

M. Laura Ramos · Geraldine Fleming
Ye Chu · Yukio Akiyama · Maria Gallo
Peggy Ozias-Akins

Chromosomal and phylogenetic context for conglutin genes in *Arachis* based on genomic sequence

Received: 3 November 2005 / Accepted: 21 February 2006 / Published online: 14 April 2006
© Springer-Verlag 2006

Abstracts Comparative genomic and cDNA sequence analysis of *ara h 2*, a major peanut allergen, and a related conglutin *ara h 6* were performed in *Arachis hypogaea* L. and its putative progenitors, *Arachis duranensis* and *Arachis ipaensis*. The complete identity between sequences encoding Ara h 2 isoforms demonstrated that these are homeologous genes and represent orthologs from diploid ancestors. Three copies of *ara h 6* were identified in *A. hypogaea*, one of them located in the A-genome and the other two in the B-genome. Expression analysis showed higher levels of *ara h 2* transcripts compared with *ara h 6*. Dual-labeled genomic in situ hybridization permitted identification of two subgenomes, each of which contained one pair of *ara h 2-ara h 6* signals localized by fluorescence in situ hybridization. Characterization of genomic clones showed close genetic linkage between Ara h 2.02 and one copy of *ara h 6* in the B-genome. The physical linkage may have arisen by tandem duplication and divergence of an ancestral gene. A gene duplication event specific to the B-genome progenitor has resulted in *ara h 6* paralogs. These data provide further evidence for progenitor relationships and genomic organization of the conglutin gene family in the genus *Arachis* and could contribute to the development of a hypoallergenic peanut.

Keywords Ara h 2 · Ara h 6 · Conglutins · Peanut allergens · Homeologous genes · Orthologous genes

Introduction

Peanut allergy is an IgE-mediated food allergy (Pumphrey et al. 1999), with clinical symptoms that include facial swelling, itching, rash, wheezing, vomiting and abdominal pain. Anaphylaxis, resulting in airway obstruction, is the cause of death in most cases (Pham and Rudner 2000). To date, eight peanut allergens have been officially named (Burks et al. 1997; Stanley et al. 1997; Kleber-Janke et al. 1999; Rabjohn et al. 1999; Chatel et al. 2003), although eleven proteins, falling into eight protein groups (7S albumin, 11S albumin, conglutin, profilin, PR-10, agglutinin, oleosin and lipid transfer protein) have been documented as allergenic (<http://www.allergome.org>). The protein structure, recognition by sera from peanut-sensitive patients, and epitope specificity have been studied in detail for the three considered the most clinically significant peanut allergens, Ara h 1, Ara h 2 and Ara h 3/4 (Burks et al. 1998). From the perspective of plant development, these allergens are abundant seed storage proteins found in the cotyledons. Most belong to the prolamin and cupin super-families of seed storage proteins (Breiteneder and Radauer 2004), which provide a reservoir of amino acids for use during germination and seedling growth.

Ara h 2, a glycoprotein with homology to the conglutin family, has a molecular mass around 17 kDa (Burks et al. 1992a; Burks et al. 1992b; de Jong et al. 1998; Kleber-Janke et al. 1999; Chatel et al. 2003). This allergen was recognized by 85–90% of patients with peanut allergy (Burks et al. 1992a; Kleber-Janke et al. 1999; Koppelman et al. 2004) establishing its importance in the etiology of the disease. Extensive characterization of purified Ara h 2 by Chatel et al. (2003) clearly showed the presence of two isoforms, Ara h 2.01 and Ara h 2.02. Ara h 2.02 has an insertion of 12 amino acids which

Electronic Supplementary Material Supplementary material is available for this article at <http://dx.doi.org/10.1007/s00438-006-0114-z> and is accessible for authorized users.

Communicated by R. Hagemann

M. L. Ramos · G. Fleming · Y. Chu · Y. Akiyama
P. Ozias-Akins (✉)
Department of Horticulture,
University of Georgia Tifton Campus,
Tifton, GA 31793, USA
E-mail: pozias@uga.edu
Tel.: +1-229-3863902
Fax: +1-229-3867371

M. Gallo
Agronomy Department, Plant Molecular and
Cellular Biology Program, and The Genetics Institute,
University of Florida, Gainesville, FL 32611, USA

contains an extra copy of the sequence DPYSPS, previously described as an immunodominant IgE epitope (Stanley et al. 1997). Along with Ara h 2, at least two other members of the conglutin gene family have been identified in peanut, specifically Ara h 6 and Ara h 7. The percentage of the patient sera that recognized these proteins was 38 and 43%, respectively (Kleber-Janke et al. 1999), leading to their classification as allergens of intermediate frequency. Ara h 6 has homology with Ara h 2, especially in the middle and C-terminal part of the protein. Almost complete inhibition of IgE-interaction with Ara h 2 by Ara h 6 demonstrated that at least some of the epitopes of Ara h 6 were cross-reactive with epitopes on Ara h 2 (Koppelman et al. 2005). Ara h 6 contains ten cysteine residues and has a molecular mass of around 15 kDa. Suhr et al. (2004) were able to identify at least three different isoforms of Ara h 6 by immunodetection, which differed by charge, not by mass. Plant genomes are rich in duplicated genes. Paralogous gene pairs may acquire different fates: (1) evolution of one copy to a pseudogene (Zhang 2003); (2) functional divergence of the two copies, one of them acquiring a new function (neofunctionalization) or simultaneous reduction of activity of both copies maintaining the total capacity of the ancestral gene (subfunctionalization) (Force et al. 1999; Gu et al. 2002) or (3) the functional preservation of both copies to increase the robustness of the genetic network (Force et al. 1999; Gu et al. 2003). Two genes with identical functions are unlikely to be stably maintained in the genome unless they differ in some aspects of their functions, which occur due to subfunctionalization (Wagner 2000; Gu et al. 2002). These changes appear to be a general rule more than the exception for duplicated genes and they occur quickly after duplication (Gu et al. 2002).

During the characterization of different isoforms of peanut allergens, no one has considered the polyploid nature of this crop. Peanut (*Arachis hypogaea* L.) is an allotetraploid ($2n=4x=40$) with an AABB genome constitution (Smartt et al. 1978), whose subgenomes may have their origins in two diploid progenitor species, *Arachis duranensis* and *Arachis ipaensis*, respectively (Kochert et al. 1996; Jung et al. 2003; Seijo et al. 2004). It has been suggested that cultivated peanut originated from a single event of allopolyploidization or, if from multiple events, they always involved the same diploid parental species (Seijo et al. 2004).

Contrary to the biochemical and immunological knowledge of individual peanut allergens there is little information about their genetic inheritance and genomic characterization. Ara h 2 is a highly reactive peanut allergen and its isoforms are targeted for the reduction of allergenicity in this important crop species. Therefore, a comparative genomic and cDNA sequence characterization of this, and related genes within peanut cultivars and their wild diploid ancestors, should enable us to determine: (1) gene copy numbers and (2) whether these copies are putative alleles at the same locus, paralogous duplications or homeologous genes derived from the

orthologs of diploid progenitor species. Furthermore, we show chromosomal localization by FISH of *ara h 2* and *ara h 6* and expression analysis of these seed-specific, developmentally regulated genes in mature and immature seeds as well as in somatic embryos of peanut. These data provide new insights about the complexity of the conglutin gene family in the genus *Arachis*.

Materials and methods

Sources of plant materials

Seeds for peanut species, *A. duranensis* and *A. ipaensis*, were obtained from Dr. Roy Pittman, USDA-ARS Germplasm Conservation Unit, Griffin, GA. Seeds for *A. hypogaea* L. cultivars Georgia Green and Florunner were obtained from Dr. Corley Holbrook, USDA-ARS Crop Genetics and Breeding, Tifton, GA. Seeds were grown in greenhouses to collect leaf material for DNA isolation and seed production. Both the peanut genomic and developing seed-specific cDNA libraries were constructed using the *A. hypogaea* breeding line, F78-1339, and were a gift of Dr. Albert Abbott, Clemson University, Clemson, SC. Their creation is described in Jung et al. (2000).

Southern blots

Genomic DNA was prepared from 1 g fresh or frozen leaf material using the 2× CTAB method described by Doyle and Doyle (1990). Genomic DNA (5–10 µg) from each species or lambda DNA (100 ng) was digested with appropriate restriction enzymes (New England Biolabs, Beverly, MA) following the manufacturer's protocol. Fragments were separated on 0.8% agarose gels run overnight at 20 V and transferred to Genescreen Plus nylon membranes (Cat. No. NEF 986; NEN Research Products, Boston, MA) using a downward alkaline transfer method described in Sambrook et al. (1989). After transfer, blots were baked for 1 h at 80°C. Blots were prehybridized in Church buffer (7% SDS; 0.25 M NaPO₄, 1 mM EDTA and 1% Bovine Serum Albumin) for 2–3 h at 65°C. Before addition of probe, the hybridization buffer was replaced with fresh Church buffer. Incubation with radioactively labeled probe was done overnight at 65°C. Washes: 20 min incubation in 2× SSPE (0.3 M NaCl, 20 mM NaH₂PO₄, 2 mM EDTA), 0.1% SDS at 65°C followed by a 30 min incubation in 0.5% SSPE, 0.1% SDS at 65°C. Hybridization signals were detected using the Cyclone Imaging System with OptiQuant software (Packard, Meriden, CT), as described in Goldman et al. (2003).

Synthesis of radioactively labeled probes for screening of Southern blots and lambda libraries

Radioactively labeled probes were synthesized using PCR as described in Goldman et al. (2003). A plasmid-encoded

copy of the previously described *ara h 2* cDNA (Stanley et al. 1997) was used as template and amplified with primers 90 and 91 (Supplemental Table 1). PCR program: 10 min denaturation step at 95°C followed by 40 cycles of 95°C for 30 s, annealing at 53°C for 30 s and elongation at 72°C for 30 s with a final elongation step at 72°C for 7 min. Synthesized product was separated from unincorporated label by passing the PCR mixture over a Sephadex G-50 column and eluting in STE (100 mM NaCl, 10 mM Tris-HCl, pH 7.5, 1 mM EDTA).

Screening of lambda genomic clones

The peanut genomic library (Lambda FIXII cloning vector, Stratagene, La Jolla, CA) was screened for inserts containing sequences homologous to the previously cloned *ara h 2* cDNA using methods described in Sambrook and Russell (2001). The library was initially plated at a density of 50,000 pfu per plate on 20 plates (150 mm in diameter) in the host strain XL1 Blue MRA (P2). Subsequent platings were on either 150 or 100 mm plates. The plates were incubated at 37°C until clear plaque formation (usually 10–12 h) and then transferred to 4°C to chill the plates before plaque lifts. Plaque lifts were performed according to manufacturer's recommendations using either 137 mm diameter filters (Cat. No. NEF-978; Perkin Elmer Life Sciences Inc., Boston, MA) or 85 mm diameter filters (Cat. No. NJOHY08550, Osmonics, <http://www.osmolabstore.com>). After baking at 80°C for 1 h, the filters were prehybridized, hybridized with radioactively labeled probe and washed as described in the Southern blots. After washing, the filters were wrapped in cellophane and exposed to X-Ray film (Super Rx A125047, Fuji Photo Film Co., Tokyo, Japan) at –80°C. Pure plaques were isolated through sequential screenings.

Screening of seed cDNA libraries

Similar methods were used for screening the seed cDNA library as those described above for the genomic library with the following changes. The host strain for the Lambda GT11 cDNA library was Y1088 (Cat. No. 200263, Stratagene, La Jolla, CA). After plating, plaques usually appeared after incubation at 37°C for 5–7 h. The initial plating of the cDNA library at 50,000 pfu per plate yielded so many positive colonies that eluents were collected from two plates (designated 7 and 12) and diluted for replating at lower densities (200 or 400 pfu per 100 mm plate). After this second plating, plaques were randomly picked and screened by PCR as described below. PCR-positive clones were replated and screened until pure.

PCR with phage eluent

Plugs containing candidate plaques were cut out of plates using Pasteur pipettes and eluted overnight at 4°C in SM [0.1 M NaCl, 8 mM MgSO₄•H₂O, 50 mM Tris-HCl (pH 7.5), 0.01% gelatin solution] and a few drops of chloroform as described in Sambrook and Russell (2001). An aliquot of the eluent was used undiluted or diluted as a source of template DNA. The aliquot was boiled for 5 min or warmed 30 s on high power in a microwave and then placed on ice before addition to a standard PCR reaction.

Preparation of phage DNA

Phage DNA was prepared from liquid lysates according to previously described methods (Sambrook and Russell 2001). DNA concentration was determined by comparison on gels to known concentrations of lambda DNA. Phage DNA was used to construct shotgun libraries.

Table 1 *Ara h 6* nucleotide sequence analysis of *A. hypogaea* cv Georgia Green (GG), *A. ipaensis* (AI) and *A. duranensis* (AD)

Source	No of clones	63	82	87	126	165	250	253	268	375	390	466	attat	529	ctag	606
A-genome type																
Genomic GG	5	g	c	a	g	a	g	c	g	t	c	g	No	t	No	a
cDNA GG	5	g	c	a	g	a	g	c	g	t	c	g	No	t	No	a
cDNA AD	6	g	c	a	g	a	g	c	g	t	c	g	No	t	No	a
B-genome type I																
Genomic GG	5	a	a	g	t	c	a	g	c	c	t	c	Yes	t	No	a
cDNA GG	1	a	a	g	t	c	a	g	c	c	t	c	Yes	t	No	a
cDNA AI	AI5,1	a	a	g	t	c	a	g	c	c	t	c	Yes	t	No	a
Lambda8A2a	2	a	a	g	t	c	a	g	c	c	t	c	Yes	t	No	a
B-genome type II																
Genomic GG	5	a	a	g	t	c	a	g	c	c	t	c	No	c	Yes	g
cDNA GG	1	a	a	g	t	c	a	g	c	c	t	c	No	c	Yes	g
cDNA AI	AI5,2+5,5	a	a	g	t	c	a	g	c	c	t	c	No	c	Yes	g
Lambda17A2a	3	a	a	g	t	c	a	g	c	c	t	c	No	c	Yes	g

Coding region: nucleotides 1–434

Shotgun library construction

Steps in the shotgun library construction were as follows: (1) Filter 8 µg λ DNA in 250 µl at 10,000 rpm for 2 min using a 0.45 µm Millipore filter tube. (2) Shear DNA using a Gene Machines HydroShear under the following conditions: 220 µl volume, 20 cycles and speed code 10. (3) Clean DNA with the Qiagen PCR purification kit. (4) End-repair in the following reaction: 36 µl Qiagen-cleaned DNA, 5 µl 10× Polynucleotide Kinase Buffer, 2 µl 25 mM ATP, 1.75 µl 1 mM dNTPs, 1 µl T4 Polynucleotide Kinase (3 U µl⁻¹), 1 µl Klenow DNA Polymerase (5 U µl⁻¹), 1 µl T4 DNA polymerase (3 U µl⁻¹). (5) Clean DNA as in step 3. (6) Remove 1 µl and check on gel. (7) Ligate into pBluescript II digested with *EcoRV*.

Shotgun library screening

Shotgun libraries were stamped onto prewashed and dried Genescreen Plus membranes (Cat. No. NEF 986; NEN Research Products, Boston, MA) with a 96-well pin and were incubated overnight at 37°C on LB + ampicillin (100 µg ml⁻¹). Filters were processed by floating for 10 min on 10% SDS, 5 min on denaturing solution (0.5 N NaOH, 1 M NaCl) and 5 min in neutralizing solution (0.5 M Tris-HCl, pH 7.5; 1.5 M NaCl). Filters were washed for 15 min in 5× SSPE, pH 7.4 and dried at 80°C for 30–60 min. They were prehybridized and hybridized as described above for Southern blots.

Sequencing and sequence analysis

Sequencing of shotgun libraries was carried out by the Pratt Lab, University of Georgia, Athens, GA (<http://www.fungen.org>) using an AB3730 capillary sequencer (Applied Biosystems). Sequencing of cDNA and genomic fragments was done in-house using a Beckman Coulter CEQ 8000 Genetic Analysis System (Beckman Coulter Inc., Fullerton, CA). Shotgun sequences were initially processed with Phred, assembled with Phrap (Ewing et al. 1998) and viewed with Consed (Gordon et al. 1998). Reassembly, editing and analysis of contigs were conducted using the Vector NTI (VNTI) Suite (InforMax, North Bethesda, MD) and included restriction site analysis, sequence alignments and primer design.

Preparations of chromosomes

Seeds were germinated on moistened filter paper. Root tips were collected and pretreated for 3 h by soaking in 2 mM 8-hydroxyquinoline at room temperature prior to fixation in 3:1 ethanol:acetic acid. The chromosome spreads were prepared by the enzymatic maceration method (Fukui 1996) by incubating for 80–85 min at 37°C.

FISH and GISH

Chromosome preparations were pre-treated with DNase-free RNase (100 µg ml⁻¹ in 2×SSC) for 1 h at 37°C. The slides were washed in 2× SSC three times for 5 min each, digested with a proteinase K solution (0.1 µg ml⁻¹ in 20 mM Tris-HCl, pH 7.5, 2 mM CaCl₂) for 10 min at 37°C followed by a wash in 1× PBS for 5 min. After this step, cytological experiments were carried out according to Akiyama et al. (2004). For FISH probes, fragment A, 0.6 kb including the ORF of *ara h 6*, and fragment B, ~3.5 kb including the 5'UTR, ORF and 3'UTR of *ara h 2*, were labeled with biotin-11-dUTP (Roche) by PCR using the primers 676/657 (Supplemental Table 1) and nick translation, respectively. The probes were mixed before hybridization. For GISH, genomic DNAs from *A. duranensis* or *A. ipaensis* were labeled with biotin-11-dUTP or digoxigenin-11-dUTP (Roche) by nick translation. Slides were incubated with denatured hybridization mixture [approximately 1 ng µl⁻¹ biotin- or digoxigenin-labeled probe, 5% dextran sulfate (MW = 500,000), 50% formamide, 2×SSC] in a humidified chamber at 37°C overnight.

Detection of signals

For GISH, the digoxigenin-labeled probes were detected with fluorescein using a fluorescent antibody enhancer set for DIG detection (Roche). Biotin-labeled probes were detected with Texas-red streptavidin (Vector Laboratories) and biotinylated anti-streptavidin (Vector Laboratories) for a second layer of Texas-red streptavidin. For FISH, the digoxigenin-labeled probes were detected with fluorescein using a fluorescent antibody enhancer set for DIG detection (Roche) and the Alexa Fluor[®]488 signal-amplification kit for fluorescein (Molecular Probes). After detection, the slides were mounted in Vectashield (Vector Laboratories) containing 1.5 µg ml⁻¹ DAPI and observed under a fluorescence microscope, Olympus BX50. Images of chromosomes were captured by a Sensys CCD camera (Photometrics) and Image Pro ver 4.1 software (Media Cybernetics).

Rehybridization

GISH analysis performed on the same spread after FISH was carried out according to Akiyama et al. (2004).

RNA isolation and analysis

RNA was isolated from *A. hypogaea* cv Georgia Green somatic embryos using the RNeasy mini kit (Qiagen). RNA isolation was carried out on several developmental seed stages from Georgia Green, *A. ipaensis* and

A. duranensis using a previously published method with slight modification (Paik-Ro et al. 2002). Immature seed stages for *A. hypogaea* are described in Paik-Ro et al. (2002). Immature seed stage 3 for *A. ipaensis* can be described as a fat seed starting to show pink on the seed coat and having a hard hull. For *A. duranensis*, immature stage 3 was a plump, elongate seed with stripes beginning to show on the seed coat.

Reverse transcriptase PCR (RT-PCR) from total RNA was performed with Superscript III first strand cDNA synthesis system (Invitrogen) and primers 185/371 to amplify *ara h 2* or 676/657 to amplify *ara h 6* (Supplemental Table 1). For northern blots, 20–40 µg of RNA was loaded on a 1.2% formaldehyde denaturing gel, subjected to electrophoresis and transferred onto nylon membrane (GeneScreen Plus). Membranes were prehybridized in 50% formamide, 5× SSC, 5× Denhart's solution, 1% SDS for 2 h at 42°C. A radioactively labeled DNA probe was added into the prehybridization solution and hybridized at 42°C overnight. The membrane was washed one time each with 2× SSC–0.1% SDS, 1× SSC–0.1% SDS and 0.5× SSC–0.1% SDS at 50°C. Blots were exposed using a Storm Phosphoimager (Amersham Biosciences).

Results

Copy number determination of *ara h 2*

Since peanut is allotetraploid and inbreeding, it is likely to contain an *ara h 2* gene in each of the A and B subgenomes and if completely inbred only one allele of each gene would be present (Jung et al. 2000). In order to determine the copy number for *ara h 2* and to predict each copy's subgenomic origin, Southern blot analyses were carried out with genomic DNAs from putative progenitor wild diploid species: *A. duranensis* Krapov & W. C. Gregory (A-genome) and *A. ipaensis* Krapov & W. C. Gregory (B-genome) and from the tetraploid species, *A. hypogaea* L. cv. Georgia Green. Wild diploid species revealed only one band hybridizing to an *ara h 2* probe in *Dra*I digests, but the bands differed slightly in size (Fig. 1a). *A. hypogaea* L. cv. Georgia Green contained two hybridizing bands, which co-migrated with each band present in the wild diploid species, suggesting two genomic copies for *ara h 2* in the tetraploid species. With the aim of determining whether the two hybridizing bands in *A. hypogaea* were likely to represent the orthologs in the diploid relatives, we first sequenced *ara h 2* PCR products from genomic DNA of the two diploid species, *A. duranensis* and *A. ipaensis*.

