Iranian J Biotech. 2019 April; 17(2): e2183 DOI: 10.21859/ijb.2183

Short Paper

A Simple Genome Walking Strategy to Isolate Unknown Genomic Regions Using Long Primer and RAPD Primer

Fajun Li [1,*], Chunpeng Fu [1], Qunfeng Li [1]

[1] Shandong Peninsula Engineering Research Center, Comprehensive Brine Utilization, Weifang University of Science and Technology, Shouguang, China

[*] **Corresponding author:** Fajun Li, Shandong Peninsula Engineering Research Center, Comprehensive Brine Utilization, Weifang University of Science and Technology, Shouguang, China; Tel/Fax:+86-5365101992; E-mail: lifajun1976@163.com

Abstract

Background: Genome walking is a DNA-cloning methodology that is used to isolate unknown genomic regions adjacent to known sequences. However, the existing genome-walking methods have their own limitations.

Objectives: Our aim was to provide a simple and efficient genome-walking technology.

Material and Methods: In this paper, we developed a novel PCR strategy (termed SLRA PCR) that uses a single long primer (SLP), a set of gene specific primers (GSP), and a random amplified polymorphic DNA (RAPD) primer for genome walking. SLRA PCR consists of two processes: the first amplification using SLP, and three successive rounds of nested PCR amplified by GSP and RAPD primer. The novelty of the approach lies in the use of long primers (SLP and GSP) and same annealing and extension temperature 68°C in combination. This method offers higher amplification efficiency, superior versatility, and greater simplicity compared with conventional randomly primed PCR methods for genome walking.

Results: The promoter regions and the first introns of the insulin-like androgenic gland hormone (IAG) gene and the *hemocyanin* gene of *Macrobrachium nipponense* were cloned using SLRA PCR, respectively.

Conclusions: This genome walking strategy can be applied to a wide range of genomes.

Keywords: Hemocyanins; Polymerase Chain Reaction; DNA Primers

1. Background

Genome walking is a DNA cloning methodology that is used to isolate unknown genomic regions adjacent to known sequences. It can be used for the identification of the integration sites of transposable elements, gene cloning for functional studies, identification of regulatory sequences outside of the coding regions, and for filling in the lingering gaps after a genome is sequenced (1). The polymerase chain reaction (PCR)-based genome walking techniques are popular as the methods are relatively fast and effortless. They can be divided into four types: inverse PCR (2-4), ligation-mediated PCR (5-8), randomly primed PCR (9-11), and overlapping primer-based PCR (12). The first two groups require a preliminary digestion of genomic DNA with a suitable restriction enzyme, which makes these methods relatively expensive and time-consuming. Since information on restriction sites is usually not available in advance, there is no guarantee that digestion with a particular restriction enzyme will be successful. This necessitates trials with several different enzymes. Therefore, the requirements of high-quality DNA for complete endonuclease digestion and efficiency of the ligation reactions have limited the utilization of these methods.

The randomly-primed PCR methods overcome the flaws mentioned above and has gained popularity for its simplicity as it needs no prior DNA manipulation (such as enzyme digestion or ligation) (13). They are relatively low-cost and relatively direct, i.e. there is no requirement for intermediate steps involving restriction enzyme digestion of the DNA template or ligation reactions. However, the conventional randomly-primed PCR often produces nonspecific amplification products due to the short-arbitrary-degenerate primers and low annealing temperatures, and may also achieve non-efficient amplification of target sequences(14). Even in improved methodologies, there are also many flaws. For instance, the improved high-efficiency thermal asymmetric interlaced (hi) TAIL-PCR still remains time-consuming, and is expensive in terms of primer

design and synthesis. Moreover, this technique poses a great challenge to the activity of DNA polymerase, and so is frequently associated with inefficient amplification results(15); The fusion primer and nested integrated PCR (FPNI-PCR) is still a time-consuming procedure that usually involves three rounds of amplification with each round consisting of multiple PCR cycles. Numerous primers (SP1, SP2, SP3, FSP1,FSP2 and degenerate primers) are applied in this process(15). Complex procedures make it difficult to be used widely. Therefore, it is highly desirable that further novel methods are developed for rapid and efficient genome walking or flanking sequence cloning. As for overlapping primer-based PCR, the walking primers needs a special design so as to they have 10 bp overlap at the 3' ends (12).

