African Journal of Biotechnology Vol. 7 (20), pp. 3582-3586, 20 October, 2008
Available online at http://www.academicjournals.org/AJB
ISSN 1684–5315 © 2008 Academic Journals

Full Length Research Paper

Cloning and analysis of the 5' and 3' flanking regions of the *Crinum asiaticum* agglutinin gene by genomic walking

Juan Lin[1]*, Yuanjie Jin[1], Junhui Wang[2] and Kexuan Tang[1]

[1]State Key Laboratory of Genetic Engineering, School of Life Sciences, Morgan-Tan International Center for Life Sciences, Fudan-SJTU-Nottingham Plant Biotechnology R and D Center, Fudan University, Shanghai 200433, China.
[2] Research Institute of Forestry, Key Laboratory of Tree Breeding and Cultivation, State Forestry Administration, Chinese Academy of Forestry, Beijing 100091, China.

Accepted 5 September, 2008

We reported a simple and efficient method, which combines restriction endonuclease digestion and adaptor ligation, for cloning unknown genomic sequences adjacent to a known sequence. After total genomic DNA is completely digested with the different sticky-end restriction endonuclease separately, the ends are full. The DNA fragments with blunt-end were then ligated separately to the adaptor. The adaptor-ligated genomic DNA fragments are used as template for cloning flanking regions from all sequence of interest. A first round PCR is performed with a gene-specific primer and the adaptor primer at its 5' and 3' end. This is followed by second PCR amplification with a nested gene-specific primer and the nested adaptor primer. Finally, the amplified products are fractionated, cloned, and sequenced. Using this method, we cloned the 5' and 3' flanking region of a mannose-binding lectin gene based upon DNA fragment obtained from China Crinum (*Crinum asiaticum* var. (Roxb. ex Herb.) Barker).

Key words: *Crinum asiaticum*, *Crinum asiaticum* agglutinin, genomic walker technology, mannose-binding lectin.

INTRODUCTION

Several polymerase chain reaction (PCR) methods have been described to amplify unknown DNA fragments adjacent to known DNA regions using only gene specific primer. The methods are of three types: inverse PCR (Huang, 1994), randomly primed PCR (Zou et al., 2003) and adaptor ligation PCR (Willems, 1998). The adaptor ligation PCR method is using adaptors or tailing to add a known DNA sequence to a restriction site of the unknown DNA. The efficiency of adaptor ligation PCR method depended on a favorable restriction map and suitable adaptor. The sequence of adaptor and primer has been improved by Siebert et al. in 1995 (Siebert et al. 1995). One end of the adaptor is blunt so that it will ligate to both ends of any DNA fragment generated by restriction endonuclease that yields blunt ends. For cloning with adaptor ligation PCR technique, it selected usually digested genomic DNA with blunt-end restriction endonuclease. But the types of blunt-end restriction endonuclease are small. The numbers of blunt-end restriction endonuclease sites are limited and its digest are ineffective to some plant species. The distributions of blunt-end restriction endonuclease sites are asymmetric; the interval between the blunt-end restriction endonuclease sites and gene specific primer is too long. So the result of PCR amplification is imperfect. It is difficult to find suitable blunt-end restriction endonuclease to some plant species. Therefore, for many especial plant materials, the selection of restriction enzyme must give greater scope.

To overcome the limitation of restriction endonuclease used, we developed a simple method for efficiently cloning genomic flanking DNA from plant using completely digested with common sticky-end restriction endonu-

*Corresponding author. E-mail: linjuan@fudan.edu.cn. Tel: +86-21-65642425. Fax: +86-21-65642425.

Abbreviation: GSP, gene specific primer; **RE**, restriction endonuclease; **CAA**, *Crinum asiaticum* agglutinin; **AP**, adaptor primer; **NAP**, nest adaptor primer.

clease. The genomic DNA was completely digested with different sticky-end restriction endonuclease separately, and the end was filled. The genomic DNA fragments with blunt-end were then ligated separately to the adaptor. The adaptor-ligated genomic DNA fragments are used as template for PCR amplification with a gene-specific primer and the adaptor primer at its 5' and 3' end. Nested PCR was performed using a second (internal) gene-specific primer and nested adaptor primer.

