Available online at www.sciencedirect.com
SCIENCE DIRECT®
Plant Science 169 (2005) 439–445
PLANT SCIENCE
www.elsevier.com/locate/plantsci

Isolation of peanut genes encoding arachins and conglutins by expressed sequence tags

Yong-Sheng Yan [a], Xiao-Dong Lin [a], Yi-Shun Zhang [a], Lei Wang [a], Keqiang Wu [b], Shang-Zhi Huang [a,*]

[a] School of Life Sciences, Zhongshan University, Guangzhou 510275, China
[b] Department of Biology, West Virginia University, Morgantown, WV 26506-6057, USA

Received 2 April 2005
Available online 24 May 2005

Abstract

Seed storage proteins commonly comprise various groups of multiple isoforms encoded by different gene families. Arachin (11S globulin), conarachin (7S globulin) and conglutin (albumin) are the three major storage proteins in peanut (*Arachis hypogaea* L.). In order to isolate peanut seed storage protein genes, we sequenced and analyzed more than 400 randomly selected clones from a cDNA library derived from mid-maturation stage cotyledons of peanut. Our analysis indicated that arachin isoform precursors were encoded by at least five genes, whereas conglutins were encoded by more than six genes. The full-length cDNAs that encode isoforms of five arachin and six conglutin precursors were obtained. Among these cDNA clones, the deduced polypeptide of Ara 1 has about 99% similarity to that of peanut allergen Ara h 3. In addition, the deduced amino acid sequences of three conglutins, AhCONG 1, AhCONG 2 and AhCONG 5, are identical to those of peanut allergens Ara h 2.02, Ara h 2 and Ara h 6, respectively. Our study identified four new members (Ara 2 to Ara 5) of arachin family and two new members (AhCONG 3 and AhCONG 6) of conglutin family in peanut.

Keywords: *Arachis hypogaea*; Expressed sequence tag (EST); Arachin; Conglutin; Peanut allergen

1. Introduction

Developing seeds accumulate massive storage proteins that are used as a source of amino acids during seed germination and seedling growth. Based on their solubility in various extraction solvents [1], seed storage proteins can be classified into four groups: soluble in water (albumins), dilute saline (globulins), alcohol/water mixtures (prolamins), and dilute acid or alkali (glutelins). Storage proteins of legume are mainly composed of 11S globulins (legumin), 7S globulins (vicilin) and albumin [2,3]. Albumins are divided into two major classes, napin and conglutin. Napin 2S albumins are a class of small (1.7–2.2S) sulphur-rich proteins. In most species, napin 2S albumins are synthesized as a single polypeptide that undergoes a number of processing steps to produced two subunits linked by disulphide bonds [4]. Napin 2S albumin exists widely in the seeds of dicotyledonous plants. Conglutins show high sequence homology to proteins of napin 2S albumin family. In contrast to napin 2S albumins, conglutins exist as a single continuous polypeptide chain in peanut seeds [5].

In peanut about 87% of the seed protein is globulin consisting of two major fractions, arachin (glycinin) and conarachin (vicilin) [6,7]. Peanuts have eminent nutritional properties, but can cause severe, type I hypersensitivity reactions. Several different peanut allergens and their genes have been isolated. The peanut allergens Ara h 1 (including Ara h 1 p41b and Ara h 1 p17), Ara h 2 (including Ara h 2.02 and Ara h 2.0), and Ara h 3/Ara h 4 could be assigned to the vicilin, conglutin, and glycinin families of seed storage proteins, respectively [8–10]. Ara h 1 and Ara h 2, both of which are major allergens, are recognized by serum IgE from over 90% of a peanut allergic patient population. In

* Corresponding author. Tel.: +86 20 84036592; fax: +86 20 84036215.
E-mail address: ls18@zsu.edu.cn (S.-Z. Huang).

0168-9452/$ – see front matter © 2005 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.plantsci.2005.04.010

contrast, Ara 3/Ara h 4, Ara h 6, and Ara h 7, which are minor allergens, are recognized by about 40–50% of the patient population. Ara h 2, and the minor allergens Ara h 6 and Ara h 7 show high sequence homology to proteins of the conglutin family from various dicotyledonous plants [9,10]. The genomic structure and organization of peanut allergens Ara h 1 p41b [11], Ara h 1 p17 (β subunit of conarachin) [12], Ara h 2.0 [13] and Ara h 3 [14] have been reported. Besides the peanut allergens, one cDNA fragment encoding an arachin isoform (accession no. AF125192) and one encoding a conglutin isoform [15] were also isolated. However, no information is available on the other members of peanut arachin and conglutin gene families.