Genomic sequence of *ara h 2* from putative progenitors *A. duranensis* and *A. ipaensis*

The sequence analysis of *ara h 2* genomic PCR products from putative progenitors of cultivated peanut showed

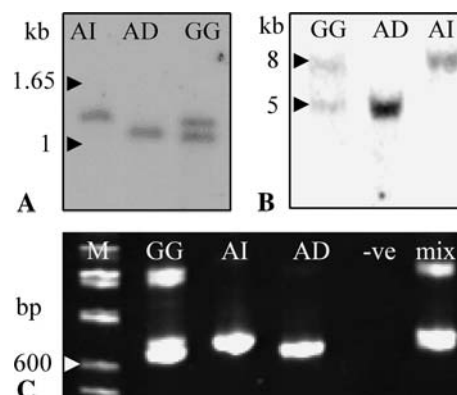


Fig. 1 **a** Southern blot hybridization of *A. ipaensis* (AI), *A. duranensis* (AD) and *A. hypogaea* cv Georgia Green (GG) to determine copy number of *ara h 2*. Each lane was loaded with 10 µg of genomic DNA digested with *Dra*I and hybridized with an *ara h 2*-specific probe. Molecular weight sizes (kb) are indicated on the left. **b** Southern blot hybridization of *A. hypogaea* cv Georgia Green, *A. duranensis* and *A. ipaensis* to determine copy number of *ara h 6*. Each lane was loaded with 5 µg of genomic DNA digested with *Nco*I and hybridized with the internal *arah6* probe. Molecular weight sizes (kb) are indicated on the left. **c** RT-PCR products from *ara h 2* subjected to 5% SDS-PAGE. *A. hypogaea* cv Georgia Green (lane 2), *A. ipaensis* (lane 1), *A. duranensis* (lane 3), negative RT-PCR (lane 4), mixture of cDNA products from *A. ipaensis* and *A. duranensis* denatured at 95°C for 5 min then cooled before loading (lane 5). The molecular weight sizes (bp) are displayed on the left

that the *A. duranensis* genomic sequence was 100% identical to *AhCONG2* (AY722689) and 99.7% identical to *Ara h 2.01* (L77197). The *A. ipaensis* genomic sequence showed 99.9, 99.6 and 99.4% nucleotide similarity to *AhCONG1* (AY581853), AY117434 and *Ara h 2.02* (AY158467), respectively (Supplemental Fig. 1). The genomic sequences from these two diploid species did not indicate the presence of an intron.

The genomic sequences from the progenitors did not share 100% sequence similarity with some of the published cDNA sequences from *A. hypogaea* which could be explained by: (1) sequence divergence since the polyploidization event gave rise to *A. hypogaea*, (2) sequence polymorphisms among the accessions used in this study and the genotypes that actually gave rise to *A. hypogaea* or (3) sequencing errors. Sequencing of genomic and cDNAs from two cultivars of *A. hypogaea* was conducted to distinguish among these possibilities.

Expressed sequences of *ara h 2* from *A. hypogaea* F78-1339

Initial screening of a cDNA library constructed from developing peanut seeds (*A. hypogaea* L. F78-1339) with an *ara h 2* probe showed that the cDNAs for *ara h 2* were extremely abundant. Forty-six randomly picked plaques were selected and tested for *ara h 2* by PCR to categorize the positive plaques as A- or B-genome derived. This random selection of plaques revealed that

ara h 2 genes from both sub-genomes were being expressed in developing peanut seeds in a ratio of 2.7:1 A- to B-genome.

After sequencing, the alignment between A- and B-genome cDNA sequences revealed the expected 36 bp indel (insertion/deletion) and a total of eight single nucleotide polymorphisms (SNPs) between the two homeologous genes (Supplemental Table 2; Supplemental Fig. 2). For our cDNA sequences of the A- and B-genome genes, only two SNPs (positions 182 and 489) were responsible for amino acid changes, a non-conservative E/61/G and a conservative D/163/E change (Supplemental Fig. 3). Our cDNA sequence for the A-genome copy (537 bp) shared 100% similarity with *AhCONG2* and did not vary from the first published cDNA sequence for Ara h 2 (L77197) except for the inclusion of the putative translational start site and an additional five

nucleotides at the 5' end. The B-genome sequence, however, showed two SNPs by comparison with cDNA AY117434, plus the inclusion of the putative translational start site and an additional six nucleotides at the 5' end. Two additional SNPs were identified by comparison of our B-genome cDNA sequence with Ara h 2.02 (AY158467) (Supplemental Table 2; Supplemental Fig. 2). The predicted amino acid sequences corresponding to the complete open reading frames of the B-genome *ara h 2* sequences revealed that only the first SNP at position 29 in AY117434 was non-silent and caused a conservative amino acid substitution at position 10 (L→P) compared with the other derived amino acid sequences. The polymorphisms at amino acid position 142 (142E) for Ara h 2.01 and 142 (142D) for Ara h 2.02 reported by Hales et al. (2004) were not present in our clones (Supplemental Fig. 3).

Table 2 Predicted amino acid sequence analysis for Ara h 6 derived from nucleotide sequences of *A. hypogaea*, *A. ipaensis*, *A. duranensis*

Nucleotide position	121	196	199	202	250	253	268	271	319	348	368	391
AA position	41	66	67	68	84	85	90	91	107	116	123	131
A-genome type ^a	R	D	I	R	D^c	Q	E	R	R	Q	N	R
B-genome type ^b	R	D	I	R	N	E	Q	R	R	Q	N	R
AY871100	R	D	I	R	N	E	Q	G	R	Q	S	R
AF092846	G	N	F	G	N	E	Q	R	G	H	N	G

The deduced amino acid sequences within either A or B-genome type shared 100% similarity

^aSequences included in A-genome type: *A. duranensis*, Psc 33 conglutin (AF366561), *A. hypogaea* cv Georgia Green genomic and cDNA clones A-genome type and *AhCONG4* (AY848699).

^bSequences included in B-genome type: *A. ipaensis*, *A. hypogaea* cv Georgia Green genomic and cDNA clones B-genome type, *AhCONG3* (AY722690) and *AhCONG5* (AY849314)

^cThe amino acid changes at positions 84, 85 and 90 between A-genome type and B-genome type are displayed in bold. Amino acid changes observed in other Ara h 6 published sequences are shown in italics (see also Supplemental Fig. 8).

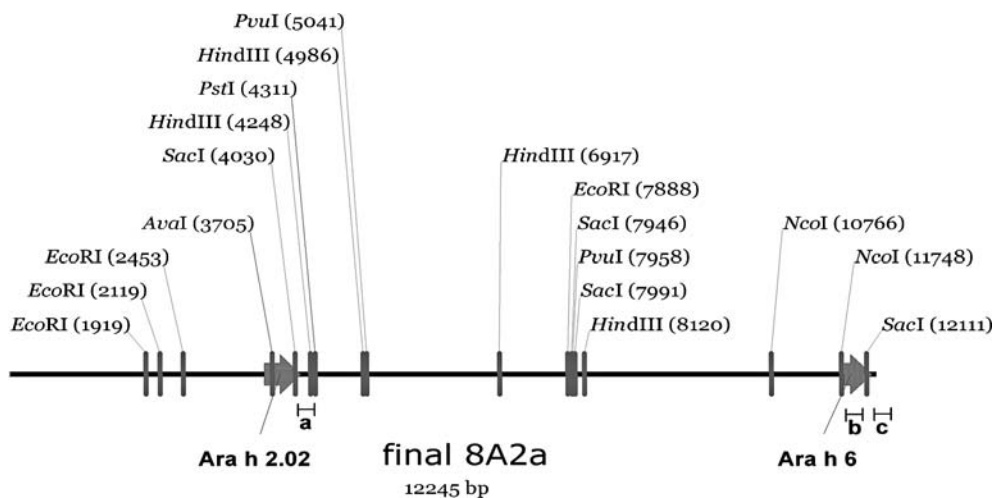


Fig. 2 Contig of 12.24 kb of genomic sequence from lambda clone 8A2a assembled with VNTI. The restriction sites displayed correspond to those used in the Southern blot analysis except for *DraI* sites. The *DraI* sites responsible for the release of bands detected by the 3'arah2 probe are located at positions 3,131 bp and 4,462 bp and by the 3'arah6 probe at 11,328 bp with the second site

in the lambda vector. The hybridization probe sites (relevant to Supplemental Fig. 5) are displayed in lowercase letters on the picture, "a" for 3'arah2, "b" for internal arah6 and "c" for 3'arah6. The ORFs corresponding to *ara h 2* and *ara h 6* are displayed as gray arrows; both of them are in the same orientation

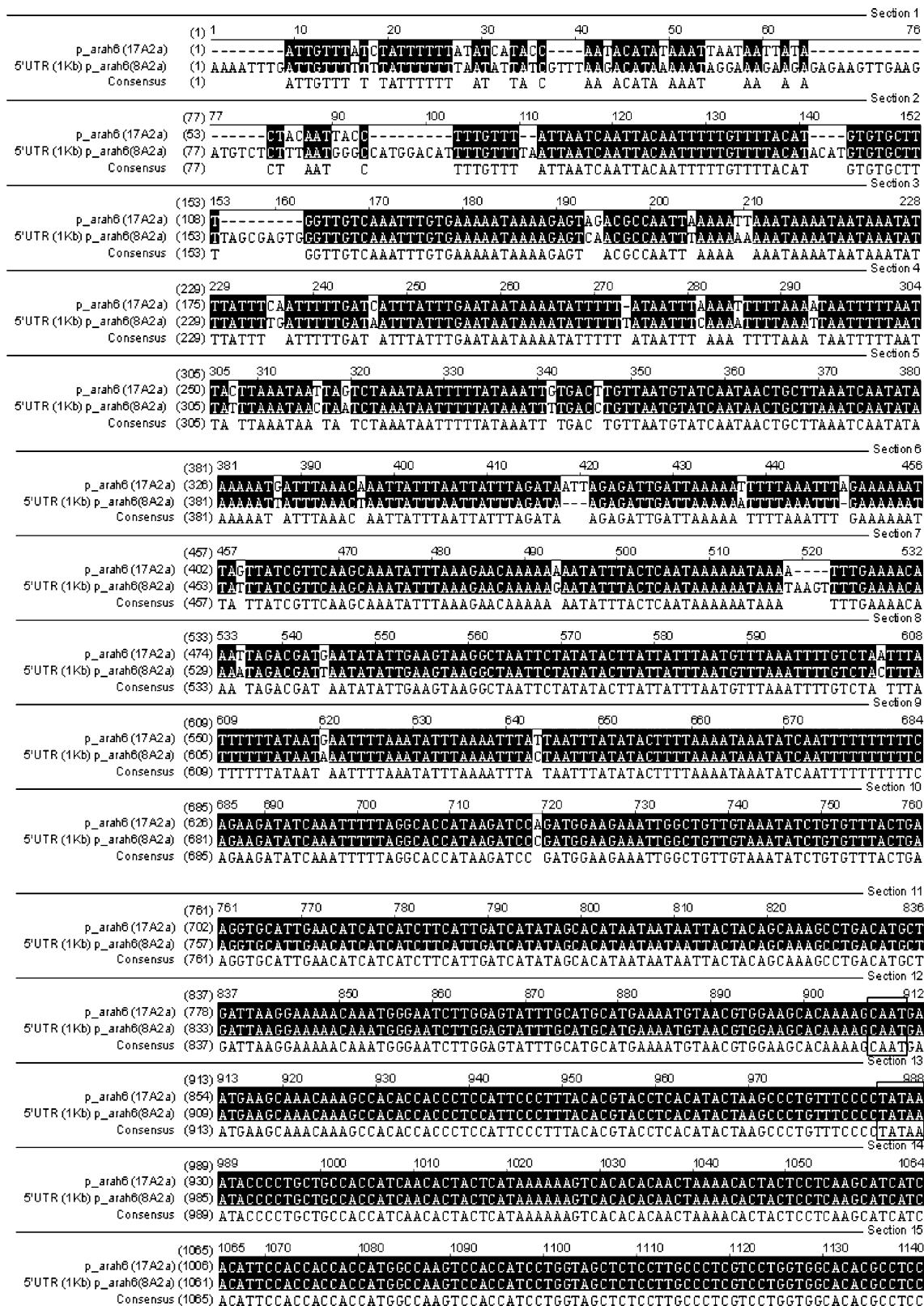


Fig. 3 Indels in the 5'UTR as shown by alignment of paralogous *ara h 6* B-genome sequences from lambda clones 17A2a and 8A2a. A start codon (ATG) is found at position 1081. Identity between

the sequences is noted by *black boxes*. TATA-box and CAAT-box motifs are outlined at positions 984 and 907 that correspond at -96 and -173 from start codons, respectively

Expressed sequences of *ara h 2* from *A. hypogaea* cv. Georgia Green, *A. duranensis* and *A. ipaensis*

Expressed sequences from *A. hypogaea* F78-1339 have been shown to align with 100% identity to genomic sequence from the proposed diploid progenitors. The presence of sequence polymorphisms was implied by comparison of our sequences with some of the previously published cDNA sequences; therefore, we sequenced cDNAs from one additional *A. hypogaea* genotype and the expressed sequences from the two diploid species using cloned RT-PCR products amplified with primers 185/371 (Supplemental Table 1). The sequence analysis of the A- and B-genome sequences of *A. hypogaea* cv. Georgia Green showed 100% sequence similarity with the *ara h 2* sequences belonging to the cDNA library of *A. hypogaea* F78-1339. Contrary to the sequence polymorphisms between published B-genome cDNAs and F78-1339, there was no polymorphism between this genotype and cv. Georgia Green. The cDNA sequences from *A. duranensis* and *A. ipaensis* exactly matched their respective genomic sequences and the A- and B-genome *ara h 2* cDNA sequences from *A. hypogaea* cv. Georgia Green. When *ara h 2* RT-PCR products from the three species were examined on a polyacrylamide gel, the diploid species, *A. duranensis* and *A. ipaensis*, showed only one band each (Fig. 1c). The product from *A. ipaensis* migrated slower than *A. duranensis*. Two doublets were observed from *A. hypogaea* cv. Georgia Green. The faster-migrating doublet comigrated with the two bands collectively present in the wild diploid species. The slower-migrating doublet seen in *A. hypogaea* most likely corresponded to heteroduplexes formed between the A- and B-genome RT-PCR products. This banding pattern was also observed when RT-PCR products of *A. ipaensis* and *A. duranensis* were mixed, denatured, reannealed and run on the polyacrylamide gel (Fig. 1c, mix).

The predicted molecular masses of the two complete encoded proteins were 18.87 and 20.22 kDa for the A- and B-genome genes and 16.64 and 17.99 kDa for the mature proteins, respectively. The calculated difference between the two theoretical mature proteins was 1.36 kDa, which was in excellent agreement with the difference of 1.380 kDa obtained by mass spectrometry (Chatel et al. 2003).

Characterization of genomic sequence of *ara h 2* and related conglutins from *A. hypogaea*

A genomic library of *A. hypogaea* F78-1339 was screened by hybridization with an *ara h 2* cDNA probe and 26 positive plaques were identified. Seven of these were purified through sequential platings and plaque lifts. Two lambda genomic clones were selected for further characterization because they yielded B-genome-size (8A2a) or A-genome-size (17A2a) PCR

products after amplification with primer sets 117/122 or 185/186 (Supplemental Table 1). The two types of clones were presumed to contain the A- and B-genome genes of *ara h 2*.

Sequencing of PCR products from these two lambda clones revealed that the selected primers were amplifying not only *ara h 2* from 8A2a, but also a sequence that had homology with *ara h 6* (AF092846) from 8A2a and 17A2a. Because cDNA sequences of Ara h 2.02 (AY158467) and Ara h 6 (AF092846) share 63% sequence similarity (Supplemental Fig. 4) and both genes can be amplified with certain primer combinations, shotgun libraries were constructed from both lambda clones and the shotgun clones that did not hybridize with lambda DNA were sequenced.

Eighty-four shotgun clones were sequenced from library 8A2a and Phred-quality sequence was assembled with Phrap, viewed with Consed and finally aligned with VNTI. A map covering 12.24 kb of lambda 8A2a is shown in Fig. 2. Only two open reading frames (ORFs) with an ATG start codon showed significant alignments based on a BLASTx query to sequences in GenBank nr. One ORF of 516 bp was identified as encoding Ara h 2.02 and the second ORF of 434 bp as *ara h 6*. Approximately, 7.6 kb of intergenic sequence separated the two ORFs. Both ORFs showed the same orientation with respect to direction of transcription. The ORF corresponding to Ara h 2.02 showed 100% sequence similarity to the B-genome cDNA and genomic sequence of *A. ipaensis*. Southern blot analysis of the lambda clone 8A2a was conducted to verify the restriction sites predicted by the sequence (Supplemental Fig. 5). *In silico* analysis of the 5'UTR region from this B-genome *ara h 2* gene sequence (PlantCARE Promoter Database; Lescot et al. 2002) showed, among other motifs, the presence of a TATA-box (TATAaat) at position -79, three CAAT-box (CAAT) motifs at -1, -30 and -166 bp and a CGTCA-motif involved in MeJA-responsiveness at around -1910 bp upstream of the start codon.

A shotgun library was also constructed with a second genomic clone, 17A2a. After screening 384 shotgun clones with three probes (3'arah6, internal arah6, lambda probe), 12 clones were selected for sequencing. Seven clones formed one contig of 1,806 bp that comprised one complete ORF of 434 bp, 1,021 bp upstream of the ATG site and 364 bp corresponding to the 3'UTR. The ORF from lambda 17A2a shared 100% homology with the ORF of *ara h 6* present in lambda 8A2a; however, the 5' and 3'UTRs showed differences. The 5'UTR of lambda 8A2a and lambda 17A2a shared 360 bp of complete homology upstream of the ATG site (Fig. 3). In this context, the TATA box (-96) and CAAT box (-173) were found at the same position (PlantCARE promoter database). Walking upstream of position -360, several SNPs and indels were identified (Fig. 3). Differences at the 3'UTR are shown in Supplemental Table 3 and Supplemental Fig. 6.

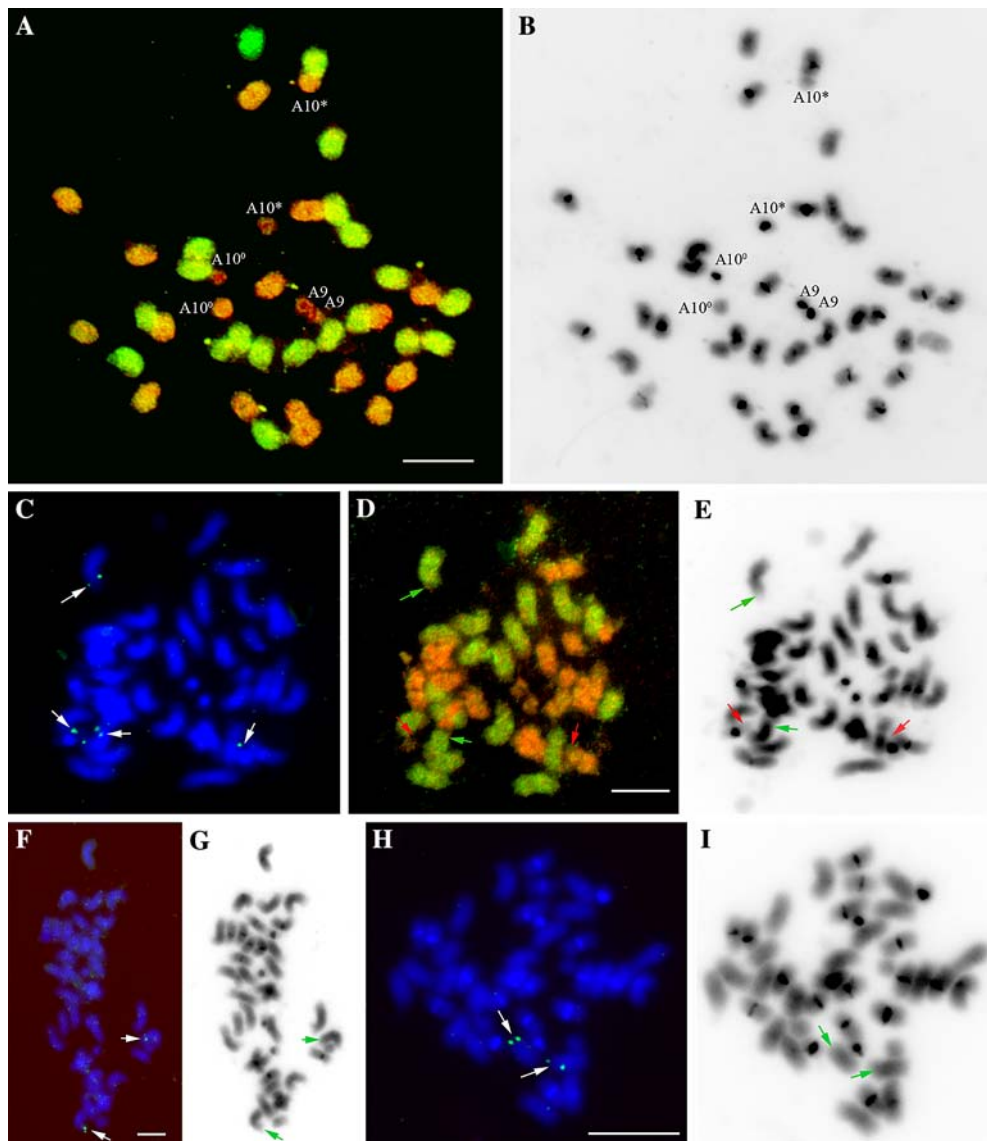


Fig. 4 Genomic in situ hybridization/Fluorescence in situ hybridization (GISH/FISH) to chromosomes of *A. hypogaea* cv Georgia Green. **a-b** Dual-labeled GISH (without use of blocking DNA). **a** Red fluorescence shows hybridization to the biotin-labeled A-genome DNA of *A. duranensis* and green fluorescence indicates hybridization to the digoxigenin-labeled genomic DNA of *A. ipaensis* (B-genome). Twenty chromosomes painted with red (A-genome) showed strong centromeric bands. Twenty chromosomes with weak centromeric bands were painted with green (B-genome). **b** The 4', 6-diamino-2-phenylidole (DAPI) counterstaining the same spread shown in (a). In (a) and in (b) the short arm of the putative chromosome pair A10 is labeled with an asterisk and the satellite by a degree sign as were observed in Seijo et al. (2004). Also, the putative small A9 pair is labeled. Scale bar = 10 μ m. **c-e** GISH/FISH. Some chromosomes are missing from this spread. **c** Green fluorescence signals (white arrows) indicate subtelomeric position of

hybridization of the digoxigenin labeled *ara h 2* and *ara h 6* probes. **d** Red fluorescence shows hybridization to the biotin-labeled A-genome DNA of *A. duranensis* and green fluorescence indicates hybridization to the digoxigenin-labeled genomic DNA of *A. ipaensis* (B-genome). **e** The 4', 6-diamino-2-phenylidole (DAPI) counterstaining of the same spread shown in (c) and in (d). In (d) and (e) red and green arrows indicate the chromosomes showing FISH signals in (C). Scale bar = 10 μ m. **f-i** FISH to highly condensed chromosomes of *A. hypogaea* cv Georgia Green. In (f) and (h) green fluorescence signals (white arrows) indicate subtelomeric position of hybridization of the digoxigenin labeled *ara h 2* and *ara h 6* probes. **g, i** 4', 6-diamino-2-phenylidole (DAPI) counterstaining the same spreads shown in (f) and (h), respectively. Green arrows show the chromosomes with weak centromeric bands that belong to the B-genome (*A. ipaensis*) as were confirmed by GISH in previous experiments. Scale bar = 10 μ m

