and societal consequences of the Human Genome Project. N. Engl. J. Med. *341*:28-37.
8. **Cox, D., C. Boillot, and F. Canzian.** 2001. Data mining: efficiency of using sequence databases for polymorphism discovery. Hum. Mutat. *17*:141-150.
9. **Gallagher, G., J. Eskdale, H.H. Oh, S.D. Richards, D.A. Campbell, and M. Field.** 1997. Polymorphisms in the TNF gene cluster and MHC serotypes in the West of Scotland. Immunogenetics *45*:188-194.
10. **Halushka, M.K., J.B. Fan, K. Bentley, L. Hsie, N. Shen, A. Weder, R. Cooper, R. Lipshutz, et al.** 1999. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. Nat. Genet. *22*:239-247.
11. **Hattori, M., S. Nanko, X.Y. Dai, R. Fukuda, and H. Kazamatsuri.** 1994. Mismatch PCR RFLP detection of DRD2 Ser311Cys polymorphism and schizophrenia. Biochem. Biophys. Res. Commun. *202*:757-763.
12. **Hiratsuka, M., T. Inoue, F. Omori, Y. Agatsuma, Y. Kishikawa, and M. Mizugaki.** 2000. Detection assay of rare variants of the *thiopurine methyltransferase* gene by PCR-RFLP using a mismatch primer in a Japanese population. Biol. Pharm. Bull. *23*:1090-1093.
13. **Irizarry, K., V. Kustanovich, C. Li, N. Brown, S. Nelson, W. Wong, and C.J. Lee.** 2000. Genome-wide analysis of single-nucleotide polymorphisms in human expressed sequences. Nat. Genet. *26*:233-236.
14. **Jurinke, C., D. van den Boom, C.R. Cantor, and H. Koster.** 2001. Automated genotyping using the DNA MassArray technology. Methods Mol. Biol. *170*:103-116.
15. **Kalinina, O., I. Lebedeva, J. Brown, and J. Silver.** 1997. Nanoliter scale PCR with TaqMan detection. Nucleic Acids Res. *25*:1999-2004.
16. **Kato, R., Y. Yamazoe, and T. Yasumori.** 1990. Pharmacogenetics and polymorphism of human P-450 which activates and detoxicates xenobiotics and carcinogens. Princess Takamatsu Symp. *21*:45-53.
17. **Kruglyak, L.** 1997. The use of a genetic map of biallelic markers in linkage studies. Nat. Genet. *17*:21-24.
18. **Lee, S.G., S. Hong, Y. Yoon, I. Yang, and K. Song.** 2001. Characterization of publicly available SNPs in the Korean population. Hum. Mutat. *17*:281-284.
19. **Marth, G., R. Yeh, M. Minton, R. Donaldson, Q. Li, S. Duan, R. Davenport, R.D. Miller, and P.Y. Kwok.** 2001. Single-nucleotide polymorphisms in the public domain: how useful are they? Nat. Genet. *27*:371-372.
20. **Marth, G.T., I. Korf, M.D. Yandell, R.T. Yeh, Z. Gu, H. Zakeri, N.O. Stitziel, L. Hillier, et al.** 1999. A general approach to single-nucleotide polymorphism discovery. Nat. Genet. *23*:452-456.
21. **Nakashima, H., T. Otsuka, Y. Ohba, M. Akahoshi, S. Nagano, E. Ogami, Y. Arinobu, K. Miyake, et al.** 1999. Two polymorphisms within interleukin-3 (hIL3) gene detected by mismatch PCR/RFLP. Genes Immun. *1*:156-158.
22. **Neve, B., P. Froguel, L. Corset, E. Vaillant, V. Vatin, and P. Boutin.** 2002. Rapid SNP allele frequency determination in genomic DNA pools by Pyrosequencing™. BioTechniques *32*:1138-1142.
23. **Niu, T., C. Chen, J. Yang, B. Wang, Z. Wang, N. Schork, Z. Fang, and X. Xu.** 1999. Blood pressure and the T174M and M235T polymorphisms of the angiotensinogen gene. Ann. Epidemiol. *9*:245-253.
24. **Okayama, H., D.T. Curiel, M.L. Brantly, M.D. Holmes, and R.G. Crystal.** 1989. Rapid, nonradioactive detection of mutations in the human genome by allele-specific amplification. J. Lab. Clin. Med. *114*:105-113.
25. **Parsons, B.L. and R.H. Heflich.** 1997. Genotypic selection methods for the direct analysis of point mutations. Mutat. Res. *387*:97-121.
26. **Stephens, J.C., J.A. Schneider, D.A. Tanguay, J. Choi, T. Acharya, S.E. Stanley, R. Jiang, C.J. Messer, et al.** 2001. Haplotype variation and linkage disequilibrium in 313 human genes. Science *293*:489-493.
27. **Taillon-Miller, P., E.E. Piernot, and P.Y. Kwok.** 1999. Efficient approach to unique single-nucleotide polymorphism discovery. Genome Res. *9*:499-505.
28. **Wasson, J., G. Skolnick, L. Love-Gregory, and M.A. Permutt.** 2002. Assessing allele frequencies of single nucleotide polymorphisms in DNA pools by Pyrosequencing™ technology. BioTechniques *32*:1144-1152.
29. **Wicks, S.R., R.T. Yeh, W.R. Gish, R.H. Waterston, and R.H. Plasterk.** 2001. Rapid gene mapping in *Caenorhabditis elegans* using a high-density polymorphism map. Nat. Genet. *28*:160-164.
30. **Youssoufian, H., H.H. Kazazian, Jr., D.G. Phillips, S. Aronis, G. Tsiftis, V.A. Brown, and S.E. Antonarakis.** 1986. Recurrent mutations in haemophilia A give evidence for CpG mutation hotspots. Nature *324*:380-382.

