Isolation and Functional Characterization of a Novel Seed-Specific Promoter Region from Peanut

Sowmini Sunkara · Pooja Bhatnagar-Mathur ·
Kiran Kumar Sharma

Abstract The importance of using tissue-specific promoters in the genetic transformation of plants has been emphasized increasingly. Here, we report the isolation of a novel seed-specific promoter region from peanut and its validation in *Arabidopsis* and tobacco seeds. The reported promoter region referred to as groundnut seed promoter (GSP) confers seed-specific expression in heterologous systems, which include putative promoter regions of the peanut (*Arachis hypogaea* L.) gene *8A4R19G1*. This region was isolated, sequenced, and characterized using gel shift assays. Tobacco transgenics obtained using binary vectors carrying *uid*A reporter gene driven by GSP and/or cauliflower mosaic virus 35S promoters were confirmed through polymerase chain reaction (PCR), RT-PCR, and computational analysis of motifs which revealed the presence of TATA, CAAT boxes, and ATG signals. This seed-specific promoter region successfully targeted the reporter *uid*A gene to seed tissues in both *Arabidopsis* and tobacco model systems, where its expression was confirmed by histochemical analysis of the transgenic seeds. This promoter region is routinely being used in the genetic engineering studies in legumes aimed at targeting novel transgenes to the seeds, especially those involved in micronutrient enhancement, fungal resistance, and molecular pharming.

Abbreviations

EMSA	Electrophoretic mobility shift assay
GUS	β-Glucuronidase
GSP	Groundnut seed promoter
CaMV	*Cauliflower mosaic virus*

S. Sunkara · P. Bhatnagar-Mathur · K. K. Sharma (✉)
Genetic Transformation Laboratory, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru 502324 Andhra Pradesh, India
e-mail: k.sharma@cgiar.org

S. Sunkara
Jawaharlal Nehru Technological University Hyderabad, Kukatpally,
Hyderabad 500085 Andhra Pradesh, India

Ⓐ Springer

Introduction

Genetic engineering has tremendous potential to transform global agriculture by providing hitherto unavailable traits into the crop gene pool. For stable expression and inheritance of the transgene, the expression cassette consists of three important components including promoter, target gene, and termination signal, of which, the promoter is essential to control the expression of the transgenes, besides providing valuable insights into their overexpression or silencing in response to the external stimuli. Till date, the most commonly employed promoters for developing transgenic plants for various traits include the constitutive promoters of the genes like 35S of the cauliflower mosaic virus (CaMV 35S) [1] or the maize ubiquitin promoter [2, 3]. However, constitutive expression of several transgenes has been reported to be associated with higher metabolic costs resulting in interference with the plant metabolic pathways leading to undesirable pleiotropic effects in transgenic plants [4, 5]. Hence, the use of inducible or tissue-specific promoters has been recognized to be an important component of plant genetic engineering to target the expression of the introduced genes in the desired tissues or under specific conditions, which not only prevents the metabolic burden to the host plant, but to a certain extent also safeguards against the regulatory concerns. In recent years, several well-characterized gene promoters have become available for transgene expression in plants [6].

A tissue-specific promoter has activity in only certain cell types. Accordingly, elements of the natural promoter region necessary for obtaining the required level of gene expression while retaining tissue specificity should be known [7]. Therefore, choosing the correct promoter, especially a tissue-specific promoter, is a major step towards achieving the desired transgene expression. For example, the use of tissue-specific promoters in developing RNAi constructs is critical to augment gene silencing strategies to avoid nontarget effects on other processes. Moreover, synthetic promoters designed with required combination of promoter motifs provide new avenues of research for tailored gene expression.

Endogenous promoters cloned from plants could direct gene expression to specific tissues, viz., root, green tissue, seed, nodule, etc., thereby avoiding transgene silencing often associated with the nonplant origin promoters. The real benefit of using an endogenous promoter is that presumably the presence of matching transcription factors so the transcription patterns can be better anticipated. In some cases, the use of endogenous regulatory regions with particular developmental expression patterns has been proven to mitigate the problem [8, 9].

Moreover, the use of tissue-specific promoters also addresses the biosafety concerns related to nonspecific expression of target genes in the transgenics. Model legumes are being rapidly developed as experimental systems to pursue a number of important biological questions unique to these plants using molecular tools including genomics and transgenics. Many achievements have been made and studied extensively to regulate the target gene expression using seed-specific promoters from the model systems during the seed development [10]. Using seed-specific promoters of legume origin has a potential to develop transgenic legume technologies for specific biotic constraints such as *Aspergillus flavus* invasion and resulting aflatoxin contamination, quality traits such as grain/seed biofortification, besides seed and oil quality improvement. It is preferable to utilize seed-specific promoters for these applications so as to limit the presence of such nontraditional products in the seed and also to minimize the metabolic costs in other plant parts.

Here, we report the isolation and characterization of a novel promoter region of legume origin—groundnut seed promoter (GSP). We cloned and sequenced the upstream regulatory region of *8A4R19G1* gene [11] using gene-specific primers and tested its tissue-specific expression pattern by sub-cloning into a binary vector upstream of *uid*A (GUS) gene encoding β-glucuronidase. The expression pattern of the cloned promoter region was

analyzed in transgenic tobacco and *Arabidopsis* plants by histochemical and fluorometric analysis of GUS and electrophoretic mobility shift assay (EMSA).

Materials and Methods

Bacterial and Plant Material

Escherichia coli strain DH5α was used for cloning and propagation of the plasmids, and disarmed *Agrobacterium tumefaciens* strain C58 was used for plant transformation. *E. coli* and *A. tumefaciens* were grown in LB and yeast extract broth (YEB) media [12] at 37 and 28 °C, respectively, with appropriate antibiotics. Seeds of peanut (*Arachis hypogaea* L. var. JL24) were sown and maintained in pots in the greenhouse. Genomic DNA was isolated from leaf tissue of 1-month-old plants using the cetyl trimethyl ammonium bromide (CTAB) method [13].

