*Bioscience Reports, Vol. 25, Nos. 5/6, October/December 2005 (© 2005)*
DOI: 10.1007/s10540-005-2895-4

# Cloning and Characterization of an Agglutinin Gene from *Arisaema lobatum*

**Juan Lin,[1,3] Xuanwei Zhou,[2] Yongzhen Pang,[1] Han Gao,[1] Jiong Fei,[2] Guo-an Shen,[2] Jin Wang,[1] Xinsheng Li,[3] Xiaofen Sun,[1,4] and Kexuan Tang[1,2,4]**

A novel agglutinin gene was cloned from *Arisaema lobatum* using SMART RACE-PCR technology. The full-length cDNA of *Arisaema lobatum* agglutinin (*ala*) was 1078 bp and contained a 774 bp open reading frame encoding a lectin precursor (proproprotein) of 258 amino acid residues with a 23 aa signal peptide. ALA contained three mannose-binding sites (QXDXNXVXY) with two-conserved domains of 45% identity, ALA-DOM1 and ALA-DOM2. The three-dimensional structure of ALA was very similar to that of GNA (*Galanthus nivalis* agglutinin). ALA shared varying identities, ranging from 40% to 85%, with mannose-binding lectins from other species of plant families, such as Araceae, Alliaceae, Iridaceae, Lillaceae, Amaryllidaceae and Bromeliaceae. Genomic sequence of *ala* was also cloned using genomic walker technology, and it was found to contain three putative TATA boxes and eight possible CAAT boxes in the 5′-flanking region. No intron was found within the region of genomic sequence. Southern blot analysis indicated that the *ala* belonged to a multi-copy gene family. Expression pattern analysis revealed that the *ala* preferentially expressed in the tissues with the higher expression being found in spadix, bud, leaf, spathe and tuber. The cloning of the *ala* gene not only provides a basis for further investigation of its structure, expression and regulation mechanism, but also enables us to test its potential role in controlling pests and fungal diseases by transferring the gene into plants in the future.

**KEY WORDS:** *Arisaema lobatum*; *Arisaema lobatum* agglutinin (ALA); cDNA cloning; genomic walker technology; mannose-binding lectin; RACE.

## INTRODUCTION

Lectins are plant proteins possessing at least one-catalytic domain that binds reversibly to a specific mono- or oligo-saccharides. One of the main physiological

[1]State Key Laboratory of Genetic Engineering, School of Life Sciences, Morgan-Tan International Center for Life Sciences, Fudan-SJTU-Nottingham Plant Biotechnology R&D Center, Fudan University, Shanghai, 200433, P.R. China.
[2]Plant Biotechnology Research Center, Fudan-SJTU-Nottingham, Plant Biotechnology R&D Center, School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai, 200030, P.R. China.
[3]Shaanxi Key Laboratory of Bio-resources, Shaanxi University of Technology, Hanzhong, 723000, P.R. China.
[4]To whom all correspondence should be addressed. E-mails: kxtang1@yahoo.com, xfsun1@sohu.com.

roles of plant lectins is to mediate defense response in plants (Van Damme *et al.*, 1998). Monocot mannose-binding lectins refer to a superfamily of strictly mannose-specific lectin. Numerous members of this superfamily have been characterized and cloned from species of the families Alliaceae, Amaryllidaceae, Orchidaceae, Bromeliaceae, Liliaceae, Araceae and Iridaceae (Van Damme *et al.*, 1998, 2000; Chai *et al.*, 2003; Fei *et al.*, 2003; Yao *et al.*, 2003a; Zhao *et al.*, 2003). The majority of the cloned lectins are homo-oligomeric lectins that are synthesized as proproteins, which are converted into the mature lectin polypeptides by the co-translational cleavage of a signal peptide and the posttranslational removal of a C-terminal peptide (Van Damme *et al.*, 1991). Aside from the homo-oligomeric lectins, there are three types of hetero-oligomeric forms of lectins. The first type is hetero-oligomeric lectin (e.g. bulb agglutinin from *Allium ursinum*, AUA) (Van Damme *et al.*, 1993). The second type is the hetero-dimer or hetero-tetramer. In some cases, these two subunit types are highly homologous (e.g. bulb agglutinin from *Allium sativum*, ASA) (Van Damme *et al.*, 1992), and in others the homology between the two domains is much lower (e.g. agglutinin from *Arum maculatum*, AMA) (Van Damme *et al.*, 1995). The third type is the hetero-octamer lectin (e.g. tulip lectin from *Tulipa* hybrid, TxLCI) (Van Damme *et al.*, 1996). Mannose-specific lectins are the most reliable ones for crop genetic modification in view of food safety (Van Damme *et al.*, 1998).

The major tuber storage protein of Araceae species is a mannose-binding lectin (Van Damme *et al.*, 1995). Except for AKA (Fei *et al.*, 2003), the lectin from *Amorphophallus konjac* being the homo-oligomeric lectin, all other mannose-binding lectins isolated from family *Araceae* belong to hetero-tetramer lectins, such as the agglutinins from *Colocasia esculenta* (CEA) (Hirai *et al.*, 1993; Bezerra *et al.*, 1995), *Amorphophallus maculatum* (AMA) (Van Damme *et al.*, 1995), *Xanthosoma sagittifolium* (XSA), *Dieffenbachia sequina* (DSA) (Van Damme *et al.*, 1995), *Arisaema heterophyllum* (AHA) (Zhao *et al.*, 2003) and *Pinellia ternata* (PTA) (Yao *et al.*, 2003a).

Many of mannose-binding lectins are the most efficient proteins conferring enhanced resistance to *Homopteran* pests. Previous studies showed that mannose-binding lectins from the family *Amaryllidaceae* had varying insecticidal activities in feeding experiments with both artificial diets and transgenic plants. Such experiments with artificial diets clearly demonstrated that the snowdrop lectin (*Galanthus nivalis* agglutinin, GNA) had detrimental effects on the development and reproduction of pests, especially sap-sucking insects (Rahbé *et al.*, 1995). Transgenic tobacco and rice expressing GNA showed significant insecticidal activity towards peach potato aphids (*Myzus persicae* Sulzer) (Hilder *et al.*, 1995) and rice planthoppers including *Nilaparvata lugens* (Rao *et al.*, 1998; Tang *et al.*, 2001) and *Laodelphax striatellus* (Wu *et al.*, 2002), while transgenic potato expressing GNA showed enhanced resistance to aphids and tomato moth (*Lacanobia oleracea*) (Down *et al.*, 1996; Fitches *et al.*, 1997). Recent studies showed that mannose-binding lectins from family Araceae were also toxic to sap-sucking insects in artificial diet assays and transgenic plants. Such as the lectins from *Pinellia pedatisecta* and *Pinellia ternata* had significant insecticidal activities towards cotton aphids (*Aphis gossypii* Glover) and peach potato aphids (*Myzus persicae* Sulzer) when incorporated into artificial diets at 1.2 g/l and 1.5 g/l, respectively (Huang *et al.*, 1997; Pan *et al.*, 1998), and the lectins from *Arisaema consarguieum* had similar anti-insect properties with GNA, with significant lethal effects to cotton aphids (*Aphis gossypii* Glover) (Li *et al.*, 2000).

Transgenic tobacco expressing PTA significantly inhibited the growth of peach potato aphid (*Myzus persicae* Sulzer) (Yao *et al*., 2003b), while transgenic rice expressing PTA also inhibited the livability and viability of rice brown planthopper (Zhang *et al*., 2003).