2. Objective

In order to overcome the technical hurdles, in this paper, we present a new PCR method for genome walking; single-long primer (SLP) and randomly-amplified-polymorphic DNA (RAPD) primer PCR (termed SLRA PCR). The basic principle of SLRA PCR is illustrated in **Figure 1**. The promoter regions and first introns of the *insulin-like androgenic gland hormone (IAG)* gene and the *hemocyanin* gene of *Macrobrachium nipponense* were cloned using this approach, respectively. The simple and efficient genome walking strategy can be applied to a wide range of genomes.

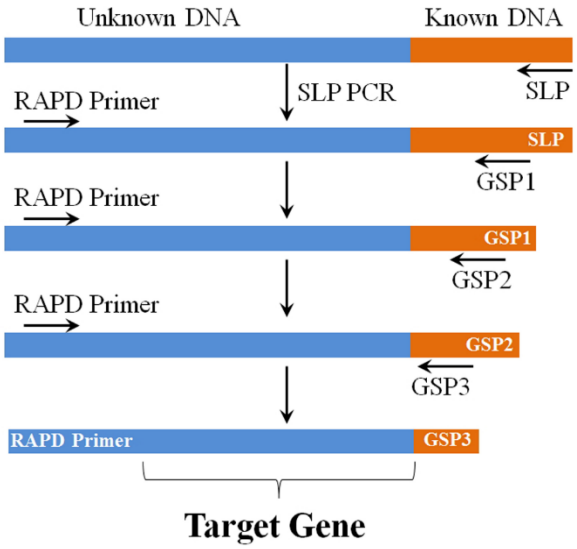

Figure 1. Schematic representation of the genome walking SLRA PCR protocol. Blue gradient block represents unknown DNA sequence, yellow gradient block represents known DNA sequence. SLP represents single long primer, GSP (1-3) represents gene specific primer, RAPD primer represents random amplified polymorphic DNA (RAPD) primer. Each arrow denotes a specific primer hybridization with the nucleotide sequence. In this SLRA PCR protocol, the system consists of two processes. The first amplification using SLP and nested PCR using GSP (1-3) and RAPD primer.

3. Materials and Methods

3.1. Materials

Adult *Macrobrachium nipponense* prawns used in this study were obtained from Tai lake in Wuxi, China (120°13'44"E, 31°28'22"N). DNA extraction kit, PrimeSTAR Max Premix, and LA polymerase were purchased from TaKaRa (Dalian, China). Genomic DNA was isolated using the DNA extraction kit according to the manufacturer's protocol.

3.2. Genome Walking by SLRA PCR

Step A: single long primer PCR (SLP PCR). The system consisted of two processes. The first amplification used a SLP, which was designed against known DNA sequence. The SLP must meet the following requirements: 30-35 nucleotides in length, GC content of 40–60%, annealing, and extension temperature of 68°C. There should be no more than three G's and C's in the last six positions at the 3' end of the primer. In this study, the SLPs of promoter regions (PIAG-SLP, PHe-SLP) and first introns (IAGI1-SLP, HeI1-SLP) of the *IAG* and *hemocyanin* gene of *M. nipponense* are listed in **Table 1**. Cycling conditions and PCR reagents are given in **Table 2**. In the SLP PCR, we employed a two-step protocol. This was possible because of the high annealing and extension temperature of 68°C, of the SLP, and the characteristics of the PrimeSTAR Max Premix, which contained a fast, high-fidelity amplification enzyme. All of these factors ensured that the SLP bound specifically to the known DNA sequence.