Isolation of Groundnut Seed Promoter Fragment

Oligonucleotide gene-specific primers for *8A4R19G1* gene (GenBank accession no. DQ450071) were synthesized using the Primer3 software [14] and used for amplifying the gene from the genomic DNA isolated from peanut and sequenced. Upon confirmation of the sequence obtained using BLAST analysis, a 523 bp of 5′ flank upstream sequences of the gene was isolated using the Genome Walker Universal Kit (DSS Takara Bio India Pvt. Ltd.) from the peanut genomic DNA and sequenced. Based on the obtained sequence, oligonucleotide primers GSP FP 5′-AAC CGG ATC CAG CTT TAA TAG CAA CTA GGC-3′ and GSP RP 5′-AACC GGA TCC GGG AAA CAG CAA CTG CTA-3′ (Table 1) were synthesized and used to amplify the putative promoter region (GSP) using polymerase chain reaction (PCR). The PCR reactions were carried out in a total volume of 25 μl that contained 200 ng of template DNA, 10 mM Tris–HCl (pH 8.4), 50 mM KCl, 1.5 mM MgCl$_2$, 0.2 μM each of forward and reverse primer, 200 μM of each dNTP, and 1.0 U of Taq DNA polymerase (Invitrogen BioServices India Pvt. Ltd). PCR was performed in a programmable thermal cycler (Eppendorf) with initial denaturation at 95 °C for 5 min followed by 35 cycles of denaturation for 1 min at 95 °C, annealing for 1 min at 59.1 °C and extension for 1 min at 72 °C, with a final extension for 10 min at 72 °C. The amplified product (~523 bp) was fractionated on 1 % agarose gel and purified using the NucleoSpin Gel Elution Kit (Bioserve Biotechnologies, India Pvt. Ltd.). Eluted bands of the PCR product were ligated into pCR-Blunt-II-TOPO vector by using Zero Blunt® TOPO® PCR Cloning Kit (Invitrogen BioServices India Pvt. Ltd.) followed by blue-white selection [12]. Plasmids isolated from the white colonies were confirmed by restriction digestion analysis using *Eco*RI and/or *Bam*HI followed by sequencing. Orientation of the

Table 1 Primers used this study

| Primer	Sequence	Amplicon size (bp)
GSP FP	5′-AAC CGG ATC CAG CTT TAA TAG CAA CTA GGC-3′	523
GSP RP	5′-AACC GGA TCC GGG AAA CAG CAA CTG CTA-3′	
GusFp	5′-TGA TCA GCG TTG GTG GGA AAG-3′	1,213
GusRp	5′-TTT ACG CGT TGC TTC CGC AG-3′	

promoter fragment was confirmed by restriction digestion with *Sph*I and *Hinc*II. The sequence has been submitted to NCBI GenBank as HM215006.

Sequence Analysis

Nucleotide sequences obtained after sequencing were analyzed using NCBI BLAST analysis [15] and "gene tool" software such as Gibb's sampling [16], Melina software [17], and multiple expectation maximization for motif elicitation [18, 19] for the presence of the promoter motifs. The GSP promoter sequence was also analyzed using various database search programs such as PlantCARE database [20, 21] and Genomatix-MatInspector software based on PLACE database [22].

Construction of Plant Expression Vectors

To confirm promoter activity in the plant system, a binary vector pPZP200>GSP:GUS (Fig. 1b) was constructed by replacing the single CaMV 35S promoter of pPZP200>35S:GUS (Fig. 1a) with the GSP promoter region of peanut at the *Bam*HI site. Subsequently, the complete cassette containing *uid*A gene driven by the GSP promoter region was subcloned into the binary vector pCAMBIA2300 using *Eco*RI and *Hind*III restriction sites, thereafter referred to as pCAMBIA2300>GSP:GUS (Fig. 1c). Restriction with *Sph*I enzyme was done to ensure that the promoter was cloned in correct orientation upstream of the *uid*A gene. These recombinant binary vector plasmids, pPZP200>35S:GUS and pCAMBIA2300>GSP:GUS, were mobilized
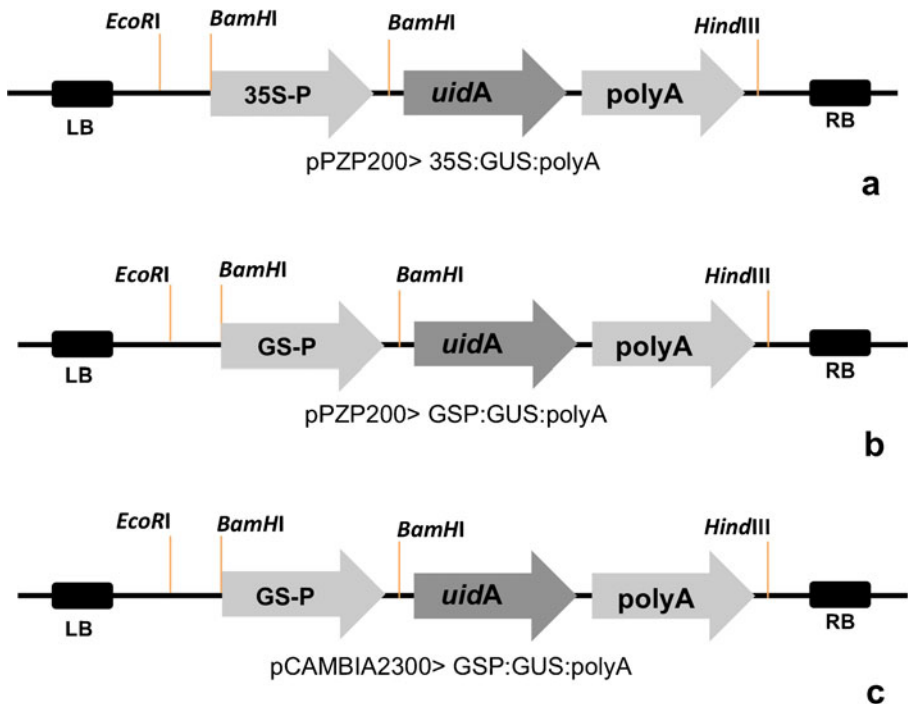
Fig. 1 Schematic representation of T-DNA region of the binary vector constructs used in this study. **a** pPZP200>35S:GUS:polyA. **b** pPZP200>GSP:GUS:polyA. **c** pCAMBIA2300>GSP:GUS:polyA

into disarmed *A. tumefaciens* strain C58 after confirmation with restriction analyses, and glycerol stocks were stored at −80 °C until further use.

Preparation of the Bacterial Culture for Agroinfection

Primary culture of *A. tumefaciens* strain C58 harboring the binary plasmids was prepared by inoculating a single colony of *Agrobacterium* in 20 ml YEB medium (with 50 mg/l each of kanamycin and rifampicin for pCAMBIA2300>GSP:GUS and 50 mg/l spectinomycin for pPZP200>35S:GUS) and grown overnight at 28 °C at 200 rpm. For floral dip transformation of *Arabidopsis*, the overnight culture (~10 %) was added to 20 ml of fresh medium with the same antibiotic and grown to the stationary phase (OD_{600} ~2.0). Cells were harvested by centrifuging at $5,500 \times g$ for 20 min, and the pellet was resuspended in 0.5× Murashige and Skoog (MS) [23], 5 % sucrose, and 0.05–0.1 % Teepol® to obtain the desired density (OD_{600} ~2.0). For tobacco transformation, 5 ml of the overnight-grown culture was pelleted at $5,500 \times g$ for 10 min, the supernatant discarded, and the pellet was resuspended in 0.5× MS so as to dilute it to an OD_{600} ~0.5. This suspension was used for the cocultivation of tobacco leaf discs using *Agrobacterium*-mediated transformation.