*Arisaema lobatum* is a traditional Chinese herb belonging to family Araceae. Under natural conditions, rare diseases and pest damage are found on *A. lobatum* plant. Whether like other Araceae species lectin genes are present in *A. lobatum* and if so, whether the anti-insect property of *A. lobatum* is related to the lectin genes is unknown. In this paper, we report on the cloning of the cDNA and genomic sequence of the lectin from *A. lobatum* using RACE-PCR technology and genomic walker method. The characterization of the cloned gene provides useful information not only on the structure of the gene, but also on the regulation of its expression.

## MATERIALS AND METHODS

### Plant Material

The tubers of *A. lobatum* were collected from Gansu province, China. The tubers were grown in pots in the greenhouse under standard conditions. Root, leaf, spadix, petiole, spathe, bud and tuber were collected and stored at −70°C until use.

### RNA Extraction

Total RNA was extracted from *A. lobatum* tissues by CTAB (cetyltrimethylammonium bromide) based RNA isolation method (Jaakola *et al*., 2001). RNA quality and concentration were checked by agarose gel electrophoresis (EC250-90, E-C Apparatus Corporation) and spectrophotometer (WFZUV-2100, Unico™ Instruments Inc.) analysis, and the RNA samples were stored at −70°C until use.

### 3′ and 5′-RACE of *ala*

The cDNA synthesis was performed with the SMART technology (SMART™ RACE cDNA Amplification Kit) for 5′- and 3′- rapid amplification of cDNA ends (RACE, CLONTECH Laboratories, Inc.). For 3′-RACE, RNA was reversely transcribed with the 3′- RACE CDS Primer A (provided in the kit), and 3′-RACE was performed with primer 1 (Table 1), designed and synthesized (Shanghai Sangon Biotechnological Company) based on the mannose-binding site sequence (MQDDCNL) conserved among most of the monocot mannose-binding lectins. For 5′-RACE, RNA was reversely transcribed with the 5′-RACE CDS Primer and SMART II A Oligonucleotide (provided in the kit). Based on the sequence of the 3′-RACE product, the specific primers 2 and 3 were designed and synthesized. The first round of PCR was performed with primer 2 and Universal Primer A Mix (UMP, provided in the kit). The PCR product was diluted 50-fold for a second round of amplification with primer 3 and Nested Universal Primer A (NUP, provided in the kit). The 5′- and 3′-RACE was performed essentially according to the SMART™ RACE cDNA Amplification Kit user manual. All the primers used in RACE were listed in Table 1.

**Table 1.** Primers Used in the Cloning of Full-Length cDNA and Genomic Sequence of *ala* by RACE and Genomic Walker Technology

| Primer | Primer sequence (5′ → 3′) |
|---|---|
| For the cloning of the full-length cDNA | |
| 3′-RACE CDS primer A | 5′-AAGCAGTGGTATCAACGCAGAGTAC(T)30$N_{-1}$N-3′ (N = A,C,G, or T; $N_{-1}$ = A, G, or C) |
| 5′-RACE CDS primer | 5′-$(T)_{25}N_{-1}N$−3′ |
| SMART II A Oligo | 5′-AAGCAGTGGTATCAACGCAGAGTA-3′ |
| UPM | 5′-CATATACGACTCACTATAGGGC-3′ (Short, 0.4 $\mu$M) and 5′-CTAATACGACTACTATAGG GCAAGCAGTGGTATCAACGCAGAGT-3′ (Long, 2 $\mu$M) |
| NUP | 5′-AAGCAGTGGTATCAACGCAGAGT-3′ |
| Primer 1 | 5′-ATGCAGGATGACTGCAACCT-3′ |
| Primer 2 | 5′-ATATGCACGATAGTTGCCAA-3′ |
| Primer 3 | 5′-GCGGAGCTGGATCGGCTGCT-3′ |
| Primer 4 | 5′-CGCCACACCATCACGGCGAA-3′ |
| Primer 5 | 5′-ATTCAAAAGACCACTTTACA-3′ |
| For the cloning of the genomic sequence | |
| AP | 5′-GTAATACGACTCACTATAGGGC-3′ |
| NAP | 5′-ACTATAGGGCACGCGTGGT-3′ |
| Primer D1 | 5′-AAGCTCCTCCTCTTCC T CCTCC-3′ |
| Primer D2 | 5′-GGAGGATGAAGACGTAGTCACC-3′ |
| Primer D3 | 5′-GCATGACCAAGTCGAAGTCGCCGTTC-3′ |
| Primer D4 | 5′-GGACAGCAGATAGTTGGTGCCG ACTG-3′ |
| Primer D5 | 5′-ACAATATCCCTA TCACGAACAACATG-3′ |
| Primer D6 | 5′-CCACAAGGGCGAGCTCGTCATCAAGG-3′ |
| Primer D7 | 5′-AAAGGA GTGTGCCCCCAGGGCGTGAT-3′ |
| Primer D8 | 5′-ATCTGAATCCGATTAATTTG TTATCGA-3′ |

## The Full-Length cDNA Amplification of *ala*

By aligning the 3′-RACE and 5′-RACE product sequences on Vector NTI™ Suite 8.0, the full-length cDNA of *ala* was deduced and amplified using gene-specific primers 4 and 5 (Table 1). The full-length cDNA amplification was performed according to the One Step RNA PCR Kit (AMV) user manual (TaKaRa, China).

## Genomic DNA Extraction

Genomic DNA was extracted from leaf materials of *A. lobatum* by Dellaporta method (Dellaporta *et al.*, 1983). Leaf materials (1 g) were homogenized in liquid nitrogen in a pre-cooled mortar, transferred to a 50-ml tube containing 5.0 ml of extraction buffer [1 volume of DNA extraction buffer (140 ml 0.25 M sorbitol, 100 ml 1 M Tris pH 8.2, 200 ml 0.25 M EDTA, 560 ml MQ water), 1 volume of nucleic lysis buffer (200 ml 1 M Tris pH 7.5, 200 ml 0.25 M EDTA, 400 ml 5 M NaCl, 20 g CTAB, 200 ml MQ water) and 0.1 volume of sarkosyl (10%) and 0.02 M Na-bisulfate] and mixed gently. The mixture was then placed in 65°C water bath for 1 h and the homogenate was extracted with 7.5 ml of the mixture of chloroform and isoamyl alcohol (24:1). The aqueous phase was removed to a new tube containing 1 volume of cold isopropanol and shaken gently until DNA precipitation. The precipitation was pelleted by centrifugation at 8000 rpm for

10 min and the pellet (DNA) was washed with 70% ethanol, dried and resuspended in 500 $\mu$l TE buffer.

## Construction of GenomeWalker DNA Libraries

GenomeWalker DNA libraries were constructed using the Universal GenomeWalker$^{TM}$ Kit (CLONTECH Laboratories, Inc.). The genomic DNA was completely digested with different blunt-end restriction enzymes (*Dra*I, *Eco*RV, *Pvu*II, *Stu*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 Genomic Sequence of *ala*

DNA fragment was amplified by PCR using primers D1 and D2, which were designed according to the *ala* cDNA sequence. The PCR amplification was performed in a PTC-100$^{TM}$ programmable (MJ Research, INC) for 30 cycles (94°C for 30 s, 55°C for 1 min, 72°C for 1 min) followed by extension for 7 min at 72°C.