We thank Dr. Mark Seielstad for careful reading of the manuscript. This study is supported in part by the National Institute of Arthritis and Musculoskeletal and Skin Diseases grant no. RO1 AR45651-OA1 and the National Institute of Environmental Health Sciences Kresge Center grant no. ES-00002.
Address correspondence to Dr. Xiping Xu, Program for Population Genetics, Harvard School of Public Health, Boston, MA, USA. e-mail: xu@hsph.harvard.edu

Received 1 May 2002; accepted 18 June 2002.

K. Hao[1], T. Niu[1], C. Sangokoya[2], J. Li[1], and X. Xu[1,3,4]
[1]Harvard School of Public Health
Boston, MA
[2]Stanford University
Stanford, CA, USA
[3]Beijing Medical University, Beijing, China
[4]Brigham and Women's Hospital
Harvard Medical School
Boston, MA, USA

Directional Genome Walking Using PCR

BioTechniques 33:830-834 (October 2002)

ABSTRACT

We describe here a PCR-based "directional genome walking" protocol. The basic procedure for the amplification consists of two rounds of PCR. A primary PCR was performed, on the genomic DNA using a biotinylated primer specific to a known sequence in the genome along with four universal walker primers that were designed with partial degeneracy. The biotinylated primary PCR products were immobilized on streptavidin-linked paramagnetic beads. This step removed all nonspecific amplification products, and the purified template was used for the second PCR using a nested primer and the walker primer-2 to increase specificity. This technique is potentially useful for cloning promoter regions and has been successfully used to isolate 5′-flanking genomic regions of many cDNA clones previously isolated by us.

INTRODUCTION

The traditional approach for "walking" from regions of known sequence into flanking DNA sequences involved the successive probing of libraries with clones obtained from prior screenings. This method, besides being laborious, was also time consuming. Though PCR is an effective method for selectively amplifying specific DNA segments, conventional PCR only allows the amplification of sequences within known boundaries. Therefore, sequence information at the extremities of the DNA fragment to be amplified is a prerequisite for selective amplification of specific DNA, thus posing a major limitation on the use of the PCR in the amplification and isolation of these unknown regions.