Copy number of *ara h 6*

Close physical linkage of two conglutin genes, *ara h 2* and *ara h 6*, prompted the generation of additional genomic and cDNA sequences of *ara h 6* and determination of its copy number. With the purpose of

determining the copy number for *ara h 6*, Southern blot analysis was carried out with *Nco*I-digested genomic DNA from wild diploid species, *A. duranensis* and *A. ipaensis*, and cultivated tetraploid, *A. hypogaea* cv Georgia Green. When probed with an internal *ara h 6* probe, only one clear band hybridized in each diploid

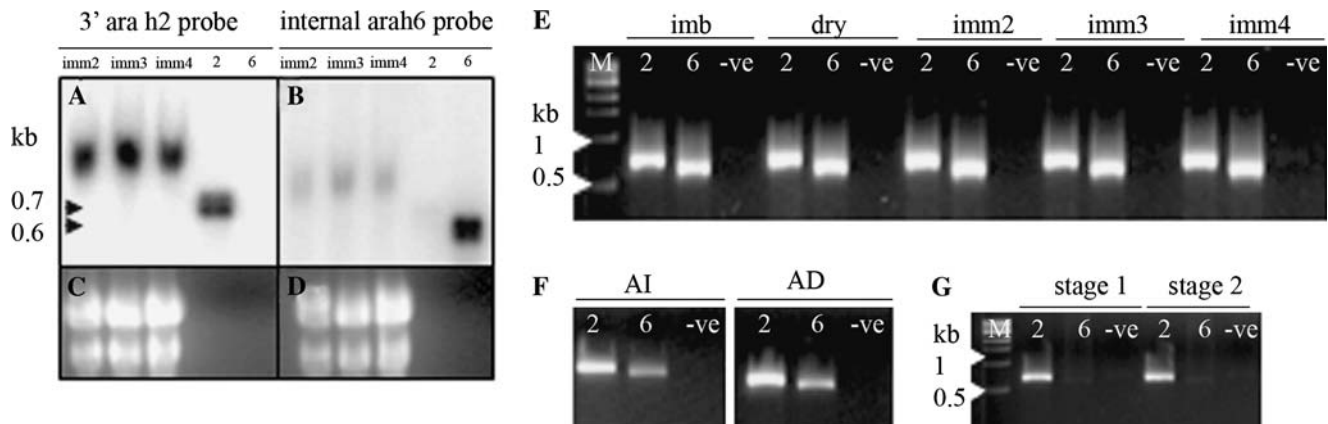


Fig. 5 a, b Northern blot analysis of *A. hypogaea* cv Georgia Green immature seed stages. Specific probes were used to detect *ara h 2* (a) and *ara h 6* transcripts (b) across the immature seed stages. c, d Stained gel as control for RNA loading in (a) and (b). e, f, g RT-PCR analysis. e RT-PCR from different developmental stages of *A. hypogaea* cv Georgia Green: imbibed seeds (imb); dry seeds (dry); immature seeds, stage 2 (imm2); immature seeds, stage 3 (imm3);

immature seeds, stage 4 (imm4). f RT-PCR from developmental stage 3 of *A. ipaensis* (AI) and *A. duranensis* (AD). g RT-PCR from somatic embryos of *A. hypogaea* cv Georgia Green, stages 1 and 2 (Supplemental Fig. 11), immature seed stage 2 (imm2), immature seed stage 3 (imm3), immature seed stage 4 (imm4) *ara h 2* RT-PCR products (2), *ara h 6* RT-PCR products (6), negative control (-ve). Molecular weight sizes (kb) on the left

species, but they differed in size (Fig. 1b). The two bands, which hybridized in the tetraploid species, were of the same size as the bands present in the diploid species, suggesting that there are at least two genomic copies for *ara h 6* in the tetraploid.

To obtain a full complement of cDNA and genomic sequences for *ara h 6*, PCR was carried out with primers 676/657 (Supplemental Table 1) using genomic and cDNAs from *A. hypogaea* cv. Georgia Green. The PCR products were then cloned and sequenced. Thirteen SNPs were found among these clones (Table 1 and Supplemental Fig. 7).

Five cDNA clones and five genomic clones from *A. hypogaea* demonstrated 100% sequence similarity with the *ara h 6* cDNA sequence from *A. duranensis* (A-genome type). One cDNA clone and five genomic clones from *A. hypogaea* L. cv Georgia Green were 100% identical to one type of *ara h 6* sequence found in *A. ipaensis* (B-genome type I). The *ara h 6* sequence belonging to the lambda 8A2a genomic clone fell in this group. One cDNA clone and five genomic clones were homologous to a second *ara h 6* cDNA sequence identified from *A. ipaensis* (B-genome type II). The *ara h 6* sequence belonging to the lambda genomic clone 17A2a fell in this group. Therefore, two copies of *ara h 6* were identified in *A. ipaensis*. The differences among these sequences were in the 3'UTR. The coding regions shared 100% sequence similarity. This result was consistent with the sequences derived from *A. hypogaea* F78-1339 and cv Georgia Green. *A. duranensis* showed only one type of *ara h 6* sequence, which exactly matched one of the *ara h 6* sequences found in cv Georgia Green. In total, three *ara h 6* sequences were identified in *A. hypogaea* cv Georgia Green. No intron was present in any of the conglutin sequences.

Translational differences in Ara h 6

From our predicted amino acid sequences for Ara h 6, three nucleotide differences at positions 250, 253 and 268 would induce amino acid changes during translation (Table 2; Supplemental Fig. 8).

All derived *ara h 6* peptide sequences that fell into the *A. duranensis* group (A-genome type) showed an amino acid sequence of DQE at these positions. The deduced sequence from clone Psc33 (AF366561) (Paik-Ro et al. 2002) and *AhCONG4* (AY848699) shared 100% amino acid sequence similarity to the present *ara h 6* *A. duranensis*-like protein.

All predicted Ara h 6 sequences that fit in the *A. ipaensis* group (B-genome type) displayed NEQ at these positions. The derived amino acid sequence of AY871100 displayed NEQ at the same three positions (250, 253 and 258), but had two other amino acid differences compared with the current *ara h 6* sequences. Unfortunately, the genotype from which this sequence originated was not reported. Derived amino acid sequence for the published *ara h 6* allergen gene (AF092846) displayed NEQ at positions 250, 253 and 268, but had eight other amino acid differences compared with our sequences (Table 2).

Our current results for *ara h 6* sequences did not detect any nucleotide differences between the two genotypes of *A. hypogaea* analyzed and their diploid putative progenitor species. Thus, the high number of mutations shown by AF092846 derived from a virginia-type cultivar of *A. hypogaea* could be due to sequencing errors rather than indicating the occurrence of a different isoform. However, none of the predicted amino acid changes affected cysteine residues, which are important to the tertiary structure of these proteins.

Signal peptide cleavage prediction for Ara h 6

Suhr et al. (2004) sequenced the N-terminus of purified Ara h 6 protein. Their peptide sequence (MRRERGRQGDS) was identical to our predicted Ara h 6 amino acid sequence starting at Met₂₂ from cDNA and genomic clones (*A. hypogaea*, *A. duranensis* and *A. ipaensis*). The N-terminal sequence of *ara h 2* (RQXWELQGDR, X = unidentified) reported by Chatel et al. (2003) showed 100% identity with the predicted amino acid sequence of *ara h 2.01* and *ara h 2.02*, beginning at Arg 22. Starting at Met₁ and using SignalP 3.0 (<http://www.cbs.dtu.dk/services/SignalP/>), we searched for the signal peptide cleavage prediction in Ara h 2 and Ara h 6 (Supplemental Fig. 9). These searches suggested that the most probable site for cleavage of Ara h 6 and Ara h 2 was between position 21 and 22 at the sites ASA-MR and ASA-RQ, respectively. Therefore, both Ara h 2 and Ara h 6 should enter the endoplasmic reticulum of the cell during translation.

Chromosomal location of *ara h 2* and *ara h 6* genes

DAPI staining revealed a differentiated pattern of heterochromatic and euchromatic regions in chromosomes of *A. hypogaea* cv Georgia Green as were reported by Raina and Mukai (1999). Dual-labeled genomic in situ hybridization (GISH) showed that 20 of 40 chromosomes with weak centromeric bands were painted by the probe containing labeled genomic DNA from *A. ipaensis*. The remaining 20 chromosomes with strong centromeric bands hybridized with the probe coming from *A. duranensis* (Fig. 4a, b).

In *A. hypogaea* cv Georgia Green, two pairs of *ara h 2-ara h 6* signals were detected by fluorescence in situ hybridization (FISH). One pair was localized on strong centromeric-band-containing chromosomes belonging to the A-genome. The other pair was detected on chromosomes belonging to the B-genome. In both A- and B-genomes, signals were localized at a subtelomeric position (Fig. 4c–e). The localization of *ara h 2-ara h 6* performed on highly condensed chromosomes showed consistently one pair of signals in a subtelomeric position on the B-genome (Fig. 4f–i).

Expression of *ara h 2* and *ara h 6* in embryos

To determine the expression of these two conglutin genes, northern blot analysis was carried out with gene-specific probes (3'arah2 and internal arah6) using total RNA extracted from three immature seed stages (Fig. 5a, b). The specificity of the 3'arah2 probe had been tested previously on a Southern blot of RT-PCR products run on a polyacrylamide gel (Supplemental Fig. 10). The level of expression for each gene was similar across these developmental stages, which was in accordance with the results reported by Paik-Ro et al.

(2002). The *ara h 2* hybridization signal was stronger than *ara h 6* suggesting indirectly that expression levels of *ara h 2* in these developmental stages were higher than *ara h 6* (Fig. 5a, b).

The levels of expression of *ara h 2* and *ara h 6* in immature seeds of *A. ipaensis*, *A. duranensis* and in somatic embryos of *A. hypogaea* cv Georgia Green were not detectable by northern blot (data not shown). Two reasons could explain this observation: (1) 20 µg of total RNA rather than 40 µg as in Fig. 5 were loaded due to a limited amount of material or (2) the level of expression in the diploid species may be lower than that from *A. hypogaea* immature seeds. A RT-PCR assay, which is more sensitive than northern blot analysis, was used to detect expression of these genes. Expression of both *ara h 2* and *ara h 6* in mature seeds (dry and imbibed) as well as immature seed stages 2, 3 and 4 of Georgia Green was observed (Fig. 5e). Expression of *ara h 2* and *ara h 6* was also detected in immature seed stage 3 (the only stage assayed) from *A. ipaensis* and *A. duranensis* (Fig. 5f). Somatic embryo stages 1 and 2 of cv Georgia Green (Supplemental Fig. 11) expressed *ara h 2*, but expression of *ara h 6* was barely detectable (Fig. 5g).

Discussion

Genes encoding the isoforms Ara h 2.01 and Ara h 2.02 are homeologous genes representing orthologs from diploid ancestors

On the basis of comparative genomic and cDNA sequence characterization of *ara h 2*, two genes encoding the two isoforms Ara h 2.01 and Ara h 2.02 were identified, and 100% sequence similarity was found among *A. hypogaea* cultivars and between *A. hypogaea* and its putative ancestors. These results, in addition to Southern blot analysis, allowed assignment of one copy corresponding to Ara h 2.01 in *A. hypogaea* to *A. duranensis*, the A-genome progenitor. The copy in *A. hypogaea* encoding Ara h 2.02 was present only in *A. ipaensis*, the B-genome progenitor. Because *A. hypogaea* has low genetic variability (Kochert 1991; Kochert et al. 1996) and no polymorphisms were identified in genes encoding Ara h 2.01 or Ara h 2.02 in the *A. hypogaea* cultivars and the respective putative ancestors that we studied, the sequence polymorphisms detected for Ara h 2.02 by Chatel et al. (2003) and Hales et al. (2004) could be due to sequencing error rather than natural variation, particularly in the case of Chatel et al. (2003) who used the same cDNA library that we used in our study.

All previously published sequences have been from cDNAs except for the “first genomic” clone for *ara h 2* reported by Viquez et al. (2001) who purportedly described the regulatory sequences of this allergen gene in *A. hypogaea* F78-1339 (AX148738). Alignment of that sequence with the B-genome sequence reported in the current paper showed that they did not share similarity upstream of the putative translational start site (position

128) nor beyond that clone's position 632, 10 bp upstream of the stop codon in the current *ara h 2* genomic sequence (Supplemental Fig. 12). The ORF reported in the current work comprised 516 bp and TAA as the stop codon in contrast with the 621 bp ORF proposed for the *ara h 2* genomic sequence and TGA as the stop codon reported in Viquez et al. (2001). Using NCBI's VecScreen BLAST function, GenBank ID AX148738 showed a strong match with vector from nt 1 to 99 and 587 to 1152, specifically with the *E. coli* lactose operon (Supplemental Fig. 13, 14). This similarity of upstream and downstream sequences with vector sequence suggests that the *ara h 2* regulatory sequences were not present in this clone.

Differences in expression of genes encoding Ara h 2.01 and Ara h 2.02

Quantification of total Ara h 2 in the extractable protein fraction of mature peanut seed has been estimated between 5.9 and 9.3% (Koppelman et al. 2001), but the ratio of Ara h 2.01 to Ara h 2.02 was not determined. In the current study a PCR analysis to specifically amplify *ara h 2* genes allowed a transcript ratio of 2.7:1, A-genome:B-genome, to be determined. We did not attempt to validate this ratio by real-time RT-PCR; however, a similar ratio was found by Yan et al. (2005) who randomly sequenced more than 400 clones from a cDNA library derived from mid-maturation stage cotyledons of peanut and found a ratio of 2:1, A- to B-genome, for *ara h 2*. In another study, Expressed Sequence Tags (ESTs) generated from a cDNA library constructed from immature pods of peanut showed that the seed storage proteins, glycinin and conglutin, each accounted for over 5% of the total ESTs, giving an idea about the redundancy of expression of these particular genes in this specific tissue (Luo et al. 2005). The number of sequences in dbEST that showed 100% identity to our sequences was too few to support our data on differential expression of A- and B-genome *ara h 2* genes. One possible explanation for the lower expression of the B-genome gene demonstrated in two of the three studies could be a negative effect of the 36 bp insertion. Nevertheless, whether the transcript levels of A- and B-genome sequences are related to *cis* and/or *trans*-regulatory differences, with protein turnover, or are influenced by stage of development will require more extensive studies.

Physical linkage of *ara h 6* and the gene encoding Ara h 2.02 in the B-genome

The linkage between the gene encoding Ara h 2.02 and *ara h 6* was shown by sequencing a 12.24 kb genomic fragment cloned from *A. hypogaea* and confirming the arrangement in the lambda clone by Southern blot analysis. Additionally, these genes were localized by

FISH in *A. hypogaea* cv Georgia Green. Two pairs of *ara h 2-ara h 6* sites were revealed, one on the A- and one on the B-genome, as distinguished by GISH. The chromosomal localization by FISH of *ara h 2-ara h 6* on highly condensed chromosomes, which reduces FISH resolution, was not the same for A- and B-genome loci indicating that the number of genes per locus is variable. The most reproducible hybridization signals were observed on the B-genome, which was consistent with the duplicated sequence information reported for *ara h 6* on the B-genome and reflects the difference in the number of gene copies per locus for both A and B-genomes.

GISH has led to improvements in chromosome identification by permitting their categorization on the basis of diploid parental ancestry and/or intergenomic exchange in tobacco (*Nicotiana tabacum*) (Kenton et al. 1993; Moscone et al. 1996; Chase et al. 2003), in the characterization of genetic stocks derived from crosses of wheat (*Triticum aestivum*) (Han et al. 2003; Fedak and Han 2005) and in the verification of hybrid status of plants in *Lilium* (Marasek et al. 2004), just to provide a few examples. In peanut, previous GISH study indicated that the *A. ipaensis* genome was similar to one of the subgenomic chromosome sets in *A. hypogaea*, whereas the *A. duranensis* genome showed weak and fragmented signals contrary to the intense and uniform signals displayed by *A. villosa* on *A. hypogaea* (Raina and Mukai 1999). In the present study, dual-labeled GISH, without the use of blocking DNA, allowed distinct and uniform hybridization along parental chromosomes in *A. hypogaea* where 20 chromosomes with a centromeric band, including the small A9 pair that characterizes this subgenome, were clearly painted with *A. duranensis* DNA, and the other 20 chromosomes lacking a clear centromeric band were painted with *A. ipaensis* DNA. The discrepancy between the results with *A. duranensis* in the two studies could be due to differences among accessions chosen for the experiments. Chromosomes with intergenomic translocations could not be identified. Thus, a combination of 5S and 18S–25S ribosomal RNA genes, dual-labeled GISH, and allergen gene probes could potentially be used to aid in the identification and study of chromosome evolution in *A. hypogaea* and its diploid relatives.

Paralogous copies of *ara h 6* in the B-genome

Sequencing results provided evidence for three different *ara h 6* genes in *A. hypogaea* cv Georgia Green. The presence of two copies of *ara h 6* in the B-genome and B-genome progenitor, but only one copy of *ara h 6* in the A-genome and A-genome progenitor, argues that a duplication of *ara h 6* occurred in *A. ipaensis* prior to the polyploidization event that founded *A. hypogaea*. Other lines of evidence support a recent origin of *A. hypogaea* from *A. duranensis* and *A. ipaensis*: (1) no evidence for chromosomal rearrangements involving 5S and 18S–25S ribosomal genes (Seijo et al. 2004); (2) 100% sequence

similarity between homeologous genes in *A. hypogaea* and their respective orthologs both in the present study and that of Jung et al. (2003) and (3) the lack of inter-genomic translocations that could be detected by GISH experiments.

Although it is probable that the three *ara h 6* copies are encoding at least two of the three isoforms reported by Suhr et al. (2004), it is not clear whether the B-genome derived isoform is being produced from one or both B-genome copies. Preserved overall similarity between promoters of duplicated genes as well as preservation of individual *cis*-elements within the respective promoters indicates that the process of transcriptional neo- and subfunctionalization is restricted to only a fraction of *cis*-elements (Harberer et al. 2004). Thus, sequence differences in the 5' and 3'UTRs as were found for *ara h 6* B-genome genes could potentially alter temporal and/or spatial expression patterns or transcriptional responses to different stimuli even though the proteins produced are expected to be identical.

Expression of *ara h 2* and *ara h 6* in somatic embryos

Our results showed reduced transcript abundance for *ara h 2* and *ara h 6* by northern blot assays in somatic embryos of *A. hypogaea* cv Georgia Green which is consistent with other studies that showed a reduction in storage protein synthesis (Morcillo et al. 1998; Morcillo et al. 2001). Taking into account that the expression level of *ara h 6* was lower than that of *ara h 2* in seeds, barely detectable levels by RT-PCR in somatic embryos could be anticipated. Nevertheless, detection of *ara h 2* by RT-PCR in stages 1 and 2 of somatic embryos makes it possible to test for silencing of these genes early during a transformation experiment and to avoid unnecessary maintenance of cultures where posttranscriptional gene silencing is not active.

The 2S albumin seed storage protein genes from *Arabidopsis* (*at2S1*, *at2S2*, *at2S3* and *at2S4*) are considered isoforms even though they share only 58–83% of similarity at the amino acid level. These genes follow similar temporal expression profiles throughout development but two of them, *at2S2* and *at2S3*, are expressed at significantly higher levels and tissue specificity differs among the four genes (Guerche et al. 1990). Similarly, *ara h 2* and *ara h 6* are members of the same gene family and their different expression levels could reflect different expression patterns and/or different functions played by these genes in seed tissues.

In the present study we report structural and functional characterization of one of the major peanut allergen genes, *ara h 2*, and a related conglutin gene, *ara h 6*, both expressed in peanut seed. We have determined that orthologs in the putative diploid progenitors are identical in nucleotide sequence to homeologous genes in *A. hypogaea*. The close physical linkage of the gene encoding Ara h 2.02 and one *ara h 6* gene of B-genome origin may have arisen by tandem duplication and

divergence of an ancestral gene. More recent gene duplication specific to the B-genome progenitor has resulted in *ara h 6* paralogs. Peanut seeds are the most economically important part of the peanut plant and are also one of the most allergenic foods. Therefore, genomic characterization of the storage protein genes encoding the triggers of allergic reactions is important from the perspective of allergy diagnosis and is an essential step in the process to develop a hypoallergenic peanut (Dodo et al. 2005; Ozias-Akins et al. 2006). The effects of a knockdown or knockout conglutin genes, particularly *ara h 2*, on the quality, fitness, viability and storage protein composition of the seed are immediate and intriguing questions waiting to be addressed.

