

Molecular characterization of major allergens Ara h 1, 2, 3 in peanut seed

Shengjuan Jiang · Songhua Wang · Yujun Sun ·
Zhengyi Zhou · Guiqin Wang

Received: 10 December 2010 / Revised: 12 January 2011 / Accepted: 21 January 2011
© Springer-Verlag 2011

Abstract Peanut is among the most commonly used dietary seeds, but peanut allergens, especially Ara h 1 (*Arachis hypogaea* allergy 1), 2 and 3, can cause severe IgE-mediated reactions. In this study, the molecular characterization and expression pattern of three allergens in peanut LUHUA 8, the representative of the cultivated lines in China, are reported. In situ hybridization and real time PCR analysis revealed high expression levels and different tissue expression patterns of the three allergens, which might be connected with many aspects, such as the strong conservation of intron phase of the allergen genes, the low energy of the mRNA's regions, and the complicated post-translational modifications. Furthermore, the different sequences between the cloned allergens and the reported sequences previously involved the charged amino acids especially in IgE epitopes, which might alter specific physicochemical and physiological properties, and thus influence the immunity of the allergens. The identification of the specific features of the allergen genes would be of considerable importance to the basic understanding of the specific characteristics of peanut seed allergens.

Keywords Allergen · Cupin · Expression · IgE epitope · Peanut

Introduction

Peanut (*Arachis hypogaea* L.), containing 45–53% oil and 24–29% protein in its seed, is among the most commonly used dietary seeds for their good nutritional value and balanced amino acid content. However, peanut is one of the most allergenic foods. It contains multiple allergens, which are responsible for triggering IgE-mediated anaphylactic reactions in both children and adults (Shreffler et al. 2006). Allergy to peanut tends to be persistent and severe, and it represents an important public health problem of increasing prevalence (Yan et al. 2005).

Although about 11 peanut allergens have been reported (<http://www.allergome.org>, Dec. 7, 2010), Ara h 1 (*Arachis hypogaea* allergy 1), 2, and 3 are classified as the major peanut allergens. Ara h 1 and Ara h 3 are both seed storage proteins, belonging to the cupin superfamily which contains structurally related proteins built from the β -barrel cupin domain (Breiteneder and Radauer 2004). Twenty-three linear IgE-binding epitopes have been mapped in Ara h 1 monomer (Burks et al. 1997; Shin et al. 1998). Four linear IgE-binding epitopes have been characterized in Ara h 3. Ara h 2 exists as a member of AAI_LTSS (Alpha-Amylase Inhibitors (AAI), Lipid Transfer (LT) and Seed Storage (SS) protein) superfamily, which is unique to higher plants (Lehmann et al. 2003). Four complete isoforms of *ara h 2* have been isolated and characterized (Li et al. 2010). Ten linear IgE-binding epitopes are characterized in Ara h 2. Resistance to enzymatic digestion with pepsin and heat is the common property of the three allergens. Besides, the IgE-binding sites which trigger the fatal allergic reactions are the structure foundation. Mutation analysis of the epitopes reveals that single or critical amino acid changes within these peptides can lead to a reduction or loss of IgE binding ability, and consequently

Communicated by P. Lakshmanan.

S. Jiang (✉) · S. Wang · Y. Sun · Z. Zhou · G. Wang
College of Life Science, Anhui Science and Technology
University, Anhui Fengyang 233100, China
e-mail: jiangsjouc@yahoo.com.cn

can affect the allergen's immunogenicity (Rabjohn et al. 1999).

By examining the timing expression and localization of the three major peanut allergen genes, Kang et al. (2007) has found that the mRNA and proteins can accumulate in the seed of the tested lines and the expression patterns of the three genes are heterogeneous, depending on the specific peanut allergen gene and the cultivar tested. The objectives of this study were to isolate, sequence, and determine the structure and organization of genomic clones encoding Ara h 1, Ara h 2 and Ara h 3 in peanut LUHUA 8, the representative of the cultivated lines in China. The expression patterns of these three major peanut allergens in the seed were also characterized and compared. The analysis of the structure and organization of the allergen genes would be helpful in understanding the specific characteristics of seed allergens.

Materials and methods

Plant materials

After sterilized with 70% (v/v) ethanol for 40–60 s and followed with treatment of 5% (v/v) sodium hypochlorite (NaOCl) for 20 min on the surface, the peanut (*Arachis hypogaea* L. LUHUA 8) seeds were washed with sterile water to remove residual NaOCl and were frozen in liquid nitrogen immediately for DNA and RNA isolation.

Gene cloning and sequence analysis

Genomic DNA of seeds was extracted using SDS method. Based on the published cDNA sequences of peanut allergen Ara h 1 (GenBank AF432231), Ara h 2 (GenBank EF609642) and Ara h 3 (GenBank ABI17154), primers of Ara h 1 (upstream primer: 5'-CCCTCCATCATTTTACATC-3', downstream primer: 5'-TGTCTCAACAAGTAT TCCATTTGAC-3'), Ara h 2 (upstream primer: 5'-ATG GCCAAGCTCACCATAC-3', downstream primer: 5'-TTA GTATCTGTCTCTGCCGC-3') and Ara h 3 (upstream primer: 5'-ATGGCCAAGCTTCTTGCG-3', downstream primer: 5'-TTAAGCCACCTCCCTCATAG-3') were constructed. The PCR amplification was carried out at 94°C for 4 min, followed by 30 cycles of 94°C for 45 s, 55–60°C (depending on the T_m value of primer pairs) for 60 s, 72°C for 90 s and a final extension step at 72°C for 10 min by using LA Taq polymerase (TaKaRa, Shiga, Japan).