Monocot mannose-binding lectins belong to the super family of plant lectin. They are widely distributed in monocotyledon plants and are believed to play a role in recognition of high-mannose type glycans of foreign micro-organisms or plant predators (Barre et al., 2001). In recent years an increasing interest is drawn to this group of agglutinins for some reasons. One is because of their unique and exclusive specificities towards mannose and these lectins have become very interesting tools in glycol-conjugate research (Haselbeck et al., 1990). The other is the discovery that this lectin exhibits striking toxicity to sap-sucking insects (Hilder et al., 1995) and microbicides (Balzarini et al., 2004). So the cloning of monocot mannose-binding lectins gene from many species of monocotyledon plants Amaryllidaceae family such as *Zephyranthes candida* (Wu et al., 2006), *Amaryllis vittata* (Wu et al., 2004), *Crinum asiaticum* (Chai et al., 2003), *Z. grandiflora* (Kai et al., 2003), *Lycoris radiata* (Zhao et al., 2003), *Clivia miniata* (Van Damme et al. 1994), *Galanthus nivalis* (Van Damme et al. 1991) were most extensively studied and well documented. But 5' and 3' flanking regions of the monocot mannose-binding lectins gene from Amaryllidaceae family have not been reported. By using the method outlined here, we successfully cloned the upstream and downstream region of a mannose-binding lectin gene based upon DNA partial sequence obtained from China Crinum (*C. asiaticum* var. (Roxb. ex Herb.) Barker). The cloning of the gene provides a basis for further study of the gene structure, expression, regulation and evolution mechanism.

MATERIALS AND METHODS

Plant material

The corms of China *C. asiaticum* var. (Roxb. ex Herb.) Barker were gathered from the Medical Plants Garden, School of Pharmacy, Second Military Medical University, Shanghai, P. R. China. The corms were grown in pots in the green house under standard conditions. Leaves were collected from two-month-old seedlings. The materials were stored at -70°C until use.

DNA isolation and purification

Genomic DNA was extracted by the method of Dellaporta (Dellaporta et al. 1983).

Genomic digests and ligations

The genomic DNA was completely digested with different sticky-end (*Hind*III, *Eco*RI, *Bam*HI, *Pst*I) and blunt-end restriction enzymes (*Dra*I, *Eco*RV, *Pvu*II, *Stu*I) (Bio-Lab) separately. The genomic DNA fragments with blunt-end were then ligated separately to the adaptor. The genomic DNA fragments with sticky-end were filled and then ligated separately to the adaptor.

PCR amplification

The amplification of upstream sequence of *Crinum asiaticum agglutinin* (CAA) genomic DNA consisted of two PCR amplifications per library. The primary PCR used the outer adaptor primer (AP) (5'-GTAATACGACTCACTATAGGGC-3') provided within the kit and an inner, gene-specific primer 5GSP1 (5'-ACTAATTACCACCTGCAGCCGTCGT-3'). The amplification was performed in a PTC-100^{TM} programmable (MJ Research, INC) for 7 cycles (25 sec at 94°C, 3 min at 72°C) and then 32 cycles (25 s at 94°C, 3 min at 67°C) followed by extension for 7 min at 67°C. The primary PCR mixture was diluted and used as a template for nested PCR with the nested adaptor primer (NAP) (5'-ACTATAGGGCACGCGTGGT-3') provided within the kit and a nested gene-specific primer 5GSP2 (5'-TGACGTTGGTGGGGCCATGAGCGGC-3'). The amplification was performed in a GeneAmp PCR System 2400 for 5 cycles (25 s at 94°C, 3 min at 72°C, and then 20 cycles (25 s at 94°C, 3 min at 67°C) followed by extension for 7 min at 67°C.

The amplification of downstream sequence of CAA genomic DNA consisted of two PCR amplifications per library. The primary PCR used the outer adaptor primer AP and an outer, gene-specific primer 3GSP1 (5'-ACCACAAGCCACCATCCGCTGC -3'). The nested PCR used the nested adaptor primer NAP and a nested gene-specific primer 3GSP2 (5'-GCCAACGTGCTGTACTCGGG-3'). The conditions of PCR reaction were the same as mentioned above.