The amplification of upstream sequence of genomic DNA of *ala* consisted of two PCR amplifications per library. In the primary PCR, the outer adaptor primer AP (provided in the kit), and an outer, gene-specific primer D3, were used. The primary PCR product mixture was diluted and used as the template for nested PCR with the nested adaptor primer NAP (provided in the kit) and the nested gene-specific primer D4. The amplification of downstream sequence of genomic DNA of *ala* consisted of two PCR amplifications per library. In the primary PCR, the outer adaptor primer AP and an outer, the gene-specific primer D5, were used (Table 1). In the nested PCR, the nested adaptor primer NAP and nested gene-specific primer D6 were used. The sequence amplification of genomic DNA of *ala* was performed essentially according to the Universal Genome Walker$^{TM}$ Kit user manual.

The sequences of 3′ and 5′ genomic DNA end products of *ala* were aligned with Vector NTI$^{TM}$ Suite 8.0 to obtain the predicted DNA fragment. Subsequently a pair of PCR primers D7 and D8 (Table 1) was designed according to the aligned sequence and synthesized for the amplification of the whole DNA sequence of *ala*. The PCR amplification was performed in a PTC-100$^{TM}$ programmable (MJ Research, INC) for 30 cycles (94°C for 30 s, 52°C for 1 min, 72°C for 3 min) followed by extension for 7 min at 72°C.

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

## Sequence Analysis

The analysis and comparison of the deduced amino acid sequence with published sequences of mannose-binding lectins were performed with Blastp (Standard Protein-Protein BLAST) on NCBI (www. ncbi.nlm.nih.gov). The conserved domains were searched with RPS-BLAST (Search the Conserved Domain Database)

on NCBI. 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). Secondary and three-dimensional structure predictions of ALA were performed with ExPASy Proteomics tools (//cn.expasy.org/tools/#secondary and//cn.expasy.org/swissmod/SWISS-MODEL.html).

## Molecular Evolution Analysis

Phylogenetic analysis was performed by aligning sequences of ALA and mannose-binding lectins from other plant species belonging to families Araceae, Iridaceae, Amaryllidaceae, Bromeliaceae, Liliaceae and Orchidaceae using the software Vector NTI$^{TM}$ Suite 8. The phylogenetic tree was constructed by the neighbor-joining (NJ) method with OMIGA 2.0 (Oxford Molecular Ltd.) using default parameters.

## Southern Blot Analyses

Aliquots of genomic DNA (20 $\mu$g/sample) were digested overnight at 37°C with *Eco*RI and *Bam*HI, respectively, which did not cut within the probe region, fractionated by 0.85% agarose gel electrophoresis and transferred onto a positively charged Hybond-N$^+$ nylon membrane (Amersham Pharmacia, England). The partial coding sequence of *ala* (953 bp) was generated by PCR using primers S1 (5′-CTAGCAGCAACCCCATTTGATTTCT-3′) and S2 (5′-CCTATATATGGC TACGAAAGCAG-3′) and used as the probe in Southern analysis. Probe labeling (biotin), hybridization and signal detection were performed using Gene images random prime labeling module and CDP-Star detection module following the manufacturer's instructions (Amersham Pharmacia, England). The membrane was washed under stringency condition (60°C) and the hybridized signals were visualized by exposure to Fuji X-ray film at room temperature for 1.5 h.

## Expression Profiling of *ala* in *A. lobatum* Tissues

RT-PCR was used to investigate the *ala* expression profiling in various tissues of *A. lobatum*. RNA was extracted separately from seven kinds of tissues including root, leaf, spadix, petiole, spathe, bud and tuber, and used in RT-PCR analysis using One Step RNA PCR Kit (TaKaRa, China). The amplification for *ala* was performed at 50°C for 30 min followed by 94°C for 2 min and by 30 cycles of amplification (94°C for 30 s, 58°C for 30 s, 72°C for 50 s) using primers S3 (5′-CACCAACTA-TCTGCTGTCCGGCGAAA-3′) and S4 (5′-GAGCGCTTCGAGCTGGTCTCC-CAGA-3′), which amplified a fragment of 690 bp from 1084 bp to 2493 bp of *ala*. At the same time, 18S ribosome RNA was also used in RT-PCR paralleling with the *ala* using primers 18SF (5′-ATGATAACTCGACGGATCGC-3′) and 18SR (5′-CTTGGATGTGGTAGCCGTTT-3′), which amplified a fragment of 126 bp from 18S rRNA, as the control under the same condition except for the amplification cycles were reduced to 20. The amplified products were analyzed with Gene analysis software package (Gene Company, USA).

## RESULTS AND DISCUSSION

### Cloning and Characterization of the Full-Length cDNA of *ala*

Based on the conserved region of plant monocot mannose-binding lectins, a specific primer (primer 1) was designed and synthesized for the amplification of 3′-end cDNA of *ala*. A fragment of about 0.5 kb was amplified in which a 3′ untranslated region of 217 bp was found downstream from the terminal codon. Subsequently, two specific primers designed according to the obtained 3′ RACE fragment sequence (primers 2 and 3) were used for the amplification of 5′-end cDNA of *ala*, resulting in a fragment of about 0.75 kb in which a 5′ untranslated region of 84 bp was identified upstream of the first ATG codon. Finally, the full-length cDNA sequence of *ala* was deduced and amplified through RT-PCR using the gene specific primers 4 and 5, followed by confirmation by sequencing. The full-length cDNA of *ala* was 1078 bp and contained a 774 bp open reading frame encoding a precursor of 258 amino acids with a calculated molecular weight of 28.3 kDa and isoelectric point of 7.23 (Fig. 1).

Like most of the mannose-binding lectins from Araceae and Amaryllidaceae species (Van Damme *et al*., 1991, 1995, 1996, 1998; Chai *et al*., 2003; Fei *et al*., 2003; Yao *et al*., 2003a; Zhao *et al*., 2003), three mannose-binding motifs (QXDXNXVXY) were also identified in *ala*. The amino acid sequences of the motifs I and II were the same as those of GNA. However, the amino acid sequences of motif III were different from those of GNA, in which Asn (N) was substituted by Phe (F). A 23-amino acid signal peptide was predicted in ALA based on the rules of predicting signal peptide (Heijne, 1986), which was in agreement with those in most reported mannose-binding lectins from family Araceae and Amaryllidaceae (Van Damme *et al*., 1998; Chai *et al*., 2003; Fei *et al*., 2003) (Fig. 1). The ALA signal peptide contained highly hydrophobic amino acid residues (69.6%), implying it was a secretary putative signal peptide. Secondary and three-dimensional structure predictions of ALA were conducted (Fig. 2). Based on the Hierarchical Neural Network method result, ALA proprotein was composed of 9.30% alpha helix, 37.21% extended strand, 17.05% beta turn and 36.43% random coil. The alpha helix lied mostly in the N-terminal and C-terminal. Penetrating through most parts of the ALA, random coil was the most abundant structural element of ALA, while extended strands were intermittently distributed in the proprotein. The amino acids (QDN) of mannose-binding motifs consisted of random coil, and the amino acid (Y) consisted of extended strand (Fig. 2a). Swiss-Model structure prediction of ALA resulted in a similar folding mode and spatial configuration to GNA and the mannose-binding lectin from *Crinum asiaticum* (CAA) (Barre *et al*., 2001; Chai *et al*., 2003) (Fig. 2b). ALA proprotein was also composed of three sub-domains, with a conserved mannose-binding motif for each domain. The first two sub-domains lied in both sides of the three-dimensional structure, and the third sub-domain lied in the middle of the structure.

