

Key techniques and efficiency analysis of amplification of flanking unknown sequences by inverse PCR

Xiaodong Xue

xxd14789632@163.com

Shangluo University

Wan Li

Shangluo University

Qinxia Li

Shangluo University

Weidong Chang

Shangluo University

Lujing Shi

Shangluo University

Zhe Wang

Shangluo University

Yongping Zhao

Shangluo University

Research Article

Keywords: Brachypodium distachyon, inverse PCR, unknown sequence, flanking sequence

Posted Date: November 29th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-3654317/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Inverse PCR (IPCR) is an accurate, simple, feasible, and ideal technique for obtaining unknown sequences. In this study, we used the model monocot *Brachypodium distachyon* (ecotype Bd21) to standardize the conditions and materials required for successfully performing the IPCR. Analysis of the amplified sequences led us to the following conclusions. First, the distance between the restriction endonuclease cleavage site and the unknown–known sequence boundary should be at least 400 bp. Second, a 6 bp restriction endonuclease such as *Nde*I produces condensed bands in a size gradient with good specificity, and therefore is a better choice than a 4 bp cutter such as *Hha*I. Third, a distance of approximately 200 bp between the second primer and the boundary sequence leads to a better amplification effect and effectively ensures the integrity of the unknown flanking sequence. The experimental conditions established in this study serve as a theoretical basis for the amplification of unknown genome sequences of Gramineae crops and other species.

Introduction

Determining the location of foreign genes and their flanking sequences in transgenic plants is of great significance for the study of gene function and genomics. A flanking sequence is a nucleotide sequence located on either side of a genomic region of interest in the chromosome [1]. After the insertion of T-DNA or Ds transposon into the host chromosome, the sequence of the flanking region usually needs to be determined by chromosome walking [2]. Chromosome walking is a method used to gradually determine the nucleotide composition of an unknown sequence adjacent to a known sequence in a biological genome or genomic library [3]. Subsequently, sequence alignment tools, such as NCBI Magic-BLAST [4], are used to quickly match the relevant sequences and locate the position of the transgene in the host genome. Chromosome walking has been used to determine sequences located upstream and downstream of a region of interest [5, 6]; identify the insertion sites of the T-DNA or transposons [7, 8]; determine the fragment overlap in PAC, YAC and BAC libraries [9]; and find the promoter and regulatory elements of known genes [10, 11].

Inverse PCR (IPCR) is one of the earliest PCR-based chromosome walking techniques. The operation process of IPCR is very simple and convenient (Fig. 1). First, the genome is digested with appropriate restriction enzymes. Then, the digested fragments are self-ligated by ligase to form a loop, and a pair of primers complementary to sequences flanking the core region are used to perform the PCR, with the ringed fragment as the template. Amplification starts from the core region and proceeds along the ring molecules to the unknown sequences on both sides. Sequences upstream and downstream the core region are obtained by cloning and sequencing the IPCR products [12]. Compared with other methods, IPCR is simple and can simultaneously obtain unknown sequences on both sides of the known sequence. To date, IPCR has been used in several studies [13–17].

At the early stage of this experiment, we used the transposon tagging method [18–20] to obtain a large number of Ds insertion lines, and planned to conduct an in-depth investigation of sequences flanking the

Ds transposon in these transgenic lines. However, considering the large number of plant materials required for this experiment and the imperfect nature of flanking unknown sequence cloning technology. In addition, the key technical points of IPCR method have not been reported yet. we decided to establish a flanking unknown sequence cloning system suitable for processing a large number of plant materials. The method established in this study provides new ideas for the analysis of foreign genes in other plant species and agriculturally important crops.

1. Materials and methods

1.1 Material

1.1.1 Plant material and plasmids

Brachypodium distachyon ecotype Bd21 was used in this study. Plants were grown at 20°C under long-day photoperiod (18 h light/6 h dark). *Escherichia coli* DH5α and *Agrobacterium tumefaciens* strain EHA105 preserved in our laboratory were used for plant transformation, along with the plasmid backbone pCAMBIA1300.

1.1.2 Enzymes and biochemical reagents

Enzymes and other reagents were purchased from commercial vendors: conventional restriction endonucleases and *Taq* polymerases from Baoshan Biological Company (Dalian); T4 DNA ligase from Quanshi Gold Company (Beijing); Green *Taq* Mix from Nuoweizan Biological Technology Co., Ltd. (Nanjing); plasmid extraction kit and fluorescence quantitative kit from Kangwei Century; genomic reverse transcription kit from Quanshi Gold Company; DNA gel recovery kit from Baitaike (Beijing); DMSO and β-mercaptoethanol from Solaibao; NaOH, NaCl, and CH₃COOK from Tianjin Kaitong; and EDTA, Tris, MgCl₂, CaCl₂ and other chemicals from Shanghai Shenggong.