Numerous modifications have been made to existing protocols for the amplification of an unknown DNA sequence that flank regions of known sequences, by PCR (5). They include inverse PCR (13), panhandle PCR (17),

vectorette PCR (1), anchored PCR (16), AP-PCR (2,19), capture PCR (9), and adapter-ligated PCR (6,15,18,20). In the methods listed above, PCR was usually carried out using restriction enzyme-digested-genomic DNA fragments as template, which was either circularized by ligation or cloned into a vector or ligated to double-stranded, partially double-stranded, or single-stranded oligonucleotide cassettes. In other cases, the amplification is carried out with locus-specific primer(s) and a vector/oligonucleotide cassette-specific primer to amplify a fragment contiguous to the known sequence. In this method, the linear amplification of other DNA fragments, due to a common vector (17) or adapter primer, results in a high background during PCR. To overcome this problem, known sequences attached/created at the ends of a region of interest were developed (4,7,11), which then were used as a template for PCR. Here the template could be used only to amplify a particular region of interest—therefore, with reasonably low noise.

The methodology was further improved to reduce the noise and to allow the amplification of the flanking region of any known sequence from that genomic source (1,8,12). In these methods, the desired specificity was attained because of the specific design of the cassette, which ensured that the cassette-specific primer is not primed in the first amplification cycle. Thus, the cassette-specific primer can only take part in the reaction after the target restriction fragment has been extended to its end in the first cycle by the locus-specific primer. These methods also have major drawbacks, the most important being the generation of nonspecific PCR products even after hot-start PCR or touchdown PCR (3) during primary amplification. Therefore, one had to ascertain the authenticity of the PCR products through Southern analysis or extensive cloning and sequencing. Besides, the presence of multiple PCR products complicates matters.

To circumvent these problems, we have introduced a biotin/streptavidin system to capture biotinylated fragments of interest before the nested PCR is carried out (14). The basic feature of this method is to start a primary PCR amplification reaction using a locus-specific biotinylated primer. The introduction of biotin, as a separation label, into the specific restriction fragment containing a part of the known locus and the flanking region helps in the process of its isolation. The biotinylated primary PCR fragments are then isolated from the complex genomic mixture using paramagnetic beads coated with streptavidin. This isolation step is very important in reducing the complexity of the template and increasing the concentration of the specific fragment by several orders of magnitude for the subsequent nested PCR (14).

All the above-mentioned methods rely on the restriction fragmentation of genomic DNA before ligation to the oligonucleotide cassette. Plant genomic DNA initially digested with a restriction enzyme might contain 10^7–10^8 (depending on the size of the genome) different fragments ranging from less than 300 bp to greater than 10 000 bp because of the uneven distribution of the restriction site. Therefore, it is essential that at least one restriction fragment be produced that contains part of the known locus, without it being too large to be efficiently amplified in vitro. In most cases, the information on the distribution of restriction enzyme sites in a region of interest is unavailable before the start of the walking experiment. Therefore, a combination of several different enzymes has to be tried to increase the probability of generating such convenient fragments. The protocol described here avoids the above limitations by using a specially designed walker primer that has degenerate and arbitrary bases towards its 3′-end. This allows the primer to anneal to different locations 5′ of the known region. It is a fast and reliable directional genome walking protocol that could be completed in a single working day using very small amounts of genomic DNA without any prior processing. We show here the successful use of this protocol to isolate 5′-flanking regions of a number of plant genes.

MATERIALS AND METHODS

Genomic DNA was isolated from plants of choice using the CTAB

method (10). Approximately 50 ng genomic DNA sample were used as a template to perform PCR amplification and later to enrich for DNA fragments of interest from the PCR fragments that are adjacent to a known sequence. This is achieved using appropriate 5′-biotinylated locus-specific primer complementary to the known region of the DNA and along with the walker primers 1, 2, 3, and 4 (Table 1) in four individual tubes, one each for each walker primer. PCR conditions were 94°C for 1 min, 47°C–50°C for 1 min, and 72°C for 4 min for 30 cycles unless otherwise specified.