Production of Transgenic Plants of *Arabidopsis*

Seeds of *Arabidopsis thaliana* (Col-1) were sown in sand/soil (1:1) mixture in 4-cm pots and kept in the culture room until germination. Plants at the four-leaf stage were transferred to the greenhouse and irrigated every 4 days until inflorescences appeared. Plants with inflorescences of about 5 cm were transformed with suspension cultures of *A. tumefaciens* harboring the binary plasmids carrying GSP or CaMV 35S promoter fragments, using floral dip protocol of Clough and Bent [24]. Plants were inoculated by direct drop-by-drop inoculation to every flower by using a micropipette [25] and covered with plastic bags and incubated in dark for 10–24 h. Inoculation with the *Agrobacterium* was repeated twice at 3-day intervals and the seeds collected when all the siliques dried.

Production of Transgenic Plants of Tobacco

Tobacco (*Nicotiana tabacum* L., var. Xanthi) seedlings were grown in Magenta boxes on 0.5× MS media under sterile-controlled environment condition for 2 weeks followed by transfer to the greenhouse. *Agrobacterium*-mediated transformation of tobacco was carried out using standard leaf disc method [26] with some modifications. The fully expanded leaves were surface sterilized by two–three sequential treatments with 70 % ethanol for 30 s followed by wiping with sterile tissue paper. These were further washed with 15 % Clorox solution for 10 min and then washed thrice with sterile water. The leaves were then cut with a sterile leaf disc borer and cultured in 9-cm diameter plastic petri dishes containing ~20 ml of MS4 medium that contained MS medium [23] supplemented with 10 μM BAP, 0.5 μM NAA, 30 g/l sucrose, and 8 g/l Bacto agar (HiMedia Laboratories Pvt. Ltd., India) at pH 5.8. The leaf discs were dipped in *Agrobacterium* inoculum that facilitated the adhesion of bacteria to the cut ends of the leaves, and then, the leaf discs were transferred to the same media with their abaxial surface in contact with the medium. Ten to twelve cocultivated leaf disc explants were plated per petri plate, sealed with parafilm, and incubated at 26±1 °C under continuous light of 100 $\mu Em^{-2} s^{-1}$ for 72 h in a 16:8 light/dark regime. At the end of this period, explants were transferred onto MS4C medium (MS4 medium supplemented with 250 mg/l cefotaxime) and subcultured onto fresh MS4C medium every 2-week interval until

shoot regeneration. The explants containing regenerated shoot buds with the plasmid pCAMBIA2300>GSP:GUS construct (containing the *npt*II gene) were subjected to selection with 50 mg/l kanamycin. The explants were subcultured onto fresh MS medium for 35–40 days at 2-week intervals for shoot elongation and rooting. The rooted shoots were transferred to the pots containing autoclaved sand and soil (1:1) mixture and maintained in a containment glasshouse until flowering and seed set.

Molecular Characterization of Putative Transgenic Plants

Genomic DNA was isolated from the leaves of T_0 and T_1 generation transgenic plants of tobacco by using the modified CTAB method [13]. PCR was set up in a total volume of 25 μl containing 10–20 μg of template DNA for amplification of the 1,213-bp *uid*A gene fragment using primers GusFp 5′-TGA TCA GCG TTG GTG GGA AAG-3′ and GusRp 5′-TTT ACG CGT TGC TTC CGC CAG-3′. The PCR conditions included initial denaturation for 5 min at 95 °C, followed by 35 cycles of denaturation for 1 min at 95 °C, annealing for 90 s at 58.8 °C, and extension at 72 °C for 90 s followed by final extension for 10 min at 72 °C. PCR products were fractionated on 1 % agarose gel. Similarly, RT-PCR analysis was carried out to confirm the integration of the *uid*A gene using the Thermoscript RT-PCR system (Invitrogen, Carlsbad, CA, USA) on total leaf RNA isolated using the TRIzol® reagent (Invitrogen, USA) and from seed tissues using RNA isolation kit (MACHEREY-NAGEL, Germany). The primer sequences for the GUS transcripts were same as those described for the PCR analysis.

EMSA

Nuclear proteins were isolated from the seeds of peanut using NE-PER Nuclear and Cytoplasmic Extraction Reagent Kit (Thermo Fisher Scientific India Pvt. Ltd., Mumbai, India). PCR-amplified product of promoter fragment of GSP was end labelled with biotin 3′ End DNA Labeling Kit (Thermo Fisher Scientific India Pvt. Ltd., Mumbai, India) and used as a probe. In vitro DNA-protein binding assay was carried out as described by Light Shift Chemiluminescent EMSA kit (Thermo Fisher Scientific India Pvt. Ltd., Mumbai, India) by combining solutions of freshly isolated nuclear proteins (~5–10 μg) and 3′ biotin-labelled nucleic acid fragments. The resulting binding mixtures were fractionated by electrophoresis on 0.8 % agarose gel for the GSP fragment. These were then transferred to Hybond-N+ nylon membrane (GE Healthcare, New Jersey, USA) and developed according to manufacturer's instructions.

GUS Assays

The harvested seeds of tobacco and *Arabidopsis* were subjected to GUS assays using X-gluc (5-bromo-4-chloro-3-indolyl-β-glucuronide in 20 mM sodium phosphate buffer (pH 7.2), 0.1 % Triton X-100, 10 mM EDTA, and 5 mM potassium ferrocyanide) as the substrate with overnight incubation at 37 °C [27]. To confirm the β-glucuronidase enzyme-specific activity in *Arabidopsis* and tobacco transformants, GUS assay was carried out in different tissues like seeds, maternal tissues, cotyledons, stem, root, leaves, and flower. For the GUS assay, tissue samples were collected in 1.5-ml Eppendorf tubes and treated with 500-μL GUS assay solution followed by incubation for 16–24 h at 37 °C. While the tissues were cleared with serial transfers in 70 % alcohol, the final samples showing blue coloration were photographed.