1.2 Method

1.2.1 DNA extraction

Fresh leaves, weighing an appropriate amount, were placed in a 1.5 mL centrifuge tube, frozen in liquid nitrogen, and ground into a fine powder using a crusher. Then, 700 μL of 2× cetyltrimethylammonium bromide (CTAB) was added to each sample and mixed. Samples were incubated in a water bath at 65°C for 20 min and then cooled to room temperature; samples could also be placed on ice for faster cooling. Then, 700 μL of chloroform was added to each sample and mixed gently by inverting the centrifuge tubes several times. Samples were then centrifuged at 5,000 × *g* for 10 min at 4°C. The supernatant was transferred to a new centrifuge tube. Then, 500 μL of isopropanol was added to each tube, mixed by inverting the tube, and incubated at -20°C for more than 1 h. Samples were then centrifuged at 15,000 × *g* for 10 min at 4°C, and the supernatant was discarded. Then, 1 mL of 75% ethanol was added to each tube, and each tube was inverted for mixing. After another centrifugation at 15,000 × *g* for 5 min at 4°C,

the supernatant was removed. The tubes were centrifuged for 1 min, and DNA deposited at the bottom of each tube was resuspended in 50 μ L of 1 $\times$ TE after drying at room temperature.

1.2.2 Primer design

Primers complimentary to the known T-DNA sequences were designed using the Serial Cloner 2-6-1 software, DNAMAN (a sequence alignment and primer design software that can be used offline) . Unknown genomic sequences were amplified from the 3' end of the T-DNA sequence. The binding site of the first primer (IPCR-R1) was located at approximately 300 bp from the 5' end of the T-DNA sequence, and that of the second primer (IPCR-R2) was located at approximately 200 bp from the 5' end of the T-DNA sequence. During amplification, primer IPCR-R1 is paired with primer IPCR-F1, and primer IPCR-R2 is paired with primer IPCR-F2. To improve product specificity, the annealing temperature of the second PCR primer was 3-5°C higher than that of the first PCR primer.

1.2.3 Enzymatic digestion, cyclization, and PCR amplification

An appropriate restriction endonuclease, selected on the basis of the known sequence, was used to prepare the following reaction:

10 $\times$ Buffer	1.5 μ L
Enzyme	0.2 μ L
Template DNA	3.5 μ L
ddH ₂ O	Up to 15 μ L

Enzymatic digestion was performed at 37°C for 2-3 h. The enzyme digestion product was transferred to a 1.5 mL centrifuge tube, washed with ddH₂O, and diluted to 250 μ L. Then, an equal volume of phenol to chloroform (1:1, v/v) mixture was added to the sample, mixed well by vortexing, and incubated at room temperature for 1 min. Centrifugation was conducted at maximum speed (13,000 rpm) for 1 min. The supernatant was transferred to another tube, and an equal volume of chloroform was added to the sample and mixed by vortexing. After incubation at room temperature for 1 min, the sample was centrifuged at maximum speed (13,000 rpm) for 2 min, and the supernatant was transferred to another tube. Then, 2.5 volumes of absolute ethanol prechilled at -20°C and 0.1 volume of 3 M sodium acetate (pH = 5.2) were added to the sample and mixed gently by shaking. After incubation at room temperature for 5 min, the sample was centrifuged at maximum speed (13,000 rpm) for 10–15 min at 4°C. The supernatant was discarded, and the liquid on the tube wall was removed as much as possible. Then, 1 mL of 70% ethanol prechilled at -20°C was added to the tube, and the tube was inverted several times. The sample was centrifuged at maximum speed (13,000 rpm) for 5 min, and ethanol was removed. To dry the sample, the tube was placed horizontally on a clean table, and the ethanol on the tube wall was allowed to volatilize. The enzyme-digested fragments were cyclized overnight at 22°C by preparing the following reaction:

T4 ligase buffer	0.5 μ L
T4 ligase	0.1 μ L
ddH ₂ O	4.4 μ L

Then, PCR was performed as follows. In the first round, the cyclized product was used as a template and amplified under the following conditions: 95°C for 5 min; 35 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 2 min; 72°C for 5 min; and 15°C hold. In the second round, the PCR product of the first round was used as the template and amplified under the following conditions: 95°C for 5 min; 35 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 2 min; 72°C for 5 min; and 15°C hold. Subsequently, a 2 μ L volume of the PCR products was analyzed by electrophoresis on 1% agarose gel at a voltage of 240 V and current of 260 mA. After electrophoresis, DNA was extracted from the gel and submitted to BGI for sequencing.

1.2.4 DNA sequence analysis

The sequence of the PCR product was analyzed using NCBI (<https://blast.ncbi.nlm.nih.gov/>), Phytozome (<https://phytozome.jgi.doe.gov/>) and Gramene (<http://ensembl.gramene.org/>) databases.

2. Results

2.1 Endonuclease selection and IPCR primer design

Theoretically, multiple randomly selected restriction enzymes can be used to perform IPCR. However, in reality, endonucleases must be selected by considering the primer binding site to avoid a situation where no suitable primer design site is possible after enzyme digestion, which would affect subsequent experiments.