Total RNAs of seeds were extracted by the CTAB/NaCl method (Azevedo et al. 2003). To eliminate the possible DNA contamination, the RNA samples were treated with DNase I (Promega, Madison, USA) and purified by ethanol precipitation. The quality of total RNA was measured by

both electrophoresis and optical absorbency. The first-strand cDNA was synthesized with the reverse transcription system (TaKaRa, Shiga, Japan) according to the manufacture's protocol. The individual gene-specific primer pairs of Ara h 1 (upstream primer: 5'-ATGA GAGGGAGGGTTTCTC-3', downstream primer: 5'-TCA GTTAAAAGCCTTCAAA-3'), Ara h 2 (upstream primer: 5'-ATGGCCAAGCTCACCATAC-3', downstream primer: 5'-TTAGTATCTGTCTCTGCCGC-3') and Ara h 3 (upstream primer: 5'-ATGGCCAAGCTTCTTGCG-3', downstream primer: 5'-TTAAGCCACCTCCCTCATAG-3') were used to amplify cDNA sequences by PCR. All the PCR products were gel-purified, ligated into pGEM-T easy vector (Promega, Madison, USA), and transformed into Top10 competent cells. Positive clones were selected and sequenced with ABI PRISM 3730 DNA sequencer.

The present location and phase of introns were evaluated for each open reading frame (ORF) by comparing DNA and cDNA sequences. Intron phase was named 0 when the intron was located in-between two consecutive codons; 1 if an intron was located between the first and second codon nucleotides and 2 if an intron was found between the second and third codon nucleotides (Kyndt et al. 2008). The deduced amino acid sequences were analyzed with the SMART program (Letunic et al. 2009) which was based on the PFAM database. The free energy of RNA structure was analyzed by using the algorithm of favorite thermodynamic by Vienna RNA Secondary Structure Prediction program (<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>). The potential serine, threonine and tyrosine phosphorylation sites were predicted with the NetPhos 2.0 Server (Blom et al. 1999). Multiple alignments were performed by Clustal_X software. Phylogenetic tree was constructed by the neighbor-joining method within the PHYLIP software package, using 1,000 bootstrap replicates.

RNA in situ hybridization

Digoxigenin-labelled riboprobes were synthesized by in vitro DIG-UTP transcription system (Roche, Mannheim, Germany) according to the manufacture's instructions. The cotyledon and the other parts of embryo (including embryonic bud, embryonic axes and embryonic root, "embryo" for short) of the mature seeds were cut to the right sizes and fixed in freshly prepared 3.7% formaldehyde, 5% acetic acid and 50% ethanol, individually, overnight at 4°C. The samples were dehydrated in an ethanol gradient, replaced with xylene, embedded in paraffin, and sectioned at 10 µm on rotary microtome (Leica, Heidelberg, Germany). The sections were mounted on poly-L-lysine-coated slides, dried at 42°C for 36 h and de-paraffinized in xylene for 20 min (two changes for 10 min each) followed by immersion in absolute ethanol

for 10 min (two changes for 5 min each). They were re-hydrated, and finally equilibrated in double-distilled water containing 0.1% DEPC. Tissue-section RNA in situ hybridization was carried out as described by Guillaumie et al. (2010) and the in situ hybridization technical instructions of Abcam Inc (Abcam, Cambridge, USA).

Real time quantitative PCR

Total RNA extraction of the embryo and cotyledon was carried out as described above and RNA was free of genomic DNA as determined by PCR. The concentration of isolated total RNA was calculated from absorbance at 260 nm with Synergy H1 Hybrid Microplate Reader (BioTek, Vermont, USA). One microgram of RNA was reverse-transcribed by using PrimeScriptTM RT reagent Kit (Perfect Real Time) (TaKaRa, Shiga, Japan) for first-strand cDNA synthesis. All cDNA samples were diluted 1:10 with RNase-free water before being used as templates in real time PCR analysis. Real time PCR was implemented by way of SYBR[®] Premix Ex Taq[™] II (Perfect Real Time) (TaKaRa, Shiga, Japan) and Applied Biosystems 7500 Real-Time PCR System (ABI, Framingham, USA). The detection was made in the MicroAmp[™] 8-strip tube with optical flat caps (ABI, Framingham, USA) in a final volume of 50 µl reaction mixture containing 25 µl SYBR[®] Premix Ex Taq[™] II solution, 2 µl primer solution with the final concentration of 0.4 µM, 1 µl ROX[™] Reference Dye II solution and 4 µl optimal RT reactant (cDNA template) and DEPC-treated water. The real time PCR amplification was carried out at 95°C for 30 s for initial denaturation, followed by 40 cycles of 95°C for 5 s, 60°C for 34 s. The individual gene-specific primer pairs of Ara h 1 (upstream primer: 5'-ACCGAGGGAGCGTGAAAGA-3', downstream primer: 5'-CGGGAAGTAGAAGGGTTGTT-3'), Ara h 2 (upstream primer: 5'-AGTAGCCCTCGCCTTTTCCTC-3', downstream primer: 5'-TCTGGCATCTTCTGTCTCCTTGT-3'), Ara h 3 (upstream primer: 5'-CAAGGAGCCATTGTTACAGTCA-3', downstream primer: 5'-CCTCTGCGGTTCTCATCGTATT-3') and actin (upstream primer: 5'-TACCAGATGGACAGGTTATCA CAAT-3', downstream primer: 5'-TGGAACCACCACTC AAGACAAT-3') were used. Dissociation curves for each amplicon were analyzed to verify the specificity of each amplification reaction and the dissociation curve was obtained by heating the amplicon from 60 to 95°C. Fluorescence detection was performed at the annealing/extension step (60°C, 34 s).