The fluorometric assay for specific GUS enzyme activity was quantified by measuring the hydrolysis rate of the fluorogenic substrate 4-methylumbelliferyl β-D-glucuronide (MUG) (HiMedia Laboratories Pvt. Ltd.) as earlier [28]. Standards were prepared in

different concentrations, i.e., 1 mM, 1 µM, and 100 nM of 4-methylumbelliferone sodium salt (HiMedia Laboratories Pvt. Ltd.) in 0.2 M sodium carbonate. 4-MUG (2 mM) was added to each sample as the substrate.

Seed extracts prepared with GUS extraction buffer (50 mM sodium phosphate, pH 7.0; 10 mM β-mercaptoethanol; 10 mM EDTA, and 0.1 % Triton X-100) were used for histochemical analysis. Plant materials (seeds, flower, and leaf) were vigorously ground to a finely pulverized powder with a pestle and mortar under liquid nitrogen in 500 µL GUS extraction buffer. The extract thus obtained was centrifuged at 8,000×g for 5 min at 4 °C. The supernatant was recovered and 50 µL of extract was added to 950 ml of 4-MUG assay buffer (2 mM) to initiate the reaction. The reaction was stopped by adding 200 µL of the reaction to 1.8 ml of 0.2 M Na_2CO_3 stop buffer at intervals of 0 min, 30 min, 60 min, and overnight, and fluorescence was measured using a DyNA Quant 200™ fluorometer (Hoefer Scientific Instruments, San Francisco, CA), following the manufacturer's instructions. The protein concentrations of each sample were determined [29] with a spectrophotometer (Shimadzu, UV-1650PC) at OD_{595} using Quick Start Bradford Protein Assay kit (BioRad), and the GUS enzyme activity was expressed as picomoles of 4-methylumbelliferone produced per milligrams of protein per minute.

Results

Sequence Analysis of GSP Fragment

Based on the results of BLAST analysis using nucleotide–nucleotide homology, the GSP promoter fragment sequence revealed similarity with aquaporin genes/tonoplast intrinsic proteins of various organisms and plants. The in silico promoter analysis revealed the presence of major motifs (Table 2) which confer seed-specific expression. The GSP promoter sequence contained transcription initiation site and some basic promoter elements such as two CAAT boxes at −287 to −283 and −233 to −229 positions and TATA box (two copies) at −469 to −462 and −479 to −474. Computational analysis of GSP sequence revealed many important promoter motifs including two AT-rich elements at −470 to −462 and −493 to −485 positions and TGCA motif at three locations (−54 to −51, −150 to −147, and −285 to −282). The RY repeat element (CATGCA) was found at −152 to −147, whereas the G box (CACGTG) was absent. A W box element TGACT that has been identified for its role in wound-induced expression was also found at −101 to −97. Similarly, GCN4 motif was present at −107 to −101 positions in the GSP promoter fragment. These observations indicated that the presence of multiple copies of the seed-specific promoter motifs in the GSP fragment may be responsible for its seed-specific promoter activity.

Transformation Studies in Tobacco

All the tobacco shoots transferred to half MS medium rooted well and were acclimatized followed by their transfer to the contained greenhouse. The plants were normal in their morphology and were comparable to the untransformed tobacco plants for all phenotypic characters.

Molecular Characterization of Putative Transgenics

A total of 25 putative transgenic tobacco plants were developed and transferred to soil. The transgenic nature of the primary transformants was confirmed with PCR using the GSP as

Table 2 Putative *cis*-acting regulatory elements identified by in silico analysis in the GSP sequence

Cis element	Consensus	Motif present in GSP sequence	Motif position	Reference
Frequent elements				
TATA box	TATAWAWA	TATAAATA	−469 to −462	[30]
	TATA	TATATA	−479 to −474	
CAAT box	CAAT	ATTG	−287 to −283	[30]
		Complimentary	−233 to −229	
Seed-specific elements				
AACA motif, endosperm specific	AACAWDD	AACAAGA	−179 to −173	[31–33]
		AACATTT	−412 to −406	
		AACAAAG	−457 to −451	
ACGT-containing element endosperm specific (ACE)	NACGTN (TGCA complimentary)	GTGCAG	−53 to −50	[31, 34]
		ATGCAG	−149 to −146	
		TTGCAG	−284 to −281	
E box, storage protein, oilrape	CANNTG	CAAGTG	−297 to −292	[35, 36]
		CAAGTG	−203 to −198	
		CAGGTG	−58 to −53	
		CAGTTG	−25 to −20	
GCN4 motif, endosperm specific, seed, storage protein	TGAGTCA	TGAGTGT	−107 to −101	[31, 32, 37]
SEF4 binding site, soybean embryo factor, seed	RTTTTTR	ATTTTTG	−83 to −77	[38]
		GTTTTTG	−257 to −251	
SEF1 binding site, soybean embryo factor, seed	ATATTAWW	ATATTATA	−436 to −429	[38]
Sh1 box, maize	TGAATG	TGAATC	−207 to −202	[39]
RY element	CATGCA	CATGCA	−152 to −147	[40–42]
AT-rich region	AT-rich region	ATATAAATA	−470 to −462	[35, 43]
		AAATTAAAT	−493 to −485	
Motif required for transcriptional regulation and seed-specific expression	AG/CCCCA	AGCTCTT	−51 to −45	[44]
		AGCTCCA	−305 to −299	
Motif responsible for activating seed-storage protein expression with one mismatch	TACACAT complimentary ATGTGTA	TACACTG	−213 to −207	[45]
		TACACTC	−277 to −271	
		TACAGAT	−475 to −469	
		ATGTTAA	−141 to −135	
		ATGTCTT	−173 to −167	
Cis element involved in seed-specific expression	AAGAA	AAAGAA	−335 to −330	[34, 46]
Other elements				
ABRE–GARE elements	ACACNNG	CTGGTGT	−243 to −250	[47, 48]
Binding core bZIP, ABA response GARE	TAACAAA/G	(complimentary) TAACAAG	−180 to −174	
Pyrimidine box	TTTTTCC	TTTTTGGC	−82 to −75	[49]
		TTTTTGC	−256 to −250	
		TTTTGG	−362 to −357	
CTCTT	CTCTT	CTCTT	−49 to −45	[50]
W box element (wound-induced expression)	TGACD	TGACA	−63 to −60	[51, 52]
		TGACT	−101 to −97	
TGCA motif	TGCA	TGCA	−54 to −51	[53]
			−150 to −147	
			−285 to −282	

Table 2 (continued)

Cis element	Consensus	Motif present in GSP sequence	Motif position	Reference
CATG containing motif	CATG	CATG	−152 to −149	[52]
TGAC-like motif	TGAGTCATCA	TGAGTCTGAC	−107 to −98	[54]
		TGAGGCTTCC	−380 to −371	

well as *uid*A gene-specific primers. Majority of the primary transformants showed the presence of an expected 1,213 bp *uid*A gene amplicon (Fig. 2). These PCR-confirmed transgenics in the T_1 generation were also confirmed for expression of the *uid*A transgene using the semiquantitative RT-PCR analysis that proved the gene integration and stable expression over generations (Fig. 3).