The partial sequencing of anonymous cDNA clones (expressed sequence tags, ESTs) has become the method of choice for the rapid and cost-effective generation of data on the coding capacity of genomes. More and more genes have been isolated using ESTs. In plants, this method was initially used for the model species *Arabidopsis thaliana* [16] and *Oryza sativa* [17], and has subsequently been used in many other plant species [18–21]. Peanut EST project has been initiated recently in USA [22]. In this study, we report the identification and analysis of new members of arachin and conglutin gene families, four complete cDNAs encoding arachin and two complete cDNAs encoding conglutin precursors though EST analysis.

2. Materials and methods

2.1. RNA isolation and cDNA library construction

Total RNA was isolated from mid-maturation stage cotyledons of peanut (*Arachis hypogaea* L. Shanyou 523) using TRIZOL Reagent (GIBCOBRL, USA) according to the manufacturer's instruction. The synthesis of cDNA and an adaptor ligation were performed using the SmartTM cDNA Library Construction Kit (CLONTECH). The resulting cDNA was directly subcloned into the λTriplEx2 vector and packaged using MaxPlaxTM lambda packaging extracts (Epicentre Technologies) according to manufacture's guidelines. The resulting primary library contained 5×10^6 pfu.

2.2. DNA template preparation and sequencing

The conversion of a TriplEx2 phagemid to a pTriplEx2 plasmid involved in vivo excision and circularization of a complete plasmid from the recombinant phage. A total of 2 μl phage (5 pfu/ml × 106 pfu/ml) was combined overnight with 200 μl of BM25.8 host cell culture and the mixture was incubated at 31 °C for 30 min without shaking. LB broth (400 μl) was added to the mixture, which was incubated at 31 °C for an additional 1 h with shaking (225 rpm). Finally, the infected cell suspension was spread on a LB/ampicillin plate to obtain isolated colonies. Well-isolated colonies were randomly picked from each clone and plasmid DNA samples were prepared separately using the E.Z.N.A. Plasmid Minipreps DNA Purification System (Omega Bio-tek, USA). The plasmid DNA was stored at −20 °C. Purified plasmid DNA was sequenced to obtain the 5′-end sequence of the insert with a modified SMART primer (5′-CTCGGGAAGCGCGCCATTGTGTT-GGT-3′) and Applied Biosystems BigDye Sequencing Mix. The sequencing reactions were separated on an Applied Biosystems 3700 DNA analyzer.

2.3. Analysis of ESTs

Sequencing with the modified SMART primer produced sequence directly from the 5′-end of the insert, hence no 5′-vector sequence trimming was required. Each sequence obtained was edited using Vecscreen (http://www.ncbi.nlm.nih.gov) to remove flanking vector sequence and assessed manually to determine sequence quality. Comparison of the DNA sequence to non-redundant protein sequence databases at the National Centre for Biotechnology (NCBI) was performed.

2.4. Cloning of complete cDNA

Complete cDNAs of storage proteins were cloned by contig or RACE. The 5′-RACE and 3′-RACE synthetic oligonucleotide primers were constructed based on the EST sequences of storage proteins determined from the analysis of ESTs. The synthetic oligonucleotides were usual 26–30 bases in order to make sure their specificity. cDNA synthesis was performed using a SMART cDNA library construction kit (Clontech) which included an Oligo sequence (5′-AAGCAGTGGTAT-CAACGCAGAGTGGCCATTATGGC-CGGG-3′) and a 5′-PCR primer (5′-AAGCA-GTGGTAT-CAACGCAGAGT-3′). The PCR products were directly subcloned into the pGEM T-Easy vector (Promega).

3. Results and discussion

3.1. Analysis of peanut storage protein ESTs

As an approach to isolate genes encoding seed proteins, over 450 random cDNA clones were sequenced and analyzed from the cDNA library which was constructed from mid-maturation stage cotyledons of peanut. Useable sequences were obtained for 414 of the clones, and seventy ESTs of arachins, conarachins and conglutins were isolated (Table 1). The abundant expressed genes were also shown in 2D PAGE (data are not shown), including *Ara 1*, *Ara 2* (encoding the isoform precursors of arachin), *Ahy α*, *Ahy β* (encoding the α and β subunits of conarachin), and *AhCONG 1*, *AhCONG 2*, *AhCONG 3* and *AhCONG 4* (encoding the isoforms of conglutin). The other arachin and conglutin genes had lower abundant expression than those mentioned above (Table 1).