Acknowledgements This work was funded by a USDA-CSREES-administered special grant (00-34420-9178), the Georgia Peanut Commodity Commission and the Peanut Foundation. The authors thank Joann Conner for assistance with sequence analysis, Gunawati Gunawan for assistance with sequencing in house, Evelyn Morgan for assistance with embryo tissues, Shailendra Goel for assistance with image analysis and Anne Bell for her technical assistance. They also thank Bert Abbott, Clemson University, for providing peanut genomic and cDNA libraries and the Pratt Lab, UGA Athens Campus, for sequencing and bioinformatics support.

References

- Akiyama Y, Conner JA, Goel S, Morishige DT, Mullet JE, Hanna WW, Ozias-Akins P (2004) High-resolution physical mapping in *Pennisetum squamulatum* reveals extensive chromosomal heteromorphism of the genomic region associated with apomixis. *Plant Physiol* 134:1733–1741
- Breiteneder H, Radauer C (2004) A classification of plant food allergens. *J Allergy Clin Immunol* 114(1):127–130
- Burks AW, Williams LW, Connaughton C, Cockrell G, O'Brien TJ, Helm RM (1992a) Identification and characterization of a second major peanut allergen, Ara h II, with use of the sera of patients with atopic dermatitis and positive peanut challenge. *J Allergy Clin Immunol* 90:962–969
- Burks AW, Williams LW, Thresher W, Connaughton C, Cockrell G, Helm RM (1992b) Allergenicity of peanut and soybean extracts altered by chemical or thermal denaturation in patients with atopic dermatitis and positive food challenges. *J Allergy Clin Immunol* 90:889–897
- 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 I, a legume vicilin protein and a major allergen in peanut hypersensitivity. *Eur J Biochem* 245:334–339
- Burks W, Sampson HA, Bannon GA (1998) Peanut allergens. *Allergy* 53:725–730
- Chase MW, Knapp S, Cox AV, Clarkson JJ, Butsko Y, Joseph J, Savolainen V, Parokonny AS (2003) Molecular systematics, GISH and the origin of hybrid taxa in *Nicotiana* (Solanaceae). *Ann Bot (Lond)* 92:107–127
- Chatel JM, Bernard H, Orson FM (2003) Isolation and characterization of two complete Ara h 2 isoforms cDNA. *Allergy Immunol* 131:14–18
- Dodo H, Konan K, Viquez O (2005) A genetic engineering strategy to eliminate peanut allergy. *Curr Allergy Asthma Rep* 5:67–73
- Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus* 12:13–15
- Ewing B, Hiller M, Wendl C, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8:175–185

- Fedak G, Han F (2005) Characterization of derivatives from wheat-*Thinopyrum* wide crosses. *Cytogenet Genome Res* 109:360–367
- Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531–1545
- Fukui K (1996) Plant chromosomes at mitosis. CRC Press, Boca, Raton, FL
- Goldman JJ, Hanna WW, Fleming G, Ozias-Akins P (2003) Fertile transgenic pearl millet [*Pennisetum glaucum* (L.) R. Br.] plants recovered through microprojectile bombardment and phosphinothricin selection of apical meristem-, inflorescence-, and immature embryo-derived embryogenic tissues. *Plant Cell Rep* 21:999–1009
- Gordon D, Abajian C, Green P (1998) Consed: a graphical tool for sequence finishing. *Genome Res* 8:195–202
- Gu Z, Nicolae D, Lu HH, Li WH (2002) Rapid divergence in expression between duplicated genes inferred from microarray data. *Trends Genet* 18:609–613
- Gu Z, Steinmetz LM, Gu X, Scharfe C, Davis RW, Li WH (2003) Role of duplicate genes in genetic robustness against null mutations. *Nature* 421:63–66
- Guerche P, Tire C, De Sa FG, De Clercq A, Van Montagu M, Krebbers E (1990) Differential expression of the Arabidopsis 2S albumin genes and the effect of increasing gene family size. *Plant Cell* 2:469–478
- Hales BJ, Bosco A, Mills KL, Hazell LA, Loh R, Holt PG, Thomas WR (2004) Isoforms of the major peanut allergen Ara h 2: IgE binding in children with peanut allergy. *Int Arch Allergy Immunol* 135:101–107
- Han FP, Fedak G, Benabdelmouna A, Armstrong K, Ouellet T (2003) Characterization of six wheat x *Thinopyrum intermedium* derivatives by GISH, RFLP, and multicolor GISH. *Genome* 46:490–495
- Harberer G, Hindemitt T, Meyers BC, Mayer KEX (2004) Transcriptional similarities, dissimilarities, and conservation of *cis*-elements in duplicated genes of Arabidopsis. *Plant Physiol* 136:3009–3022
- de Jong EC, Van Zijverden M, Spanhaak S, Koppelman SJ, Pellegroni H, Penninks AH (1998) Identification and partial characterization of multiple major allergens in peanut proteins. *Clin Exp Allergy* 28:743–751
- Jung S, Swift D, Sengoku E, Patel M, Teule F, Powell G, Moore K, Abbott A (2000) The high oleate trait in the cultivated peanut [*Arachis hypogaea* L.]. I. Isolation and characterization of two genes encoding microsomal oleoyl-PC desaturases. *Mol Gen Genet* 263:796–805
- Jung S, Tate PL, Horn R, Kochert G, Moore K, Abbott AG (2003) The phylogenetic relationship of possible progenitors of the cultivated peanut. *J Hered* 94:334–340
- Kenton A, Parokony AS, Gleba YY, Bennett MD (1993) Characterization of the *Nicotiana tabacum* L. genome by molecular cytogenetics. *Mol Gen Genet* 240:159–169
- Kleber-Janke T, Cramer R, Appenzeller U, Schlaak M., Becker WM (1999) Selective cloning of peanut allergens, including profilin and 2S albumins, by phage display technology. *Int Arch Allergy Immunol* 119:265–274
- Kochert G (1991) Restriction fragment length polymorphism in plants and its implications. *Subcell Biochem* 17:167–190
- Kochert G, Stalker HT, Gimenes M, Galgalo L, Lopes CR, Moore K (1996) RFLP and cytogenetic evidence of the origin and evolution of allotetraploid domesticated peanut, *Arachis hypogaea* (Leguminosae). *Am J Bot* 83:1282–1291
- Koppelman SJ, Vlooswijk RA, Knippels LM, Hessing M, Knol EF, van Reijssen FC, Bruijnzeel-Koomen CA (2001) Quantification of major peanut allergens Ara h 1 and Ara h 2 in the peanut varieties Runner, Spanish, Virginia, and Valencia, bred in different parts of the world. *Allergy* 56:132–137
- Koppelman SJ, Wensing M, Ertmann M, Knulst AC, Knol EF (2004) Relevance of Ara h1, Ara h2 and Ara h3 in peanut-allergic patients, as determined by immunoglobulin E Western blotting, basophil-histamine release and intracutaneous testing: Ara h2 is the most important peanut allergen. *Clin Exp Allergy* 34:583–590
- Koppelman SJ, de Jong GA, Laaper-Ertmann M, Peeters KA, Knulst AC, Hefle SL, Knol EF (2005) Purification and immunoglobulin E-binding properties of peanut allergen Ara h 6: evidence for cross-reactivity with Ara h 2. *Clin Exp Allergy* 35:490–497
- Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P, Rombauts S (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res* 30:325–327
- Luo M, Dang P, Guo Z, He G, Holbrook CC, Bausher MG, Lee RD (2005) Generation of expressed sequence tags (ESTs) for gene discovery and marker development in cultivated peanut. *Crop Sci* 45:346–353
- Marasek A, Hasterok R, Wiejacha K, Orlikowska T (2004) Determination by GISH and FISH of hybrid status in Lilium. *Hereditas* 140:1–7
- Morcillo F, Aberlenc-Bertossi F, Hamon S, Duval Y (1998) Differential accumulation of storage proteins, 7S globulins, during zygotic and somatic embryos development in oil palm (*Elaeis guineensis* Jacq.). *Plant Physiol Biochem* 36:509–514
- Morcillo F, Hartmann C, Duval Y, Tregear JW (2001) Regulation of 7S globulin gene expression in zygotic and somatic embryos of oil palm. *Physiol Plant* 112:233–243
- Moscone EA, Matzke MA, Matzke AJ (1996) The use of combined FISH/GISH in conjunction with DAPI counterstaining to identify chromosomes containing transgene inserts in amphidiploid tobacco. *Chromosoma* 105:231–236
- Ozias-Akins P, Ramos ML, Chu Y (2006) Hypoallergenic foods beyond infant formulas. In: Food allergy: comprehension, treatment, and prevention. ASM Press, Herndon, VA (in press)
- Paik-Ro OG, Seih JC, Smith RL (2002) Seed-specific, developmentally regulated genes of peanut. *Theor Appl Genet* 104:236–240
- Pham TS, Rudner EJ (2000) Peanut allergy. *Cutis* 65:285–289
- Pumphrey RSH, Wilson PB, Faragher EB, Edwards SR (1999) Specific immunoglobulin E to peanut, hazelnut and brazil nut in 731 patients: similar patterns found at all ages. *Clin Exp Allergy* 29:1256–1259
- Rabjohn P, Helm EM, Stanley JS, West CM, Sampson HA, Burks AW, Bannon GA (1999) Molecular cloning and epitope analysis of the peanut allergen Ara h 3. *J Clin Invest* 103:535–542
- Raina SN, Mukai Y (1999) Genomic in situ hybridization in *Arachis* (Fabaceae) identifies the diploid wild progenitors of cultivated (*A. hypogaea*) and related wild (*A. monticola*) peanut species. *Plant Syst Evol* 214:251–262
- Sambrook J, Russell DW (2001) Molecular cloning, a laboratory manual, 3rd edn. Cold Spring Harbor Press, Cold Spring Harbor
- Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning, a laboratory manual. 2nd edn. Cold Spring harbor Press, Cold Spring Harbor, NY
- Seijo JG, Lavia GI, Fernandez A, Krapovickas A, Ducasse D, Moscone EA (2004) Physical mapping of the 5S and 18S-25S rRNA genes by FISH as evidence that *Arachis duranensis* and *A. ipaensis* are the wild diploid progenitors of *A. hypogaea* (Leguminosae). *Am J Bot* 91(9):1294–1303
- Smartt J, Gregory WC, Gregory MP (1978) The genomes of *Arachis hypogaea*. 1. Cytogenetics studies of putative genome donors. *Euphytica* 27:665–675
- Stanley JS, King N, Burks AW, Huang SK, Sampson H, Cockrell G, Helm RM, West CM, Bannon GA (1997) Identification and mutational analysis of the immunodominant IgE binding epitopes of the major peanut allergen Ara h 2. *Arch Biochem Biophys* 342:244–253
- Suhr M, Wicklein D, Lepp U, Becker W-M (2004) Isolation and characterization of natural Ara h 6: Evidence for a further peanut allergen with putative clinical relevance based on resistance to pepsin digestion and heat. *Mol Nutr Food Res* 48:390–399

- Viquez OM, Summer CG, Dodo HW (2001) Isolation and molecular characterization of the first genomic clone of a major peanut allergen, Ara h 2. *J Allergy Clin Immunol* 107:713–717
- Wagner A (2000) Decoupled evolution of coding region and mRNA expression patterns after gene duplication: implications for the neutralist-selectionist debate. *Proc Natl Acad Sci U S A* 97:6579–6584
- Yan YS, Lin XD, Zhang YS, Wang L, Wu K, Huang SZ (2005) Isolation of peanut genes encoding arachins and conglutins by expressed sequence tags. *Plant Sci* 169:439–445
- Zhang J (2003) Evolution by gene duplication: an update. *Trends Ecol Evol* 18:292–298

Supplemental Table 1

List of primers					
Primer Name	Target Sequence	Position of 3' nucleotide	Primer Length (nt)	Direction	Use (second primer/annealing temp)
#90	A- or B-genome <i>ara h 2</i>	170 ^a	18	sense	<i>ara h 2</i> probe (#91/53°C)
#91	A- or B-genome <i>ara h 2</i>	677 or 713	22	antisense	<i>ara h 2</i> probe (#91/53°C)
#117	A- or B-genome <i>ara h 2</i>	28	20	sense	PCR (#122/55°C)
#122	A- or B-genome <i>ara h 2</i>	346 or 382	18	antisense	PCR (#117/55°C)
#185	A or B-genome <i>ara h 2</i>	8	25	sense	PCR (#186/60°C) full arah2 (#371/58°C)
#186	A- or B-genome <i>ara h 2</i>	267 or 303	26	antisense	PCR (#185/60°C)
#370	A- or B-genome <i>ara h 2</i>	478 or 514	22	sense	3'arah2 probe (#371/ 55°C)
#371	A- or B-genome <i>ara h 2</i>	654 or 690	24	antisense	3'arah2 probe (#370/ 55°C)
#676	A- or B-genome <i>ara h 6</i>	34	34	sense	full arah6 probe (#657/58°C)
#657	A- or B-genome <i>ara h 6</i>	603 or 608	27	antisense	full arah6 (#676/58°C) 3'arah6 probe (#656/56°C)
#441	A- or B-genome <i>ara h 6</i>	101 ^b	23	sense	internal arah6 probe (#658/67°C)
#658	A or B-genome <i>ara h 6</i>	275 ^b	20	antisense	internal arah6 probe (#441/67°C)
#656	A or B-genome <i>ara h 6</i>	432	20	sense	3'ara h6 probe (#657/56°C)

^aOne mismatch at 5' end primer in B-genome.

^bTwo mismatches at 5' end primer in A-genome.

Supplemental Table 2

Nucleotide positions of SNPs and indels between *ara h 2* cDNA sequences from *A. hypogaea* L. F78-1339 and published sequences.

position	1-8 indel	9	29	81	126	156	182	227-262 indel	243	272	288	300	372	462	489
<i>AhCONG2</i> (AY722689)	8bp	G	T	C	G	G	A	-	-	C	C	A	C	T	C
Ara h 2.01 (L77197)	-	G	T	C	G	G	A	-	-	C	C	A	C	T	C
cDNA <i>ara h 2</i> (A-genome)	8bp	G	T	C	G	G	A	-	-	C	C	A	C	T	C
cDNA <i>ara h 2</i> (B-genome)	8bp	G	T	A	T	A	G	36bp	T	A	C	A	A	C	A
AY117434	-	-	C	A	T	A	G	36bp	C	A	C	A	A	C	A
Ara h 2.02 (AY158467)	8bp	G	T	C	G	A	G	36bp	T	A	T	A	A	C	A
<i>AhCONG1</i> (581853)	8bp	G	T	A	T	A	G	36bp	T	A	C	G	A	C	A

SNPs that induce changes at the amino acid level are displayed in black boxes, and in bold those that do not induce changes (see also supplemental Fig. 2). Position numbers begin at the putative translational start site (ATG) and extend to the stop codon (TAA) at position 517.

Supplemental Table 3

SNPs and indels present in the 3'UTRs of two <i>ara h 6</i> genes, one from lambda 8A2a (B-genome type I) and one from lambda 17A2a (B-genome type II)					
3'UTR <i>ara h 6</i> sequence	485-489 TTATA site	529	557-560 CTAG site	606	651
Lambda 8A2a	Yes	T	No	A	28bp
Lambda 17A2a	No	C	Yes	G	54bp

Position numbers correspond to the 3'UTR. The stop codon is located at position 436. Starting at position 651 and downstream the sequences diverge with no homology (see also supplemental Fig. 4).

Supplemental Fig. 1 Nucleotide sequence alignment between *ara h 2* genomic sequences derived from *A. duranensis* (AD), *A. ipaensis* (AI), published cDNA sequences of *A. hypogaea* that shared the higher levels of similarity and the *ara h 2* genomic sequences from *A. hypogaea* cv Georgia Green. Putative start codon (ATG) at position 1 and stop codon (TAA) at position 517.

Supplemental Fig. 2 Alignment of *ara h 2* cDNA sequences belonging to the A- and B-genomes of *A. hypogaea* F78-1339 and previously published cDNA sequences. A putative start codon (ATG) is shown at position 1 and a stop codon (TAA) at position 517. Sequences without the 36 bp insertion belong to the A-genome and encode Ara h 2.01 isoform. Sequences that belong to the B-genome show the 36 bp insertion (position 227-262) and encode the Ara h 2.02 isoform.

Supplemental Fig. 3 Alignment of deduced amino acid sequences from cDNAs of *ara h 2* belonging to A-genome, B-genome, and unassigned published sequences. Identity among all sequences is noted by black boxes, arrowhead indicates non-conservative changes and asterisks show conservative amino acid changes.

Supplemental Fig. 4 Nucleotide alignment of the *ara h 6* cDNA sequence published by Kleber-Janke et al. (1999) and the cDNA that encodes for the Ara h 2.02 isoform from Chatel et al. (2003). The ORF of these genes share a 63% of sequence identity and the homology is especially in the middle and 3' end region.

Supplemental Fig. 5 Southern blot analysis with the lambda 8A2a genomic clone. (A), (B) and (C) were hybridized with the 3'arah2 probe, (D) hybridized with internal arah6 probe, and (E) with the 3'arah6 probe. The enzymes used in each digestion are displayed above each lane. Molecular weight sizes (kb) are displayed on the left.

Using an *ara h 2*-specific cDNA fragment (3'arah2) as a probe, one hybridizing band of almost 4 kb was observed in the *SacI* digest, one hybridizing band of 500 bp in the *SacI-PstI* digest (this band was slightly higher than the 300 bp expected), and one band approximately 10 kb in the *NcoI* digest (A). The *NcoI-SacI* double-digest looked the same as the *SacI* digest which indicated that there was no *NcoI* site within the *SacI* band recognized by 3'arah2. In the *DraI* digestion, one hybridizing band of 1.3 kb was detected (B). This band was the same size as the larger of the two hybridizing *DraI* fragments shown in the genomic Southern blot (Figure 1A), which corresponds to the B-genome. Two hybridizing bands of ~10 kb and 0.5 kb were observed in the *HindIII* digestion (B) which was consistent with the presence of a *HindIII* site within the probe sequence. *EcoRI* digestion showed only one band of approximately 5.5 kb that agreed with the expected size (C). The 9 kb band released by *AvaI* (C) was in accordance with only one *AvaI* site present in the genomic fragment, indicating that the other *AvaI* site is present in the lambda vector. Southern blot analysis of lambda 8A2a also was carried out with a probe that would specifically detect the second ORF contained in this clone, *ara h 6* (D). *SacI* digestion showed one hybridizing band of around 4 kb (D). The same band was identified in the *SacI-PstI* double digest, indicating that a *PstI* site was not present within the *SacI* fragment that hybridized with the *ara h 6* ORF-specific probe. In the *NcoI* digestion one band of around 6 kb was detected and in the *NcoI-SacI* digestion one

hybridizing band slightly smaller than 500 bp. This size was a little larger than the expected 363 bp size predicted from the sequence.

Results using a third probe designed to recognize the 3' region of the *ara h 6* gene are shown in E. In the *Dra*I digestion, one strongly hybridizing band of almost 4 kb was identified (E). The 3'arah6 probe also weakly recognized a band similar in size to the one labeled with the 3'arah2 probe, suggesting a weak cross-hybridization of the 3' regions of *ara h 2* and *ara h 6* genes (compare *Dra*I lanes in B and E). One band was detected in the *Hind*III digestion and two bands in the *Nco*I, but the upper, weaker band probably resulted from cross-hybridization with *ara h 2* sequence. The hybridization patterns revealed by Southern blot analysis confirmed the predicted restriction sites in this genomic clone plus the presence of *ara h 2* and the related conglutin *ara h 6* on this genomic fragment.

Supplemental Fig. 6 ORF and 3'UTR alignment from two *ara h 6* sequences belonging to the B-genome. A putative start codon (ATG) is shown at position 1 and a stop codon (TAG) at position 436. The TTATA indel at position 485-489 is present only in the *ara h 6* sequence belonging to lambda 8A2a (B-genome type I), and the CTGA indel at position 557-560 is only in lambda 17A2a (B-genome type II).

Supplemental Fig. 7 ORF and 3'UTR alignment from *ara h 6* sequences coming from *A. duranensis*, *A. ipaensis* and the B-genome of *A. hypogaea* F78-1339. The *ara h 6* sequence that belongs to *A. duranensis* shows no TTATA and CTGA indels in the 3'UTR.

Supplemental Fig. 8 Amino acid alignment from the deduced sequences of *ara h 6*. Identity between all sequences is noted by yellow shading, identity between three or four of them by light blue, letters in green shading indicate conservative amino acid changes, and white shading non-conservative changes. The first 49 amino acid sequences starting at Met₁ were used in the prediction of a putative signal peptide cleavage site. The amino acid changes between these sequences did not affect the cysteine residues located at positions 35, 47, 79, 80, 92, 94, 105, 128, 136 and 145.

Supplemental Fig. 9 Signal peptide prediction of *ara h 6* and *ara h 2* sequences. (A and C) SignalP 3.0 using the neural networks (NN) model showed the most likely cleavage site between positions 21 and 22 for *ara h 6* (ASA-MR) and *ara h 2* (ASA-RQ). (B and D) SignalP 3.0 using the hidden Markov model pointed out the maximum cleavage site probabilities 0.756 and 0.560 between positions 21 and 22 for *ara h 6* and *ara h 2* respectively.