Gel Shift Assay (EMSA)

To examine binding of the nuclear proteins to the regulatory elements of the peanut seed (GSP), gel mobility shift assay was carried out using nuclear extracts from different plant tissues. The biotin-labelled PCR amplicon of GSP was used as DNA probe. The EMSA of GSP fragment exhibited stronger affinity showing distinctly shifted bands with peanut seed nuclear extracts, compared to the nuclear extracts of other tissues and GSP promoter DNA probe alone (Fig. 4, lanes 2, 3, and 4).

Histochemical and Fluorometric GUS Analysis

*uid*A gene expression was observed in both *Arabidopsis* and tobacco transformants obtained following the *Agrobacterium* floral dip and modified leaf disc methods, respectively. While the seeds of the untransformed controls did not take up the histochemical stain, transgenic seeds showed a distinct blue color. Moreover, a localized GUS expression was observed in seeds of transgenic *Arabidopsis*s and tobacco transformed with pCAMBIA2300>GSP:GUS and pPZP200>35S:GUS constructs (Fig. 5b, c, e, f), whereas no GUS staining was observed in other plant parts such as leaves, roots, and shoots.

The fluorometric analysis of leaves, flower, and seeds from the transgenic tobacco plants in T_1 generation along with their untransformed counterparts indicated fluorescence activity in transgenic tobacco plants (Fig. 6a, b). The histochemical and fluorometric results clearly confirmed the observations of gel retardation assays, thereby indicating that the *uid*A gene driven by GSP promoter was specifically expressed in the seed tissues of the transgenic *Arabidopsis* and tobacco plants.

Discussion

Tissue-specific expression of a gene is regulated by the presence of specific motifs in DNA sequence of the promoter region. In the present study, we observed two TATA boxes and one CAAT box in the promoter region of groundnut seed-specific promoter (GSP). These results are in accordance with the previous study, where the number and spacing of TATA box elements have been reported to be essential for high levels of transcription initiation in phaseolin gene in the developing embryos [55], potentially affecting the tissue-specific
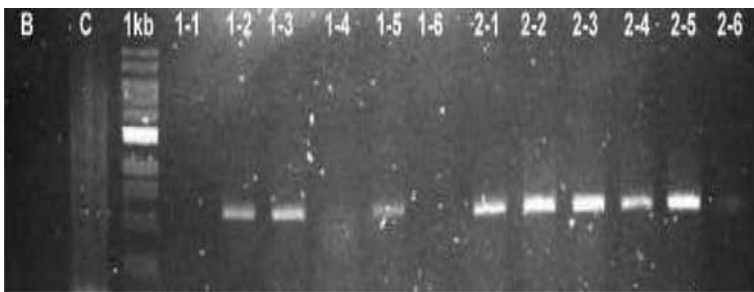
Fig. 2 PCR analysis of the genomic DNA of T1 generation transgenics of tobacco transformed with the binary plasmid pCAMBIA2300>GSP:*uidA*:polyA amplifying the 1,213-bp fragment of *uidA* (GUS) gene. *Lane B* blank; *Lane C* untransformed tobacco control

expression [56]. Interestingly, GSP promoter fragment contained CTCTT motif at −49 to −45 position that might have contributed to the seed-specific expression. A nodule lectin gene promoter reportedly also had CTCTT motif instead of TATA and CAAT elements [50].

The sequence analysis of the GSP promoter region fragment revealed the presence of several seed-specific promoter motifs such as A/T-rich motifs that are important for binding of transcriptional factors and also required for seed-specific expression [35]; RY repeat element (or legumin box), which is also a sequence motif required for seed-specific expression [42]; AGCCCA motif, which is a sequence motif required for transcriptional regulation and seed-specific expression [44]; TACACAT motif with one base to mismatch which is responsible for activating seed storage protein expression [45]; E-box (CANNTG) in the presence of other seed-specific promoter motifs which help in the activation of seed-specific promoters for heterologous expression [36]; and ACGT motif that is also required for seed-specific expression [34].

The RY repeat element (CATGCA) was found at −152 to −147 positions in GSP fragment. This RY element (CATGCA (C/T) [57] is widely distributed in seed-specific promoters of monocots and dicots, including the legumin and USP genes of *Vicia faba* [58, 59], the napin genes of *Brassica napus* [60, 61], and the maize C1 gene [62] and is also known as the legumin box [63] or the Sph element [62]. These conserved RY repeats of seed-specific genes are important regulatory factors for seed and plant development [41] and are also one of the important conserved motif in seed storage proteins of three plant families including Brasssicaceae, Fabaceae, and Poaceae [64], besides being present in promoter (−500 to −1000) regions of genes regulating the embryonic pathways [65].
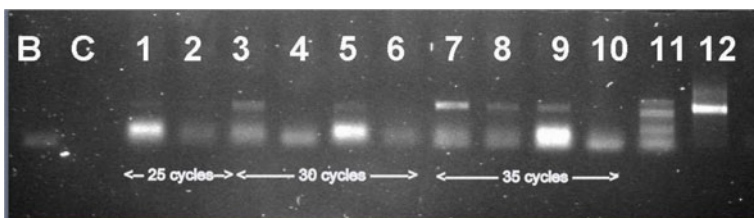

Fig. 3 Semiquantitative RT-PCR of transgenic tobacco plants expressing *uid*A gene driven by legume seed-specific promoters. *Lanes 1, 3, 5, 7*, and *9* carry cDNA from seeds of pCAMBIA2300>GSP:GUS. *Lanes 2, 4, 6, 8*, and *10* carry cDNA from the leaves of pCAMBIA2300>GSP>GUS. *Lane 11* carries 100 bp ladder. *Lane 12* has plasmid as positive control. *B* and *C* represent blank and untransformed control
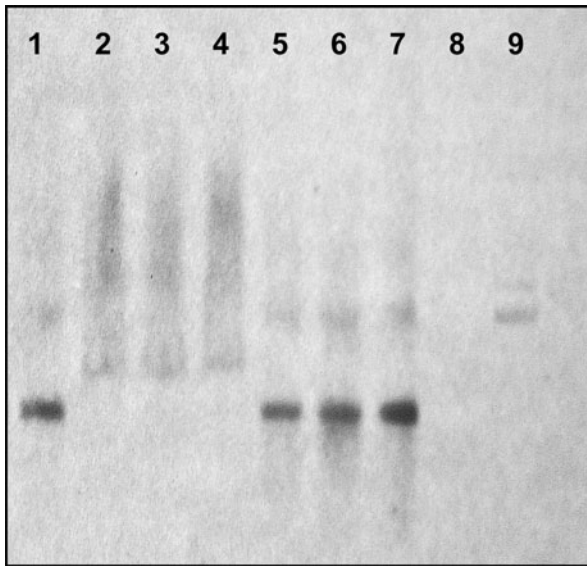
Fig. 4 EMSA for the confirmation of promoter region from peanut-groundnut seed-specific promoter (GSP) binding assay on 0.8 % agarose gel. *Lane 1* contains unbound GSP fragment. *Lanes 2–4* carry GSP fragment bound with peanut seed nuclear extracts in the presence of EDTA and KCl. *Lanes 5–7* carry GSP fragment bound with leaf, immature seed, and testa extracts, respectively. *Lane 8* is blank and the *lane 9* carries the 100-bp ladder