A total of 50 µL streptavidin-linked paramagnetic beads (Roche Applied Science, Mannheim, Germany) was washed three times in 1× PCR buffer (10 mM Tris-HCl, pH 8.8, 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% gelatin, 0.005% Tween® 20, and 0.005% Nonidet™ P-40), the amplified biotinylated PCR products were immobilized on streptavidin-linked paramagnetic beads, and the non-biotinylated DNA was washed off.

Nested PCR was carried out using 1 µL immobilized primary PCR products, separately purified from the previous step, as template along with nested locus-specific primer-2 and the walker primer-2 (5′-CTAATACGACTCACTATAGGG-3′) in four separate tubes. The primer concentration and the PCR cycling conditions were the same as mentioned earlier. Five microliters of the amplification product were electrophoresed on 1% agarose gel and visualized after ethidium bromide staining. The amplified nested PCR products were cloned into a TA cloning vector as per the manufacturer's instructions (Promega, Madison, WI, USA).

RESULTS AND DISCUSSION

In this protocol, the genomic DNA is used without any further modifications such as restriction digestion of genomic DNA and/or ligation of special adapters. The four universal walker primers used separately for amplification along with the biotinylated locus-specific primer reduces the complexity of the primary PCR products. The four universal walker primers differ from each other at their 3′-end. The initial four bases at the 3′-end were arbitrarily fixed for each universal walker primer, and the next four bases were completely degenerate (to reduce the permutations and combinations, universal bases can also be used instead of introducing degeneracy). The remaining 5′ portion of these universal walker primers is identical and arbitrarily fixed as (5′-CTAATACGACTCACTATAGGG-3′). Theoretically, the four-base sequence combination present on the 3′-end of each of the universal walker primers can exist once in every 256 bp in the genomic DNA. The next four bases being degenerate can also form a perfect complement on the same template DNA. Additional base pairing also can occur between the 5′ arbitrary sequence of the walker primer and the corresponding region of the genomic DNA. Overall, these universal walker primers can anneal to the template at 47°C–55°C in the PCR depending on the base composition of the template. The biotinylated locus-specific primer will determine the position and direction of the genome walk during the PCR. The primary PCR may amplify multiple products because of multiple priming of the universal walker primer. The immobilization of the primary PCR products onto streptavidin-linked paramagnetic beads will help to remove all the nonspecific amplification products and enriches for the biotinylated specific products. This enrichment step is very important in reducing the complexity of the template and increasing the concentration of the specific fragment by several orders of magnitude for the subsequent nested PCR amplification. Thus, the use of the nested locus-specific primer 2 ensures the selective amplification of the desired fragment in the second round of PCR.

To test the efficacy of the methodology, we PCR amplified 5′-flanking regions of different genes isolated in our laboratory. Using gene-specific pri-

Table 1. Walker Primers (5′ → 3′)

1.	CTAATACGACTCACTATAGGGNNNNATGC
2.	CTAATACGACTCACTATAGGGNNNNGATC
3.	CTAATACGACTCACTATAGGGNNNNTAGC
4.	CTAATACGACTCACTATAGGGNNNNCTAG