In addition, the length, annealing temperature, GC content, and binding specificity of conventional primers, the distance between primers and the unknown–known sequence boundary should all be considered during the design process of IPCR primers. In this study, we found that amplification is not ideal when the distance between the primer and unknown–known sequence boundary is too short or too long; it is not only difficult to design primers, but also impossible to amplify the target sequence. Therefore, to ensure that sufficient sequence is available for designing primers, the cleavage site of a restriction endonuclease should be at least 400 bp away from the boundary. Considering this rule, we selected five restriction endonucleases (from 5' to 3' end): *HhaI*, *AclII*, *SacII*, *NheI*, and *NdeI* (Fig. 2A).

2.2 Effect of different endonucleases on PCR amplification

Through multiple experiments using five pre-selected restriction enzymes, we found that the final effects of *AclII*, *SacII*, and *NheI* were not ideal, and the effect of PCR amplification using the product digested by *AclII* and *NheI* enzyme is unstable. When PCR was amplified by 96 samples and gel electrophoresis was performed, sometimes only 3 samples had obvious bands, sometimes about 20 samples had obvious

bands. *SacII* produced miscellaneous bands when amplified. Only *HhaI* and *NdeI* produced ideal results. Generally, 60–87 out of 96 samples (per plate) digested with *NdeI* or *HhaI* could be successfully amplified by IPCR. Therefore, *HhaI* and *NdeI* were used in all subsequent analyses.

Using *HhaI*-digested samples as the template, IPCR amplification generated relatively short (150–750 bp) and consistent bands, with poor specificity (Fig. 3A). After *NdeI* digestion, IPCR produced large but mostly inconsistent bands, with good specificity (Fig. 3B). Gel electrophoresis strip gradient is diversified, and the size of the bands ranged from 150 to 2,000 bp.

2.3 Effect of primer–boundary distance on PCR amplification

Through a large number of experiments on *HhaI* and *NdeI* restriction endonucleases, we found that the distance between the nearest primer and the unknown–known sequence boundary is key to determining whether the target sequence could be expanded in the second round of IPCR. Because of the limitations of the current Sanger sequencing method [21], dozens of base pairs, or even more than 100 bp, may be mismatched or missing during the sequencing process. When primers were located too close to the boundary (less than 100 bp) in the second round of PCR (IPCR-R2), it was difficult to obtain the complete, or even partial, sequence of the boundary region adjacent to the Ds transposon (Fig. 4A). Therefore, when designing the second round of nested PCR primers (IPCR-R2), the distance between the primers and the boundary region should be greater than 100 bp (ideally ~200 bp; Fig. 4B), which can effectively expand to the unknown sequence of the flanking.

2.4 Analysis of sequencing results

A total of 327 sequences were amplified: 118 from *HhaI*-digested samples and 209 from *NdeI*-digested samples (Table 1). We examined the length distribution range of the obtained sequences. The results showed that sequences generated by the 4 bp cutter *HhaI* were mainly concentrated in the range of 100–200 bp (57 out of 118 sequences; 48%), and the lowest distribution range was in the ≥500 bp interval, with only 4 out of 118 sequences (3%). Among the sequences obtained using the 6 bp cutter *NdeI*, most were concentrated in the range of 200–500 bp (94 out of 209 sequences; 45%), and the lowest distribution range was in the ≤30 bp interval (8 out of 209 sequences; 4%). The final sequencing results showed that sequences amplified from *HhaI*-digested samples were short, while those amplified from *NdeI*-digested samples were long. Thus, if the objective is to amplify longer sequences or with high accuracy, endonucleases with 6 bp restriction sites should be selected.

Table 1 Length distribution of the amplified sequences

Number of sequences within a specific length distribution range						
	≤30 bp	30–100 bp	100–200 bp	200–500 bp	≥500 bp	Total
<i>HhaI</i>	13	32	57	12	4	118
<i>NdeI</i>	8	34	63	94	10	209

2.5 PCR validation of flanking sequences

Some amplified sequences could not be successfully aligned to sequences available in the target species database (Fig. 5A). Therefore, to verify whether the sequence in the target DNA is correct, we designed primers relevant to the sample and verified all sequences obtained by sequencing. One of the primers was specific to the Ds transposon (IPCR-R1 or IPCR-R2), while the other primer was specific to the unknown flanking sequence obtained by amplification. After amplification verification, all verification sequences could again be amplified to the expected target band. Sequence alignment showed that some of the sequences were amplified from the Ds transposon, and some were amplified from genomic DNA (Fig. 5B), thus proving that IPCR is a feasible technique that can be effectively used to obtain unknown DNA sequences.

3. Discussion

Many different kinds of PCR-based chromosome walking techniques have been developed to date [22-26]. The IPCR method established in this study is ideal for determining unknown sequences as it is a relatively simple, low-input, and economical method that involves convenient primer design and can be used on a large number of experimental materials simultaneously.