Data were analyzed with 7500 SDS 1.4.0 software (ABI, Framingham, USA). To generate a baseline-subtracted plot of the logarithmic increase in fluorescence signal (ΔR_n) versus cycle number, baseline data were collected between cycles 3 and 15. All amplification plots were analyzed with

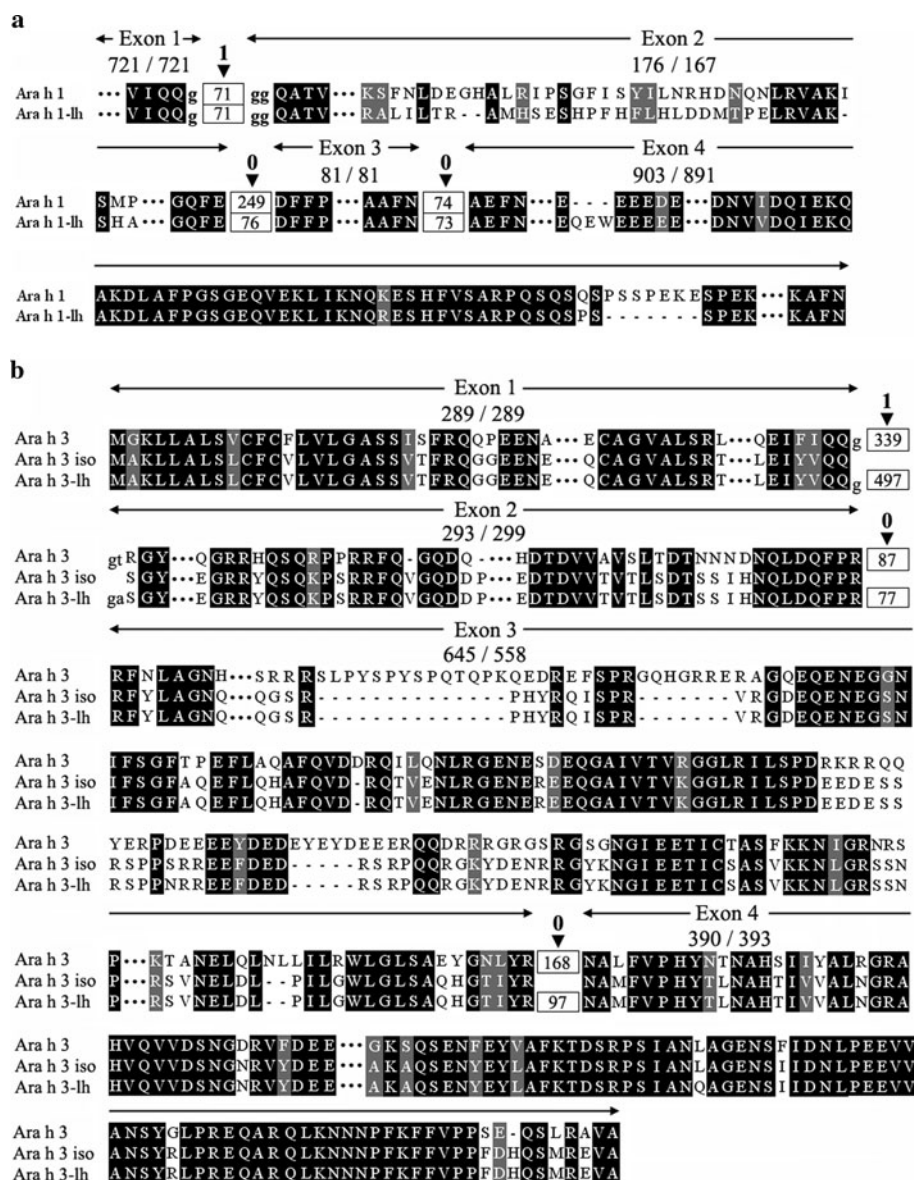
an R_n threshold of 0.2 to obtain C_t values. Expression levels were determined as the number of amplification cycles needed to reach a fixed threshold in the exponential phase of the PCR reaction (C_t). Based on five dilutions of template ranging from 1 to 2^{-4} , the amplification efficiency for each gene was calculated. The mRNA abundance for each candidate gene was calculated by $2^{-\Delta\Delta C_t}$ method (Livak and Schmittgen 2001). Actin gene in peanut was used as the reference to normalize the amount of total transcripts in each sample. The real time quantitative PCR experiments were repeated at least 3 times independently with a new RNA extraction from different plant material each time. No-template controls (NTC) were included in each primer pair PCR reaction. All data were given in terms of relative mRNA amount.

Results

Molecular characterization of peanut allergen genes and proteins

The sequences of *ara h 1*, 2 and 3 were named *ara h 1-lh*, *ara h 2-lh* and *ara h 3-lh* in this research, respectively. The obtained genomic DNA sequence of *ara h 1-lh* (GenBank HM640249) was 2,425 bp long. The cDNA was 1,860 bp in length which encoded a protein of 619 amino acids with a predicted molecular mass of approximately 70.8 kDa. Analysis of *ara h 1-lh* gene structure showed that it consisted of four exons and three introns. The four exons of 721, 167, 81 and 891 bp were, respectively, interspaced by the three introns of 71, 76 and 73 bp, which all began with GT and ended with an AG dinucleotide, sequences which were thought of being necessary for correct RNA splicing of various eukaryotic genes (Breathnach et al. 1978). Genomic sequence comparison with previously cloned *ara h 1* (GenBank AF432231) demonstrated that exon 1 and 3 were similar, and exon 2 and 4 of *ara h 1-lh* missed 9 and 12 bp, respectively (Fig. 1a). The intron 2 was short of 173 bp particularly. Further analysis of the intron phase distributions stated that the three introns were phase 1 intron, phase 0 intron and phase 0 intron. The BLASTp searching at NCBI showed that the *Ara h 1-lh* shared 91% (580/632) identity with *Ara h 1* and the divergency was mainly caused by the amino acids coded by exon 2 and 4. Analysis of the IgE epitopes showed that IgE-binding peptides (including 11, 15, 20, 22 and 23 epitopes) in *Ara h 1*, were distinct from the corresponding sequences in *Ara h 1-lh*, and the different sites were mainly the charged amino acids (Table 1a). Phosphorylation sites were predicted by NetPhos 2.0 Server and the results presented that there were 25 phosphorylation sites for serine,

Fig. 1 **a** Alignment of Ara h 1 and Ara h 1-lh. **b** Alignment of Ara h 3, Ara h 3 isoform and Ara h 3-lh. The boundary of each exon is marked (← →). The numbers above the sequences mean the length of the exon for each protein. The length of each intron is boxed. The closed inverted triangle marks the type of the intron phase. The same sequences are omitted and indicated by dots. The amino acids are shaded in black and dark gray for identification. Gaps introduced into sequences to optimize alignment are represented by (-)



9 phosphorylation sites for threonine and 3 phosphorylation sites for tyrosine in Ara h 1-lh.

