restriction enzyme treatment
Cassette ligation
Denaturation/Annealing
DNA polymerase
PCR
Non-specific amplification
Genomic DNA
Restriction enzyme treatment
Known region
Cassette ligation
First reaction from gene specific primer
GSP
Second reaction from cassette primer
PCR
Amplification of only the desired region

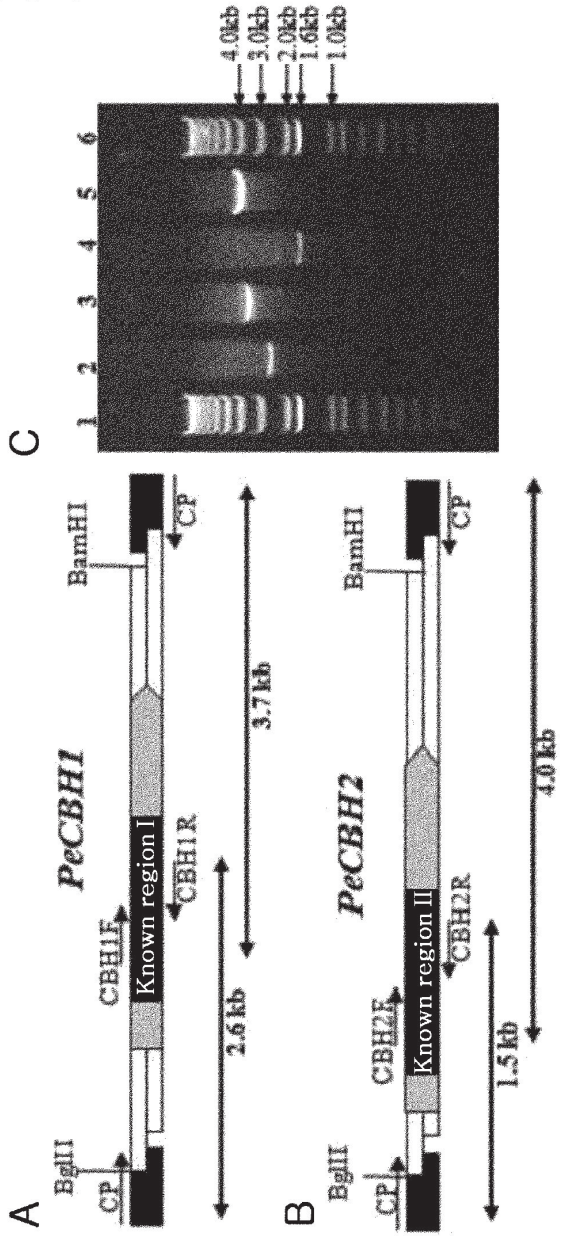
[Fig. 2]


Genomic DNA
BamHI, BglII digestion
Klenow fragment with ddGTP
GATC
ddG
ddG
CTAG
Ligation with cassette
Denaturation
Annealing
ddC
ddG
No fill in
CP
GSP
CP
Amplification of only the desired region
Genomic DNA
BamHI, BglII digestion
Klenow fragment with ddGTP
GATC
ddG
ddG
CTAG
Ligation with cassette
Denaturation
Annealing
ddC
ddG
No fill in
CP
CP
No non-specific amplification

[Fig. 3]


ddGTPTreatment
dGTPTreatment
URA3
1
2
3
4
5
6
7
8
9
10
3.0 kb
2.2 kb
CP
URA3
GSP
2.2 kb
CP
pBluescriptII
3 kb
CP

[Fig. 4]


A
BglII
CP
PeCBH1
CBH1F
Known region I
CBH1R
BamHI
CP
2.6 kb
3.7 kb
B
BglII
CP
PeCBH2
CBH2F
Known region II
CBH2R
BamHI
CP
1.5 kb
4.0 kb
C
1
2
3
4
5
6
4.0kb
3.0kb
2.0kb
1.6kb
1.0kb