Step B: Nested PCR. The second amplification was three-step nested PCR. Three gene-specific primers (GSP1, GSP2 and GSP3) were designed against the known sequence, from outside to inside (**Fig. 1**), and were located inside the SLP (not contain all 3 GSP primers.). Each GSP was 26-28 nucleotides in length to ensure a high annealing and extension temperature. The GC content of GSP is same as SLP. Other conditions were the same as that used for the SLP PCR. RAPD primers were randomly selected. The GSPs of promoter regions (PIAG-GSP1-3, PHe-GSP1-3) and first introns (IAGI1-GSP1-3, HeI1-GSP1-3) of the IAG and the hemocyanin gene of *M. nipponense* are shown in **Table 1**. Cycling conditions and PCR reagents are given in **Table 2**. In the nested amplification, we used 5 µlitr of the SLP PCR product (no dilution and purification) as the first nested PCR template. 1 µlitr of the first round of nested PCR product was diluted 50-100 times and used as the template for the second round of nested circulation, and so on for subsequent rounds. Given the characteristics of the RAPD primer (10 nucleotides), we applied a low annealing temperature of 40°C to allow the RAPD primer to bind to the DNA. An annealing and extension temperature of 68°C was used to allow the GSP to bind effectively to the corresponding sequence.

The LA polymerase was used to ensure extension efficiency.

3.3. Cloning and Sequencing

5 µlitr of the PCR products were used in electrophoresis in a 1.5% agarose gel in a TBE buffer stained with ethidium bromide. The DNA was visualized under a UV light. Bands corresponding to the largest products were excised from the agarose gels, purified with Agarose Gel DNA Recovery Kit (TaKaRa, China), and cloned into a pMD18-T TA cloning vector (TaKaRa, China), and then sequenced with M13 forward or reverse primers.

3.4. Sequence Analysis

The sequences were analyzed by computational methods. Promoter was predicted by NNPP version (http://www.fruitfly.org/seq_tools/promoter.html), Transcription factors binding sites in the promoter sequence of gene were predicted by TFSEARCH (http://www.cbrc.jp/htbin/nph-tfsearch).

Table 1. Primer sequences of promoter regions and first introns of the insulin-like androgenic gland hormone (*IAG*) gene and hemocyanin gene of *M. nipponense* used in SLRA PCR

Primer Name	Nucleotide Sequence (5′→3′)
Promoter *IAG*	
PIAG-SLP	CCTTGAGAGAAGAATTTGAAGTCTTGAGATGAAGG
PIAG-GSP1	CCAGAAAGCAGGAGAGCGAATTTGGAGG
PIAG-GSP2	GAAACACCGGTTAAACGGACAAGGGGGC
PIAG-GSP3	GACACAGCCGTTCAGGGTATGATCCCAG
PIAG-RAPD	GGCACGTAAG
First intron of *IAG*	
IAGI1-SLP	AAGACTCTGGGATCATACCCTGAACGGCTGTGTCC
IAGI1-GSP1	GCCCCCTTGTCCGTTTAACCGGTGTTTC
IAGI1-GSP2	CCTCCAAATTCGCTCTCCTGCTTTCTGG
IAGI1-GSP3	TCTCAAGACTTCAAATTCTTCTCTCAAGG
IAGI1-RAPD	CAGCGCTACG
Promoter of Hemocyanin	
PHe-SLP	ATCGCTCTGGAAGCCGTCCAGAGAGTGACTCTCG
PHe-GSP1	GCTAAAGCTGGGCCAGGCTGCGGCAG
PHe-GSP2	CAGCAGAGCGCACAAGACAAACACCTT
PHe-GSP3	TGTGTGCTGTGCCGGACCAGGGGAATC
PHe-RAPD	CCACGGGAAG
First intron of Hemocyanin	
HeI1-SLP	GCCAATAGGTATGTAGTGTCTGCTGATTTCACG
HeI1-GSP1	GGCGACGGACTTTGCCTCGGCCTTGC
HeI1-GSP2	CACAGCACACAATGAAGGTGTTTGTC
HeI1-GSP3	GCCCAGCTTTAGCTTCGAGAGTCACTC
HeI1-RAPD	TAGCCACTGG

Table 2. Thermal cycling conditions and PCR reagents for the SLRA PCR

Program	No. of cycles	Cycling conditions	PCR reagents
SLP PCR	30	98°C for 10 s ; (98°C for 10 s; annealing and extension at 68°C for 1 min); termination extension at 68°C for 3 min	PrimeSTARMax Premix 25 µl, SLP 0.2-0.3 µM, DNA 20 ng, ddH2O to 50 µl
Nested PCR	30	94°C for 90 s; (94°C for 30 s; annealing at 40°C for 30 s and extension at 68°C for 3 min); termination extension at 68°C for 5 min	Buffer(Mg+) 2.5 µl, dNTP 200 µM, RAPD 0.5 µM, GSP 0.5 µM, LA polymerase 0.5 U, PCR product 1 µl, ddH2O to 25 µl