Through the application of the IPCR technology in *Brachypodium distachyon*, reached the following conclusions. First, the genomic DNA extraction method affects the results of IPCR and the self-ligation of enzyme-digested DNA fragments [27]. DNA extracted using the traditional plant DNA extraction method (improved CTAB method) was ideal as a template for IPCR as it was of appropriate purity and could be effectively digested by different restriction endonucleases and effectively cyclized in subsequent experiments. Second, the selection of restriction endonucleases is critical for successful IPCR amplification. This is because the distance between the cleavage site of the restriction enzyme and the left and right boundaries affects DNA amplification; a large distance reduces the amplification efficiency of ordinary polymerases, and a short distance leads to either multiple amplification products (because of random binding to fragments) or no amplification. Other factors such as price, availability, and potential star activity should also be considered while selecting the restriction enzyme. Third, primer design and site selection are key to the success of IPCR. Nested PCR is usually used for chromosome walking, and the credibility of IPCR products can be greatly increased through two or three rounds of nested PCR and by comparing the PCR product size and concentration. During the primer design phase, in addition to the annealing temperature, length, GC content, and other attributes of conventional primers, the distance between two amplification primers should also be fully considered.

In the current study, we established a convenient and feasible IPCR method through the amplification and analysis of 327 *B. distachyon* sequences. The screening of different restriction endonucleases and primers and the analysis of sequencing results revealed that *HhaI* and *NdeI* restriction endonucleases were suitable for the digestion of *B. distachyon* DNA, and *NdeI* was a better choice than *HhaI* as it

produced condensed bands of different sizes with good specificity. The experimental method and technical scheme developed in this study could be used to design IPCR primers in other studies, and will serve as a reference for the acquisition of unknown sequences and gene mapping in other species.

Declarations

Funding statement

This work was supported by the Shaanxi Provincial Department of Education Key Scientific Research Project (21JY008) ; by Shangluo University Science and Technology Research Project (20SKY010); Natural Science Basic Research Program of Shaanxi (2022JQ-223); Supported By Young Talent fund of University Association for Science and Technology in Shaanxi, China(20220213); Shangluo Science and Technology Plan Project(2022-Y-0051).

Acknowledgments

We are grateful to Dr. Hailong An (Shandong Agricultural University) for providing us the Ds vector and PCR test equipment. Thank you very much for Taniya Dhillon's careful polishing of this paper.

Declaration of competing interest

The authors declare no conflict of interest.

Author Contribution

X.X. carried out the experimental design and thesis writing, W.L. and Q.L. carried out the experimental operation, W.C. and L.S. revised and improved the thesis, and Z.W. and Y.Z. carried out the picture processing and table organization.

References

1. Jeltsch A, Adam S, Dukatz M et al (2021) Deep Enzymology Studies on DNA Methyltransferases Reveal Novel Connections between Flanking Sequences and Enzyme Activity. *Journal of Molecular Biology* 433, 1-15.
2. Wang JY , Bi XT , Chen W et al (2021) Identification of the insertion site of transgenic DNA based on cyclization of the target gene with the flanking sequence and nested inverse PCR. *Talanta Open* 3: 1-6.
3. Rishi AS, Nelson ND, Goyal A (2004) Genome walking of large fragments: An improved method. *Journal of Biotechnology* 111 (1): 9-15.
4. Boratyn GM, Thierry-Mieg J, Thierry-Mieg D et al (2019) Magic-BLAST, an accurate RNA-seq aligner for long and short reads. *BMC Bioinformatics* 20: 1-19.