Supplemental Fig. 10 Southern blot analysis showing *ara h 2* probe specificity after separation of RT-PCR products using 5% SDS-PAGE. Two *ara h 2* probes were checked for specificity against *ara h 2* (2) and *ara h 6* (6) RT-PCR products from *A. hypogaea* cv Georgia Green (GG), *A. ipaensis* (AI), and *A. duranensis* (AD). The arah2 full probe (generated with primers 185/371) recognizes the ORF and part of the 3'UTR of *ara h 2*, and the 3'arah2 probe (generated with primers 370/371) encompasses the 3' end of the ORF and the same part of the 3'UTR as the arah2 full probe. After hybridization,

the membranes were washed at 65°C under “normal” stringency (15 min in 1XSSC twice and once in 0.5X SSC) followed by “high” stringency (20 min in 0.1X SSC; stringent wash). Since there was no signal from *ara h 6* after probing with 3'arah2 probe and washing at high stringency, 3'arah2 probe was chosen to perform northern blot assays.

Supplemental Fig. 11 Somatic embryos of *A. hypogaea* cv Georgia Green. **A** Stage 1: less differentiated embryos. Only protruding embryos with a smooth surface were collected for RNA extraction. **B** Stage 2: clearly differentiated embryos. Only protruding embryos with defined structure were collected for RNA extraction.

Supplemental Fig. 12 Nucleotide sequence alignment between our B-genome *ara h 2* sequence, the ‘first genomic clone’ for *ara h 2* published by Vizquez et al. (2001) and the *Lac Z* gene (Accession number V00296). The start codon (ATG) is at position 127. A TAA stop codon is present in our sequence at position 644 whereas a TGA stop codon was reported for AX148738 at position 789. A red asterisk point out the divergence in sequence identity between the *ara h 2* sequence reported by Vizquez et al. (2001) and our *ara h 2* sequence. A green arrow head shows the start point where the *Lac Z* gene matches with *ara h 2* sequence from Vizquez et al. (2001)

Supplemental Fig. 13 Nucleotide sequence alignment between the 3' end region of *Lac Z* gene (Accession number V00296), the ‘first genomic’ *ara h 2* sequence reported by Vizquez et al (2001) and our B-genome *ara h 2* sequence. 56 bp from the 3' region of *Lac Z* gene, starting at position 46-106, matched with the putative promoter region reported by Vizquez et al. (2001). Putative start codon at position 127.

Supplemental Fig. 14 Nucleotide sequence alignment between the ‘first genomic’ *ara h 2* sequence reported by Vizquez et al (2001) and the 3' end region of *Lac Z* gene (Accession number V00296). Putative start codon at position 111.

Supplemental Fig. 1

		Section 1									
		(1)	1	10	20	30	40	50	52		
L77197 (Stanley, J.S.)	(1)	-----	G	CTCACCATACTAGTAGCCCT	CGCCCTTTTTCCTCCTCGCTGCC						
Arah2 A.duranensis	(1)	ATGGCCAAG	CTCACCATACTAGTAGCCCT	CGCCCTTTTTCCTCCTCGCTGCC							
arah2.01 GG	(1)	ATGGCCAAG	CTCACCATACTAGTAGCCCT	CGCCCTTTTTCCTCCTCGCTGCC							
AY722689 (CON G2-Arah2)	(1)	ATGGCCAAG	CTCACCATACTAGTAGCCCT	CGCCCTTTTTCCTCCTCGCTGCC							
AY117434 (Becker et al)	(1)	-----	CTCACCATACTAGTAGCCCT	CGCCCTTTTTCCTCCTCGCTGCC							
AY158467 (Chatel et al.)	(1)	ATGGCCAAG	CTCACCATACTAGTAGCCCT	CGCCCTTTTTCCTCCTCGCTGCC							
Arah2 A. ipaensis	(1)	ATGGCCAAG	CTCACCATACTAGTAGCCCT	CGCCCTTTTTCCTCCTCGCTGCC							
arah2.02 GG	(1)	ATGGCCAAG	CTCACCATACTAGTAGCCCT	CGCCCTTTTTCCTCCTCGCTGCC							
AY581853(CON G1-Arah2.02)	(1)	ATGGCCAAG	CTCACCATACTAGTAGCCCT	CGCCCTTTTTCCTCCTCGCTGCC							
Consensus	(1)	ATGGCCAAG	CTCACCATACTAGTAGCCCT	CGCCCTTTTTCCTCCTCGCTGCC							
		Section 2									
		(53)	53	60	70	80	90	100	104		
L77197 (Stanley, J.S.)	(46)	ACGCATCTGCGAGGCGAGCAGTGGGAACT	CAAGGAGACAGAAGATGCCAGAG								
Arah2 A.duranensis	(53)	ACGCATCTGCGAGGCGAGCAGTGGGAACT	CAAGGAGACAGAAGATGCCAGAG								
arah2.01 GG	(53)	ACGCATCTGCGAGGCGAGCAGTGGGAACT	CAAGGAGACAGAAGATGCCAGAG								
AY722689 (CON G2-Arah2)	(53)	ACGCATCTGCGAGGCGAGCAGTGGGAACT	CAAGGAGACAGAAGATGCCAGAG								
AY117434 (Becker et al)	(44)	ACGCATCTGCGAGGCGAGCAGTGGGAACT	CAAGGAGACAGAAGATGCCAGAG								
AY158467 (Chatel et al.)	(53)	ACGCATCTGCGAGGCGAGCAGTGGGAACT	CAAGGAGACAGAAGATGCCAGAG								
Arah2 A. ipaensis	(53)	ACGCATCTGCGAGGCGAGCAGTGGGAACT	CAAGGAGACAGAAGATGCCAGAG								
arah2.02 GG	(53)	ACGCATCTGCGAGGCGAGCAGTGGGAACT	CAAGGAGACAGAAGATGCCAGAG								
AY581853(CON G1-Arah2.02)	(53)	ACGCATCTGCGAGGCGAGCAGTGGGAACT	CAAGGAGACAGAAGATGCCAGAG								
Consensus	(53)	ACGCATCTGCGAGGCGAGCAGTGGGAACT	CAAGGAGACAGAAGATGCCAGAG								
		Section 3									
		(105)	105	110	120	130	140	150	156		
L77197 (Stanley, J.S.)	(97)	CCAGCTCGAGAGGGCGAACCT	GAGGCCCTGCGAGCAACATCTCATGCAGAAAG								
Arah2 A.duranensis	(105)	CCAGCTCGAGAGGGCGAACCT	GAGGCCCTGCGAGCAACATCTCATGCAGAAAG								
arah2.01 GG	(105)	CCAGCTCGAGAGGGCGAACCT	GAGGCCCTGCGAGCAACATCTCATGCAGAAAG								
AY722689 (CON G2-Arah2)	(105)	CCAGCTCGAGAGGGCGAACCT	GAGGCCCTGCGAGCAACATCTCATGCAGAAAG								
AY117434 (Becker et al)	(96)	CCAGCTCGAGAGGGCGAACCT	TAGGCCCTGCGAGCAACATCTCATGCAGAAA								
AY158467 (Chatel et al.)	(105)	CCAGCTCGAGAGGGCGAACCT	GAGGCCCTGCGAGCAACATCTCATGCAGAAA								
Arah2 A. ipaensis	(105)	CCAGCTCGAGAGGGCGAACCT	TAGGCCCTGCGAGCAACATCTCATGCAGAAA								
arah2.02 GG	(105)	CCAGCTCGAGAGGGCGAACCT	TAGGCCCTGCGAGCAACATCTCATGCAGAAA								
AY581853(CON G1-Arah2.02)	(105)	CCAGCTCGAGAGGGCGAACCT	TAGGCCCTGCGAGCAACATCTCATGCAGAAA								
Consensus	(105)	CCAGCTCGAGAGGGCGAACCT	TAGGCCCTGCGAGCAACATCTCATGCAGAAA								
		Section 4									
		(157)	157	170	180	190	200				
L77197 (Stanley, J.S.)	(149)	ATCCAACGTGACGAGGATTTCATATG	ACGGGACCCGTACAGCCCTAGTCAGG								
Arah2 A.duranensis	(157)	ATCCAACGTGACGAGGATTTCATATG	ACGGGACCCGTACAGCCCTAGTCAGG								
arah2.01 GG	(157)	ATCCAACGTGACGAGGATTTCATATG	ACGGGACCCGTACAGCCCTAGTCAGG								
AY722689 (CON G2-Arah2)	(157)	ATCCAACGTGACGAGGATTTCATATG	ACGGGACCCGTACAGCCCTAGTCAGG								
AY117434 (Becker et al)	(148)	ATCCAACGTGACGAGGATTTCATATG	ACGGGACCCGTACAGCCCTAGTCAGG								
AY158467 (Chatel et al.)	(157)	ATCCAACGTGACGAGGATTTCATATG	ACGGGACCCGTACAGCCCTAGTCAGG								
Arah2 A. ipaensis	(157)	ATCCAACGTGACGAGGATTTCATATG	ACGGGACCCGTACAGCCCTAGTCAGG								
arah2.02 GG	(157)	ATCCAACGTGACGAGGATTTCATATG	ACGGGACCCGTACAGCCCTAGTCAGG								
AY581853(CON G1-Arah2.02)	(157)	ATCCAACGTGACGAGGATTTCATATG	ACGGGACCCGTACAGCCCTAGTCAGG								
Consensus	(157)	ATCCAACGTGACGAGGATTTCATATG	ACGGGACCCGTACAGCCCTAGTCAGG								
		Section 5									
		(209)	209	220	230	240	250	260			
L77197 (Stanley, J.S.)	(201)	ATCC	-----	-----	-----	-----	GTACAGCCCTAG				
Arah2 A.duranensis	(209)	ATCC	-----	-----	-----	-----	GTACAGCCCTAG				
arah2.01 GG	(209)	ATCC	-----	-----	-----	-----	GTACAGCCCTAG				
AY722689 (CON G2-Arah2)	(209)	ATCC	-----	-----	-----	-----	GTACAGCCCTAG				
AY117434 (Becker et al)	(200)	ATCC	GTACAGCCCTAGTCAGGACCCGGACAGACG	GATCC	GTACAGCCCTAG						
AY158467 (Chatel et al.)	(209)	ATCC	GTACAGCCCTAGTCAGGACCCGGACAGACG	GATCC	GTACAGCCCTAG						
Arah2 A. ipaensis	(209)	ATCC	GTACAGCCCTAGTCAGGACCCGGACAGACG	GATCC	GTACAGCCCTAG						
arah2.02 GG	(209)	ATCC	GTACAGCCCTAGTCAGGACCCGGACAGACG	GATCC	GTACAGCCCTAG						
AY581853(CON G1-Arah2.02)	(209)	ATCC	GTACAGCCCTAGTCAGGACCCGGACAGACG	GATCC	GTACAGCCCTAG						
Consensus	(209)	ATCC	GTACAGCCCTAGTCAGGACCCGGACAGACG	GATCC	GTACAGCCCTAG						
		Section 6									
		(261)	261	270	280	290	300	312			
L77197 (Stanley, J.S.)	(217)	TCCATATGATCGGAGAGG	CGCTGGATC	TCTCAGCACCA	AGAGAGGTGTTGC						
Arah2 A.duranensis	(225)	TCCATATGATCGGAGAGG	CGCTGGATC	TCTCAGCACCA	AGAGAGGTGTTGC						
arah2.01 GG	(225)	TCCATATGATCGGAGAGG	CGCTGGATC	TCTCAGCACCA	AGAGAGGTGTTGC						
AY722689 (CON G2-Arah2)	(225)	TCCATATGATCGGAGAGG	CGCTGGATC	TCTCAGCACCA	AGAGAGGTGTTGC						
AY117434 (Becker et al)	(252)	TCCATATGATCGGAGAGG	CGCTGGATC	TCTCAGCACCA	AGAGAGGTGTTGC						
AY158467 (Chatel et al.)	(261)	TCCATATGATCGGAGAGG	CGCTGGATC	TCTCAGCACCA	AGAGAGGTGTTGC						
Arah2 A. ipaensis	(261)	TCCATATGATCGGAGAGG	CGCTGGATC	TCTCAGCACCA	AGAGAGGTGTTGC						
arah2.02 GG	(261)	TCCATATGATCGGAGAGG	CGCTGGATC	TCTCAGCACCA	AGAGAGGTGTTGC						
AY581853(CON G1-Arah2.02)	(261)	TCCATATGATCGGAGAGG	CGCTGGATC	TCTCAGCACCA	AGAGAGGTGTTGC						
Consensus	(261)	TCCATATGATCGGAGAGG	CGCTGGATC	TCTCAGCACCA	AGAGAGGTGTTGC						

Section 7							
	(313)	313	320	330	340	350	364
L77197 (Stanley, J.S.) (269)		AATGAGCTGAACGAGTTT	GAGAACAACCAAAGGTGCATGTGCGAGGCATTGC				
Arah2 A.duranensis (277)		AATGAGCTGAACGAGTTT	GAGAACAACCAAAGGTGCATGTGCGAGGCATTGC				
arah2.01 G-G (277)		AATGAGCTGAACGAGTTT	GAGAACAACCAAAGGTGCATGTGCGAGGCATTGC				
AY722689 (CON G2-Arah2) (277)		AATGAGCTGAACGAGTTT	GAGAACAACCAAAGGTGCATGTGCGAGGCATTGC				
AY117434 (Becker et al) (304)		AATGAGCTGAACGAGTTT	GAGAACAACCAAAGGTGCATGTGCGAGGCATTGC				
AY158467 (Chatel et al.) (313)		AATGAGCTGAACGAGTTT	GAGAACAACCAAAGGTGCATGTGCGAGGCATTGC				
Arah2 A. ipaensis (313)		AATGAGCTGAACGAGTTT	GAGAACAACCAAAGGTGCATGTGCGAGGCATTGC				
arah2.02 G-G (313)		AATGAGCTGAACGAGTTT	GAGAACAACCAAAGGTGCATGTGCGAGGCATTGC				
AY581853(CON G1-Arah2.02) (313)		AATGAGCTGAACGAGTTT	GAGAACAACCAAAGGTGCATGTGCGAGGCATTGC				
Consensus (313)		AATGAGCTGAACGAGTTT	GAGAACAACCAAAGGTGCATGTGCGAGGCATTGC				
Section 8							
	(365)	365	370	380	390	400	416
L77197 (Stanley, J.S.) (321)		AACAGATC	CATGGAGAACCAGAGCGATAGGTTG	CAGGGGAGGCAACAGGAGCA			
Arah2 A.duranensis (329)		AACAGATC	CATGGAGAACCAGAGCGATAGGTTG	CAGGGGAGGCAACAGGAGCA			
arah2.01 G-G (329)		AACAGATC	CATGGAGAACCAGAGCGATAGGTTG	CAGGGGAGGCAACAGGAGCA			
AY722689 (CON G2-Arah2) (329)		AACAGATC	CATGGAGAACCAGAGCGATAGGTTG	CAGGGGAGGCAACAGGAGCA			
AY117434 (Becker et al) (356)		AACAGATC	CATGGAGAACCAGAGCGATAGGTTG	CAGGGGAGGCAACAGGAGCA			
AY158467 (Chatel et al.) (365)		AACAGATC	CATGGAGAACCAGAGCGATAGGTTG	CAGGGGAGGCAACAGGAGCA			
Arah2 A. ipaensis (365)		AACAGATC	CATGGAGAACCAGAGCGATAGGTTG	CAGGGGAGGCAACAGGAGCA			
arah2.02 G-G (365)		AACAGATC	CATGGAGAACCAGAGCGATAGGTTG	CAGGGGAGGCAACAGGAGCA			
AY581853(CON G1-Arah2.02) (365)		AACAGATC	CATGGAGAACCAGAGCGATAGGTTG	CAGGGGAGGCAACAGGAGCA			
Consensus (365)		AACAGATC	CATGGAGAACCAGAGCGATAGGTTG	CAGGGGAGGCAACAGGAGCA			
Section 9							
	(417)	417	430	440	450	468	
L77197 (Stanley, J.S.) (373)		ACAGTTCAAGAGGGAGCTCAGGAACTTGCCTCAACAGTGC	GGCCTT	AGGGCA			
Arah2 A.duranensis (381)		ACAGTTCAAGAGGGAGCTCAGGAACTTGCCTCAACAGTGC	GGCCTT	AGGGCA			
arah2.01 G-G (381)		ACAGTTCAAGAGGGAGCTCAGGAACTTGCCTCAACAGTGC	GGCCTT	AGGGCA			
AY722689 (CON G2-Arah2) (381)		ACAGTTCAAGAGGGAGCTCAGGAACTTGCCTCAACAGTGC	GGCCTT	AGGGCA			
AY117434 (Becker et al) (408)		ACAGTTCAAGAGGGAGCTCAGGAACTTGCCTCAACAGTGC	GGCCTT	AGGGCA			
AY158467 (Chatel et al.) (417)		ACAGTTCAAGAGGGAGCTCAGGAACTTGCCTCAACAGTGC	GGCCTT	AGGGCA			
Arah2 A. ipaensis (417)		ACAGTTCAAGAGGGAGCTCAGGAACTTGCCTCAACAGTGC	GGCCTT	AGGGCA			
arah2.02 G-G (417)		ACAGTTCAAGAGGGAGCTCAGGAACTTGCCTCAACAGTGC	GGCCTT	AGGGCA			
AY581853(CON G1-Arah2.02) (417)		ACAGTTCAAGAGGGAGCTCAGGAACTTGCCTCAACAGTGC	GGCCTT	AGGGCA			
Consensus (417)		ACAGTTCAAGAGGGAGCTCAGGAACTTGCCTCAACAGTGC	GGCCTT	AGGGCA			
Section 10							
	(469)	469	480	490	500	510	520
L77197 (Stanley, J.S.) (425)		CCACAGCGTTGCGACTTGGAC	GTGCGAAAGTGGCGGCAGAGACAGATACTAA				
Arah2 A.duranensis (433)		CCACAGCGTTGCGACTTGGAC	GTGCGAAAGTGGCGGCAGAGACAGATACTAA				
arah2.01 G-G (433)		CCACAGCGTTGCGACTTGGAC	GTGCGAAAGTGGCGGCAGAGACAGATACTAA				
AY722689 (CON G2-Arah2) (433)		CCACAGCGTTGCGACTTGGAC	GTGCGAAAGTGGCGGCAGAGACAGATACTAA				
AY117434 (Becker et al) (460)		CCACAGCGTTGCGACTTGGAC	GTGCGAAAGTGGCGGCAGAGACAGATACTAA				
AY158467 (Chatel et al.) (469)		CCACAGCGTTGCGACTTGGAC	GTGCGAAAGTGGCGGCAGAGACAGATACTAA				
Arah2 A. ipaensis (469)		CCACAGCGTTGCGACTTGGAC	GTGCGAAAGTGGCGGCAGAGACAGATACTAA				
arah2.02 G-G (469)		CCACAGCGTTGCGACTTGGAC	GTGCGAAAGTGGCGGCAGAGACAGATACTAA				
AY581853(CON G1-Arah2.02) (469)		CCACAGCGTTGCGACTTGGAC	GTGCGAAAGTGGCGGCAGAGACAGATACTAA				
Consensus (469)		CCACAGCGTTGCGACTTGGAC	GTGCGAAAGTGGCGGCAGAGACAGATACTAA				
Section 11							
	(521)	521	530	540	550	560	572
L77197 (Stanley, J.S.) (477)		CACCTATCTCAAAAAAGAAAAAGAAAAAGAAAAATAGCTTATATATAAG					
Arah2 A.duranensis (485)		CACCTATCTCAAAAAAGAAAAAGAAAAAGAAAAATAGCTTATATATAAG					
arah2.01 G-G (485)		CACCTATCTCAAAAAAGAAAAAGAAAAAGAAAAATAGCTTATATATAAG					
AY722689 (CON G2-Arah2) (485)		CACCTATCTCAAAAAAGAAAAAGAAAAAGAAAAATAGCTTATATATAAG					
AY117434 (Becker et al) (511)		-----					
AY158467 (Chatel et al.) (520)		-----					
Arah2 A. ipaensis (521)		CACCTATCTCAAAAAAGAAAAAGAAAAAGAAAAATAGCTTATATATAAG					
arah2.02 G-G (521)		CACCTATCTCAAAAAAGAAAAAGAAAAAGAAAAATAGCTTATATATAAG					
AY581853(CON G1-Arah2.02) (521)		CACCTATCTCAAAAAAGAAAAAGAAAAAGAAAAATAGCTTATATATAAG					
Consensus (521)		CACCTATCTCAAAAAAGAAAAAGAAAAAGAAAAATAGCTTATATATAAG					
Section 12							
	(573)	573	580	590	600	610	624
L77197 (Stanley, J.S.) (529)		CTATTATCTATGGTTATGTT	AGTTTTGGTAATAATAAAGATCATCACTATA				
Arah2 A.duranensis (537)		CTATTATCTATGGTTATGTT	AGTTTTGGTAATAATAAAGATCATCACTATA				
arah2.01 G-G (537)		CTATTATCTATGGTTATGTT	AGTTTTGGTAATAATAAAGATCATCACTATA				
AY722689 (CON G2-Arah2) (537)		CTATTATCTATGGTTATGTT	AGTTTTGGTAATAATAAAGATCATCACTATA				
AY117434 (Becker et al) (511)		-----					
AY158467 (Chatel et al.) (520)		-----					
Arah2 A. ipaensis (573)		CTATTATCTATGGTTATGTT	AGTTTTGGTAATAATAAAGATCATCACTATA				
arah2.02 G-G (573)		CTATTATCTATGGTTATGTT	AGTTTTGGTAATAATAAAGATCATCACTATA				
AY581853(CON G1-Arah2.02) (573)		CTATTATCTATGGTTATGTT	AGTTTTGGTAATAATAAAGATCATCACTATA				
Consensus (573)		CTATTATCTATGGTTATGTT	AGTTTTGGTAATAATAAAGATCATCACTATA				