### Homology Analysis

The database search and analysis with Blastp showed that the deduced amino acid sequences of ALA had higher homologies with hetero-oligomeric mannose-

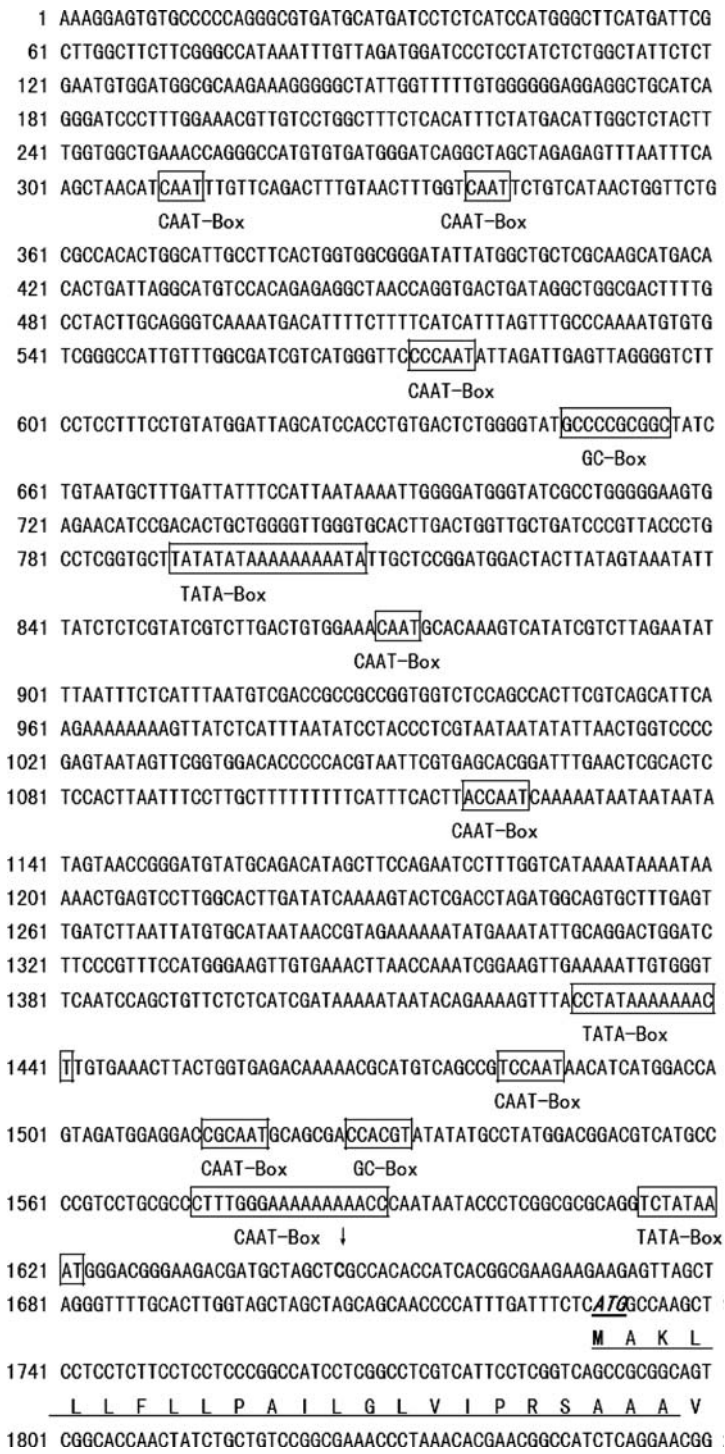


**Fig. 1.** (Continued)

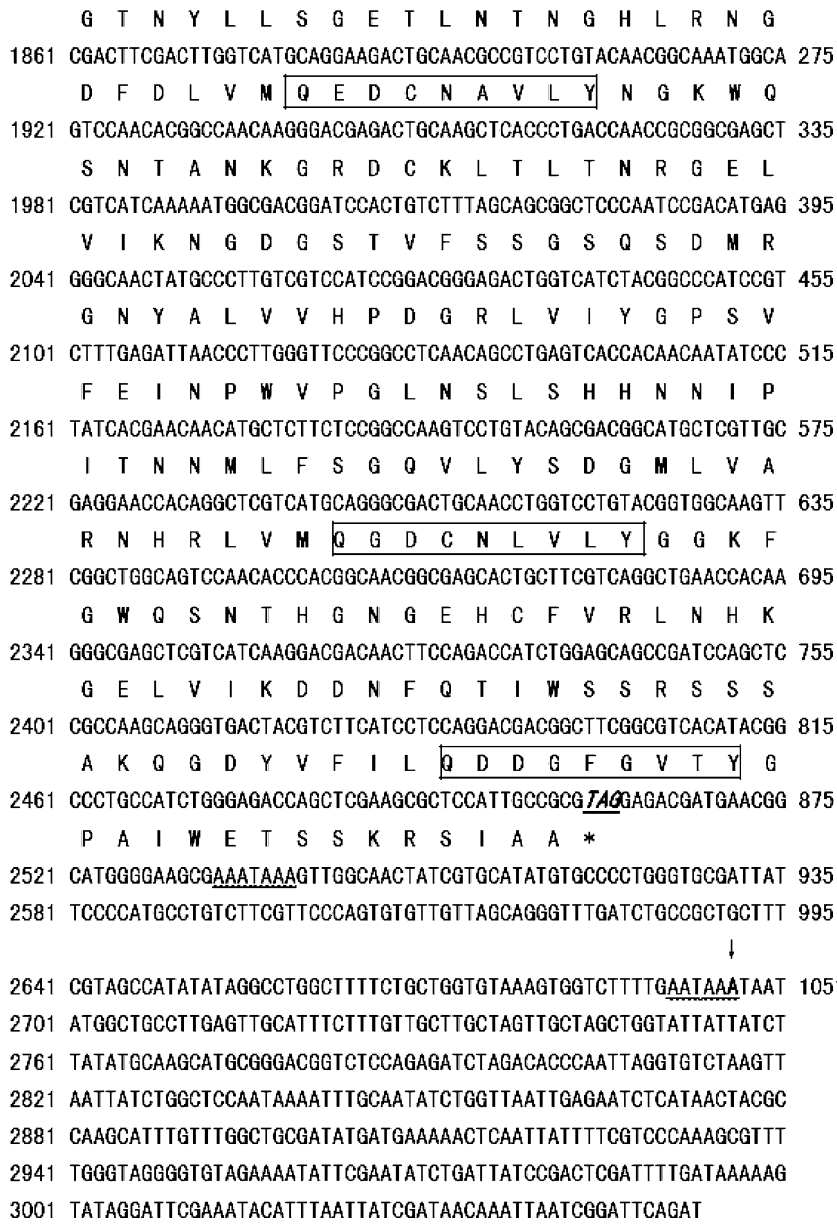


**Fig. 1.** The full-length cDNA and genomic DNA sequences and the predicted amino acid sequence of *A. lobatum* agglutinin gene (*ala*). The nucleotide sequence of cDNA is numbered in the right side, and the nucleotide sequence of genomic DNA is numbered in the left side. The start codon (ATG) and the stop codon (TAG) are italicized in bold. Mannose-binding motifs (QXDXN/FXVXY), the putative TATA-box, CAAT-boxes and GC-boxes are boxed. The predicted single peptide sequence is underlined. The polyadenylation signals are wave underlined. The upright arrowheads show the start site of transcription (first letter C) and the site of poly A, respectively.