5. Antal Z, Rascle C, Fevre M, Bruel C (2004) Single oligonucleotide nested PCR: A rapid method for the isolation of genes and their flanking regions from expressed sequence tags. *Current Genetics* 46(4): 240-246.
6. Tanksley SD, Ganai MW, Martin GB (1995) Chromosome landing: A paradigm for map-based gene cloning in plants with large genomes. *Trends in Genetics* 11(2): 63-68.
7. Kim SR, Jeon JS, An G (2011) Development of an efficient inverse PCR method for isolating gene tags from T-DNA insertional mutants in rice. *Methods in Molecular Biology* 678: 139-146.
8. Rogers YC, Munk AC, Meincke LJ, Han CS (2005) Closing bacterial genomic sequence gaps with adaptor-PCR. *Biotechniques* 39(1): 31-33.
9. Liu YG, Whittier RF (1995) Thermal asymmetric interlaced PCR: Automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. *Genomics* 25(3): 674-681.
10. Dalal M, Chinnusamy V, Bansal K (2010) Isolation and functional characterization of lycopene beta-cyclase (CYC-B) promoter from *Solanum habrochaites*. *BMC Plant Biology* 10 (1): 61-75.
11. Fauteux F, Stromvik MV (2009) Seed storage protein gene promoters contain conserved DNA motifs in Brassicaceae, Fabaceae and Poaceae. *BMC Plant Biology* 9: 126.
12. Ochman H, Gerber AS, Hartl DL (1988) Genetic applications of an inverse polymerase chain reaction. *Genetics* 120(3): 621-623.
13. Abelleyro MM, Marchione VD, Palmitelli M et al (2019) Inverse PCR to perform long-distance haplotyping: main applications to improve preimplantation genetic diagnosis in hemophilia. *European Journal of Human Genetics* 27(4): 603-611.
14. Chen L, Tu Z, Hussain J et al (2010) Isolation and heterologous transformation analysis of a pollen-specific promoter from wheat (*Triticum aestivum* L.). *Molecular Biology Reports* 37(2): 737-744.
15. Tsaftaris A, Pasentzis K, Argiriou A (2010) Rolling circle amplification of genomic templates for inverse PCR (RCA-GIP): A method for 5'- and 3'-genome walking without anchoring. *Biotechnology Letters* 32(1): 157-161.
16. Carmen RL , Pamela RC , Beatriz LI et al (2005) Genotyping the hemophilia inversion hotspot by use of inverse PCR. *Clin Chem* 51: 1154-8.
17. Yu QG, Hu NJ, Lu YN, Vivek R (2001) Nerurkar and Richard Yanagihara. Rapid acquisition of entire DNA polymerase gene of a novel herpesvirus from green turtle fibropapilloma by a genomic walking technique. *Journal of Virological Methods* 91(2): 183-195.
18. Mine A, Matsumoto A, Schlueter T et al (2022) A versatile Tn7 transposon-based bioluminescence tagging tool for quantitative and spatial detection of bacteria in plants. *Plant Communications* 3(1): 1-15.
19. Qu S , Desai A , Wing R et al (2008) A versatile transposon-based activation tag vector system for functional genomics in cereals and other monocot plants. *Plant Physiology* 146(1): 189-199.

20. Park SH, JunN S, Kim CM et al (2007) Analysis of gene-trap Ds rice populations in Korea. *Plant Molecular Biology* 65(4): 373-384.
21. Anna enovská a, Eva Drozdová a, Kristna Brzobohatá a et al (2021) Sanger sequencing of mitochondrial hypervariable region of ancient samples for DNA authentication and screening before high-throughput sequencing. *Journal of Archaeological Science: Reports* 40: 1-5.
22. Li FJ, Bai HK, Xiong YW et al (2015) Molecular characterization of insulin-like androgenic gland hormone-binding protein gene from the oriental river prawn *Macrobrachium nipponense* and investigation of its transcriptional relationship with the insulin-like androgenic gland hormone gene. *General and Comparative Endocrinology* 216: 152-160
23. Song J, Liu J, Weng M et al (2013) Cloning of galactinol synthase gene from *Ammopiptanthus mongolicus* and its expression in transgenic *Photinia serrulata* plants. *Gene* 513(1): 118-127.
24. Yan YX, An CC, Li L et al (2003) 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* 31: e68.
25. Terauchi R, Kahl G (2000) Rapid isolation of promoter sequences by TAIL-PCR: The 5-flanking regions of Pal and Pgi genes from yams(*Dioscorea*). *Mol Gen Genet* 263: 554-560.
26. Harreither W, Sygmund C, Dunhofen E et al (2009) Cellobiose dehydrogenase from the ligninolytic basidiomycete *Ceriporiopsis subvermispota*. *Applied and Environmental Microbiology* 75(9): 2750-2757.
27. Pang KM, Knecht DA (1997) Partial inverse PCR: a technique for cloning flanking sequences. *BioTechniques* 22(6): 1046-1048