Cloning and sequencing of PCR products

All the PCR products were purified using Gel Extraction Mini Kit (Watson, China), ligated to pMD18-T vectors (TaKaRa, P. R. China), transformed into *E. coli* strain DH5α and then sequenced with DYEnamic Direct dGTP Sequencing Kit (Amersham) and a 373A DNA sequencer.

Sequence analysis

The analysis and comparison of 5' and 3' flanking regions sequence of the *C. asiaticum* agglutinin gene using published sequences of MBL were performed with BLASTN (Standard Nucleotide-Nucleotide BLAST) on NCBI (www.ncbi.nlm.nih. gov) and Vector NTI Suite 8.0. Promoter motifs and transcription start site of 5' upstream were analyzed using the PlantCARE database (a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences).

RESULTS AND DISCUSSION

Construction of genome walker DNA libraries

We selected four blunt-end restriction enzymes (*Dra*I, *Eco*RV, *Pvu*II, *Stu*I) and four sticky-end restriction enzymes (*Hind*III, *Eco*RI, *Bam*HI, *Pst*I) to digest the genomic DNA of *C. asiaticum.* The results showed among blunt-end restriction enzymes, *Pvu*II only can completely digest

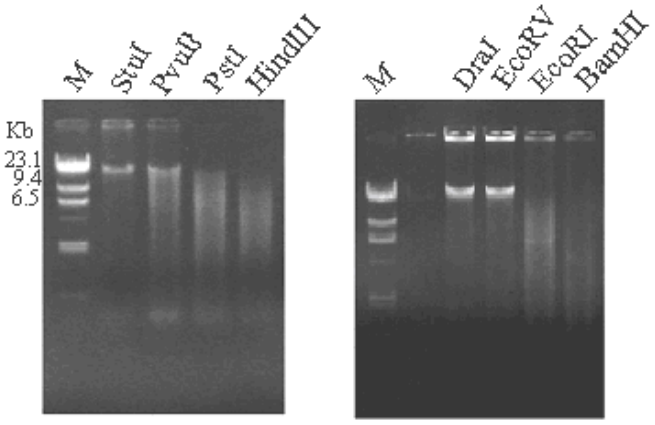

Figure 1. Electrophoresis photos of digested DNA of *Crinum asiaticum* leaves. Left: Lane 1 shows DNA marker *Hin*dIII (TaKaRa Biotechnology Co. Ltd., Dalian of P. R. China); lanes 2 - 5 shows the digested DNA using four restriction enzymes (*Stu*I, *Pvu*II, *Pst*I, *Hin*dIII); Right: Lane 1 shows DNA marker *Hin*dIII (TaKaRa Biotechnology Co. Ltd., Dalian of P. R. China); lanes 2-5 shows the digested DNA using four restriction enzymes (*Dra*I, *Eco*RV, *Eco*RI, *Bam*HI).

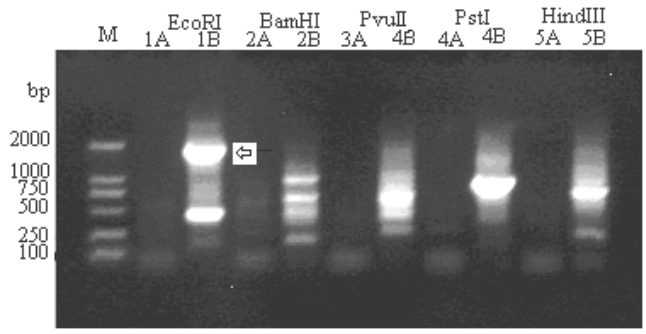

Figure 2. Electrophoresis photos of upstream sequence amplification of *Crinum asiaticum* agglutinin gene. Lane 1 shows DNA marker DL2000 (TaKaRa Biotechnology Co. Ltd., Dalian of P. R. China); Lanes 2 to 10 shows the amplification of upstream sequence; Lanes 1A, 2A, 3A, 4A, 5A shows the amplification of primary PCR; lanes 1B, 2B, 3B, 4B, 5B shows the amplification using nested PCR of digested genomic DNA with five restriction enzymes (*Eco*RI, *Bam*HI, *Pvu*II, *Pst*I, *Hin*dIII) (the arrowhead shows the upstream amplification sequence).