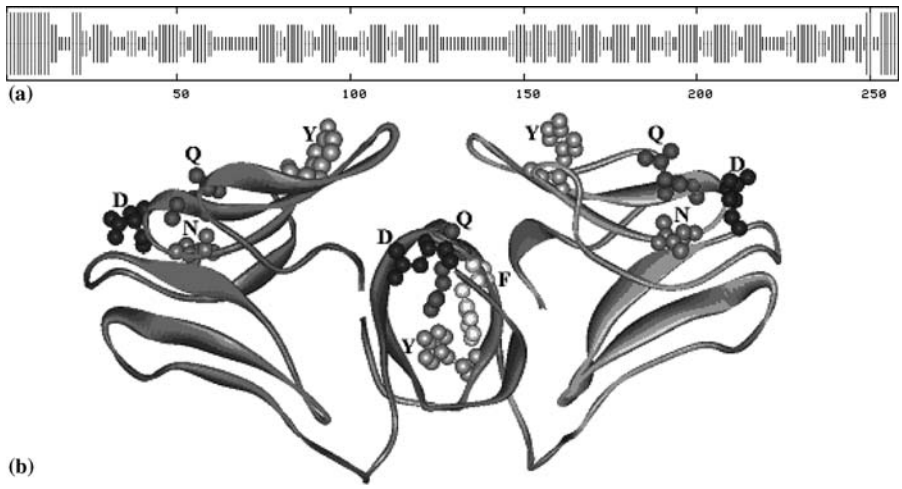


**Fig. 2.** The two- and three-dimensional structures of the predicted ALA polypeptide. (a) The two-dimensional structure. $\alpha$-helix and extended strand are denoted as vertical long bars and vertical short bars respectively, with the horizontal line presenting the random coil running through the whole molecule. (b) The three-dimensional structure. $\beta$-sheets and random coils are indicated in dark and light patches, respectively. The amino acid residues QXDXN/FXVXY constituting the three mannose-binding sites are signified with line-bead spatial configurations.

binding lectins than with homo-oligomeric mannose-binding lectins. The comparison of deduced amino acid sequence of ALA with those of hetero-tetramer mannose-binding lectins through search with Blastp showed ALA had the highest identity with the mannose-binding lectins from other Arareae species (Hirai *et al*., 1993; Bezarra *et al*., 1995; Van Damme *et al*., 1995; Yao *et al*., 2003a; Zhao *et al*., 2003). The percentages of identity were from 83% to 73%, such as 83% (203/243), 80% (201/250), 73% (185/252) and 74% (180/241) identities to AHA (AAP50524), PTA (AAR27794), CEA (BAA03722) and AMA (AAC48997), respectively (Fig. 3). ALA had also identities with mannose-binding lectins from other plant families such as Alliaceae, Hyacinthaceae, Iridaceae and Lillaceae, and the percentages of identity were from 42% to 30%, such as 42% (103/245), 41% (102/247), 36% (92/255), 30% (82/270) and 32% (78/240) identities to CSA (*Crocus sativus* agglutinin) (AAK29007), CVA (*Crocus vernus* agglutinin) (AAG10404), THA (*Tulipa* hybrid agglutinin) (S62647), ASA (S23496) and HHA (*Hyacinthoides hispanica* agglutinin) (AAD16404), respectively.

According to the NCBI conserved domain search, the predicted ALA possessed two-conserved domains, called ALA-DOM1 and ALA-DOM2. ALA-DOM1 was between $T_{27}$ and $W_{130}$ amino acid and ALA-DOM2 was between $D_{146}$ and $S_{250}$. They were 45% (48/107) identity. One belonged to agglutinin family (pfam01453) (probable mannose binding) (CD-Length = 110 residues) and the members of this family were plant lectins that also contained a number of S-locus glycoproteins. The other belonged to B-lectin (cd00028) (Bulb-type mannose-specific lectin) (CD-Length = 116 residues) and the members of this family were involved in a-D-mannose recognition and contained a consensus sequence motif (QXDXNXVXY).

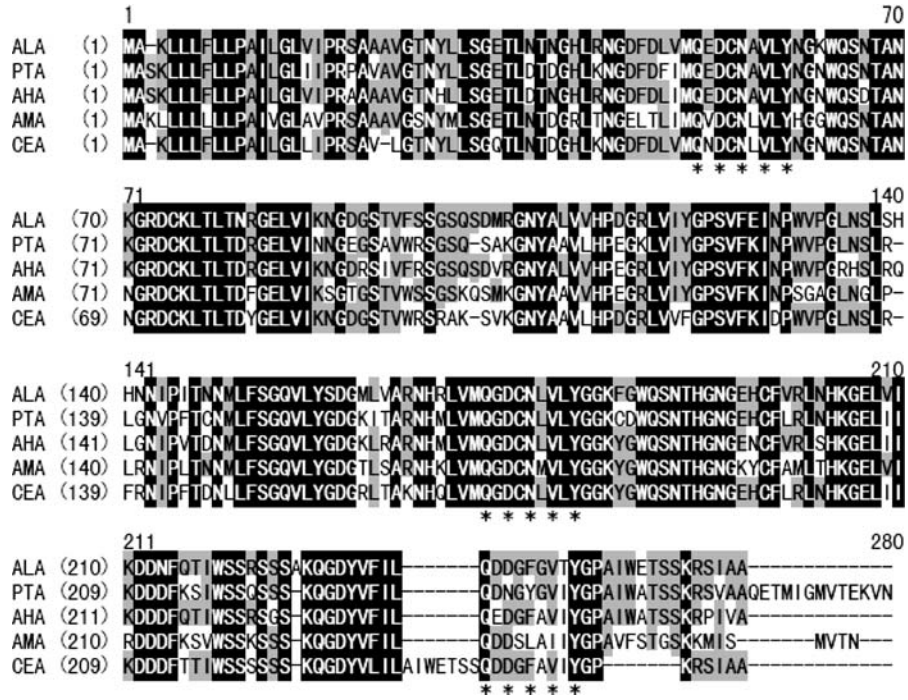

**Fig. 3.** Multiple alignment of amino acid sequence of ALA with those of other mannose-binding lectins from Araceae. The alignment is performed with the Vector NTI Suite 8.0 by using the published mannose-binding lectin sequences from Araceae, e.g. PTA (*Pinellia ternata*) (AAP20876), AHA (*Arisaema heterophyllum*) (AAP50524), AMA (*Arum maculatum*) (AAC48997), CEA (*Colocasia esculenta*) (BAA03722) and ALA (*Arisaema lobatum*) (AAS60304). Gaps are introduced for optimal alignment and maximum similarity between all compared sequences. The identical amino acids among all the aligned sequences are shown in black background and the identical amino acids with ALA are shown in gray background. Mannose binding motif is shown with '*'.

The comparison of the sequence of ALA-DOM1 and ALA-DOM2 with those conserved domains through search from NCBI and GNA showed the higher identity in a consensus sequence motif (QXDXNXVXY). One consensus sequence motif quite matched the QXDXNXVXY sequence in ALA-DOM1 and ALA-DOM2. The function of Val was to interact with C3 and C4 of mannose through hydrophobic interactions.