Figures

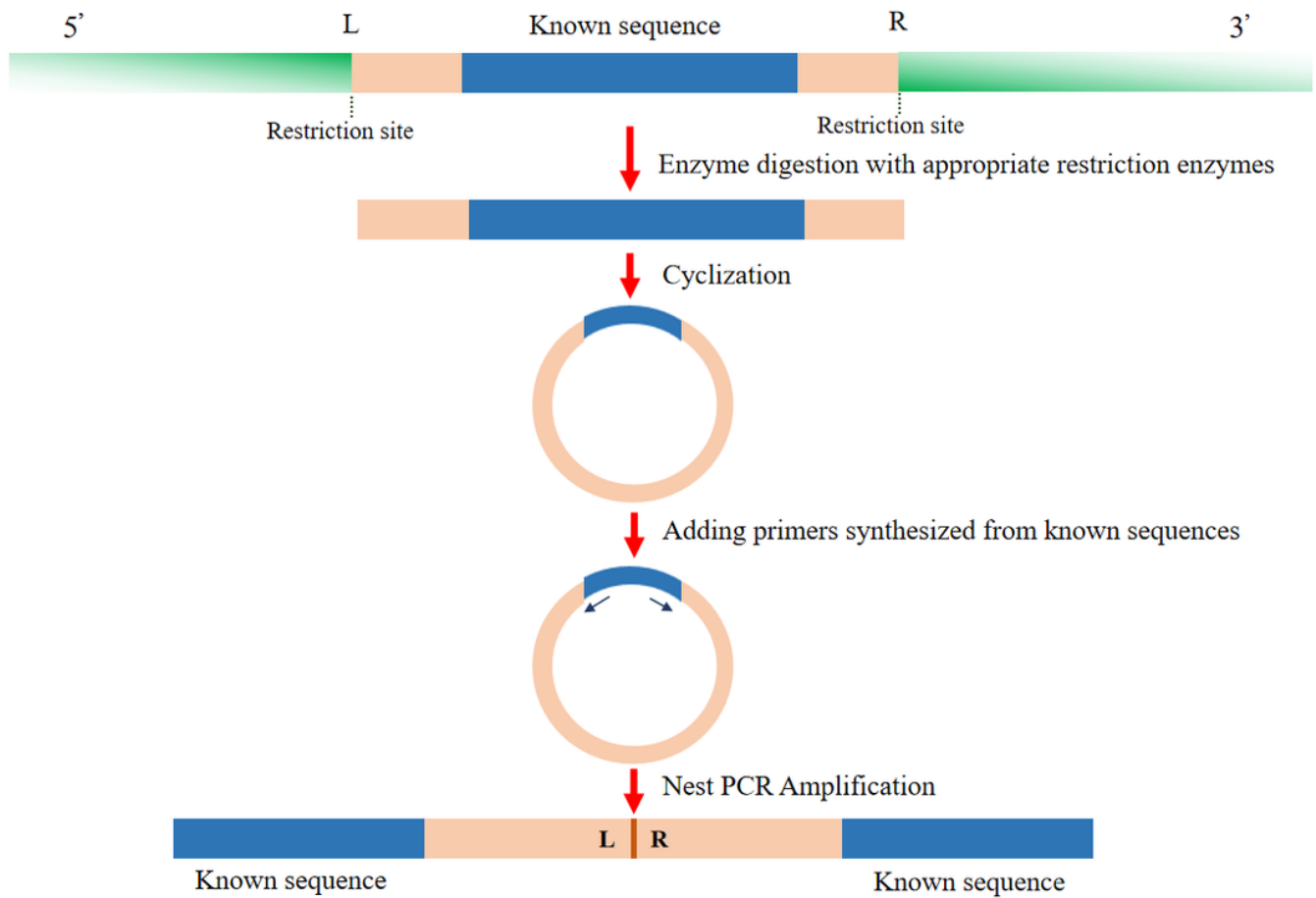


Figure 1

Schematic diagram of Inverse PCR principle.

GTAGGATGGAAAATCC is the outermost end of the known Ds sequence, and the left outer DNA sequence of the known Ds sequence is unknown sequence of IPCR pre-amplification.