		Section 13					
		(625)	625	630	640	650	660 676
L77197 (Stanley, J.S.) (581)			TGAATGTGTTGATCGTGTTAACTAAGGCAAGCTTAGGTTATATGAGCACCTT				
Arah2 A.duranensis (588)			TGAATGTGTTGATCGTGTTAACTAAGGCAAGCTTAGGTTATATGAGCACCTT				
arah2.01 G-G (588)			TGAATGTGTTGATCGTGTTAACTAAGGCAAGCTTAGGTTATATGAGCACCTT				
AY722689 (CON G2-Arah2) (588)			TGAATGTGTTGATCGTGTTAACTAAGGCAAGCTTAGGTTATATGAGCACCTT				
AY117434 (Becker et al) (511)			-----				
AY158467 (Chatel et al.) (520)			-----				
Arah2 A. ipaensis (624)			TGAATGTGTTGATCGTGTTAACTAAGGCAAGCTTAGGTTATATGAGCACCTT				
arah2.02 G-G (624)			TGAATGTGTTGATCGTGTTAACTAAGGCAAGCTTAGGTTATATGAGCACCTT				
AY581853(CON G1-Arah2.02) (624)			TGAATGTGTTGATCGTGTTAACTAAGGCAAGCTTAGGTTATATGAGCACCTT				
Consensus (625)			TGAATGTGTTGATCGTGTTAACTAAGGCAAGCTTAGGTTATATGAGCACCTT				
		Section 14					
		(677)	677	690	700	710	728
L77197 (Stanley, J.S.) (633)			TAGAGTGCTTTTATGGCGTTGCTCTATGTTTTGTTGCTG			CAGAGTTGTAACCA	
Arah2 A.duranensis (640)			TAGAGTGCTTTTATGGCGTTGCTCTATGTTTTGTTGCTG			-----	
arah2.01 G-G (640)			TAGAGTGCTTTTATGGCGTTGCTCTATGTTTTGTTGCTGA			-----	
AY722689 (CON G2-Arah2) (640)			TAGAGTGCTTTTATGGCGTTGCTCTATGTTTTGTTGCTG			CAGAGTTGTAACCA	
AY117434 (Becker et al) (511)			-----				
AY158467 (Chatel et al.) (520)			-----				
Arah2 A. ipaensis (676)			TAGAGTGCTTTTATGGCGTTGCTCTATGTTTTGTTGCTG			-----	
arah2.02 G-G (676)			TAGAGTGCTTTTATGGCGTTGCTCTATGTTTTGTTGCTGA			-----	
AY581853(CON G1-Arah2.02) (676)			TAGAGTGCTTTTATGGCGTTGCTCTATGTTTTGTTGCTG			CAGAGTTGTAACCA	
Consensus (677)			TAGAGTGCTTTTATGGCGTTGCTCTATGTTTTGTTGCTG				
		Section 15					
		(729)	729	740	750	760	770 780
L77197 (Stanley, J.S.) (685)			TCTTGAAATAATATAAAAAAGATCATGTTTTGTT			-----	
Arah2 A.duranensis (678)			-----				
arah2.01 G-G (679)			-----				
AY722689 (CON G2-Arah2) (692)			TCTTGAAATAATATAAAAAAGATCATGTTTTGTTCAAAAAAAAAAAAAAAAAA				
AY117434 (Becker et al) (511)			-----				
AY158467 (Chatel et al.) (520)			-----				
Arah2 A. ipaensis (714)			-----				
arah2.02 G-G (715)			-----				
AY581853(CON G1-Arah2.02) (728)			TCTTGAAATAATATAAAAAAGATCATGTTTTGTTTGGCGACGAAAAAAAAA				
Consensus (729)							

Supplemental Fig. 2

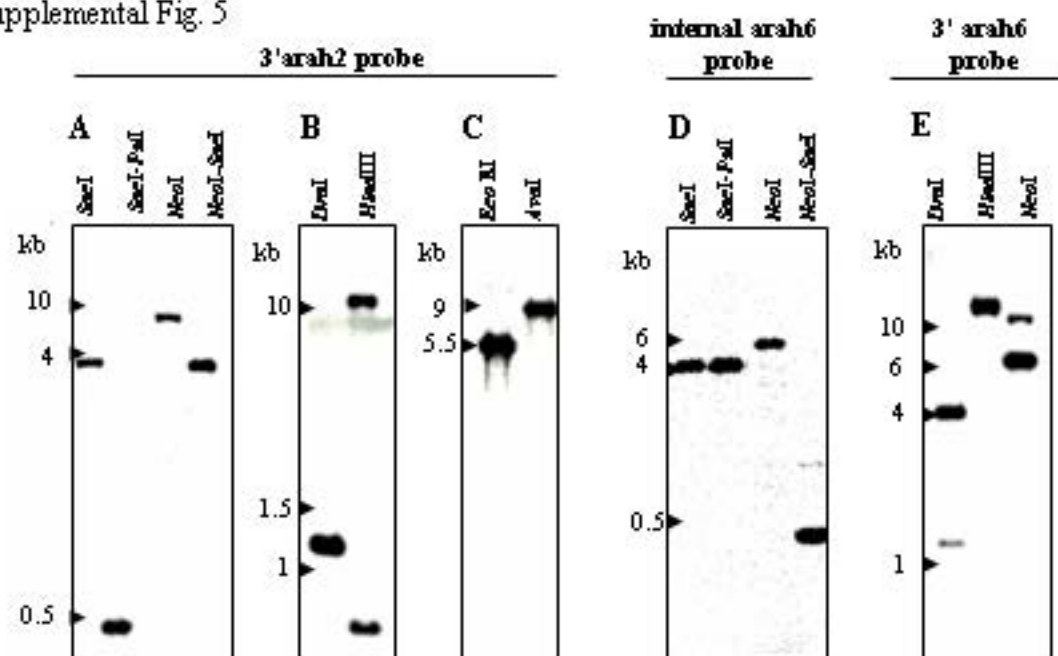
		Section 9					
	(417)	417	430	440	450	468	
L77197 (Stanley, J.S.) (373)		ACAGTTCAAGAGGGAGCTCAGGAACTTGCCTCAACAGTGC	GGCCTT	AGGGCA			
AY722689 (CON G2-Arah2) (381)		ACAGTTCAAGAGGGAGCTCAGGAACTTGCCTCAACAGTGC	GGCCTT	AGGGCA			
Ara h2 cDNA (A-genome) (381)		ACAGTTCAAGAGGGAGCTCAGGAACTTGCCTCAACAGTGC	GGCCTT	AGGGCA			
AY158467 (Chatel et al.) (417)		ACAGTTCAAGAGGGAGCTCAGGAACTTGCCTCAACAGTGC	GGCCTT	AGGGCA			
Ara h2 cDNA (B-genome) (417)		ACAGTTCAAGAGGGAGCTCAGGAACTTGCCTCAACAGTGC	GGCCTT	AGGGCA			
AY581853 (CON G1-Arah2.02) (417)		ACAGTTCAAGAGGGAGCTCAGGAACTTGCCTCAACAGTGC	GGCCTT	AGGGCA			
AY117434 (Becker et al.) (408)		ACAGTTCAAGAGGGAGCTCAGGAACTTGCCTCAACAGTGC	GGCCTT	AGGGCA			
C Consensus (417)		ACAGTTCAAGAGGGAGCTCAGGAACTTGCCTCAACAGTGC	GGCCTT	AGGGCA			
		Section 10					
	(469)	469	480	490	500	510	520
L77197 (Stanley, J.S.) (425)		CCACAGCGTTGCGACTTTGGA	C	GTCTGAAAGTGGCGGCAGAGACAGATACTAAA			
AY722689 (CON G2-Arah2) (433)		CCACAGCGTTGCGACTTTGGA	C	GTCTGAAAGTGGCGGCAGAGACAGATACTAAA			
Ara h2 cDNA (A-genome) (433)		CCACAGCGTTGCGACTTTGGA	C	GTCTGAAAGTGGCGGCAGAGACAGATACTAAA			
AY158467 (Chatel et al.) (469)		CCACAGCGTTGCGACTTTGGA	A	GTCTGAAAGTGGCGGCAGAGACAGATACTAAA			
Ara h2 cDNA (B-genome) (469)		CCACAGCGTTGCGACTTTGGA	A	GTCTGAAAGTGGCGGCAGAGACAGATACTAAA			
AY581853 (CON G1-Arah2.02) (469)		CCACAGCGTTGCGACTTTGGA	A	GTCTGAAAGTGGCGGCAGAGACAGATACTAAA			
AY117434 (Becker et al.) (460)		CCACAGCGTTGCGACTTTGGA	A	GTCTGAAAGTGGCGGCAGAGACAGATACTAAA			
C Consensus (469)		CCACAGCGTTGCGACTTTGGA	A	GTCTGAAAGTGGCGGCAGAGACAGATACTAAA			
		Section 11					
	(521)	521	530	540	550	560	572
L77197 (Stanley, J.S.) (477)		CACCTATCTCAAAAAAGAAAAAGAAAAAGAAAAAGAAATAGCTTATATATAAG					
AY722689 (CON G2-Arah2) (485)		CACCTATCTCAAAAAAGAAAAAGAAAAAGAAAAAGAAATAGCTTATATATAAG					
Ara h2 cDNA (A-genome) (485)		CACCTATCTCAAAAAAGAAAAAGAAAAAGAAAAAGAAATAGCTTATATATAAG					
AY158467 (Chatel et al.) (520)		-----					
Ara h2 cDNA (B-genome) (521)		CACCTATCTCAAAAAAGAAAAAGAAAAAGAAAAAGAAATAGCTTATATATAAG					
AY581853 (CON G1-Arah2.02) (521)		CACCTATCTCAAAAAAGAAAAAGAAAAAGAAAAAGAAATAGCTTATATATAAG					
AY117434 (Becker et al.) (511)		-----					
C Consensus (521)		CACCTATCTCAAAAAAGAAAAAGAAAAAGAAAAAGAAATAGCTTATATATAAG					
		Section 12					
	(573)	573	580	590	600	610	624
L77197 (Stanley, J.S.) (529)		CTATTATCTATGGT	TATGTT	AGTTTTGGTAATAATAAAGATCATCACTATA			
AY722689 (CON G2-Arah2) (537)		CTATTATCTATGGT	TATGTT	AGTTTTGGTAATAATAAAGATCATCACTATA			
Ara h2 cDNA (A-genome) (502)		-----					
AY158467 (Chatel et al.) (520)		-----					
Ara h2 cDNA (B-genome) (573)		CTATTATCTATGGT	AATGTT	AGTTTTGGTAATAATAAAGATCATCACTATA			
AY581853 (CON G1-Arah2.02) (573)		CTATTATCTATGGT	AATGTT	AGTTTTGGTAATAATAAAGATCATCACTATA			
AY117434 (Becker et al.) (511)		-----					
C Consensus (573)		CTATTATCTATGGT	ATGTT	AGTTTTGGTAATAATAAAGATCATCACTATA			
		Section 13					
	(625)	625	630	640	650	660	676
L77197 (Stanley, J.S.) (581)		TGAATGTGTTGATCGTGTTAACTAAGGCAAGCTTAGGTTATATGAGCACCTT					
AY722689 (CON G2-Arah2) (588)		TGAATGTGTTGATCGTGTTAACTAAGGCAAGCTTAGGTTATATGAGCACCTT					
Ara h2 cDNA (A-genome) (502)		-----					
AY158467 (Chatel et al.) (520)		-----					
Ara h2 cDNA (B-genome) (624)		TGAATGTGTTGATCGTGTTAACTAAGGCAAGCTTAGGTTATATGAGCACCTT					
AY581853 (CON G1-Arah2.02) (624)		TGAATGTGTTGATCGTGTTAACTAAGGCAAGCTTAGGTTATATGAGCACCTT					
AY117434 (Becker et al.) (511)		-----					
C Consensus (625)		TGAATGTGTTGATCGTGTTAACTAAGGCAAGCTTAGGTTATATGAGCACCTT					
		Section 14					
	(677)	677	690	700	710	728	
L77197 (Stanley, J.S.) (633)		TAGAGTGCTTTTATGGCGTTGTCTATGTTTTGTTGCTGCAGAGTTGTAACCA					
AY722689 (CON G2-Arah2) (640)		TAGAGTGCTTTTATGGCGTTGTCTATGTTTTGTTGCTGCAGAGTTGTAACCA					
Ara h2 cDNA (A-genome) (502)		-----					
AY158467 (Chatel et al.) (520)		-----					
Ara h2 cDNA (B-genome) (676)		TAGAGTGCTTTTATGGCGTTGTCTATGTTTTGTTGCTGCAGAGTTGTAACCA					
AY581853 (CON G1-Arah2.02) (676)		TAGAGTGCTTTTATGGCGTTGTCTATGTTTTGTTGCTGCAGAGTTGTAACCA					
AY117434 (Becker et al.) (511)		-----					
C Consensus (677)		TAGAGTGCTTTTATGGCGTTGTCTATGTTTTGTTGCTGCAGAGTTGTAACCA					
		Section 15					
	(729)	729	740	750	760	770	780
L77197 (Stanley, J.S.) (685)		TCTTGAAATAATATAAAAAAGATCATG	TTTTGTT	-----			
AY722689 (CON G2-Arah2) (692)		TCTTGAAATAATATAAAAAAGATCATG	TTTTGTT	CAAAAAAAAAAAAAAAAAAAAA			
Ara h2 cDNA (A-genome) (502)		-----					
AY158467 (Chatel et al.) (520)		-----					
Ara h2 cDNA (B-genome) (728)		TCTTGAAATAATATAAAAAAGATCATG	-----				
AY581853 (CON G1-Arah2.02) (728)		TCTTGAAATAATATAAAAAAGATCATG	TTTTGTTT	TGGCGACGAAAAAAAAAAAA			
AY117434 (Becker et al.) (511)		-----					
C Consensus (729)		TCTTGAAATAATATAAAAAAGATCATG					
		Section 16					
	(781)	781	798				
L77197 (Stanley, J.S.) (718)		-----					
AY722689 (CON G2-Arah2) (744)		AAAAAAAA	-----				
Ara h2 cDNA (A-genome) (502)		-----					
AY158467 (Chatel et al.) (520)		-----					
Ara h2 cDNA (B-genome) (754)		-----					
AY581853 (CON G1-Arah2.02) (780)		AAAAAAAAAAAAAAAAAAAA					
AY117434 (Becker et al.) (511)		-----					
C Consensus (781)		-----					

Supplemental Fig. 3

		Section 1				
		①	10	20	30	44
Tsl L77 197 (Starky, J.S.)	①	---	LTILVAL	LALFL	LAANA	JARQQWELQGD
TslcDNA Ara 12 (Ageiome)	①	MAXL	TILVAL	LALFL	LAANA	JARQQWELQGD
Translation of AY7 22689 (CONG2-Ara12)	①	MAXL	TILVAL	LALFL	LAANA	JARQQWELQGD
TslcDNA Ara 12 (B-geiome)	①	MAXL	TILVAL	LALFL	LAANA	JARQQWELQGD
TslcDNA AY117434 (Becker et al.)	①	---	LTILVAL	LALFL	LAANA	JARQQWELQGD
TslcDNA AY158467 (Catelet et al.)	①	MAXL	TILVAL	LALFL	LAANA	JARQQWELQGD
Translation of AY581853 (CONG1-Ara12D2)	①	MAXL	TILVAL	LALFL	LAANA	JARQQWELQGD
Consensus	①	MAXL	TILVAL	LALFL	LAANA	JARQQWELQGD
		Section 2				
		(45)	45	50	60	70
Tsl L77 197 (Starky, J.S.)	(42)	CEQHLNQKI	QRDED	SYGRDP	YSP	QDPYSF
TslcDNA Ara 12 (Ageiome)	(45)	CEQHLNQKI	QRDED	SYGRDP	YSP	QDPYSF
Translation of AY7 22689 (CONG2-Ara12)	(45)	CEQHLNQKI	QRDED	SYGRDP	YSP	QDPYSF
TslcDNA Ara 12 (B-geiome)	(45)	CEQHLNQKI	QRDED	SYGRDP	YSP	QDPYSF
TslcDNA AY117434 (Becker et al.)	(42)	CEQHLNQKI	QRDED	SYGRDP	YSP	QDPYSF
TslcDNA AY158467 (Catelet et al.)	(45)	CEQHLNQKI	QRDED	SYGRDP	YSP	QDPYSF
Translation of AY581853 (CONG1-Ara12D2)	(45)	CEQHLNQKI	QRDED	SYGRDP	YSP	QDPYSF
Consensus	(45)	CEQHLNQKI	QRDED	SYGRDP	YSP	QDPYSF
		Section 3				
		(89)	89	100	110	120
Tsl L77 197 (Starky, J.S.)	(74)	YDRRGAGSS	QHQCRC	CNELNE	FENN	QRCMC
TslcDNA Ara 12 (Ageiome)	(77)	YDRRGAGSS	QHQCRC	CNELNE	FENN	QRCMC
Translation of AY7 22689 (CONG2-Ara12)	(77)	YDRRGAGSS	QHQCRC	CNELNE	FENN	QRCMC
TslcDNA Ara 12 (B-geiome)	(89)	YDRRGAGSS	QHQCRC	CNELNE	FENN	QRCMC
TslcDNA AY117434 (Becker et al.)	(86)	YDRRGAGSS	QHQCRC	CNELNE	FENN	QRCMC
TslcDNA AY158467 (Catelet et al.)	(89)	YDRRGAGSS	QHQCRC	CNELNE	FENN	QRCMC
Translation of AY581853 (CONG1-Ara12D2)	(89)	YDRRGAGSS	QHQCRC	CNELNE	FENN	QRCMC
Consensus	(89)	YDRRGAGSS	QHQCRC	CNELNE	FENN	QRCMC
		Section 4				
		(133)	133	140	150	160
Tsl L77 197 (Starky, J.S.)	(118)	QGRQ	QEQQ	FKREL	RNL	PQQC
TslcDNA Ara 12 (Ageiome)	(121)	QGRQ	QEQQ	FKREL	RNL	PQQC
Translation of AY7 22689 (CONG2-Ara12)	(121)	QGRQ	QEQQ	FKREL	RNL	PQQC
TslcDNA Ara 12 (B-geiome)	(133)	QGRQ	QEQQ	FKREL	RNL	PQQC
TslcDNA AY117434 (Becker et al.)	(130)	QGRQ	QEQQ	FKREL	RNL	PQQC
TslcDNA AY158467 (Catelet et al.)	(133)	QGRQ	QEQQ	FKREL	RNL	PQQC
Translation of AY581853 (CONG1-Ara12D2)	(133)	QGRQ	QEQQ	FKREL	RNL	PQQC
Consensus	(133)	QGRQ	QEQQ	FKREL	RNL	PQQC

		Section 1									
AF092846 (Kleber-Janke et al.)	(1)	1	10	20	30	40	50	60	75		
AY158467 (Chatel et al.)	(1)	-----					GCACACGCCCTCCGC	AATGAGGCGCGAG			
Consensus	(1)	ATGGCCAAGCTCACCATACTAGTAGGCCCTGCGCCTTTTCTCTCTCGCT					GCCACACGCATCTGC	---GAGGCGAGCAG			
		Section 2									
AF092846 (Kleber-Janke et al.)	(76)	76	90	100	110	120	130	140	150		
AY158467 (Chatel et al.)	(78)	AGGGGGGAGA	CAGGGG	GACTCATCA	AGCTGCGAGAGGCAGGT	TGACGGGGT	TAACTTCA	AGCCCTGCGGAGCAGCAC			
Consensus	(76)	GGG	CA	GG	GAC	A	AG	TGC	AGAG	CAG	T
		Section 3									
AF092846 (Kleber-Janke et al.)	(151)	151	160	170	180	190	200	210	225		
AY158467 (Chatel et al.)	(103)	ATAATGCGAGAGGATC	-----			ATGGGCGAG	CAA-----	GAGCGACTAC	---		
Consensus	(151)	T	ATGCAGA	ATC		ATGG	CG	G	CA		
		Section 4									
AF092846 (Kleber-Janke et al.)	(226)	226	240	250	260	270	280	290	300		
AY158467 (Chatel et al.)	(220)	-----G-AC	TCTTACA	-----		ATTTTGGAGTA	---CTCGATCTC	CCGAC			
Consensus	(226)	CCTAGTCAG	GACCGGAC	AGACGTGATCCGTACAGCCCTAGTCCAT		ATGATC	GGAGAGGCG	CTGGATCTCTC	AG		
		Section 5									
AF092846 (Kleber-Janke et al.)	(301)	301	310	320	330	340	350	360	375		
AY158467 (Chatel et al.)	(295)	CAGCAACAGAGGTG	TGCG	ATGAGCTGAACGAG	ATG	GAGAAACACA	CAGAG	ATGCATGTGCGAGGCATTGCA	GC	AA	
Consensus	(301)	CA	CAA	AGAGGTG	TGC	ATGAGCTGAACGAG	T	GAGAA	CA	AG	TGCATGTGCGAGGCATTGCA
		Section 6									
AF092846 (Kleber-Janke et al.)	(376)	376	390	400	410	420	430	440	450		
AY158467 (Chatel et al.)	(370)	ATAATGGAGAACCACT	GCGAT	GGGTGCGAGGAC	AGGCCAAAT	GGTGCAGCACT	TTCAGAGAGAGGCTCAT	GAACCTTG			
Consensus	(376)	ATAATGGAGAACCA	CGCAT	AGGTGCGAGG	AGGCCAAC	AGCAACAG	TTCAGAGAGGAGCTCAG	GAACCTTG			
		Section 7									
AF092846 (Kleber-Janke et al.)	(461)	461	460	470	480	490	500	510	525		
AY158467 (Chatel et al.)	(445)	CCC	CAACAGTGTAACT	TTTGGG	CACCCACAGCGTTGCGAT	TTGGACGT	---GAGTGGCGGCAGAT	GTCTAGACTCA	AA		
Consensus	(461)	CC	CAACAGTG	C	T	GGGCACCCACAGCGTTGCGA	TTGGA	GT	AGTGGCGGCAG	C	AGA
		Section 8									
AF092846 (Kleber-Janke et al.)	(526)	526	540	550	560	570	580	590	600		
AY158467 (Chatel et al.)	(519)	AAATTATAATCTGTGCCAAAACAA	ACTAGTAGGAAGTAGCTGATATTATGAGCTATTATGTATGCTTGTTCGTT								
Consensus	(526)	A	-----								
		Section 9									
AF092846 (Kleber-Janke et al.)	(601)	601	610	620	630	640	650	660	675		
AY158467 (Chatel et al.)	(520)	GATAATAAATATCATCACTGTATGAATGTGGTGATAGGTTAGGTTATATGAGCACCTTCGGTGTGCTCTTATGCG									
Consensus	(601)	-----									
		Section 10									
AF092846 (Kleber-Janke et al.)	(676)	676	690	700	710	720	730	740	750		
AY158467 (Chatel et al.)	(520)	TTTACCTATTTTTTGCTACTGCGAAGTTTT									