In the past few years, many mannose-binding lectins were isolated and characterized. According to the previous reports, homo-oligomeric lectins contained three, two or one conserved mannose-binding motif (QXDXNXVXY) (Ramachandraiah and Chandra, 2000). Until now, only nine monocot mannose-binding lectins composed of either intact or cleaved two-domain protomers have been identified, that were ASA of Alliaceae (Van Damme *et al*., 1992), HHA of Lillaceae (AAD16404), CVA of Iridaceae (AAG10404), CSA of Iridaceae (AAK29007), THA of Lillaceae (S62647), AMA of Araceae (Van Damme *et al*., 1995), CEA of Araceae

(Hirai *et al.*, 1993; Bezerra *et al.*, 1995), PTA of Araceae (Yao *et al.*, 2003a) and AHA of Araceae (Zhao *et al.*, 2003). These lectins possessed two domains, called DOM1 and DOM2. DOM1 contained one consensus mannose-binding motif and DOM2 contained two, one or no consensus mannose-binding motif. For example, HHA and ASA contained two consensus mannose-binding motifs. AHA, PTA, CEA, AMA, CSA and CVA contained one consensus mannose-binding motif. THA contained no consensus mannose-binding motif (data no shown). These motifs were essential for mannose binding property. Therefore, according the sequence analysis, it is suggested that ALA belongs to the two-domain containing homo-oligomeric lectin family with mannose binding property.

## Cloning of Genomic Sequence of *ala*

Based on the *ala* cDNA sequence, primers D1 and D2 were designed and synthesized for the amplification of genomic sequence of *ala*. A fragment of about 0.7 kb (ALAD1) was obtained which was the same as the cDNA sequence of *ala*. No intron was found within the region of the genomic sequence.

Four specific primers (D3, D4, D5 and D6) were designed and synthesized according to the obtained genomic DNA fragment. The DNA fragments of upstream and downstream sequences were amplified using GenomeWalker DNA libraries as the template. A specific band of about 1.8 kb was amplified by primers NAP and D4 using *Dra*I digested genomic DNA as the template. Sequence analysis showed that a fragment of 87 bp was overlapped to ALAD1 5′-end sequence. A specific band of about 0.7 kb was amplified by primers NAP and D6 using *Eco*RV digested genomic DNA as the template. Sequence analysis showed that a fragment of 98 bp was overlapped to ALAD1 3′-end sequence.

Based on the aligned sequences of 3′ downstream and 5′ upstream, two specific primers (D7 and D8) were designed and synthesized. A specific band of about 3 kb was amplified and sequence analysis revealed that this sequence completely coincided with the deduced DNA sequence. The entire ALA genomic DNA was 3053 bp containing a 777 bp gene-coding region, a 1729 bp of 5′-upstream and 574 bp of 3′ downstream regions (Fig. 1).

## Characterization of 3′- and 5′-Franking Regions of *ala*

The 3′-flanking region of the gene contained 60% of A + T. This value was slightly lower than those of other lectin genes, such as ricin gene (70% of A + T). There were two polyadenylation signals, the first (AAATAAA) was located from the 27th to 33th bp downstream from the stop codon and the second (AATAAA) was located from the 185th to 190th bp downstream from the stop codon.

The 5′-flanking region of the ALA had high content of A + T (56%), which was also found in other mannose-binding lectin genes, such as 5′ regulatory region of CEA gene (58% of A + T) (AF178113). Analysis of the promoter sequence of ALA using the PlantCARE database identified a conserved transcription start site at −84 bp positions upstream from the start codon ATG. Usually, the structure of 5′ gene flanking region of eukaryotes comprises four parts: the site of start

transcription, TATA box, CAAT box and GC box. TATA box is generally located at $-32 \pm 7$ bp positions upstream from the start of transcription. The consensus sequence for the TATA box is [T (CG) TATA (TA) $A_{1-3}$(CT) A] that is important for eukaryotic transcription (Joshi, 1987). A TATAA sequence was found to be located at −24 bp positions upstream from the start of transcription in *ala*, which might be important for the transcription control as well. The consensus sequence for TATA box was TCTATAAAT.

Eight CAAT boxes were identified at the −56, −126, −159, −522, −772, −1067, −1304 and −1332 bp positions upstream from the start of transcription (Fig. 1). The CAAT box was sometimes important for the efficiency of eukaryotic transcription (Benoist *et al*., 1980). Usually, CAAT box is found at the $-77 \pm 10$ bp positions upstream from the start of transcription, although longer intervals are also found. Two GC boxes (GCCCCGCGGC and CCACGT) were found to be located at −990 and −104 bp positions respectively upstream from the start of transcription in *ala* (Fig. 1).

## Molecular Evolution Analysis

Phylogenetic tree analysis revealed that all the aligned mannose-binding lectins originated from an ancestor and evolute into three groups (Fig. 4). Two groups contained mannose-binding lectins belonging to homo-oligomeric lectin (upper and lower branches in Fig. 4), and the other one contained mannose-binding lectin belonging to hetero-oligomeric lectin (middle branch in Fig. 4). Among the hetero-oligomeric lectin group, *A. lobatum* and other Araceae species, such as *P. ternata*, *P. pedatisecta C. esculenta*, *A. maculatum* and *A. heterophyllum* were clustered into one sub-group. According to the phylogenetic tree, ALA and other reported hetero-oligomeric lectins formed a group, suggesting that ALA belonged to hetero-oligomeric lectin.

## Southern Blot Analysis

The presence of multicopies of mannose-binding lectin gene in the genome has been reported in many plant species, particularly those belonging to Amaryllidaceae and Araceae families (Van Damme *et al*., 1991, 1992, 1995; Hirai *et al*., 1993; Hester *et al*., 1995). To investigate if the *ala* belonged to a multi-copy gene family, genomic DNA of *A. lobatum* was digested with *Eco*RI and *Bam*HI respectively, followed by hybridization with partial coding sequence of *ala* as the probe. Although there might exist partial digestion of the samples, reflected by the presence of high molecular weight bands and various intensities of the hybridization bands, multiple hybridization bands were present in each lane, indicating that the *ala* belonged to a multi-copy gene family (Fig. 5).

## Expression Profiling of *ala* in Various Tissues

To investigate the *ala* expression pattern in various tissues of *A. lobatum*, total RNA was isolated from root, leaf, spadix, petiole, spathe, bud and tuber respectively, and subjected to RT-PCR analysis using the primers S3 and S4. The result