It has also been reported that CATG-containing motif is enriched in early/transient genes in seed development, and CATG interacting factors may control physiological responses directly affecting pathogen viability, such as camalexin biosynthesis and cell wall modifications [52].

In the GSP fragment, AACA motifs were present at −179 to −173, −412 to −406, and −457 to −451 positions. Previous report by Takaiwa et al. [33] also details the importance of AACA motifs in driving high level of gene expression in tobacco endosperms. Similarly, two AT-rich regions that were observed in GSP fragment regulated seed-specific expression, and the CATGCAT/A sequence did not affect the concavallin expression in seeds [43].
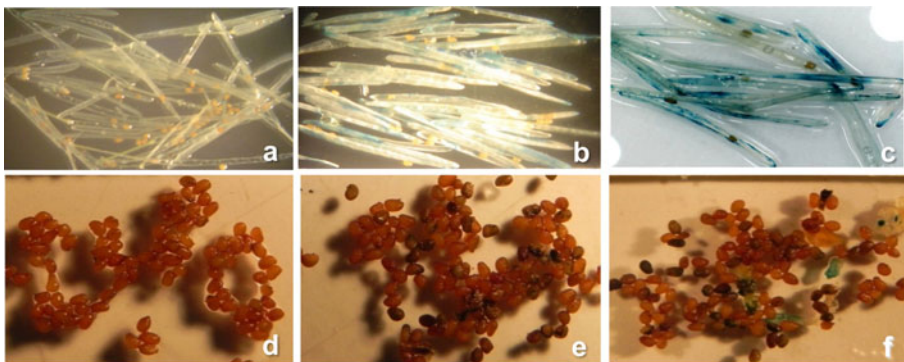
Fig. 5 Validation of seed-specific promoters in transgenic *Arabidopsis* (**a–c**) and tobacco (**d–f**). **a, d** Histochemical GUS assay in untransformed control showing negative. **b, e** GUS expression in transgenic seeds carrying the plasmid pPZP200>35S:GUS:polyA. **c, f** GUS expression in transgenic seeds carrying the plasmid pCAMBIA2300>GSP:GUS:polyA
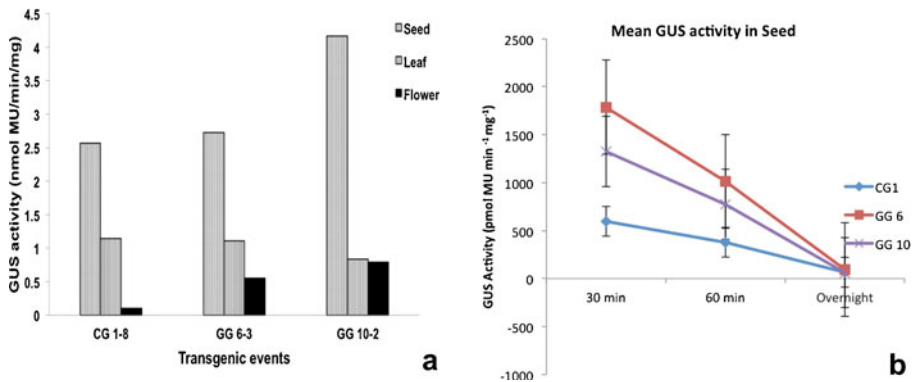
Fig. 6 Expression patterns of transgenics expressing seed-specific promoter GSP based on GUS activity using MUG assay. **a** Comparative expression patterns in seed, leaf, and flower tissues. **b** Mean GUS activity ± SE measured at different time intervals (30 min, 60 min, and overnight assay) with five replicates in the seeds of independent transformants with promoter construct (GG = GSP)

In the present study, we also found TGAG occurring twice in groundnut seed (GSP) promoter region. TGAC-like motifs, including a TGAGTCATCA region in a 22 bp, have also been observed to play a crucial role in seed-specific expression of a pea seed lectin [54]. Similarly, the conserved GCN4 motif that has been reported to be involved in controlling seed-specific expression in a number of seed-storage protein genes [46] was also present at −107 to −101 in the GSP fragment.

The BLAST analysis of GSP fragment revealed similarity with many tonoplast intrinsic proteins (TIPs) and aquaporin genes. Aquaporin α-TIP is specifically expressed in the membrane of protein storage vacuoles in seeds of many plant species [66]. Studies showing the late accumulation of α-TIP during seed maturation and its disappearance during germination and seedling growth [67–69] suggest that the protein may play a key role at the early stages of seedling growth. Rice TIP expression patterns under various abiotic stress conditions including dehydration, high salinity, abscisic acid (ABA), and seed germination were investigated by real-time PCR analysis, where OsTIP1s (OsTIP1;1 and OsTIP1;2) were found to be highly expressed during seed germination, whereas OsTIP3s (OsTIP3;1 and OsTIP3;2) specifically expressed in mature seeds with a decrease in expression levels upon germination [70]. Besides, the TGACG motif, which is a W box element important for high-level gene expression of *A. tumefaciens* T-DNA octopine synthase (*ocs*) [71], nopaline synthase (*nos*) [72], CaMV 35S [73], and rice *GOS2* [74], was also found to be present within the GSP.

Based on the above results and the literature available, it is indicated that groundnut seed-specific (GSP) promoter fragment possesses promoter activity and promotes the seed-specific expression of heterologous gene. The molecular characterization studies such as PCR and RT-PCR analyses of the tobacco transgenics containing pCAMBIA2300>GSP:GUS in T_0 and T_1 generations confirmed the stable integration of the transgene (*uid*A) over generations. The presence of three TGCA motifs observed at −54 to −51, −150 to −147, and −285 to −282 positions in the GSP fragment is crucial since the number and spacing of TGCA motif reportedly affect the binding of nuclear DNA binding protein [53] which substantiates the strong binding affinity of the promoter with the seed nuclear proteins.