Supplemental Fig. 5

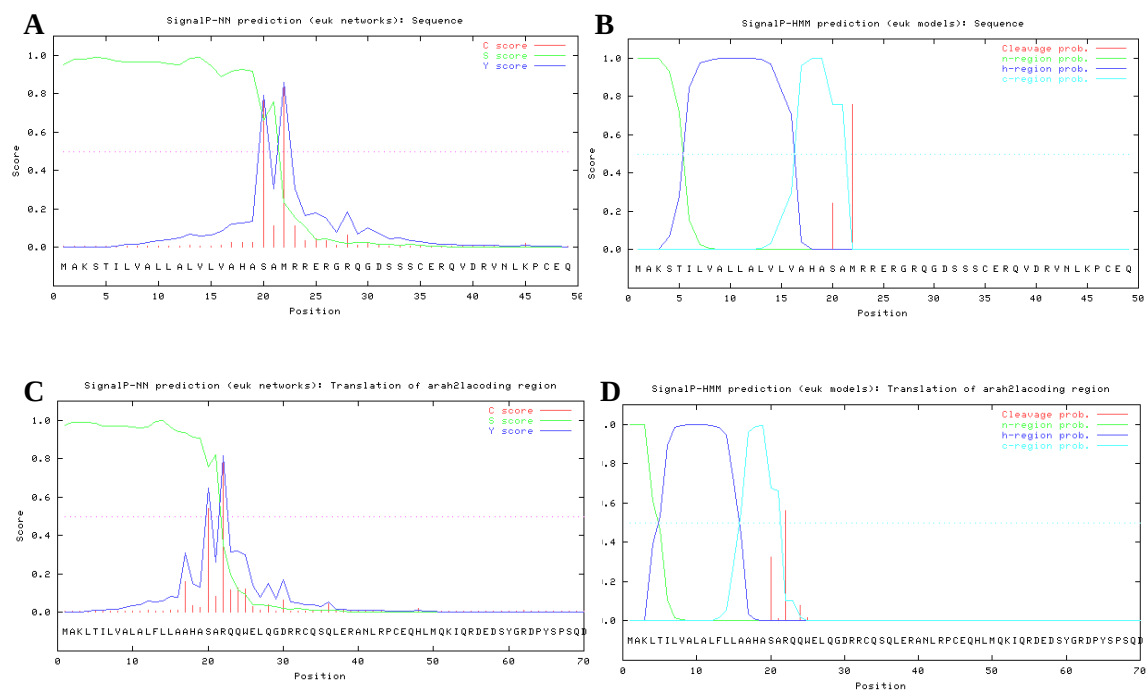


	(1)	10	20	30	40	50	60	70	80	Section 1
AD 6.1,1arab6	(1)	ATGGGCCAAAGTCCACCATTCTGGTAGCTCTCCTTGGCCCTCGTCTGGTGGGCACACGCGCTCCGC	ATGAGGCGCGGAGAGGGG							
AI 5.1,1arab6	(1)	ATGGGCCAAAGTCCACCATTCTGGTAGCTCTCCTTGGCCCTCGTCTGGTGGGCACACGCGCTCCGC	AATGAGGCGCGGAGAGGGG							
AI 5.2+5.5arab6	(1)	ATGGGCCAAAGTCCACCATTCTGGTAGCTCTCCTTGGCCCTCGTCTGGTGGGCACACGCGCTCCGC	AATGAGGCGCGGAGAGGGG							
Arah6 ORF+3' (8A2a)	(1)	ATGGGCCAAAGTCCACCATTCTGGTAGCTCTCCTTGGCCCTCGTCTGGTGGGCACACGCGCTCCGC	AATGAGGCGCGGAGAGGGG							
Arah6 ORF+3' (17A2a)	(1)	ATGGGCCAAAGTCCACCATTCTGGTAGCTCTCCTTGGCCCTCGTCTGGTGGGCACACGCGCTCCGC	AATGAGGCGCGGAGAGGGG							
Consensus	(1)	ATGGGCCAAAGTCCACCATTCTGGTAGCTCTCCTTGGCCCTCGTCTGGTGGGCACACGCGCTCCGC	AATGAGGCGCGGAGAGGGG							Section 2
	(81)	90	110	120	130	140	150	160		
AD 6.1,1arab6	(81)	GCGACAGGGGGACTCATCAAGCTGCGAGAGGCGAGGTAGACAGAGGGTAAACCTCAAGCCCTGCGAGCAGCACAATAATGACAGA								
AI 5.1,1arab6	(81)	GACACAGGGGGACTCATCAAGCTGCGAGAGGCGAGGTAGACAGAGGGTAAACCTCAAGCCCTGCGAGCAGCACAATAATGACAGA								
AI 5.2+5.5arab6	(81)	GACACAGGGGGACTCATCAAGCTGCGAGAGGCGAGGTAGACAGAGGGTAAACCTCAAGCCCTGCGAGCAGCACAATAATGACAGA								
Arah6 ORF+3' (8A2a)	(81)	GACACAGGGGGACTCATCAAGCTGCGAGAGGCGAGGTAGACAGAGGGTAAACCTCAAGCCCTGCGAGCAGCACAATAATGACAGA								
Arah6 ORF+3' (17A2a)	(81)	GACACAGGGGGACTCATCAAGCTGCGAGAGGCGAGGTAGACAGAGGGTAAACCTCAAGCCCTGCGAGCAGCACAATAATGACAGA								
Consensus	(81)	GACACAGGGGGACTCATCAAGCTGCGAGAGGCGAGGTAGACAGAGGGTAAACCTCAAGCCCTGCGAGCAGCACAATAATGACAGA								Section 3
	(161)	170	180	190	200	210	230	240		
AD 6.1,1arab6	(161)	GGATAATGGGCGAGCAAGAGCAGTAGCACTCTACGATATTAGGAGTACTCGATCTCTCCGACCAAGCAACAGAGGTGCTGCG								
AI 5.1,1arab6	(161)	GGATCATGGGCGAGCAAGAGCAGTAGCACTCTACGATATTAGGAGTACTCGATCTCTCCGACCAAGCAACAGAGGTGCTGCG								
AI 5.2+5.5arab6	(161)	GGATCATGGGCGAGCAAGAGCAGTAGCACTCTACGATATTAGGAGTACTCGATCTCTCCGACCAAGCAACAGAGGTGCTGCG								
Arah6 ORF+3' (8A2a)	(161)	GGATCATGGGCGAGCAAGAGCAGTAGCACTCTACGATATTAGGAGTACTCGATCTCTCCGACCAAGCAACAGAGGTGCTGCG								
Arah6 ORF+3' (17A2a)	(161)	GGATCATGGGCGAGCAAGAGCAGTAGCACTCTACGATATTAGGAGTACTCGATCTCTCCGACCAAGCAACAGAGGTGCTGCG								
Consensus	(161)	GGATCATGGGCGAGCAAGAGCAGTAGCACTCTACGATATTAGGAGTACTCGATCTCTCCGACCAAGCAACAGAGGTGCTGCG								Section 4
	(241)	250	260	270	280	290	300	310	320	
AD 6.1,1arab6	(241)	GATGAGCTGGACAGATGGAGAACACAGAGAGATGCATGTGCGAGGCCATTGCAGCAGATAATGGAGAACCAAGTGGCGATAG								
AI 5.1,1arab6	(241)	GATGAGCTGGAACAGATGGAGAACACACAGAGATGCATGTGCGAGGCCATTGCAGCAGATAATGGAGAACCAAGTGGCGATAG								
AI 5.2+5.5arab6	(241)	GATGAGCTGGAACAGATGGAGAACACACAGAGATGCATGTGCGAGGCCATTGCAGCAGATAATGGAGAACCAAGTGGCGATAG								
Arah6 ORF+3' (8A2a)	(241)	GATGAGCTGGAACAGATGGAGAACACACAGAGATGCATGTGCGAGGCCATTGCAGCAGATAATGGAGAACCAAGTGGCGATAG								
Arah6 ORF+3' (17A2a)	(241)	GATGAGCTGGAACAGATGGAGAACACACAGAGATGCATGTGCGAGGCCATTGCAGCAGATAATGGAGAACCAAGTGGCGATAG								
Consensus	(241)	GATGAGCTGGAACAGATGGAGAACACACAGAGATGCATGTGCGAGGCCATTGCAGCAGATAATGGAGAACCAAGTGGCGATAG								
	(321)	330	340	350	360	370	380	390	400	Section 5
AD 6.1,1arab6	(321)	GTTCGAGGACAGGCCAAATGGTGCAGCAGTTTCAAGAGAGAGGCTCATGAACTTGCC	CAACAGTGTAACTTTAGGGCACCAAC							
AI 5.1,1arab6	(321)	GTTCGAGGACAGGCCAAATGGTGCAGCAGTTTCAAGAGAGAGGCTCATGAACTTGCC	CAACAGTGTAACTTTAGGGCACCAAC							
AI 5.2+5.5arab6	(321)	GTTCGAGGACAGGCCAAATGGTGCAGCAGTTTCAAGAGAGAGGCTCATGAACTTGCC	CAACAGTGTAACTTTAGGGCACCAAC							
Arah6 ORF+3' (8A2a)	(321)	GTTCGAGGACAGGCCAAATGGTGCAGCAGTTTCAAGAGAGAGGCTCATGAACTTGCC	CAACAGTGTAACTTTAGGGCACCAAC							
Arah6 ORF+3' (17A2a)	(321)	GTTCGAGGACAGGCCAAATGGTGCAGCAGTTTCAAGAGAGAGGCTCATGAACTTGCC	CAACAGTGTAACTTTAGGGCACCAAC							
Consensus	(321)	GTTCGAGGACAGGCCAAATGGTGCAGCAGTTTCAAGAGAGAGGCTCATGAACTTGCC	CAACAGTGTAACTTTAGGGCACCAAC							Section 6
	(401)	410	420	430	440	450	460	470	480	
AD 6.1,1arab6	(401)	AGCGTTGCGATTGGACGTGAGTGGCGGCGAGATGCTAGACTCAAAAATAAATAATCTGTGCGCAAAAC	AAACTAGTAGGAAG							
AI 5.1,1arab6	(401)	AGCGTTGCGATTGGACGTGAGTGGCGGCGAGATGCTAGACTCAAAAATAAATAATCTGTGCGCAAAAC	AAACTAGTAGGAAG							
AI 5.2+5.5arab6	(401)	AGCGTTGCGATTGGACGTGAGTGGCGGCGAGATGCTAGACTCAAAAATAAATAATCTGTGCGCAAAAC</								

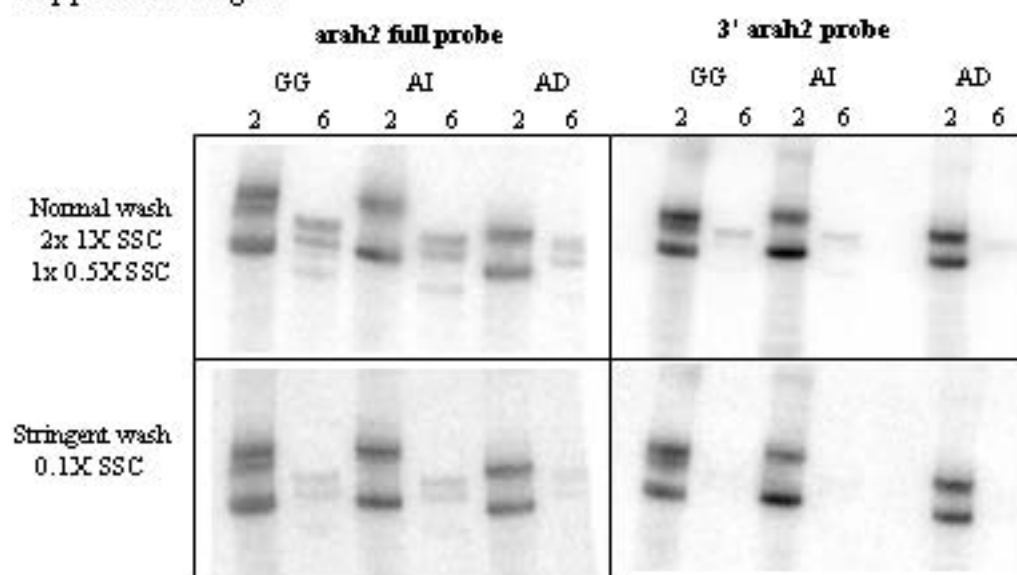
Supplemental Fig. 8

		Section 1					
		(1)	1	10	20	30	45
PSC33 conglutin		(1)	-AKSTILVALLALVLVAHASAMRRERGRQGDSSSCERQVDRVNLK				
Translation of AY848699 (CONG4-Psc33)		(1)	MAKSTILVALLALVLVAHASAMRRERGRQGDSSSCERQVDRVNLK				
Translation of AD 6,1 arah6		(1)	MAKSTILVALLALVLVAHASAMRRERGRQGDSSSCERQVDRVNLK				
Translation of 4,10 arah6		(1)	MAKSTILVALLALVLVAHASAMRRERGRQGDSSSCERQVDRVNLK				
Translation of AY849314 (CONG5-Arah6)		(1)	MAKSTILVALLALVLVAHASAMRRERGRQGDSSSCERQVDRVNLK				
Translation of AY722690 (CONG3)		(1)	MAKSTILVALLALVLVAHASAMRRERGRQGDSSSCERQVDRVNLK				
protein arah6		(1)	-----AHASAMRRERGRQGDSSSCERQVDRVNLK				
Translation of AY871100		(1)	-----MRRERGRQGDSSSCERQVDRVNLK				
Translation of AI 5,1 arah6		(1)	MAKSTILVALLALVLVAHASAMRRERGRQGDSSSCERQVDRVNLK				
Translation of 4,6 arah6		(1)	MAKSTILVALLALVLVAHASAMRRERGRQGDSSSCERQVDRVNLK				
Consensus		(1)	MAKSTILVALLALVLVAHASAMRRERGRQGDSSSCERQVDRVNLK				
		Section 2					
		(46)	46	60	70	80	90
PSC33 conglutin		(46)	PCEQHIMQRIMGEQEYDSYDIRSTRSSDQQQRCCDELDQMENTE				
Translation of AY848699 (CONG4-Psc33)		(46)	PCEQHIMQRIMGEQEYDSYDIRSTRSSDQQQRCCDELDQMENTE				
Translation of AD 6,1 arah6		(46)	PCEQHIMQRIMGEQEYDSYDIRSTRSSDQQQRCCDELDQMENTE				
Translation of 4,10 arah6		(46)	PCEQHIMQRIMGEQEYDSYDIRSTRSSDQQQRCCDELDQMENTE				
Translation of AY849314 (CONG5-Arah6)		(46)	PCEQHIMQRIMGEQEYDSYDIRSTRSSDQQQRCCDELDNEMENTQ				
Translation of AY722690 (CONG3)		(46)	PCEQHIMQRIMGEQEYDSYDIRSTRSSDQQQRCCDELDNEMENTQ				
protein arah6		(30)	PCEQHIMQRIMGEQEYDSYDFGSTRSSDQQQRCCDELDNEMENTQ				
Translation of AY871100		(25)	PCEQHIMQRIMGEQEYDSYDIRSTRSSDQQQRCCDELDNEMENTQ				
Translation of AI 5,1 arah6		(46)	PCEQHIMQRIMGEQEYDSYDIRSTRSSDQQQRCCDELDNEMENTQ				
Translation of 4,6 arah6		(46)	PCEQHIMQRIMGEQEYDSYDIRSTRSSDQQQRCCDELDNEMENTQ				
Consensus		(46)	PCEQHIMQRIMGEQEYDSYDIRSTRSSDQQQRCCDELDNEMENTQ				
		Section 3					
		(91)	91	100	110	120	135
PSC33 conglutin		(90)	RCMCEALQQIMENQCDRLQDRQMVGQFKRELMNLPQQCNFRAPQR				
Translation of AY848699 (CONG4-Psc33)		(91)	RCMCEALQQIMENQCDRLQDRQMVGQFKRELMNLPQQCNFRAPQR				
Translation of AD 6,1 arah6		(91)	RCMCEALQQIMENQCDRLQDRQMVGQFKRELMNLPQQCNFRAPQR				
Translation of 4,10 arah6		(91)	RCMCEALQQIMENQCDRLQDRQMVGQFKRELMNLPQQCNFRAPQR				
Translation of AY849314 (CONG5-Arah6)		(91)	RCMCEALQQIMENQCDRLQDRQMVGQFKRELMNLPQQCNFRAPQR				
Translation of AY722690 (CONG3)		(91)	RCMCEALQQIMENQCDRLQDRQMVGQFKRELMNLPQQCNFRAPQR				
protein arah6		(75)	RCMCEALQQIMENQCDGLQDRQMVGQFKRELMNLPQQCNFGAPQR				
Translation of AY871100		(70)	RCMCEALQQIMENQCDRLQDRQMVGQFKRELMNLPQQCNFRAPQR				
Translation of AI 5,1 arah6		(91)	RCMCEALQQIMENQCDRLQDRQMVGQFKRELMNLPQQCNFRAPQR				
Translation of 4,6 arah6		(91)	RCMCEALQQIMENQCDRLQDRQMVGQFKRELMNLPQQCNFRAPQR				
Consensus		(91)	RCMCEALQQIMENQCDRLQDRQMVGQFKRELMNLPQQCNFRAPQR				
		(136)	136	145			
PSC33 conglutin		(135)	CDLDVSGGRC				
Translation of AY848699 (CONG4-Psc33)		(136)	CDLDVSGGRC				
Translation of AD 6,1 arah6		(136)	CDLDVSGGRC				
Translation of 4,10 arah6		(136)	CDLDVSGGRC				
Translation of AY849314 (CONG5-Arah6)		(136)	CDLDVSGGRC				
Translation of AY722690 (CONG3)		(136)	CDLDVSGGRC				
protein arah6		(120)	CDLDVSGGRC				
Translation of AY871100		(115)	CDLDVSGGRC				
Translation of AI 5,1 arah6		(136)	CDLDVSGGRC				
Translation of 4,6 arah6		(136)	CDLDVSGGRC				
Consensus		(136)	CDLDVSGGRC				