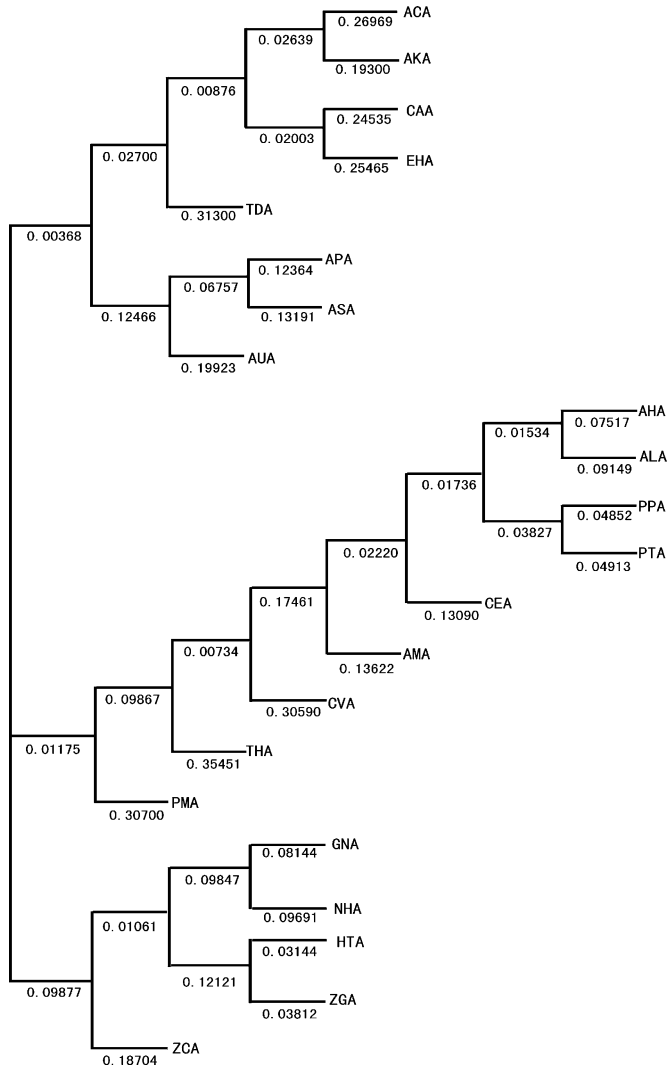


**Fig. 4.** The phylogeny of mannose-binding lectins from Araceae, Alliaceae, Iridaceae, Amaryllidaceae, Bromeliaceae, and Orchidaceae family. A phylogenetic tree is drawn using the CLUSTAL W program. The mannose-binding lectin sequences are downloaded from NCBI, and the accession numbers are bellow: ACA (*Ananas comosus*) from Bromeliaceae (AAM28277), AKA (*Amorphophallus konjac*) from Araceae (AAP22169), CAA (*Crinum asiaticum*) from Amaryllidaceae (AAO59506), EHA (*Epipactis helleborine*) from Orchidaceae (AAC48927), TDA (*Typhonium divaricatum*) from Araceae (AAQ55289), APA (*Allium porrum*) from Alliaceae (AAC37361), ASA (*Allium sativum*) from Alliaceae (AAA32643), AUA (*Allium ursinum*) from Alliaceae (AAC37358), AHA (*Arisaema heterophyllum*) from Araceae (AAP50524), ALA (*Arisaema lobatum*) from Araceae (AAS60304), PPA (*Pinellia pedatisecta*) from Araceae (AAR27793), PTA (*Pinellia ternata*) from Araceae (AAO20876), CEA (*Colocasia esculenta*) from Araceae (BAA03722), AMA (*Arum maculatum*) from Araceae (AAC48997), CVA (*Crocus vernus*) from Iridaceae (AAG10402), THA (*Tulipa* hybrid cultivar) from Liliaceae (S62647), PMA (*Polygonatum multiflorum*) from Ruscaceae (AAC49413), GNA (*Galanthus nivalis*) from Amaryllidaceae (AAL07474), NHA (*Narcissus hybrid*) from Amaryllidaceae (M88117), HTA (*Hippeastrum* sp. tetrameric) from *Amaryllidaceae* (AAA33362), ZGA (*Zephyranthes grandiflora*) from Amaryllidaceae (AAP37975) and ZCA (*Zephyranthes candida*) from Amaryllidaceae (AAM94385).

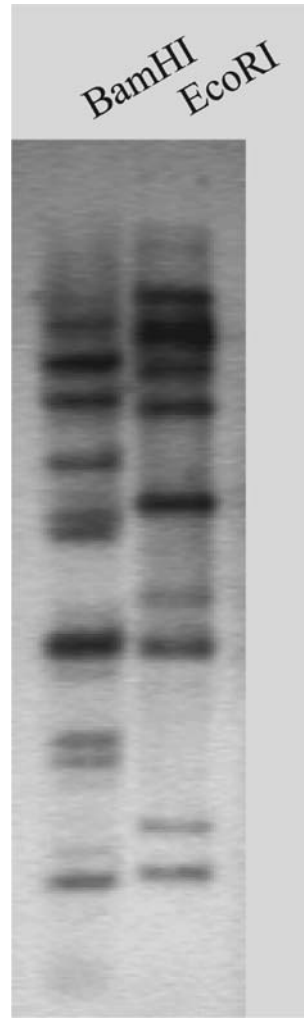


**Fig. 5.** Southern blot analysis. Genomic DNA was extracted from leaves of *A. lobatum* and digested with *Eco*RI and *Bam*HI respectively followed by hybridization with partial *ala* coding fragment as the probe.

showed that, like lectin genes from some Araceae species such as *aha* (Zhao *et al.*, 2003), the *ala* preferentially expressed in the tissues with different levels (Fig. 6). Higher expression was found in spadix, bud, leaf, spathe and tuber, while the relatively lower expression was found in root and petiole. The expression pattern of *ala* is different from that of *aka* gene from *A. konjac*, another Araceae species, in which the highest expression was found in root (Fei *et al.*, 2003), suggesting that the expression of the lectin gene is species-specific.

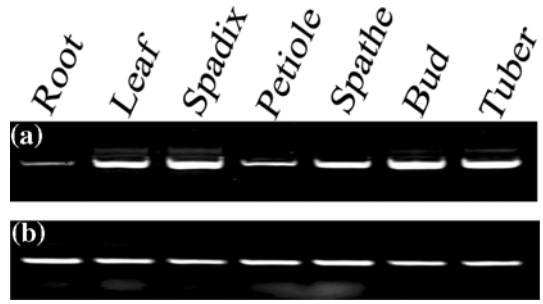


**Fig. 6.** RT-PCR analysis. Total RNA was isolated separately from root, leaf, spadix, petiole, spathe, bulb and tuber of *A. lobatum* followed by reverse transcription PCR analysis. The 30 and 20 PCR cycles are used respectively for the amplification of *ala* (a) and 18S rRNA (b, as the control).

Earlier studies showed that many mannose-binding lectins from Araceae had similar insecticidal activities with GNA in feeding experiments with both artificial diets and transgenic plants. Their insecticidal activities were mostly towards sap-sucking insects, such as cotton aphids (*Aphis gossypii* Glover) and peach potato aphids (*Myzus persicae* Sulzer) (Li *et al.*, 2000; Yao *et al.*, 2003a, b). The cloning of *ala* enables the subsequent anti-insect identification of *ala* through both transgenic technology and protein expression in the future.

## ACKNOWLEDGMENTS

This research was funded by China National Transgenic Plant Research and Commercialization Project, China Ministry of Education, Shanghai Agricultural Committee and Shanghai Science and Technology Committee.

## REFERENCES

Barre, A., Bourne, Y., Van Damme, E. J. M., Peumans, W. J., and Rouge, P. (2001) Mannose-binding plant lectins: different structural scaffolds for a common sugar-recognition process. *Biochimie* **83:**645–651.

Benoist, C., O'Hare, K., Breathnach, R., and Chambon, P. (1980) The ovalbumin gene-sequence of putative control regions. *Nucleic Acids Res.* **8:**127–142.

Bezerra, I. C., Castro, L. A., Neshich, G., de Almeida, E. R., de Sa, M. F., Mello, L. V., and Monte-Neshich, D. C. (1995) A corm-specific gene encodes tarina major globulin of taro (*Colocasia esculenta* L. Schott). *Plant Mol. Biol.* **28:**137–144.