The obtained sequence of *ara h 2-lh* was 519 bp long and the longest ORF encoded a protein of 172 amino acids which was in accordance with Ara h 2.02 (GenBank EF609642), one isoform of Ara h 2. Analysis of the cloned genomic sequence revealed the absence of intron. Phosphorylation sites predictions reported that Ara h 2-lh protein had 9 phosphorylation sites for serine, 4 phosphorylation sites for tyrosine, and no phosphorylation sites for threonine.

The cloned genomic sequence of *ara h 3-lh* (GenBank HM640250) was 2,210 bp long and the cDNA was 1,539 bp in length. Its longest ORF coded a protein of 512 amino acids with a predicted molecular mass of approximately 58.3 kDa. Ara h 3-lh gene consisted of four exons

and three introns. The four exons of 289, 299, 558 and 393 bp were, respectively, interspaced by the three introns of 497, 77 and 97 bp, with the splice junctions between exons and introns respecting the GT-AG consensus sequence rule. Three introns were phase 1 intron, phase 0 intron and phase 0 intron (Fig. 1b). The BLASTp searching at NCBI showed that Ara h 3-lh shared 99% (510/512) identity with Ara h 3 isoform (GenBank ABI17154) and shared 99% (508/510) identity with glycinin (GenBank AAT39430). It had only 71% (368/517) identity with Ara h 3 (GenBank ACH91862). And all the four IgE-binding peptides in Ara h 3 were distinct from the corresponding sequences in Ara h 3-lh, and especially the identity of epitope 4 was only 23%. The substituted amino acids were mainly involved in the charged residues, such as E, D, K and R etc. (Table 1b). Ara h 3 protein had 19

Table 1 Pairwise comparisons at the vital IgE-binding peptides

Peptide/exon	Ara h 1		Ara h 1-lh		Different amino acids
	Sequences	Position	Sequences	Position	
(a) Ara h 1-lh with Ara h 1					
11/3	<u>SYLQ</u> <u>E</u> <u>F</u> SRNT	311–320	SYLQ <u>G</u> FSRNT	308–317	E → G
15/4	NNFGK <u>L</u> FEVK	409–418	<u>D</u> NFG <u>R</u> LFEVK	406–415	N → D, K → R
20/4	IDQIEKQ <u>A</u> <u>K</u> <u>D</u>	551–560	<u>V</u> DQIEKQAKD	551–560	I → V
22/4	KESHFV <u>S</u> ARP	578–587	<u>R</u> ESHFVSARP	578–587	K → R
23/4	PEKESPEKED	597–606	PEKED <u>Q</u> <u>E</u> <u>E</u> <u>E</u> <u>N</u>	595–604	SP → DQ, K → E, D → N
Peptide/exon	Ara h 3		Ara h 3-lh		Different amino acids
	Sequences	Position	Sequences	Position	
(b) Ara h 3-lh with Ara h 3					
1/1	IETWNP <u>N</u> NQEFECAG	33–47	IETWNPNNQEF <u>Q</u> CAG	53–67	E → Q
2/3	GNIFSGFTPE <u>F</u> LAQA	240–254	<u>S</u> NIFSGFA <u>Q</u> EFL <u>Q</u> HA	238–252	G → S, TP → AQ, AQ → QH
3/3	VTVRGGLR <u>L</u> ILSPDRK	279–293	VT <u>V</u> KGGRLILSPD <u>E</u> <u>E</u>	276–290	R → K, RK → EE
4/3	DEDEYE <u>Y</u> DEE <u>D</u> RR	303–315	DED <u>R</u> SRP <u>Q</u> QRG <u>K</u> <u>Y</u>	305–317	EYEYD → RSRPQ, EEDRR → QRGKY

The different IgE epitopes in primary structure are marked by the single-letter amino acids. The position of each peptide and the exon to which the peptide belongs is indicated. The bold, underlined amino acids are the vital sites in Ara h 1 and Ara h 3. The bold, doublelined sites are the substituted amino acids in Ara h 1-lh and Ara h 3-lh. The different substituted amino acids are also shown

phosphorylation sites for serine, 9 phosphorylation sites for tyrosine and 4 phosphorylation sites for threonine.

Phylogenetic analysis

Using the deduced Ara h 1-lh and Ara h 3-lh as templates, conserved domain search in CDD database (Marchler-Bauer et al. 2009) revealed that Ara h 1-lh and Ara h 3-lh belonged to the bi-cupin superfamily. The phylogenetic tree was constructed by using the sequences with single cupin domain [including germin (GER) and germin-like protein (GLP)], plant seed storage proteins (SSPs) with bi-cupin domain, Ara h 1, Ara h 3 and the related 7S and 11S proteins. The results presented that Ara h 1-lh and Ara h 3-lh proteins were clustered together with the SSPs and formed a monophyletic group, separating from GLPs and GERs proteins (Fig. 2). In the subtree of SSPs, Ara h 1-lh and Ara h 3-lh protein belonged to different sub-branch, implying that they followed distinct evolutionary patterns. Multiple alignments of all proteins in the phylogenetic tree demonstrated that GLPs and GERs had the same conserved active motifs of cupin domain that comprised **G(X)₅H(X)₃E(X)₆G** and **G(X)₅P(X)G(X)₂H(X)₃N** (Khuri et al. 2001), whereas SSPs, Ara h 1-lh and Ara h 3-lh were short of the conserved motifs. In addition, there was single cupin domain in Ara h 1-lh protein by SMART prediction, because Ara h 1-lh missed some sequences when compared to Ara h 1 protein.