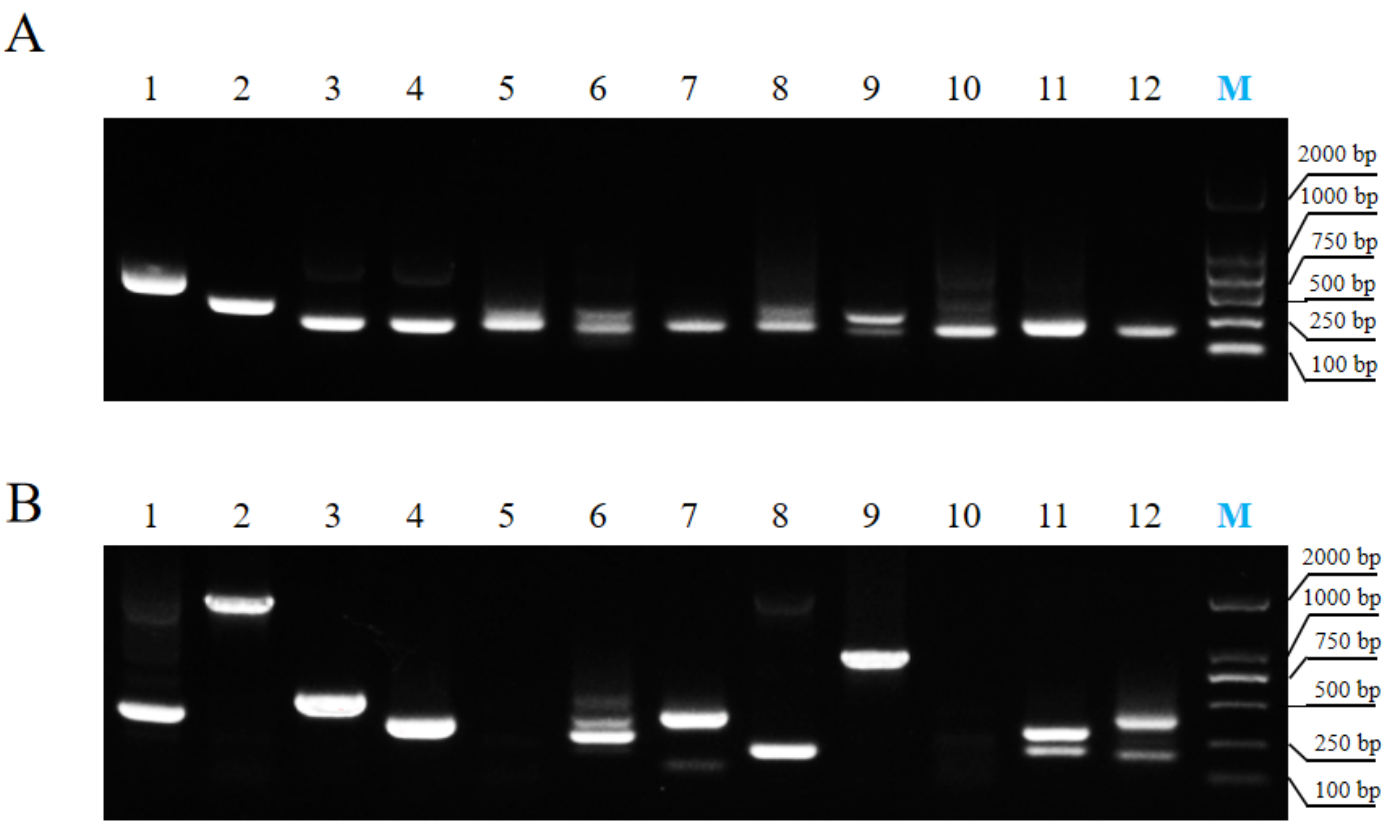


Figure 3

Effect of different restriction endonucleases on amplification. A, B: Effect of endonucleases with 4 bp restriction site (A) and 6 bp restriction site (B) on IPCR amplification. Lane M indicates the 2 kb DNA marker.