The histochemical staining of seeds of stable transgenic *Arabidopsis* and tobacco plants carrying GSP-*uid*A construct further supports the tissue specificity of this promoter region. These results are in line with the previous reports where *uid*A expression was reported in the

developing seeds of transgenic tobacco within 12–21 days after flowering [75]. Quantitative fluorometric studies indicated a higher GUS activity in the seeds of transgenic plants when compared to the untransformed plants.

Our study attempted isolation, cloning, and sequencing of a groundnut seed-specific promoter region from peanut. In silico characterization of the promoter region was done followed by functional characterization in the heterologous model plant *A. thaliana* and tobacco plants. The sequence of the groundnut seed-specific promoter region was deposited in GenBank with accession no. HM215006.

In conclusion, the seed-specific sequence motifs, binding assay with nuclear proteins of seeds by gel shift data, and molecular characterization followed by *uid*A reporter gene expression studies of the transgenic *Arabidopsis* and tobacco plants support the potential of the groundnut seed promoter for seed-specific transgene expression. This seed-specific promoter region successfully targeted the transgene to the seed tissues that were confirmed by histochemical and fluorometric analysis and also by electrophoretic mobility shift assay. This promoter sequence can be potentially used for modification of seed phenotypes in agronomically important crops, and further promoter function characterization might be useful to elucidate its role for enhanced transcriptional activity.

References

1. Odell, J. T., Nagy, F., & Chua, N. H. (1985). *Nature, 313*, 810–812.
2. Christensen, A. H., Sharrock, R. A., & Quail, P. H. (1992). *Plant Molecular Biology, 18*, 675–689.
3. Cornejo, M. J., Luth, D., Blankenship, K. M., Anderson, O. D., & Blechl, A. E. (1993). *Plant Molecular Biology, 23*, 567–581.
4. Hsieh, T. H., Lee, J. T., Charng, Y. Y., & Chan, M. T. (2002). *Plant Physiology, 130*, 618–626.
5. Kasuga, M., Liu, Q., Miura, S., Yamaguchi-Shinozaki, K., & Shinozaki, K. (1999). *Nature Biotechnology, 17*, 287–291.
6. Park, S. H., Yi, N., Kim, Y. S., Jeong, M. H., Bang, S. W., Choi, Y. D., & Kim, J. K. (2010). *Journal of Experimental Botany, 61*, 2459–2467.
7. Zheng, C., & Baum, B. J. (2008). *Methods in Molecular Biology, 434*, 205–219.
8. Kasuga, M., Miura, S., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2004). *Plant Cell Physiology, 45*, 346–350.
9. Lee, J. T., Prasad, V., Yang, P. T., Wu, J. F., Ho, T. H. D., Charng, Y. Y., & Chan, M. T. (2003). *Plant Cell and Environment, 26*, 1181–1190.
10. Devi, I. S., Dinesh Kumar, V., Ansari, N. A., & Sivasankar, A. (2010). *International Journal of Environmental Science and Development, 1*, 20–23.
11. Zhuang, W. J., Zhao, Y-Li., Chen, H., & Zheng, Y. (2008). In: 3rd International Conference for Peanut Genomics and Biotechnology on Advances in Arachis through Genomics and Biotechnology (AAGB-2008), ICRISAT, Hyderabad (AP), India; 4–8 November, pp 71.
12. Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular cloning: a laboratory manual*. New York: Cold Spring Harbor.
13. Porebski, S., Bailey, L. G., & Baum, B. R. (1997). *Plant Molecular Biology Reports, 15*, 8–15.
14. Rozen, S., & Skaletsky, H. (2000). In: Misener, S., & Krawetz S. A. (eds.) *Bioinformatics methods and protocols*: methods in molecular biology, Humana, Totowa, pp 365–386.
15. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). *Nucleic Acids Research, 25*, 3389–3402.
16. Lawrence, C. E., Altschul, S. F., Boguski, M. S., Liu, J. S., Neuwald, A. F., & Wootton, J. C. (1993). *Science, 262*, 208–214.
17. Poluliakh, N., Takagi, T., & Nakai, K. (2003). *Bioinformatics, 19*, 423–424.