4. Results

DNA quality plays a critical role for ligation-mediated PCR genome walking. However, there is no special requirement for SLRA PCR strategy, DNA obtained by general DNA extraction kit can meet the requirements. Furthermore, minute traces of genomic DNA (20 ng) was used in SLRA PCR walking.

Using the above protocol, we cloned the promoter sequences and the first introns of the *IAG* gene (GenBank: KF811212) and hemocyanin gene (GenBank: KF887993) from *M.nipponense* (**Fig. 2**). TATA box, CAAT box, and the transcriptional start site were found in the *IAG* promoter sequence (about 700bp) (**Supplementary Fig. 1**). TATA box, CAAT box, and the transcriptional start site were found in the *hemcoyanin* promoter sequence (about 600bp) (**Supplementary Fig. 2**). The first introns of *IAG* (1795bp) (**Fig. 2**) and hemocyanin (700bp) were also isolated, the exon-intron splice junctions in *IAG* and the hemcoyanin gene followed the GT-AG pattern.

5. Discussion

5.1. Principle of the Current Technique

The principle of genome walking with SLRA PCR is illustrated in **Figure 1**. The novelty of the approach lies in the use of long primers (SLP or GSP) and identical annealing and extension temperatures (68°C) in combination. Such high annealing and extension temperature is essential for this protocol. At 68°C, SLP or GSP can hybridize with the known DNA strictly,

which ensures the PCR amplification stringency, in parallel, the amplification efficiency of DNA polymerase is also ensured. A series of key technical points of this strategy are as follow: (i) Compared with the general primers (20-22 nucleotides), SLP (30-35 nucleotides) and GSP (26-28 nucleotides) play a key role in the process, which decreases the binding rate of primers to nonspecific sites and allows to capture the target sequence reliably.

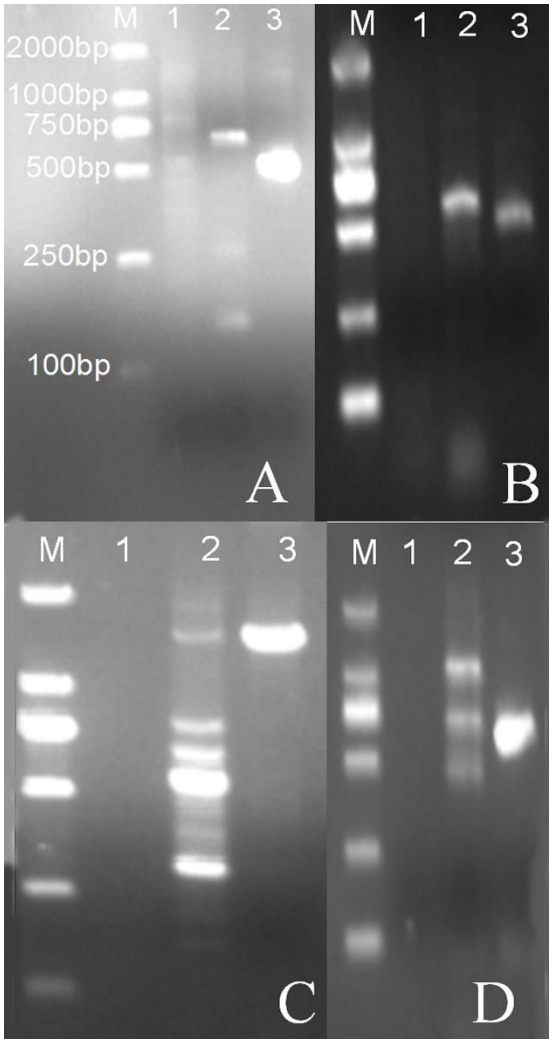

Figure 2. Agarose gel electrophoresis analysis of the nested PCR products. PCR products (5μl) are loaded on a 1.5% agarose gel. A: promoter of IAG gene; B: promoter of hemocyanin gene; C: The first introns of IAG gene; D: The first introns of hemocyanin gene. M, DNA ladder, six bands from top to bottom is 2000 bp, 1000 bp, 750 bp, 500 bp, 250 bp, 100 bp, respectively. Lane 1-3 are products amplified by the first, second, third nested PCR, respectively.