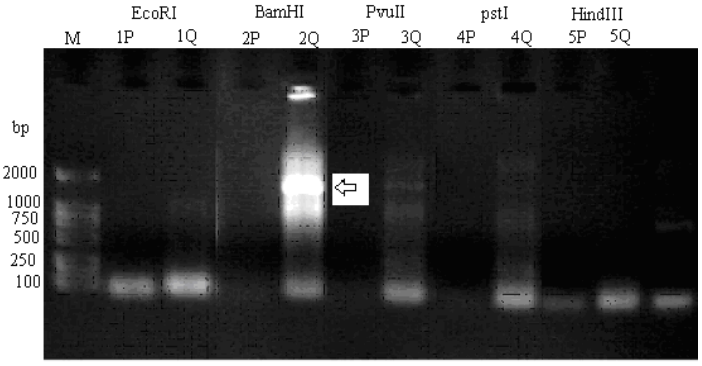

Figure 3. Electrophoresis photos of downstream sequence amplification of *Crinum asiaticum* agglutinin gene. Lane 1 shows DNA marker DL 2000 (TaKaRa Biotechnology Co. Ltd., Dalian of P. R. China); lanes 2 to 10 shows the amplification of downstream sequence; Lanes 1P, 2P, 3P, 4P, 5P shows the amplification of primary PCR; lanes 1Q, 2Q, 3Q, 4Q, 5Q shows the amplification using nested PCR of digested genomic DNA with five restriction enzymes (*Eco*RI, *Bam*HI, *Pvu*II, *Pst*I, *Hin*dIII) (the arrowhead shows the upstream amplification sequence).

the genomic DNA, while *Eco*RV and *Dra*I, *Stu*I cannot completely digest the genomic DNA. And among sticky-end restriction enzymes, *Hind*III, *Eco*RI, *Bam*HI and *Pst*I can completely digest the genomic DNA (Figure 1). The genomic DNA was completely digested with five restriction enzymes (*Pvu*II, *Hind*III, *Eco*RI, *Bam*HI and *Pst*I) separately and the DNA fragments were then ligated separately to the GenomeWalker adaptor. The adaptor-ligated genomic DNA fragments were referred for convenience as GenomeWalker 'libraries'.

Amplification of the 5' and 3' flanking region nucleotide sequence

According to the sequence of the cDNA sequence and DNA partial sequence (GenBank Acc. No. AY212158, AY212159), four specific primers (5GSP1 and 5GSP2, 3GSP1 and 3GSP2) were designed and synthesized, and GenomeWalker DNA libraries were constructed. Following PCR amplification, the DNA fragment of upstream sequence showed many bands in different template in 1.0% agarose gel (Figure 2). There is a specific band of about 1900 bp in lane 3 using *Eco*RI digested genomic DNA as template. Sequence analysis showed that 535 bp of the fragments is overlapped to CAA cDNA 5' end sequence. The DNA fragment of downstream sequence also showed many bands using different template in 1.0% agarose gel (Figure 3). There is a specific band of about 1600 bp in lane 5 using *Bam*HI digested genomic DNA as template. Sequence analysis showed that 676 bp of 5' end of the fragments is overlapped to CAA cDNA 3' end sequence.

According to the sequence of the upstream and down-stream of genomic DNA, two specific primers (CAAF2: 5'-AAAGATAAGCAAGATAACAATGTATTTAAT-3' and CAAR2:5'-ATCTTTTGTATTAAAATATGGAACTTTCA-3') were designed and synthesized. Sequence analysis revealed that the obtained entire CAA genomic DNA was 3027 bp containing a 528 bp complete ORF encoding a polypeptide of 175 amino acids, a 1403 bp of 5'-upstream sequences and a 1096 bp of 3'-downstream sequences (Figure 4) (GenBank Acc. No. AY452806).

Analysis of transcription start sites

Analysis of transcription start sites using Neural Network

Promoter Prediction (http://promotor.biosino.org/) found a core promoter sequence: GGCAAAATGCTATAAATGGTGGAGGCTCCAAGGCTAA AACCACCACAAGC (score=0.97). The box letter C shows the site of transcription start. The prediction transcription start sites are same with the 5' end of cDNA full-length sequence.