Free energy of RNA regions

It has been found that secondary structure of the sequences expanding Kozak sequence and the downstream sequence of initial codon, and secondary structure of 3'-end of mRNA are closely related with the translation efficiency (Zhang et al. 2006). So we explored the minimum free energy of sequences of RNA, including the 53 nt expanding the start codon ("5' region" for short), and 53 nt of 3'-end of mRNA before the polyA sequence ("3' region" for short). The analysis with Vienna online program displayed that the minimum free energies of 5' region and 3' region in the three genes were −12.42 and −12.8 kcal/mol (*ara h 1-lh*), −2.3 and −11 kcal/mol (*ara h 2-lh*) and −4.5 and −13.7 kcal/mol (*ara h 3-lh*).

mRNA localization in mature seed

The in situ hybridization results showed that the three genes were widely expressed in the tissues of the mature seed, including the embryo and cotyledon (Fig. 3), which was in accord with the results reported from Kang et al. (2007). To further determine relative expression amount of the same gene in different parts of the seed, real-time quantitative PCR was employed with actin as internal reference. Dissociation curves of target genes and reference gene exhibiting single peak at corresponding melting temperature were indicative of the specific amplifications.

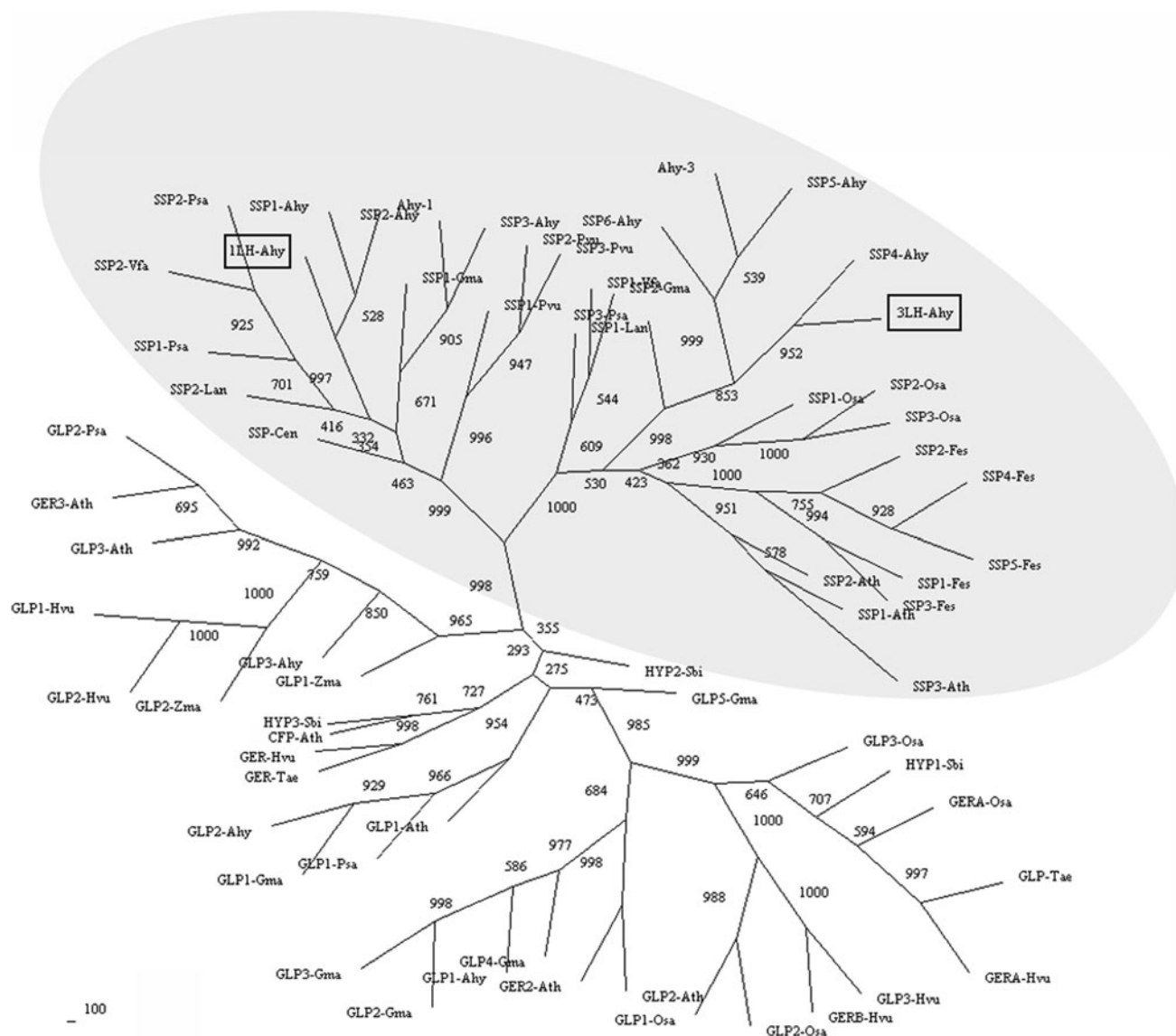


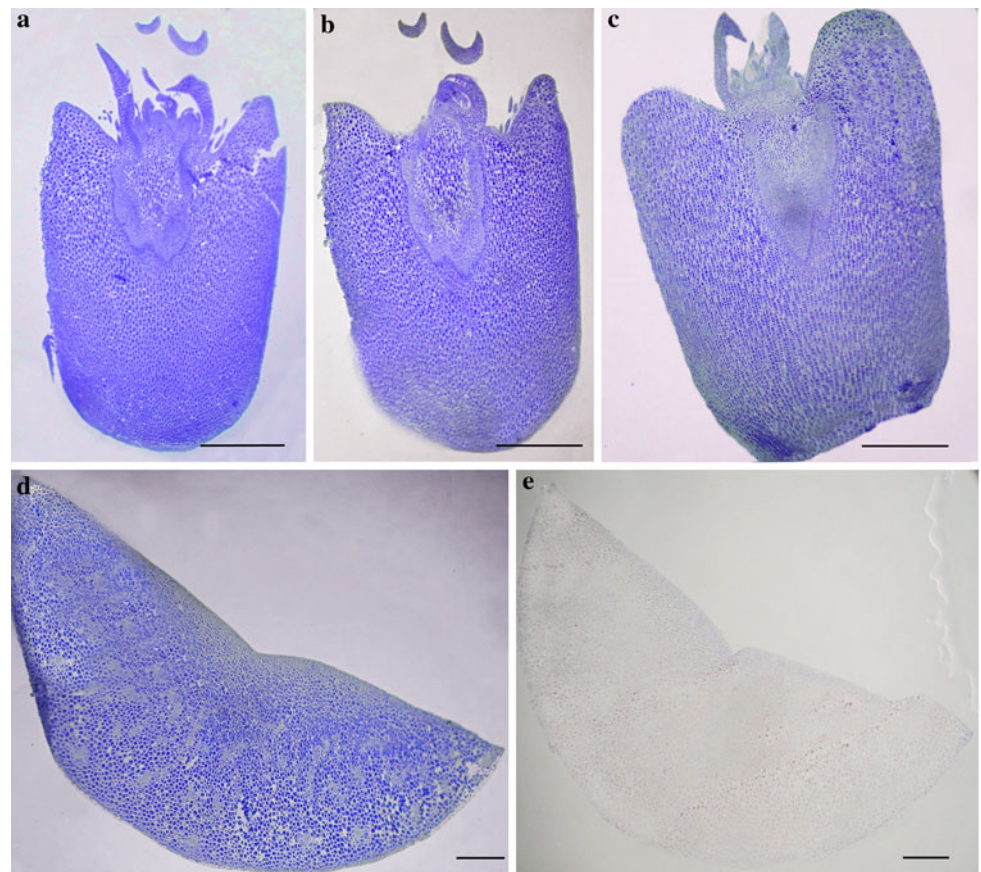
Fig. 2 Unrooted phylogenetic tree constructed by the neighbor-joining method by using the sequences with cupin domain, including germin (GER), germin-like protein (GLP), plant seed storage proteins (SSPs) with bi-cupin domain, Ara h 1, Ara h 3 and the related 7S and 11S proteins. Bootstrap value on 1,000 replicates is indicated at each branch point. The sequences are: *Arabidopsis thaliana*: GER2-Ath (NP_001119334), GER3-Ath (NP_197563), GLP1-Ath (BAD44168), GLP2-Ath (NP_198734), GLP3-Ath (AAG51848), SSP1-Ath (CAB81440), SSP2-Ath (AAD10679), SSP3-Ath (AAD10680), CFP-Ath (NP_177620); *Arachis hypogaea*: GLP1-Ahy (ADD71877), GLP2-Ahy (ADD71876), GLP3-Ahy (ADD71878), SSP1-Ahy (AAT00597), SSP2-Ahy (PDB: 2EAA_A), SSP3-Ahy (PDB: 2EA7_A), SSP4-Ahy (AAT39430), SSP5-Ahy (AAU21490), SSP6-Ahy (AAR02860), 1LH-Ahy (HM640249), Ahy-1 (AAL27476), 3LH-Ahy (HM640250), Ahy-3 (AAM46958); *Canavalia ensiformis*: SSP-Cen (P50477); *Fagopyrum esculentum*: SSP1-Fes (BAA21760), SSP2-Fes (O23878), SSP3-Fes (O23880), SSP4-Fes (Q9XFM4), SSP5-Fes (AAD32713); *Glycine max*: GLP1-Gma (ACG69497), GLP2-Gma