This template-capture step can reduce the template complexity and increase the specificity of the following nested PCR(1). Under higher stringent conditions, the single specific primer binds to complementary sites within captured DNA templates, and the target DNA fragments containing known and adjacent unknown sequences will be amplified. (ii) To reduce false positive products, nested PCR can be employed. The successive PCR rounds using different GSP can significantly reduce the non-specific amplification because the spurious PCR products are unlikely to contain binding sites for the inner specific primers(16). (iii) In order to avoid the randomness of RAPD primer, we use the SLP PCR product as the first nested PCR template. In the nested PCR, the RAPD primers hybridize with the rigorous and precise template so as to improve the precision of the experiment.

Notably, the Tm of the SLP and GSP should be designed to be higher than 68°C. Such a high annealing temperature could ensure the binding of specific primers to the correct locus at stringent conditions in the following amplifications(17).

5.2. Advantages and Potential Applications SLRA PCR

Compared with conventional randomly-primed PCR methods for genome walking, the SLRA PCR method has several advantages: (i) Higher efficiency: conventional randomly primed PCR methods tend to produce small amplification products or nonspecific amplification products (14, 18). In our preliminary experiments, owing to nonspecific amplification, we failed to clone the promoter sequence of IAG using (hi) TAIL-PCR and FPNI-PCR protocol, respectively. In SLRA PCR strategy, we use the SLP PCR product as the first nested PCR template, which overcomes the problem of nonspecific amplification produced by the randomness of RAPD primer. Furthermore, we use long primer (SLP and GSP) and same annealing and extension temperature of 68°C together, which not only improves the accuracy of the experiment but also boosts its efficiency, an approximate 1.8 Kb sequence of the first intron of IAG was acquired using this protocol. (ii) Superior versatility: owing to the diversity of RAPD, we can always find one hybridize with the unknown sequence, this method can effectively avoid the experimental failure caused by false amplification. Thus, the SLRA PCR can be performed on a wide range of genomes. (iii) Greater simplicity: degenerate primers and complex PCR procedures are unused in SLRA PCR strategy. Simple PCR progresses avoid the problem of decreased activity of DNA polymerase produced by complex program in hi TAIL-PCR technique. This entire process can be completed in a single working day.

6. Conclusions

The advantages of the novel SLRA PCR method make it useful for obtaining a full gene sequence, including detecting promoters and regulatory elements in the genome. This method is also an efficient tool to identify transposon integration sites and sequence-tagged sites in genomic DNA for gene functional studies and to obtain operons from a known short DNA sequence. In

summary, we have developed an efficient and versatile genome walking method, and this method can be applied to a wide range of genomes for various purposes.

Acknowledgements

This work was funded by the Shandong Provincial Natural Science Foundation of China (ZR2016CM12); a project of Shandong Province Higher Educational Science and Technology Program (J17KB112 and J16LE59); the Project of Shandong Peninsula Engineering Research Center of Comprehensive Brine Utilization (2018LS009, 2018LS015 and 2018LS019); the Doctoral Fund of Weifang University of Science and Technology (2017BS03).

Conflict of Interest

There is no conflict of interest.

Authors' Contribution

These authors contributed to this work equally.