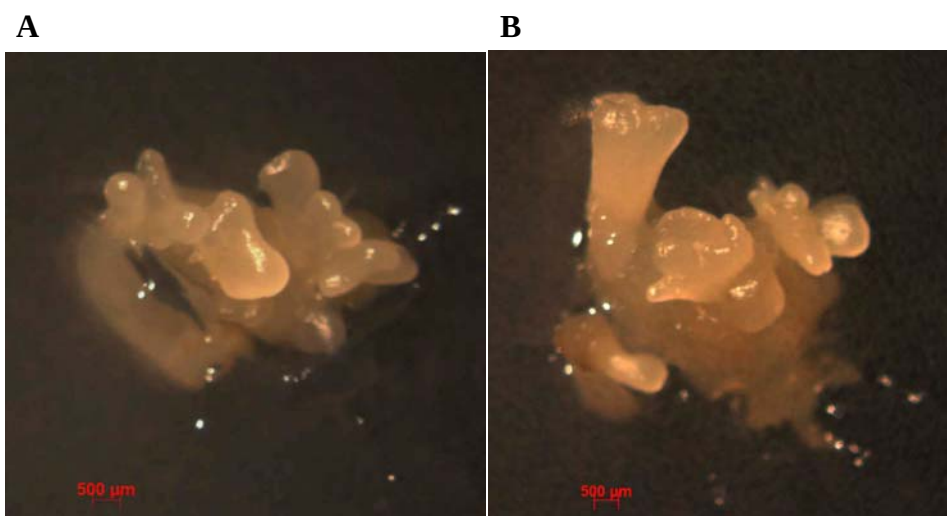
Supplemental Fig. 9



Supplemental Fig. 10



Supplemental Fig. 11



Supplemental Fig. 12

		Section 1									
V00296.1 ELACZ 3020-2480bp	(1)	1	10	20	30	40	50				
AX148738(Arah2Vizquez)	(1)	-----	-----	-----	-----	-----	-----				
5'-ORF-3'Arah2(B-genome)	(1)	CTTTCCCT	TTTCC	ACGTACCTC	ACAT	GCAAAATCCCT	CCTTT	CACCTATA			
Consensus	(1)	T	ACG	AC	GCA	A	C	TAT			
		Section 2									
V00296.1 ELACZ 3020-2480bp	(51)	51	60	70	80	90	100				
AX148738(Arah2Vizquez)	(43)	ATTATTTT	TGAC	CAC	AGA	CCAACTGGT	AATGGT	AGCGAC	CGGCCT	TCAGCT	
5'-ORF-3'Arah2(B-genome)	(51)	AATACCAT	CAC	CAC	CAT	CCACACTCA	AAAAAC	ACCACAC	ACTCTT	TCAATA	
Consensus	(51)	A	TA	T	CAC	CCA	AA	A	C	C	TCA
		Section 3									
V00296.1 ELACZ 3020-2480bp	(101)	101	110	120	130	140	150				
AX148738(Arah2Vizquez)	(93)	GGAATTC	G	GGG	CCG	CAATGGCC	AAAGCTC	ACCATACT	AGTA		
5'-ORF-3'Arah2(B-genome)	(101)	CACATTC	CATAA	CCACC	CAACAA	CAATGGCC	AAAGCTC	ACCATACT	AGTA		
Consensus	(101)	ATTC		C	CC	C	CAATGGCC	AAAGCTC	ACCATACT	AGTA	
		Section 4									
V00296.1 ELACZ 3020-2480bp	(151)	151	160	170	180	190	200				
AX148738(Arah2Vizquez)	(134)	GCCCTCGCCCTTTT	CCTCCTCGCT	GCCCA	CGCATCT	GCGAGGC	CAGCAGTG				
5'-ORF-3'Arah2(B-genome)	(151)	GCCCTCGCCCTTTT	CCTCCTCGCT	GCCCA	CGCATCT	GCGAGGC	CAGCAGTG				
Consensus	(151)	GCCCTCGCCCTTTT	CCTCCTCGCT	GCCCA	CGCATCT	GCGAGGC	CAGCAGTG				
		Section 5									
V00296.1 ELACZ 3020-2480bp	(201)	201	210	220	230	240	250				
AX148738(Arah2Vizquez)	(184)	GGAACCT	CAAGGAGAC	AGAAGAT	GCCAGAGCC	AGCTCGAGAGGG	CGAACC				
5'-ORF-3'Arah2(B-genome)	(201)	GGAACCT	CAAGGAGAC	AGAAGAT	GCCAGAGCC	AGCTCGAGAGGG	CGAACC				
Consensus	(201)	GGAACCT	CAAGGAGAC	AGAAGAT	GCCAGAGCC	AGCTCGAGAGGG	CGAACC				
		Section 6									
V00296.1 ELACZ 3020-2480bp	(251)	251	260	270	280	290	300				
AX148738(Arah2Vizquez)	(234)	TGAGGCCCT	GCGAGCA	ACATCTCAT	GCAAGAA	ATCCAAC	GTGACGAGGAT				
5'-ORF-3'Arah2(B-genome)	(251)	TTAGGCCCT	GCGAGCA	ACATCTCAT	GCAAGAA	ATCCAAC	GTGACGAGGAT				
Consensus	(251)	T	AGGCCCT	GCGAGCA	ACATCTCAT	GCAAGAA	ATCCAAC	GTGACGAGGAT			
		Section 7									
V00296.1 ELACZ 3020-2480bp	(301)	301	310	320	330	340	350				
AX148738(Arah2Vizquez)	(284)	TCATATG	AACGGG	ACCCGTAC	AGCCCTAG	TCAGGATCC	GTACAGCCCTAG				
5'-ORF-3'Arah2(B-genome)	(301)	TCATATG	AACGGG	ACCCGTAC	AGCCCTAG	TCAGGATCC	GTACAGCCCTAG				
Consensus	(301)	TCATATG	ACGGG	ACCCGTAC	AGCCCTAG	TCAGGATCC	GTACAGCCCTAG				
		Section 8									
V00296.1 ELACZ 3020-2480bp	(351)	351	360	370	380	390	400				
AX148738(Arah2Vizquez)	(334)	TC					CATATGATCGGA				
5'-ORF-3'Arah2(B-genome)	(351)	TC					CATATGATCGGA				
Consensus	(351)	TC					CATATGATCGGA				
		Section 9									
V00296.1 ELACZ 3020-2480bp	(401)	401	410	420	430	440	450				
AX148738(Arah2Vizquez)	(348)	GAGGCGCTGGAT	CCTCTCAG	CACCAAGAGAGGT	GTTGCAAT	GAGCTGAAC					
5'-ORF-3'Arah2(B-genome)	(401)	GAGGCGCTGGAT	CCTCTCAG	CACCAAGAGAGGT	GTTGCAAT	GAGCTGAAC					
Consensus	(401)	GAGGCGCTGGAT	CCTCTCAG	CACCAAGAGAGGT	GTTGCAAT	GAGCTGAAC					
		Section 10									
V00296.1 ELACZ 3020-2480bp	(451)	451	460	470	480	490	500				
AX148738(Arah2Vizquez)	(398)	GAGTTTGAGA	ACAACCAAGGT	GCATGTGCGAGGC	ATTGCAACAGAT	CAT					
5'-ORF-3'Arah2(B-genome)	(451)	GAGTTTGAGA	ACAACCAAGGT	GCATGTGCGAGGC	ATTGCAACAGAT	AAT					
Consensus	(451)	GAGTTTGAGA	ACAACCAAGGT	GCATGTGCGAGGC	ATTGCAACAGAT	AT					
		Section 11									
V00296.1 ELACZ 3020-2480bp	(501)	501	510	520	530	540	550				
AX148738(Arah2Vizquez)	(448)	GGAGAACCAGAG	CGATAGGTT	GCAGGGGAGG	CAACAGGAGCA	ACAGTTCA					
5'-ORF-3'Arah2(B-genome)	(501)	GGAGAACCAGAG	CGATAGGTT	GCAGGGGAGG	CAACAGGAGCA	ACAGTTCA					
Consensus	(501)	GGAGAACCAGAG	CGATAGGTT	GCAGGGGAGG	CAACAGGAGCA	ACAGTTCA					
		Section 12									
V00296.1 ELACZ 3020-2480bp	(551)	551	560	570	580	590	600				
AX148738(Arah2Vizquez)	(498)	AGAGGGAGCTC	AGGAAC	TGCCTCAAC	AGTGCGGCCT	TAGGGCACCACAG					
5'-ORF-3'Arah2(B-genome)	(551)	AGAGGGAGCTC	AGGAAC	TGCCTCAAC	AGTGCGGCCT	TAGGGCACCACAG					
Consensus	(551)	AGAGGGAGCTC	AGGAAC	TGCCTCAAC	AGTGCGGCCT	AGGGCACCACAG					
		Section 13									
V00296.1 ELACZ 3020-2480bp	(601)	601	610	620	630	640	650				
AX148738(Arah2Vizquez)	(548)	CGTTGCGACTT	GGA	GTGAAAGT	GCGGCAG	CGCGCGCAATTC	CGC				
5'-ORF-3'Arah2(B-genome)	(601)	CGTTGCGACTT	GGA	GTGAAAGT	GCGGCAG	CGCGCGCAATTC	CGC				
Consensus	(601)	CGTTGCGACTT	GGA	GTGAAAGT	GCGGCAG	G C G	GAATTC	CGC			
		Section 14									
V00296.1 ELACZ 3020-2480bp	(651)	651	660	670	680	690	700				
AX148738(Arah2Vizquez)	(597)	CGATACTGAC	CGGGCTCC	AGGAGTC	GTCGCC	ACCAATCC	CCATATGGAAAC				
5'-ORF-3'Arah2(B-genome)	(650)	CGATACTGAC	CGGGCTCC	AGGAGTC	GTCGCC	ACCAATCC	CCATATGGAAAC				
Consensus	(651)	CGATACTGAC	CGGGCTCC	AGGAGTC	GTCGCC	ACCAATCC	CCATATGGAAAC				

		Section 15									
	(701)	701	710	720	730	740	750				
V00296.1 ELACZ3020-2480bp	(60)	CGTCGATATTCAGCCATGTGCC	TTCTTCCGCGTGCAGCAGATGGCGATGG								
AX148738(Arah2Vizquez)	(647)	CGTCGATATTCAGCCATGTGCC	TTCTTCCGCGTGCAGCAGATGGCGATGG								
5'-ORF-3'Arah2(B-genome)	(700)	TATT-ATCTATGGAATGTAGTT	-TTGGTAATAATAAAGATC	---ATCA							
Consensus	(701)	CGTCGATATTCAGCCATGTGCC	TTCTTCCGCGTGCAGCAGATGGCGATGG								
		Section 16									
	(751)	751	760	770	780	790	800				
V00296.1 ELACZ3020-2480bp	(110)	CTGGTTTCCATCAGTTGCT	---GTTGACTGTAGCG	-GCTGATGTTGAACT							
AX148738(Arah2Vizquez)	(697)	CTGGTTTCCATCAGTTGCT	---GTTGACTGTAGCG	-GCTGATGTTGAACT							
5'-ORF-3'Arah2(B-genome)	(746)	CTA-TATGAATGTGTTGAT	CGTGTTAACTAAGGCAAGCTTAGGTTATATG								
Consensus	(751)	CTGGTTTCCATCAGTTGCT	GTTGACTGTAGCG	GCTGATGTTGAACT							
		Section 17									
	(801)	801	810	820	830	840	850				
V00296.1 ELACZ3020-2480bp	(156)	GGAAG--TCGCCGCGCCACT	--GGTGTGGGCCATAATTCAAATTCGCGCGT								
AX148738(Arah2Vizquez)	(743)	GGAAG--TCGCCGCGCCACT	--GGTGTGGGCCATAATTCAAATTCGCGCGT								
5'-ORF-3'Arah2(B-genome)	(794)	AGCACCTTAGAGTGCTTTTAT	GGCGTTGCTATGTTTTGTTGTCGAGA								
Consensus	(801)	GGAAG	TCGCCGCGCCACT	GGTGTGGGCCATAATTCAAATTCGCGCGT							
		Section 18									
	(851)	851	860	870	880	890	900				
V00296.1 ELACZ3020-2480bp	(202)	CCCGCAGCGC	-----AGACCGT	-TTTCGCTCGGGAAG							
AX148738(Arah2Vizquez)	(789)	CCCGCAGCGC	-----AGACCGT	-TTTCGCTCGGGAAG							
5'-ORF-3'Arah2(B-genome)	(844)	GTTGTAACCATCTTGAAATAATATAAAA	AGATCATGTTTTGTTTGGCGA								
Consensus	(851)	CCCGCAGCGC	AGACCGT	TTTCGCTCGGGAAG							
		Section 19									
	(901)	901	910	920	930	940	950				
V00296.1 ELACZ3020-2480bp	(233)	ACGTACGGGGTATACATGTCTG	GACAATGGCAGATCCCAGCGGTCAAAAACA								
AX148738(Arah2Vizquez)	(820)	ACGTACGGGGTATACATGTCTG	GACAATGGCAGATCCCAGCGGTCAAAAACA								
5'-ORF-3'Arah2(B-genome)	(894)	GTATATTATGTTTT-ATTTC	CATTTTGGTTTTTCTGATTGATGATT								
Consensus	(901)	ACGTACGGGGTATACATGTCTG	GACAATGGCAGATCCCAGCGGTCAAAAACA								
		Section 20									
	(951)	951	960	970	980	990	1000				
V00296.1 ELACZ3020-2480bp	(283)	GGCGGCAGTAAGGCGGT	CGGGATAGTTTTCTTGCGGCCCTAAATCCGAGCC								
AX148738(Arah2Vizquez)	(870)	GGCGGCAGTAAGGCGGT	CGGGATAGTTTTCTTGCGGCCCTAAATCCGAGCC								
5'-ORF-3'Arah2(B-genome)	(943)	AGTTGAATAAAAG--GTATT	GAAGTTTTTTTCC--CCTAAACTTCATC								
Consensus	(951)	GGCGGCAGTAAGGCGGT	CGGGATAGTTTTCTTGCGGCCCTAAATCCGAGCC								
		Section 21									
	(1001)	1001	1010	1020	1030	1040	1050				
V00296.1 ELACZ3020-2480bp	(333)	AGTTTACCCGCTCTGCTAC	CTGCGCCAGCTGGCAGTTC	---AGGCCAA							
AX148738(Arah2Vizquez)	(920)	AGTTTACCCGCTCTGCTAC	CTGCGCCAGCTGGCAGTTC	---AGGCCAA							
5'-ORF-3'Arah2(B-genome)	(988)	TATTTAAAAATTTAGAT-CAAG	TAAAATTTTAAATTTTTTAAAGTTAT								
Consensus	(1001)	AGTTTACCCGCTCTGCTAC	CTGCGCCAGCTGGCAGTTC	AAGCCAA							
		Section 22									
	(1051)	1051	1060	1070	1080	1090	1100				
V00296.1 ELACZ3020-2480bp	(378)	--TCCGCGCCGGATGCGGTGTAT	---CGCTCGCCACTTCAACATCAACGG								
AX148738(Arah2Vizquez)	(965)	--TCCGCGCCGGATGCGGTGTAT	---CGCTCGCCACTTCAACATCAACGG								
5'-ORF-3'Arah2(B-genome)	(1037)	ATTATTTGGTGGATAGGGTGT	TTTTCACGCTAAAATTTTGAATA-CAGGTG								
Consensus	(1051)	TCCGCGCCGGATGCGGTGTAT	CGCTCGCCACTTCAACATCAACGG								
		Section 23									
	(1101)	1101	1110	1120	1130	1140	1150				
V00296.1 ELACZ3020-2480bp	(423)	TAATCGCCATT--TGACCACTACC	---ATCAATCCGGTAGGTTTTCCGG-								
AX148738(Arah2Vizquez)	(1010)	TAATCGCCATT--TGACCACTACC	---ATCAATCCGGTAGGTTTTCCGG-								
5'-ORF-3'Arah2(B-genome)	(1086)	TTAGAGAAAAGAA	TGACGCGAAGAGAGATGGATGTGCTTCTTTGTTAGAG								
Consensus	(1101)	TAATCGCCATT	TGACCACTACC	ATCAATCCGGTAGGTTTTCCGG							
		Section 24									
	(1151)	1151	1160	1170	1180	1190	1200				
V00296.1 ELACZ3020-2480bp	(467)	-----CTGATAAATAA-GGTTTTCCCC	-TGATGCTGCCACGCGTGAGCGG								
AX148738(Arah2Vizquez)	(1054)	-----CTGATAAATAAAGGTTTTCCCC	-TGATGCTGCCACGCGTGAGCGG								
5'-ORF-3'Arah2(B-genome)	(1136)	ACGCACTAGTCAATT	CAGGACAACATTATGTCAATGTCAATTAGGTTTG								
Consensus	(1151)	CTGATAAATAAAGGTTTTCCCC	TGATGCTGCCACGCGTGAGCGG								
		Section 25									
	(1201)	1201	1210	1220	1230	1240	1250				
V00296.1 ELACZ3020-2480bp	(510)	TCGTAATCAGCACCGCATCAG	---CAAGTGAT	-----CT-GC							
AX148738(Arah2Vizquez)	(1098)	TCGTAATCAGCACCGCATCAA	---CAAGTGAT	-----TTTGC							
5'-ORF-3'Arah2(B-genome)	(1186)	ACAAAAAATAAATAATTCG	AGAGACAATTATAACATAAGAAAAATTTTA								
Consensus	(1201)	TCGTAATCAGCACCGCATCAA	CAAGTGAT	TTTGC							
		Section 26									
	(1251)	1251	1260	1270	1280	1290	1300				
V00296.1 ELACZ3020-2480bp	(544)	CGTGCACTGCA--ACAACGCTG	CTTCGGCTGGTAATGGCCCGCGCCTT								
AX148738(Arah2Vizquez)	(1133)	CGTGCACTGCA--ACAACGCTG	CTTCGGCTGGTAATGGCCCGCGCCTT								
5'-ORF-3'Arah2(B-genome)	(1236)	AAATGCACTGGAACACATTGGTG	-TTCTAGTTGTTTAAACCGTTGATTTC								
Consensus	(1251)	CGTGCACTGCA	ACAACGCTG	CTTCGGCTG	T	C					
		Section 27									
	(1301)	1301	1306								
V00296.1 ELACZ3020-2480bp	(592)	CCAGCG									
AX148738(Arah2Vizquez)	(1163)	-----									
5'-ORF-3'Arah2(B-genome)	(1285)	ATTAAT									
Consensus	(1301)										

Supplemental Fig. 13

Figure 1 displays a multiple sequence alignment of the ELACZ 3078-3019 bp region. The alignment is divided into three sections, each showing the sequence of V00296.1 ELACZ, AX148738(Arah2Vizquez), and 5'-ORF-3'/Arah2(B-genome). The alignment is color-coded by nucleotide: A (green), C (blue), G (red), and T (yellow). Consensus sequences are shown at the bottom of each section.

Section 1 (Positions 1-70):

V00296.1 ELACZ 3078-3019 bp (1) 1 10 20 30 40 50 60 70
 AX148738(Arah2Vizquez) (1) -----TTATTATTATTTTGTGACACAGACCC
 5'-ORF-3'/Arah2(B-genome) (1) -----TCCTTACGCGAAATACGGCGACATGGCCGTGCCCGGTTATTATTATTTTGTGACAC-AGACAC
 Consensus (1) CTTTCCCTTTTCCACGTACCTCACATGCAAAATCCCTCCTTTCACCTATAAATACCATCACACCATCTCA

Section 2 (Positions 71-140):

V00296.1 ELACZ 3078-3019 bp (71) 71 80 90 100 110 120 130 140
 AX148738(Arah2Vizquez) (28) AACTGGTAATGGTAGCGAC---CGGCGCT-CAGCTG---
 (62) AACTGGTAATGGTAGCGAC---CGGCGCT-CAGCTGGAATTCGGGG---CGC CAATGGCCAAAGCTCAC
 5'-ORF-3'/Arah2(B-genome) (71) CACTCAA AAAACACCCACACACTCTTCAATACACATTCCATAACCAACCAACCAATGGCCAAAGCTCAC
 Consensus (71) AACTGGTAATGGTAGCGAC CGGCGCT CAGCTG AT C C C CAATGGCCAAAGCTCAC

Section 3 (Positions 141-210):

V00296.1 ELACZ 3078-3019 bp (141) 141 150 160 170 180 190 200 210
 AX148738(Arah2Vizquez) (58) -----
 (124) CATACTAGTAGCCCTCGCCCTTTTCTCTCTCGCTGCCACGCATCTGCGAGGCGAGCAGTGGGAACTCCAA
 5'-ORF-3'/Arah2(B-genome) (141) CATACTAGTAGCCCTCGCCCTTTTCTCTCTCGCTGCCACGCATCTGCGAGGCGAGCAGTGGGAACTACAA
 Consensus (141) CATACTAGTAGCCCTCGCCCTTTTCTCTCTCGCTGCCACGCATCTGCGAGGCGAGCAGTGGGAACTCAA

Section 4 (Positions 211-280):

V00296.1 ELACZ 3078-3019 bp (211) 211 220 230 240 250 260 270 280
 AX148738(Arah2Vizquez) (58) -----
 (194) GGAGACAGAAAGATGCCAGAGCCAGCTCGAGAGGGCGAACTGAGGCCCTGCGAGGCAACATCTCATGCAGA
 5'-ORF-3'/Arah2(B-genome) (211) GGAGACAGAAAGATGCCAGAGCCAGCTCGAGAGGGCGAACTGAGGCCCTGCGAGGCAACATCTCATGCAGA
 Consensus (211) GGAGACAGAAAGATGCCAGAGCCAGCTCGAGAGGGCGAACTAGGCCCTGCGAGGCAACATCTCATGCAGA

		Section 1					
AX148738(Arah2Vizquez)	(1)	1	10	20	30	40	53
	(1)	TCCTTACGCGAAATACGGGCAGACATGGCCTGCCCGGT					TATTATTATTTTGA
	(1)	-----					TATTATTATTTTGA
	(1)						TATTATTATTTTGA
		Section 2					
AX148738(Arah2Vizquez)	(54)	54	60	70	80	90	106
	(54)	CAC-AGACCAACTGGTAATGGTAGCGACCGGCGCTCAGCTG					GGAATTCGCGGCC
	(16)	CACAGACCAACTGGTAATGGTAGCGACCGGCGCTCAGCTG					-----
	(54)	CAC AGACCAACTGGTAATGGTAGCGACCGGCGCTCAGCTG					
		Section 3					
AX148738(Arah2Vizquez)	(107)	107	120	130	140		159
	(106)	GCCAATGGCCAAGCTCACCATACTAGTAGCCCTCGCCCTTTTCCTCCTCGCTG					
	(57)	-----					
	(107)						
		Section 4					
AX148738(Arah2Vizquez)	(160)	160	170	180	190	200	212
	(159)	CCCACGCATCTGCGAGGCAGCACTGGGAAGCTCCAAGGAGACAGAAGATGCCAG					
	(57)	-----					
	(160)						
		Section 5					
AX148738(Arah2Vizquez)	(213)	213	220	230	240	250	265
	(212)	AGCCAGCTCGAGAGGGCGAACCTGAGGCCCTGCGAGCAACATCTCATGCAGAA					
	(57)	-----					
	(213)						
		Section 6					
AX148738(Arah2Vizquez)	(266)	266	280	290	300		318
	(265)	GATCCAACGTGACGAGGATTCATATGAACGGGACCCGTACAGCCCTAGTCAGG					
	(57)	-----					
	(266)						
		Section 7					
AX148738(Arah2Vizquez)	(319)	319	330	340	350	360	371
	(318)	ATCCGTACAGCCCTAGTCCATATGATCGGAGAGGCGCTGGATCCTCTCAGCAC					
	(57)	-----					
	(319)						
		Section 8					
AX148738(Arah2Vizquez)	(372)	372	380	390	400	410	424
	(371)	CAAGAGAGGTGTTGCAATGAGCTGAACGAGTTTGAGAACCAACCAAGGTGCAT					
	(57)	-----					
	(372)						