(ACG69478), GLP3-Gma (ACG69479), GLP4-Gma (ACG69480), GLP5-Gma (ACG69481), SSP1-Gma (AAB01374), SSP2-Gma (BAA19058); *Hordeum vulgare*: GERA-Hvu (AAG00425), GERB-Hvu (AAG00426), GER-Hvu (PDB: 1FI2_A), GLP1-Hvu (CAA75907), GLP2-Hvu (ABG46233), GLP3-Hvu (ABG46234); *Lupinus angustifolius*: SSP1-Lan (ACN39600), SSP2-Lan (ACB05815); *Oryza sativa*: GLP1-Osa (Q6YZZ6), GERA-Osa (BAD05773), GLP2-Osa (Q6YZZ7), GLP3-Osa (Q6YZA9), SSP1-Osa (NP_001046769), SSP2-Osa (AAX85990), SSP3-Osa (CAA32706); *Phaseolus vulgaris*: SSP1-Pvu (AAC23610), SSP2-Pvu (P07219), SSP3-Pvu (AAC04316); *Pisum sativum*: GLP1-Psa (CAB65369), GLP2-Psa (CAC34417), SSP1-Psa (CAA47814), SSP2-Psa (CAB82855), SSP3-Psa (PDB: 3KSC_A); *Sorghum bicolor*: HYP1-Sbi (XP_002443983), HYP2-Sbi (XP_002461995), HYP3-Sbi (XP_002466260); *Triticum aestivum*: GER-Tae (AAA34270), GLP-Tae (CAB55559); *Vicia faba*: SSP1-Vfa (CAA81262), SSP2-Vfa (CAP06335); *Zea mays*: GLP1-Zma (NP_001148746), GLP2-Zma (AAQ95582). Ara h 1-lh and Ara h 3-lh are boxed

Fig. 3 Localization of *ara h 1-lh*, *ara h 2-lh* and *ara h 3-lh* transcripts in different tissues of peanut seeds. **a–c** The *ara h 1-lh*, *ara h 2-lh* and *ara h 3-lh* transcripts were observed in embryo, individually.

d Cotyledon view of the *ara h 1-lh* transcript. The views of the other two genes transcripts in cotyledon were omitted because of the similar expression pattern with *ara h 1-lh* transcript.

e Cotyledon control hybridized with the *ara h 1-lh* sense probe. There was no positive signal. Each gene had the appropriate control hybridized with corresponding sense probe. Other controls were identical and were omitted. Bars represent 1,000 μm



Moreover, based on five dilutions of template ranging from 1 to 2^{-4} , the amplification results stated that the efficiencies of the targets and reference gene were similar, so $2^{-\Delta\Delta\text{Ct}}$ method could be used to analyze the data. The transcript level of the *ara h 3-lh* was higher in the cotyledons than in the embryo, with mRNA of *ara h 3-lh*

in cotyledons reaching about 5.13 times compared with that of the embryo. The mRNA of *ara h 1-lh* was in the similar levels in cotyledons and the embryo (about 1.23 times). But mRNA of *ara h 2-lh* in cotyledons was only about one-third of the mRNA in the embryo (Table 2).