Chai, Y., Pang, Y., Liao, Z., Zhang, L., Sun, X., Lu, Y., Wang, S., and Tang, K. (2003) Molecular cloning and characterization of a mannose-binding lectin gene from *Crinum asiaticum*. *J. Plant Physiol.* **160:**913–920.

Dellaporta, S. L., Wood, J. M., and Hicks, J. B. (1983) A plant DNA minipreparation. *Plant Mol. Biol. Rep.* **1:**19–21.

Down, R. E., Gatehouse, A. M. W., Hamilton, W. D. O., and Gatehouse, J. A. (1996) Snowdrop lectin inhibits development and decreases fecundity of the glasshouse potato aphid (*Aulacorthum solani*) when administered *in vitro* and via transgenic plants both in laboratory and glasshouse trials. *J. Insect Physiol.* **42:**1035–1045.

Fei, J., Liao, Z., Chai, Y., Pang, Y., Yao, J., Sun, X., and Tang, K. (2003) Molecular cloning and characterization of a novel mannose-binding lectin gene from *Amorphophallus konjac*. *Mol. Biol. Rep.* **30:**177–183.

Fitches, E., Gatehouse, A. M. W., and Gatehouse, J. A. (1997) Effects of snowdrop lectin (GNA) delivered via artificial diet and transgenic plants on the development of tomato moth (*Lacanobia oleracea*) larvae in laboratory and glasshouse trials. *J. Insect Physiol.* **43:**727–739.

Heijne, G. V. (1986) A new method for predicting signal sequence cleavage sites *Nucleic Acids Res.* **14:**4683–4690.

Hester, G., Kaku, H., Goldstein, I. J., and Wright, C. S. (1995) Structure of Mannose-specific snowdrop (*Galanthus nivalis*) lectin is representative of a new plant lectin family. *Nat. Struct. Biol.* **2:**472–479.

Hilder, V. A., Powell, K. S., Gatehouse, A. M. R., and Gatehouse, J. A. (1995) Expression of snowdrop lectin in transgenic tobacco plants results in added protection against aphids. *Transgenic Res.* **4:**18–25.

Hirai, M., Nakamura, K., Imai, T., and Sato, T. (1993) cDNAs encoding for storage proteins in the tubers of taro (*Colocasia esculenta* Schott). *Jpn. J. Genet.* **68:**229–236.

Huang, D., Pan, Y., Zhang, S., Cao, J., Yang, X., Zhang, J., and Yi, W. (1997) The discovery of insecticidal protein against aphids from *Pinellia pedatisecta* and *P. ternata*. *Sci. Agri. Sin.* **30:**94.

Jaakola, L., Pirttila, A. M., Halonen, M., and Hohtola, A. (2001) Isolation of high quality RNA from bilberry (*Vaccinium myrtillus* L.) fruit. *Mol. Biotechnol.* **19:**201–203.

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

Li, R. Z., Gao, W. J., Ji, D. F., and Li, C. X. (2000) Isolation of agglutinin from *A. consarguieum* and analysis on its resistance to *Aphis gossypii*. *Acta Gossypii Sin.* **12:**54–56.

Pan, Y. H., Zhang, S. X., Cao, J. P., Yang, X. M., Zhang, J., Ying, W. Z., and Huang, D. F. (1998) The isolation, purification of *Pinellia pedatisecta* lectin and its activity on aphid-resistance. *Prog. Nat. Sci.* **8:**502–505.

Ramachandraiah, G. and Chandra, N. R. (2000) Sequence and structural determinants of mannose recognition. *Proteins* **39:**358–364.

Rahbé, Y., Sauvion, N., Febvay, G., Peumans, W. J., and Gatehouse, A. M. (1995) Toxicity of lectins and processing of ingested proteins in the pea aphid *Acyrthosiphon pisum*. *Entomol. Exp. Appl.* **76:**143–155.

Rao, K. V., Rathore, K. S., Hodges, T. K., Fu, X., Stoger, E., Sudhakar, D., Williams, S., Christou, P., Bharathi, M., Bown, D. P., Powell, K. S., Spence, J., Gatehouse, A. M. R., and Gatehouse, J. A. (1998) Expression of snowdrop lectin (GNA) in transgenic rice plants confers resistance to rice brown planthopper. *Plant J.* **15:**469–477.

Tang, K., Zhao, Z., Sun, X., Wan, B., Qi, H., and Lu, X. (2001) Production of transgenic rice homozygous lines with enhanced resistance to rice brown plant hopper. *Acta Biotechnol.* **21:**117–128.

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

Van Damme, E. J. M., Smeets, K., Torrekens, S., Van Leuven, F., Goldstein, I. J., and Peumans, W. J. (1992) The closely related homomeric and heterodimeric mannose-binding lectins from garlic are encoded by one-domain and two-domain lectin genes, respectively. *Eur. J. Biochem.* **206:**413–420.

Van Damme, E. J. M., Smeets, K., Torrekens, S., Van Leuven, F., and Peumans, W. J. (1993) The mannose-specific lectins from ramsons (*Allium ursinum* L.) are encoded by three sets of genes. *Eur. J. Biochem.* **217:**123–129.

Van Damme, E. J. M., Goossens, K., Smeets, K., Van Leuven, F., Verhaert, P., and Peumans, W. J. (1995) The major tuber storage protein of *Araceae* species is a lectin: characterization and molecular cloning of the lectin from *Arum maculatum*. *Plant Physiol.* **107:**1147–1158.

Van Damme, E. J. M., Briké, F., Winter, H. C., Van Leuven, F., Goldstein, I. J., and Peumans, W.J. (1996) Molecular cloning of two different mannose-binding lectins from tulip bulbs. *Eur. J. Biochem.* **236:**419–427.

Van Damme, E. J. M., Peumans, W. J., Barre, A., and Rougé, P. (1998) Plant lectins: a composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. *Crit. Rev. Plant Sci.* **17:**575–692.

Van Damme, E. J. M., Astoul, C. H., Barre, A., Rougé, P., and Peumans, W. J. (2000) Cloning and characterization of a monocot mannose-binding lectin from *Crocus vernus* (family *Iridaceae*). *Eur. J. Biochem.* **267:**5067–5077.

Wu, A., Sun, X., Pang, Y., and Tang, K. (2002) Homozygous transgenic rice lines expressing GNA with enhanced resistance to the rice sap-sucking pest *Laodelphax striatellus*. *Plant Breed.* **21:**93–95.

Yao, J., Zhao, X., Zhang, H., Liao, Z., Chen, F., Song, J., Sun, X., and Tang, K. (2003a) Molecular cloning and characterization of a novel lectin gene from *Pinellia ternata*. *Cell Res.* **13:**301–308.

Yao, J., Pang, Y., Qi, H., Wan, B., Zhao, X., Kong, W., Sun, X., and Tang, K. (2003b) Transgenic tobacco expressing *Pinellia ternata* agglutinin confers enhanced resistance to aphids. *Transgenic Res.* **12:**715–722.

Zhang, H., Wu, X., Tang, K., Wang, X., Sun, X., and Zhou, K. (2003) A primary study of transferring the *Pinellia ternata* agglutinin (*pta*) gene into rice and expression. *Acta Genet. Sin.* **30:**1013–1019.

Zhao, X., Yao, J., Liao, Z., Zhang, H., Chen, F., Lu, Y., Song, J., Sun, X., Yu, S., and Tang, K. (2003) Molecular cloning of a novel mannose-binding lectin gene from *Arisaema heterophyllum*. *Plant Sci.* **165:**55–60.