Analysis of the 5' and 3' flanking region nucleotide sequence

The 5' flanking region of the CAA had high content of A+T (58.87%), which was also found in other mannose-binding lectin genes, such as 5' regulatory region of CEA gene (58% of A+T) (GenBank Acc. No. AF178113). Analysis of the promoter sequence of CAA using the PlantCARE database identified a conserved transcription start site at -75 bp positions upstream from the start codon ATG. Usually, the structures of 5' gene flanking region of eukaryotes comprised four parts: the site of start transcription, TATA box, CAAT box and GC box. TATA boxes were generally located at -32±7 bp positions upstream from the start of transcription. The consensus sequence for the TATA box was [T(CG)TATA (TA)A_{1-3}(CT)A] that was important for eukaryotic transcription (Joshi et al., 1987). A TATAA sequence was found to be located at -23 bp positions upstream from the start of transcription in CAA, which might be important for the transcription control as well. The consensus sequence for TATA box was TATAAAT (Figure 4).

Nine CAAT boxes were identified at the -130, -224, -488, -523, -613, -755, -1038, -1232 and -1306 bp positions upstream from the start of transcription (Figure 4). The CAAT box is sometimes important for the efficiency of eukaryotic transcription (Kozak et al., 1981). Usually, CAAT boxes were found at the -77±10 bp positions upstream from the start of transcription, although longer intervals have also been found.

There were two GC boxes (CACGTT) located at -237 and -347 bp positions respectively upstream from the start of transcription. Two inverted repeats, ATAACAATGTA and ATAACAATGTA, were identified at -1303 and -1229 bp positions respectively upstream from the start of transcription (Figure 3).

In addition to cis-acting element, some other elements were also identified in the 5' flanking region of the CAA, which included inducible elements by physiological and environmental factors, e.g. abscisic acid responsiveness (ABRE: CGTACGTGC A), MeJA-responsiveness (CGTCA-motif: CGTCA, TGACG-motif: TGACG), and the tissue or developmental stage specific factors, e.g. endosperm expression (Skn-1-like motif: GTCAT; GCN4 motif: CAAGCCA).

The 3'- flanking region (1096 bp long) of the gene contained 63.96% of A+T. This value was slightly lower than those of other lectin genes, such as ricin gene (70% of A+T) (Halling et al., 1985). A 3'-untranslated region

Figure 4. The genomic DNA sequence and the predicted amino acid sequence of the *Crinum asiaticum* agglutinin gene (*caa*). The start codon (ATG) and the stop codon (TAG) are italicized and in bold. Mannose-binding motifs (QXDXN/FXVXY) are boxed. The putative TATA-box, CAAT-boxes, GC-boxes and polyadenylation signals were underlined with gray background. The upright arrowheads showed the start site of transcription (first letter A) and the site of plus poly (A) (letter T).

(214 bp) followed the CAA- coding region and the comparison analysis revealed this region was perfectly matched in the genomic sequence and cDNA sequence (Cai et al., 2003). There were two sequences in the 3'-untranslated region of the CAA gene that resembled the dual plant gene polyadenylation signals reported before (Halling et al., 1985). The canonical polyadenylation signals $AATAA_{1-3}$, was always found from the 10^{th} to 35^{th} bp upstream from the polyadenylation site in plant genes. The two sequences were found to be located at the 42^{th} bp downstream from the stop codon and at the 14^{th} bp upstream from polyadenylation site in the CAA cDNA clone. Besides cis-acting element, some other elements belonging to tissue or developmental stage specific factors, e.g. the root-specific expression (As1: TGACGTAA), were also identified in the 3' flanking region of the CAA genomic sequence. Three inverted repeats, ACTTTCGACGAAAAAACAAAAATAT, AAAAGGCATATGGGACTACATG and AAGGCATATGGGACTACATGCG, were identified at the 519^{th}, 738^{th} and 930^{th} bp positions respectively downstream from the stop codon (Figure 4).

ACKNOWLEDGEMENTS

This research is financially supported by the National Eleven-Five Key Technology R&D Program of China (2006BAD24B08), National Science Fund for Fostering Talents in Basic Science, Nation Science Education Base at Fudan University (J0630643).