Table 2 Expression of *ara h-lh* genes as measured by real-time quantitative PCR

Tissues	<i>ara h 1-lh</i> Ct	<i>ara h 2-lh</i> Ct	<i>ara h 3-lh</i> Ct	<i>actin</i> Ct	$N_{ara h 1-lh}$	$N_{ara h 2-lh}$	$N_{ara h 3-lh}$
Embryo	17.88	20.34	28.51	29.46	1	1	1
	17.81	20.29	28.48	29.34	(0.97–1.04)	(0.93–1.08)	(0.92–1.09)
	17.69	20.26	28.46	29.17			
Cotyledon	16.6	21.08	25.22	28.44	1.23	0.31	5.13
	16.61	21.13	25.25	28.48	(1.20–1.27)	(0.29–0.33)	(4.69–5.62)
	16.57	21.05	25.20	28.36			

The three Ct values of the allergen genes in embryo and cotyledon are shown. $N_{ara h 1-lh}$, $N_{ara h 2-lh}$ and $N_{ara h 3-lh}$ mean normalized *ara h 1-lh*, *ara h 2-lh* and *ara h 3-lh* amount in cotyledon relative to embryo by using $2^{-\Delta\Delta\text{Ct}}$ method, $n = 3$. The transcript levels of the *ara h 1-lh*, *ara h 2-lh* and *ara h 3-lh* in cotyledon were about 1.23, 0.31 and 5.13 times compared with those in the embryo

Discussion

There are mounting concerns regarding the introduction of allergic peanut food into the human food supply, since the consequences of allergic reactions can be severe or even fatal to susceptible individuals. The allergens of peanut are the main seed storage proteins. They are synthesized in the endoplasmic reticulum, and then are sorted, transported, packed and stored in the seed (Tandang-Silvas et al. 2010). By a big affectionate kiss, peanut allergens may be transferred to the skin of an infant, and this may be a route for peanut sensitization (Nolan et al. 2007). In view of the negative roles of peanut allergens in human health, the identification of specific biochemical and physiological properties of allergens is of great importance to the basic understanding of allergens.

As members of the cupin superfamily, Ara h 1 and Ara h 3 belong to 7S and 11S globulin family, respectively. In legumes, 11S and 7S globulins constitute about 50–90% of the total seed proteins. In our paper, RNA in situ hybridization results revealed that all three genes were widely expressed in the tissues of the mature seeds, including the embryo and cotyledon. The results of real-time PCR also showed that the three allergens had low Ct values and their Ct values were even lower than that of the internal reference gene, which indicated the high mRNA amounts. Furthermore, Kang et al. (2007) has found that *ara h 1*, 2 and 3 mRNA accumulated in both embryonic axes and cotyledons. Based on all the findings, the high expression levels of the three allergens in seeds can be concluded. So we speculate that some aspects might be correlated with their abundant expressions.

The first aspect involves the stability of the mRNA structure. The low energy of the 3' region in the three mRNA implied that secondary structure at the mRNA 3'-end was stable. De Smit and van Duin (1994) has found that the structure at mRNA 5'-end has no effect on the translation efficiency unless its free energy is <-6.0 kcal/mol. The free energy of 5' region of *ara h 1-lh* was <-6.0 kcal/mol, probably suggesting that the stable 5' region of *ara h 1-lh* might have contributed to higher translation, thus making Ara h 1-lh protein the most abundant allergen in peanut.

The second aspect relates to the intron phase. Noteworthy, our analysis presented a strong conservation of phase 1, 0 and 0 within Ara h 1-lh and Ara h 3-lh genes. These findings show that the degree of gene's conservation extends beyond the primary sequence identity. Here we suppose the conservation of the introns is likely to be a feature of these genes and it might be related to the high and specific expression of the allergen genes. However, little attention and data have been devoted to this research.

The third aspect deals with the gene expression regulation. It was hard to determine the ascertained copy numbers of the three allergens because there were several genes with high homology to the three allergens in peanut genome. However, high accumulation of the three allergens in seeds highlights the effects of the *cis*-acting elements and *trans*-acting factors in the expression regulations of the mRNA transcription. Unfortunately, data are still scarce in the signal transduction networks which regulate the expression of the allergens. Furthermore, there were many phosphorylation sites in the Ara h-lh proteins (37 sites in Ara h 1-lh, 13 sites in Ara h 2-lh and 32 sites in Ara h 3-lh). We speculate that the complicated phosphorylation regulation might be helpful to maintain the proteins' stability or allow these proteins to survive or function under a great variety of extreme conditions, such as the dry store condition.

The question "What makes an allergen an allergen?" has puzzled generations of researchers, and now we still do not have a conclusive answer. Peanut allergens can trigger IgE-mediated allergic reactions, so IgE epitopes in the allergens might be the first possible consideration. The pairwise comparisons of Ara h 1-lh with Ara h 1 and Ara h 3-lh with Ara h 3 indicated that the different sites were mainly involved in the charged amino acids, such as E, D, K and R. Barre et al. (2007) has put forward that the different distribution of charged residues on the molecular surface of the allergens, which significantly occurs in IgE-binding epitopes, is apparently responsible for the discrepancy of their overall three-dimensional conformations. The variations might alter specific physicochemical, functional and physiological properties, thus influence the immunity of the allergens. It has been a successful example that Ara h 3 isoform displays potentially decreased allergenicity (Kang and Gallo 2007). So a peanut where the IgE-binding epitopes of the major allergens have been removed or substituted may prevent the development of peanut hypersensitivity in individuals genetically predisposed to peanut allergy.