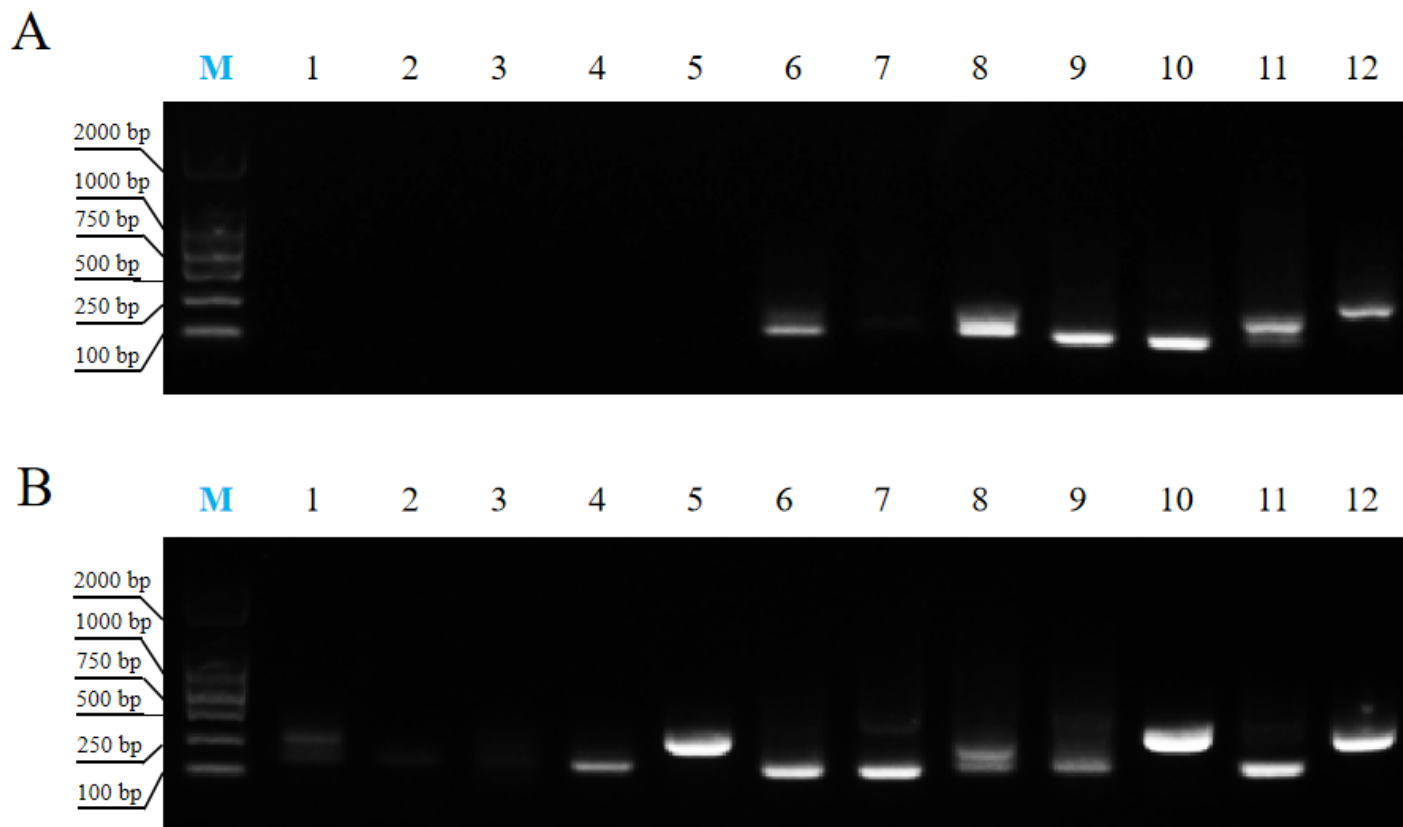


Figure 4

Effect of the distance between the primer binding site and the boundary region on IPCR amplification. A,B: Amplification using primers located approximately 100 bp from the boundary sequence (A) and approximately 200 bp from the boundary sequence (B). Lane M shows the 2 kb DNA marker.

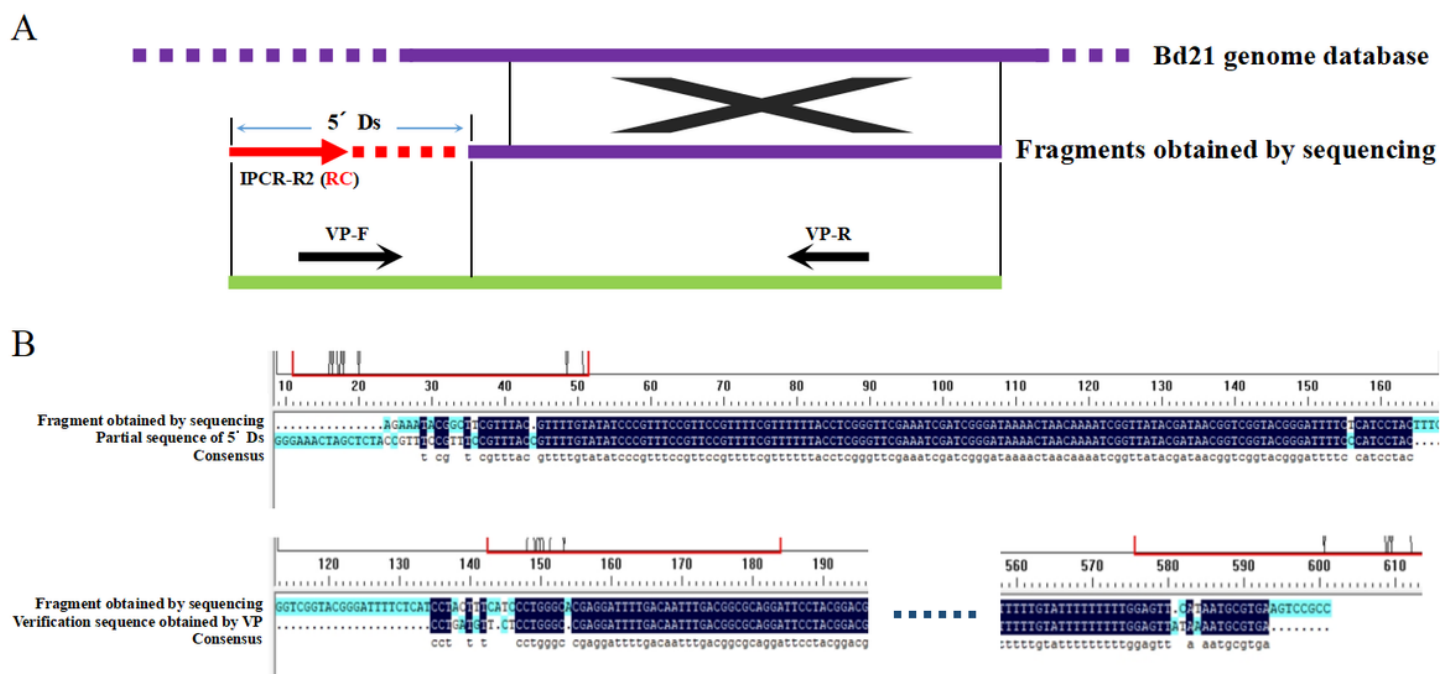


Figure 5

Comparison and verification of amplified sequences. A: Analysis of the amplified sequence using Gramene BLAST (length, 450 bp; score, 438; e-value, 0.0; sequence identity, 99.3%). RC: Reverse complementarity. B: Validation of the amplified sequence. Alignment of the amplified sequence with the Ds element (top) and with the verified sequence (bottom). VP: Validation primer.