REFERENCES

Balzarini J, Hatse S, Vermeire K, Princen K, Aquaro S, Perno CF, Clercq ED, Egberink H, Mooter GV, Peumans W, Van Damme E, Schols D (2004). Mannose-specific plant lectins from the *Amaryllidaceae* family qualify as efficient microbicides for prevention of human immunodeficiency virus infection. Antimicrob. Agents Chemother. 48: 3858-3870.

Barre A, Bourne Y, Van Damme EJM, Peumans WJ, Rouge P (2001). Mannose-binding plant lectins: different structural scaffolds for a common sugar-recognition process. Biochimie. 83: 645-651.

Chai YR, Pang YZ, Liao ZH, Zhang L, Sun XF, Lu YQ, Wang S, Tang KX. (2003). Molecular cloning and characterization of a mannose-binding lectin gene from *Crinum asiaticum*. J. Plant Physiol. 160: 913-920.

Dellaporta SL, Wood JM, Hicks JB (1983). A plant DNA minipreparation. Plant Mol. Biol. Rep. 1: 19-21.

Halling KC, Halling AC, Murray EE, Ladin BF, Houston LL, Weaver RF (1985). Genomic cloning and characterization of a ricin gene from *Ricinus communis*. Nucl. Acids Res. 13: 8019-8033.

Haselbeck A, Schickaneder E, Vondereltz H, Hosel W (1990). Structural characterization of glycoprotein carbohydrate chains by using digoxigenin-labeled lectins on blots. Anal. Biochem. 191: 25-30.

Huang SH (1994). Inverse polymerase chain reaction: An efficient approach to cloning cDNA ends. Mol. Biotechnol. 2: 15-22.

Hilder VA, Powell KS, Gatehouse AMR, Gatehouse JA (1995). Expression of snowdrop lectin in transgenic tobacco plants results in added protection against aphids. Transgenic Res. 4: 18-25.

Joshi CP (1987). An inspection of the domain between putative TATA box and translation start site in 79 plant genes. Nucleic Acids Res. 15: 6643-6653.

Kai GY, Zheng JG, Zhang L, Pang YZ, Liao ZH, Li ZG, Zhao LX, Sun XF, Tang KX (2003). Molecular cloning of a new lectin gene from *Zephyrathes grandiflora*. DNA Seq. 14: 335-338.

Kozak M (1981). An analysis of 5'-noncoding sequence from 699 vertebrate messenger RNAs. Nucleic Acids Res. 9: 5233-5252.

Siebert PD, Chenchik A, Kellogg DE, Lukyanov KA, Lukyanov SA (1995). An improved PCR method for walking in uncloned genomic DNA. Nucleic Acids Res. 23: 1087-1088.

Van Damme EJM, Smeets K, Van Leuven F, Peumans WJ (1994). Molecular cloning of mannose-binding lectins from *Clivia miniata*. Plant Mol. Biol. 24: 825-830.

Van Damme EJM, Kaku H, Perini F, Goldstein IJ, Peeters B, Yagi F, Decock B, Peumans WJ (1991). Biosynthesis, primary structure and molecular cloning of snowdrop (*Galanthus nivalis* L.) lectin. Eur. J. Biochem. 202: 23-30.

Willems H (1998). Adaptor PCR for the specific amplification of unknown DNA fragments. Biotechniques, 24: 26-28.

Wu CF, An J, He XJ, Deng J, Hong ZX, Liu C, Lu HZ, Li YJ, Wang CJ, Chen F, Bao JK (2004). Molecular cloning of a novel mannose-binding lectin gene from bulbs of *Amaryllis vittata* (Amaryllidaceae). Acta Bot. Sin. 46: 1301-1306.

Wu CF, Li J, An J, Chang LQ, Chen F, Bao JK (2006). Purification, biological activities, and molecular cloning of a novel mannose-binding lectin from bulbs of *Zephyranthes candida* Herb (Amaryllidaceae). J. Integr. Plant Biol. 48: 223-231.

Zhao XY, Yao JH, Sun XF, Tang KX (2003). Molecular cloning and characterization of a novel lectin gene from *Lycoris radiata*. DNA Seq. 14: 223-226.

Zou N, Ditty S, Li B, Lo SC (2003). Random priming PCR strategy to amplify and clone trace amounts of DNA. Biotechniques, 35: 758-760, 762-765.