Overall, it appears likely that many factors can contribute to the allergenicity and the specific expression patterns of the allergens. However, signal molecules controlling the accumulation of the allergens in the seed are likewise poorly understood. Discovery of the new regulators and understanding the signal networks at a molecular level would be helpful in reducing allergenicity of peanut allergens.

Acknowledgments The work was supported by Key Program of Excellent Talents Foundation of the Education Department of Anhui Province in China (2009SQZR112ZD) and Key Construction Subject of Anhui Science and Technology University (AKXK20102-1). The work described has not been published before and its publication has been approved by all co-authors. The authors have no conflict of

interest. We thank professor Guangming Jin, professor Xunhai Zhang and Mr. Xuan Wang for their assistance. We also appreciate comments and suggestions made by editors and anonymous reviewers.

References

- Azevedo H, Lino-Neto T, Tavares RM (2003) An improved method for high-quality RNA isolation from needles of adult maritime pine trees. *Plant Mol Biol Rep* 21:333–338
- Barre A, Jacquet G, Sordet C, Culerrier R, Rougé P (2007) Homology modelling and conformational analysis of IgE-binding epitopes of Ara h 3 and other legumin allergens with a cupin fold from tree nuts. *Mol Immunol* 44:3243–3255
- Blom N, Gammeltoft S, Brunak S (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol* 294:1351–1362
- Breathnach R, Benoist C, O'Hare K, Cannon F, Chambon P (1978) Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. *Proc Natl Acad Sci USA* 75:4853–4857
- Breiteneder H, Radauer C (2004) A classification of plant food allergens. *J Allergy Clin Immunol* 113:821–830
- Burks AW, Shin D, Cockrell G, Stanley JS, Helm RM, Bannon GA (1997) Mapping and mutational analysis of the IgE-binding epitopes on Ara h 1, a legume vicilin protein and a major allergen in peanut hypersensitivity. *Eur J Biochem* 245:334–339
- De Smit MH, van Duin J (1994) Control of translation by mRNA secondary structure in *Escherichia coli*. A quantitative analysis of literature data. *J Mol Biol* 244:144–150
- Guillaumie S, Mzid R, Méchin V, Léon C, Hichri I, Destrac-Irvine A, Trossat-Magnin C, Delrot S, Lauvergeat V (2010) The grapevine transcription factor WRKY2 influences the lignin pathway and xylem development in tobacco. *Plant Mol Biol* 72:215–234
- Kang IH, Gallo M (2007) Cloning and characterization of a novel peanut allergen Ara h 3 isoform displaying potentially decreased allergenicity. *Plant Sci* 172:345–353
- Kang IH, Srivastava P, Ozias-Akins P, Gallo M (2007) Temporal and spatial expression of the major allergens in developing and germinating peanut seed. *Plant Physiol* 144:836–845
- Khuri S, Bakker FT, Dunwell JM (2001) Phylogeny, function, and evolution of the cupins, a structurally conserved, functionally diverse superfamily of proteins. *Mol Biol Evol* 18:593–605
- Kyndt T, Haegeman A, Gheysen G (2008) Evolution of GHF5 endoglucanase gene structure in plant-parasitic nematodes: no evidence for an early domain shuffling event. *BMC Evol Biol* 8:305
- Lehmann K, Hoffmann S, Neudecker P, Suhr M, Becher WM, Rösch P (2003) High-yield expression in *Escherichia coli*, purification, and characterization of properly folded major peanut allergen Ara h 2. *Protein Expr Purif* 31:250–259
- Letunic I, Doerks T, Bork P (2009) SMART 6: recent updates and new developments. *Nucleic Acids Res* 37:D229–D232
- Li J, Shefcheck K, Callahan J, Fenselau C (2010) Primary sequence and site-selective hydroxylation of prolines in isoforms of a major peanut allergen protein Ara h 2. *Protein Sci* 19:174–182
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* 25:402–408
- Marchler-Bauer A, Anderson JB, Chitsaz F et al (2009) CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res* 37:205–210
- Nolan RC, de Leon MP, Rolland JM, Loh RKS, O'Hehir RE (2007) What's in a kiss: peanut allergen transmission as a sensitizer? *J Allergy Clin Immunol* 119:755
- Rabjohn P, Helm EM, Stanley JS, West CM, Sampson HA, Burks AW, Bannon GA (1999) Molecular cloning and epitope analysis of peanut allergen Ara h 3. *J Clin Invest* 103:535–542
- Shin DS, Compadre CM, Maleki SJ, Kopper RA, Sampson H, Huang SK, Burks AW, Bannon GA (1998) Biochemical and structural analysis of the IgE binding sites on Ara h 1, an abundant and highly allergic peanut protein. *J Biol Chem* 273:13753–13759
- Shreffler WG, Castro RR, Kucuk ZY, Charlop-Powers Z, Grishina G, Yoo S, Burks AW, Sampson HA (2006) The major glycoprotein allergen from *Arachis hypogaea*, Ara h 1, is a ligand of dendritic cell-specific ICAM-grabbing nonintegrin and acts as a Th2 adjuvant in vitro. *J Immunol* 177:3677–3685
- Tandang-Silvas MR, Fukuda T, Fukuda C, Prak K, Cabanos C, Kimura A, Itoh T, Mikami B, Utsumi S, Maruyama N (2010) Conservation and divergence on plant seed 11S globulins based on crystal structures. *Biochim Biophys Acta* 1804:1432–1442
- Yan YS, Lin XD, Zhang YS, Wang L, Wu KQ, Huang SZ (2005) Isolation of peanut genes encoding arachins and conglutins by expressed sequence tags. *Plant Sci* 169:439–445
- Zhang W, Xiao W, Wei H, Zhang J, Tian Z (2006) mRNA secondary structure at start AUG codon is a key limiting factor for human protein expression in *Escherichia coli*. *Biochem Biophys Res Commun* 349:69–78