References

1. Wang H, Yao T, Cai M, Xiao X, Ding X, Xia L. A genome walking strategy for the identification of nucleotide sequences adjacent to known regions. *Biotechnol Lett*. 2013;**35**(2):279-284. doi: 10.1007/s10529-012-1076-3 pmid: 23108875
2. Benkel BF, Fong Y. Long range-inverse PCR (LR-IPCR): extending the useful range of inverse PCR. *Genet Anal*. 1996;**13**(5):123-127. pmid: 9021401
3. Ren M, Chen Q, Li L, Zhang R, Guo S. Successive chromosome walking by compatible ends ligation inverse PCR. *Mol Biotechnol*. 2005;**30**(2):95-102. doi: 10.1385/MB:30:2:095 pmid: 15920279
4. Tsaftaris A, Pasentzis K, Argiriou A. Rolling circle amplification of genomic templates for inverse PCR (RCA-GIP): a method for 5'- and 3'-genome walking without anchoring. *Biotechnol Lett*. 2010;**32**(1):157-161. doi: 10.1007/s10529-009-0128-9 pmid: 19760115
5. Arnold C, Hodgson IJ. Vectorette PCR: a novel approach to genomic walking. *PCR Methods Appl*. 1991;**1**(1):39-42. pmid: 1842919
6. Tsuchiya T, Kameya N, Nakamura I. Straight walk: a modified method of ligation-mediated genome walking for plant species with large genomes. *Anal Biochem*. 2009;**388**(1):158-160. doi: 10.1016/j.ab.2009.02.002 pmid: 19454221
7. Bae JH, Sohn JH. Template-blocking PCR: an advanced PCR technique for genome walking. *Anal Biochem*. 2010;**398**(1):112-116. doi: 10.1016/j.ab.2009.11.003 pmid: 19903447
8. Burke E. High-throughput TAIL-PCR as a tool to identify DNA flanking insertions. Plant Functional Genomics: Springer; 2003. p. 241-271.
9. Ji J, Braam J. Restriction site extension PCR: a novel method for high-throughput characterization of tagged DNA fragments and genome walking. *PLoS One*. 2010;**5**(5):e10577. doi: 10.1371/journal.pone.0010577 pmid: 20485508
10. Reddy PS, Mahanty S, Kaul T, Nair S, Sopory SK, Reddy MK. A high-throughput genome-walking method and its use for cloning unknown flanking sequences. *Anal Biochem*. 2008;**381**(2):248-253. pmid: 18674512
11. Liu YG, Chen Y. High-efficiency thermal asymmetric interlaced PCR for amplification of unknown flanking sequences. *Biotechniques*. 2007;**43**(5):649-650, 652, 654 passim. doi: 10.2144/000112601 pmid: 18072594
12. Li H, Ding D, Cao Y, Yu B, Guo L, Liu X. Partially overlapping primer-based PCR for genome walking. *PLoS One*. 2015;**10**(3):e0120139. doi: 10.1371/journal.pone.0120139 pmid: 25811779
13. Wang S, He J, Cui Z, Li S. Self-formed adaptor PCR: a simple and efficient method for chromosome walking. *Appl Environ Microbiol*. 2007;**73**(15):5048-5051. doi: 10.1128/AEM.02973-06 pmid: 17483260
14. Yuanxin Y, Chengcai A, Li L, Jiayu G, Guihong T, Zhangliang C. T-linker-specific ligation PCR (T-linker PCR): an advanced PCR technique for chromosome walking or for isolation of tagged DNA ends. *Nucleic Acids Res*. 2003;**31**(12):e68. doi: 10.1093/nar/gng068 pmid: 12799457
15. Wang Z, Ye S, Li J, Zheng B, Bao M, Ning G. Fusion primer and nested integrated PCR (FPNI-PCR): a new high-efficiency strategy for rapid chromosome walking or flanking sequence cloning. *BMC Biotechnol*. 2011;**11**(1):109. doi: 10.1186/1472-6750-11-109 pmid: 22093809
16. Guo H, Xiong J. A specific and versatile genome walking technique. *Gene*. 2006;**381**:18-23. doi: 10.1016/j.gene.2006.06.002 pmid: 16914272
17. Deng J, Wei M, Yu B, Chen Y. Efficient amplification of genes involved in microbial secondary metabolism by an improved genome walking method. *Appl Microbiol Biotechnol*. 2010;**87**(2):757-764. doi: 10.1007/s00253-010-2569-4 pmid: 20376630
18. Luo P, Su T, Hu C, Ren C. A novel and simple PCR walking method for rapid acquisition of long DNA sequence flanking a known site in microbial genome. *Mol Biotechnol*. 2011;**47**(3):220-228. doi: 10.1007/s12033-010-9332-z pmid: 20878504