272 J. Agric. Food Chem. 2010, 58, 272–281
DOI:10.1021/jf902970p

JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY
ARTICLE

Characterization of the 11S Globulin Gene Family in the Castor Plant *Ricinus communis* L.

Tarik Chileh, Belén Esteban-García, Diego López Alonso, and Federico García-Maroto*

Grupo de Biotecnología de Productos Naturales (BIO279), Facultad de Ciencias Experimentales, Universidad de Almería, Campus de La Cañada 04120, Spain

The 11S globulin (legumin) gene family has been characterized in the castor plant *Ricinus communis* L. Phylogenetic analysis reveals the presence of two diverged subfamilies (*RcLEG1* and *RcLEG2*) comprising a total of nine genes and two putative pseudogenes. The expression of castor legumin genes has been studied, indicating that it is seed specific and developmentally regulated, with a maximum at the stage when cellular endosperm reaches its full expansion (around 40−45 DAP). However, conspicuous differences are appreciated in the expression timing of individual genes. A characterization of the 5′-proximal regulatory regions for two genes, *RcLEG1-1* and *RcLEG2-1*, representative of the two legumin subfamilies, has also been performed by fusion to the *GUS* reporter gene. The results obtained from heterologous expression in tobacco and transient expression in castor, indicating seed-specific regulation, support the possible utility of these promoters for biotechnological purposes.

KEYWORDS: 11S globulins; legumins; castor bean; *Ricinus communis*; seed development; storage proteins; crystalloid proteins

INTRODUCTION

The castor plant (*Ricinus communis* L.) is an important oil-seed crop that produces a very unique oil, enriched in ricinoleic acid (12-hydroxyoleate). Castor oil commonly comprises up to 50−60% of the seed weight, making it one of the highest yielding oil-seed crops (*1*). Castor oil contains over 85% of ricinoleate, conferring to it adequate technological properties for diverse industrial uses, serving as a source for paints and varnishes, nylon-type plastics, hydraulic fluids and lubricants, cosmetics, fungicides, etc. (*1*). Both composition and high oil content of the castor seed also makes it an attractive source for biodiesel production (*2*).

However, the castor bean represents both a potent occupational sensitizer and an unusual environmental allergen, which pose health threats for workers during plant cultivation and seed processing (*3*). 11S globulins are recognized as common strong allergens present in the seeds of diverse plant species, such as sesame, peanut, soya, treenut, hazelnut, cashew nut, mustard, etc. (*4*). Accordingly, 11S globulin proteins have also been identified as one of the allergenic components in castor seeds, besides the 2S albumin proteins (*5*).

The 11−13S globulins, also called legumins, form the major storage reserve of dicotyledonous seeds (*6*). They were first characterized in Fabaceae but are widespread through both angiosperms and gymnosperms (*6*). In all species investigated so far, legumin genes are organized as a small gene family which can be divided into several subfamilies (*6*).

*To whom correspondence should be addressed. Tel: +34950015033. Fax: +34950015008. E-mail: fgmaroto@ual.es.

Legumins are specifically expressed in the storage tissues of seeds, i.e., the endosperm or the embryo (*7*). In the castor bean, the endosperm forms the major storage tissue, surrounding the thin papery cotyledons of a centrally located embryo (*8*). 11S globulins are localized in the protein bodies as highly insoluble inclusions called crystalloids, making up some 75% of the protein reserve in the mature castor seed endosperm (*9*). The other important reserve proteins, the 2S albumins, are mainly contained in the matrix of protein bodies besides the 7S lectins including the toxic ricine (*9*).

Previous biochemical work showed that similar to other plant species, castor bean contains at least six legumin isoforms that are assembled, in a random combination, to form hexamers of ca. 330 kDa (*10*, *11*). Each monomeric isoform consists of a large acidic subunit or α-chain (29−34 kDa isoforms) and a small basic subunit or β-chain (20.5−23.5 kDa) linked by a single disulfide bond (*11*). The two subunits are derived from a single precursor protein, which is cleaved post-translationally.