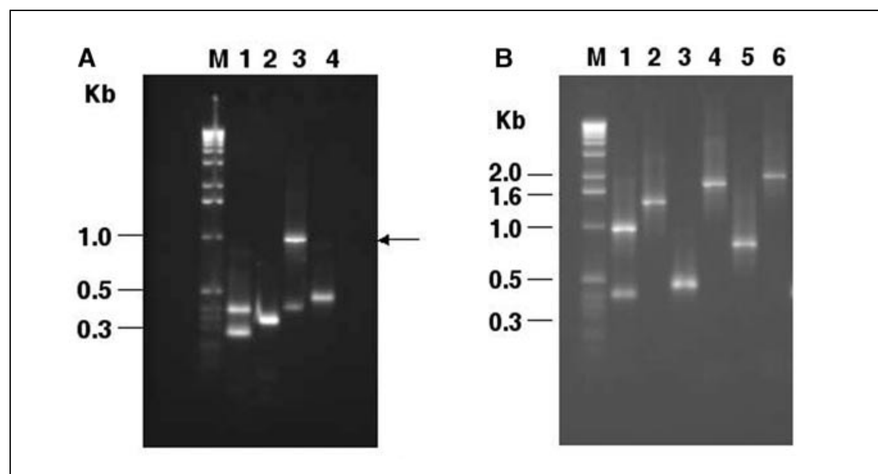

Figure 1. PCR amplification products. (A) Products generated after two rounds of PCR from regions upstream of the glyoxalase I gene from *Brassica juncea*. In the first round of PCR, genomic DNA was amplified using a biotinylated gene-specific primer and four sets of walker primers. Biotinylated strands of these PCR products were re-amplified using a nested gene-specific primer and a primer that is complementary to the walker primers used in the first round of PCR. The different products generated after the second round of PCR are shown in lanes 1–4. Note the different product sizes. The reaction that produced the biggest molecule was further analyzed (arrow). (B) Products of regions upstream of *B. juncea* glyoxalase gene (lane 1), *Pennisetum* Na^+/H^+ antiporter gene (lane 2), pea *tef*A (lane 3), *Pennisetum* ATPase C subunit (lane 4), *Pennisetum* phosholipase D (lane 5), and *Pennisetum* porin gene (lane 6). Note that these are all PCR products after the second round of PCR as described in part A, and only the products of one of the four walker reactions, that which yielded the largest product after second round of PCR for each gene, have been shown. M, molecular weight marker (Invitrogen, Carlsbad, CA, USA).

mers synthesized in the antisense orientation towards the 5′-end of the cDNA(s), we were able to amplify 5′-flanking regions of all the genes we attempted. Figure 1, A and B, shows that the PCR amplification products were discrete fragments in all the cases after nested PCR. The PCR fragments were gel-purified and cloned. To verify the authenticity of the cloned fragments, the insert DNA was sequenced completely and compared with the corresponding cDNA sequence for the 5′-end overlap because the nested locus-specific primer was synthesized such that there was at least a 100-bp overlap with the 5′-end of the known cDNA sequence. The detailed analysis of these promoter regions will be published elsewhere.

Unlike earlier methods, the method presented here does not rely on the restriction of genomic DNA and therefore does not suffer because of the uneven distribution of the restriction enzyme sites, in the region of interest, to generate convenient fragments for PCR amplification. The utilization of biotinylated locus-specific primer 1 and the purification of primary PCR-amplified products on streptavidin-linked paramagnetic beads before nested PCR reduce the template complexity and increase the desired template concentration by several orders of magnitude. For the easy, efficient, and reliable directional genome walking, we have developed a protocol in which genomic DNA required is as little as 50 ng and with as little as 30 bp sequence information to isolate 5′- and 3′-flanking regions of interest. The effectiveness of the rapid PCR-based DNA walking method described here will be valuable in isolating promoters and regulatory elements from sequences obtained from cloned cDNAs. This technique can also be used for directional genome walking from known regions to unknown regions.