18. Bailey, T. L., & Elkan, C. (1994). *Proceedings of the Second International Conference on Intelligent Sys for Molecular Biology AAAI Press, Menlo Park, California,* pp. 28–36.
19. Bailey, T. L., Williams, N. M. C., & Li, W. W. (2006). *Nucleic Acids Research, 34,* 369–373.
20. Lescot, M., Dehais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouze, P., & Rombauts, S. (2002). *Nucleic Acids Research, 30,* 325–332.
21. Rombauts, S., Dehais, P., Montagu, M. V., & Rouze, P. (2002). *Nucleic Acids Research, 27,* 295–296.
22. Higo, K., Ugawa, Y., Iwamoto, M., & Korenaga, T. (1999). *Nucleic Acids Research, 27,* 297–300.
23. Murashige, T., & Skoog, F. (1962). *Physiologia Plantarum, 15,* 473–497.
24. Clough, S. J., & Bent, A. F. (1998). *Plant Journal, 16,* 735–743.
25. Trujillo, M. M., Briones, V. L., Ponce, J. L. C., & Estrella, L. H. (2004). *Plant Molecular Biology Reporter, 22,* 63–70.
26. Horsch, R. B., Fry, J., Hoffmann, N., Neidermeyer, J., Rogers, S. G., & Fraley, R. T. (1988). In: Gelvin, S. B., & Schilperoort, R. A. (Eds.) *Plant molecular biology manual.* Kluwer, Dordrecht, pp 1–9.
27. Jefferson, R. A. (1987). *Plant Molecular Biology Reports, 5,* 387–405.
28. Jefferson, R. A., Kavanagh, T. A., & Bevan, M. W. (1987). *EMBO Journal, 6,* 3901–3907.
29. Bradford, M. M. (1976). *Analytical Biochemistry, 72,* 248–254.
30. Forde, B., Heyworth, A., Pywell, J., & Kreis, M. (1985). *Nucleic Acids Research, 13,* 7327–7339.
31. Wu, C., Washida, H., Onodera, Y., Harada, K., & Takaiwa, F. (2000). *Plant Journal, 23,* 415–421.
32. Yoshihara, T., Washida, H., & Takaiwa, F. (1996). *FEBS Letters, 383,* 213–218.
33. Takaiwa, F., Yamanouchi, U., Yoshihara, T., Washida, H., Tanabe, F., Kato, A., & Yamada, K. (1996). *Plant Molecular Biology, 30,* 1207–1221.
34. Vincentz, M., Leite, A., Neshich, G., Vriend, G., Matter, C., Barros, L., Weinberg, D., de Almeida, E. R., de Carvalho, M. P., Aragao, F., & Gander, E. S. (1997). *Plant Molecular Biology, 34,* 879–889.
35. Stalberg, K., Ellerstom, M., Ezcurra, I., Ablov, S., & Rask, L. (1996). *Planta, 199,* 515–519.
36. Kawagoe, Y., & Murai, N. (1992). *Plant Journal, 2,* 927–936.
37. Muller, M., & Knudsen, S. (1993). *Plant Journal, 4,* 343–355.
38. Allen, R. D., Bernier, F., Lessard, P. A., & Beachy, R. N. (1989). *Plant Cell, 1,* 623–631.
39. D'Aoust, M. A., Nguyen-Quoc, B., Le, V. Q., & Yelle, S. (1999). *Plant Cell Reports, 18,* 803–808.
40. Bobb, A. J., Eiben, H. G., & Bustos, M. M. (1995). *Plant Journal, 8,* 331–343.
41. Bobb, A. J., Chern, M. S., & Bustos, M. M. (1997). *Nucleic Acids Research, 25,* 641–647.
42. Chamberland, S., Daile, N., & Bernier, F. (1992). *Plant Molecular Biology, 19,* 937–949.
43. Yamamoto, S., Nishihara, M., Morikawa, H., Yamauchi, D., & Minamikawa, T. (1995). *Plant Molecular Biology, 27,* 729–741.
44. Chen, Z. L., Schuler, M. A., & Beachy, R. N. (1986). *Proceedings of the National Academy of Sciences of the United States of America, 83,* 8560–8564.
45. Josefsson, L. G., Lenman, M., Ericson, M. L., & Rask, L. (1987). *The Journal of Biological Chemistry, 262,* 12196–12201.
46. Wu, C. Y., Suzuki, A., Washida, H., & Takaiwa, F. (1998). *Plant Journal, 14,* 673–683.
47. Kim, S. Y., Chung, H. J., & Thomas, T. L. (1997). *Plant Journal, 11,* 1237–1251.
48. Ogawa, M., Hanada, A., Yamauchi, Y., Kuwahara, A., Kamiya, Y., & Yamaguchi, S. (2003). *Plant Cell, 15,* 1591–1604.
49. Cercos, M., Gomez-Cadenas, A., & Ho, T.-H. D. (1999). *Plant Journal, 19,* 107–118.
50. Law, I. J. (2000). *Plant Science, 153,* 43–54.
51. Nishiuchi, T., Shinshi, H., & Suzuki, K. (2004). *Journal of Biological Chemistry, 279,* 55355–55361.
52. Eulgem, T., Weigman, V. J., Chang, H. S., McDowell, J. M., Holub, E. B., Glazebrook, J., Zhu, T., & Dangl, J. L. (2004). *Plant Physiology, 135,* 1129–1144.
53. Howley, P. M., & Gatehouse, J. A. (1997). *Plant Molecular Biology, 33,* 175–180.
54. De Pater, S., Pham, K., Chua, N. H., Memelink, J., & Kijne, J. (1993). *Plant Cell, 5,* 877–886.
55. Grace, M. L., Chandrasekharan, M. B., Hall, T. C., & Crowe, A. J. (2004). *Journal of Biological Chemistry, 279,* 8102–8110.
56. Kloeckner-Gruissem, B., Vogel, J. M., & Freeling, M. (1992). *EMBO Journal, 11,* 157–166.
57. Dickinson, C. D., Evans, R. P., & Nielsen, N. C. (1988). *Nucleic Acids Research, 16,* 371.
58. Baumlein, H., Wobus, U., Pustell, J., & Kafatos, F. C. (1986). *Nucleic Acids Research, 14,* 2707–2720.
59. Baumlein, H., Boerjan, W., Nagy, I., Panitz, R., Inze, D., & Wobus, U. (1991). *Molecular and General Genetics, 225,* 121–128.
60. Ellerstrom, M., Stalberg, K., Ezcurra, I., & Rask, L. (1996). *Plant Molecular Biology, 32,* 1019–1027.
61. Stalberg, K., Ellerstrom, M., Josefsson, L. G., & Rask, L. (1993). *Plant Molecular Biology, 23,* 671–683.
62. Kao, C. Y., Cocciolone, S. M., Vasil, I. K., & McCarty, D. R. (1996). *Plant Cell, 8,* 1171–1179.
63. Gatehouse, J. A., Evans, I. M., Croy, R. R. D., & Boulter, D. (1986). Philosophical transactions of the Royal Society of London. *Series B, Biological sciences, 314,* 367–384.

64. Fauteux, F., & Stromvik, M. V. (2009). *BMC Plant Biology, 9*, 126.
65. Suzuki, M., Wang, H. H. Y., & McCarty, D. R. (2007). *Plant Physiology, 143*, 902–911.
66. Maurel, C., Chrispeels, M. J., Lurin, C., Tacnet, F., Geelen, D., Ripoche, P., & Guern, J. (1997). *Journal of Experimental Botany, 48*, 421–430.
67. Johnson, K. D., Herman, E. M., & Chrispeels, M. J. (1989). *Plant Physiology, 91*, 1006–1013.
68. Melroy, D. L., & Herman, E. M. (1991). *Planta, 184*, 113–122.
69. Hoh, B., Hinz, G., Jeong, B. K., & Robinson, D. G. (1995). *Journal of Cell Science, 108*, 299–310.
70. Li, G. W., Peng, Y. H., Yu, X., Zhang, M. H., Cai, W. M., Sun, W. N., & Su, W. A. (2008). *Journal of Plant Physiology, 165*, 1879–1888.
71. Ellis, J., Llewellyn, D., Walker, J., Dennis, E., & Peacock, W. (1987). *EMBO Journal, 6*, 3203–3208.
72. Mitra, A., & An, G. (1989). *Molecular and General Genetics, 215*, 294–299.
73. Fang, R. X., Nagy, F., Sivasubramaniam, S., & Chua, N. H. (1989). *Plant Cell, 1*, 141–150.
74. De Pater, B. S., van der Mark, F., Rueb, S., Katagiri, F., Chua, N. H., Schilperoort, R. A., & Hensgens, L. A. (1992). *Plant Journal, 2*, 837–844.
75. Philip, R., Darnowski, D. W., Sundararaman, V., Cho, M. J., & Vodkin, L. O. (1998). *Plant Science, 137*, 191–204.