REFERENCES

1. **Arnold, C. and I.J. Hodgson.** 1991. Vectorette PCR: a novel approach to genomic walking. PCR Methods Appl. *1*:39-42.
2. **Dominguez, O. and C. Lopez-Larrea.** 1994. Gene walking by unpredictably primed PCR. Nucleic Acids Res. *22*:3247-3248.
3. **Don, R.H., P.T. Cox, B.J. Wainwright, K. Baker, and J.S. Mattick.** 1991. “Touchdown” PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res. *19*:4008.
4. **Fors, L., R.A. Saavedra, and L. Hood.** 1990. Cloning of the Shark Po promoter using a genomic walking technique based on the polymerase chain reaction. Nucleic Acids Res. *18*:2793-2799.
5. **Hui, E.K., P.C. Wang, and S.J. Lo.** 1998. Strategies for cloning unknown cellular flanking DNA sequences from foreign integrants. Cell Mol. Life Sci. *54*:1403-1411.
6. **Iwahana, H., T. Tsujisawa, R. Katashima, K. Yoshimoto, and M. Itakura.** 1994. PCR with end trimming and cassette ligation: a rapid method to clone exon-intron boundaries and a 5′-upstream sequences of genomic DNA based on a cDNA sequence. PCR Methods Appl. *4*:19-25.
7. **Jones, D.H. and S.C. Winistorfer.** 1997. Amplification of 4–9 kb human genomic DNA flanking a known site using a panhandle PCR variant. BioTechniques *23*:132-138.
8. **Kilstrup, M. and K.N. Kristiansen.** 2000. Rapid genome walking: a simplified oligo-cassette mediated polymerase chain reaction using a single genome-specific primer. Nucleic Acids Res. *28*:e55.
9. **Lagerstrom, M., J. Parik, H. Malmgren, J. Stewart, U. Pettersson, and U. Landegren.** 1991. Capture PCR: efficient amplification of DNA fragments adjacent to a known sequence in human and YAC DNA. PCR Methods Appl. *1*:111-119.
10. **Lodhi, M.A., G.N. Ye, N.F. Weeden, and B.I. Reisch.** 1994. A simple and efficient method for DNA extraction from grapevine cultivars, *Vitis* species and *Ampelopsis*. Plant Mol. Biol. Rep. *12*:6-13.
11. **Mueller, P.R. and B. Wold.** 1989. In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. Science *246*:780-786.
12. **Niu, L.L. and A.M. Fallon.** 1999. Retrieval of flanking DNA using a PCR-based approach with restriction enzyme-digested genomic DNA template. BioTechniques *26*:630-634.
13. **Ochman, H., A.S. Gerber, and D.L. Hartl.** 1988 Genetic applications of an inverse polymerase chain reaction. Genetics *120*:621-623.
14. **Reddy, M.K., S. Nair, and S.K. Sopory.** 2002. A new approach for efficient directional genome walking using polymerase chain reaction. Anal. Biochem. *306*:154-158.
15. **Riley, J., R. Butler, D. Ogilvie, R. Finnear, D. Jenner, S. Powell, R. Anand, J.C. Smith, and A.F. Markham.** 1990. A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. Nucleic Acids Res. *18*:2887-2890.
16. **Roux, K.H. and P. Dhanarajan.** 1990. A strategy for single site PCR amplification of dsDNA: priming digested cloned or genomic DNA from an anchor-modified restriction and a short internal sequence. BioTechniques *8*:48-57.
17. **Shyamala, V. and G.F.L. Ames.** 1989. Genome walking by single-specific primer polymerase chain reaction: SSP-PCR. Gene *84*:1-8.
18. **Siebert, P.D., A. Chenchik., D.E. Kellogg., K.A. Lukyanov, and S.A. Lukyanov.** 1995. An improved PCR method for walking in uncloned genomic DNA. Nucleic Acids Res. *23*:1087-1088.
19. **Trueba, G.A. and R.C. Johnson.** 1996. Random primed genome walking PCR: a simple procedure to retrieve nucleotide fragments adjacent to known DNA sequences. BioTechniques *21*:20.
20. **Willems, H.** 1998. Adaptor PCR for the specific amplification of unknown DNA fragments. BioTechniques *24*:26-28.

Address correspondence to Dr. M.K. Reddy, International Centre for Genetic Engineering and Biotechnology (ICGEB), Aruna Asaf Ali Marg, New Delhi 110 067, India. e-mail: reddy@icgeb.res.in

Received 10 May 2002; accepted 17 July 2002.

R.N. Mishra, S.L. Singla-Pareek, S. Nair, S.K. Sopory, and M.K. Reddy
International Centre for Genetic Engineering and Biotechnology (ICGEB)
New Delhi, India

For reprints of this or any other article, contact *Reprints@BioTechniques.com*