Although genes encoding components of the castor seed such as lectins and 2S albumins have been characterized to some detail (*12*, *13*), information about genomic structure, organization, and regulation of legumins is still missing for this plant. However, regulatory sequences of the storage protein genes are valuable tools because of the practical utility of seed specific promoters in crop improvement, production of valuable proteins, or modification of seed storage compounds including oil composition.

Here, we report the study of the 11S globulin family in *R. communis*, along with a characterization of the regulatory regions from representative genes of two legumin subfamilies.

 Published on Web 11/12/2009

Table 1. Legumin Gene Family in *R. communis*[a]

gene name	locus	GenBank acc. no.	genomic scaffold	TIGR transcript assemblies	exon length (nts)	intron length (nts)
RcLEG1-1	EEF38213	gil223536537	scf_1106159304412	TA1039_3988	292, 245, 498, 396	108, 103, 81
RcLEG1-2	EEF38210	gil223536537	scf_1106159304412	TA1037_3988 TA1030_3988	292, 245, 498, 396	107, 103, 81
RcLEG1-3	EEF37813	gil223536104	scf_1106159292040	TA1035_3988	289, 245, 510, 393	110, 112, 97
RcLEG1-4	EEF37814	gil223536104	scf_1106159292040	TA1043_3988	289, 245, 510, 393	110, 112, 97
RcLEG1-5	EEF30344	gil223528296	scf_1106159302472	TA1038_3988	277, 236, 486, 387	102, 240, 101
RcLEG1-6	EEF37812	gil223536104	scf_1106159292040	TA1046_3988	292, 245, 501, 390	78, 74, 95
RcLEG2-1	EEF28918	gil223526676	scf_1106159306026	TA1029_3988 TA1032_3988	328, 260, 507, 378	76, 97, 105
RcLEG2-2	EEF28917	gil223526676	scf_1106159306026	TA1027_3988	316, 260, 507, 408	94, 99, 76
RcLEG2-3	EEF39577	gil223537958	scf_1106159305582	N.A.	1 nt deletion in exon 1; PTC in exon 2	
RcLEG2-4	EEF32199	gil223530296	scf_1106159304964	TA2458_3988	322, 251, 489, 381	112, 93, 112
RcLEG2-5	EEF35875	gil223534137	scf_1106159298126	N.A.	19 nts deletion in exon 1, PTC in exon 1	

[a] GenBank data available through the NCBI are provided next to the references for the castor bean transcript assemblies (TAs) annotated by The Institute for Genomic Research (TIGR; http://plantta.tigr.org), supporting each of the functional genes. Intron/exon structures were deduced by comparison to their respective TAs. Two putative pseudogenes *RcLEG2-3* and *RcLEG2-5* are also included (see explanation in the text) that were not represented by any EST in the dBEST database. N.A., not available; PTC, premature termination codon.

Data obtained from heterologous expression in tobacco and transient expression in castor is provided supporting the possible utility of these promoters for biotechnological purposes.

MATERIALS AND METHODS

Plant Materials. Seeds of *R. communis* L. var. IN15 were kindly provided by Dr. Leonardo Velasco (Instituto de Agricultura Sostenible, CSIC, Córdoba, Spain). Plants cultivated in a greenhouse were used as a source of seeds at different developmental stages for RT-PCR analysis and SAAT agroinfiltration experiments. Developing seeds were classified into defined stages, S1 to S7, according to morphological characteristics of seeds and capsules, as follows. S1, light green seeds, still growing (4−5 mm size); S2, light green seeds, almost full size (about 6−7 mm); S3, milky white seeds reaching maximum size (8−9 mm); S4, ivory-white seed with hardening testa, translucent endosperm, and soft capsule; S5, seeds with a hard testa that becomes light brown, expanding opaque endosperm still not filling the seed, with the capsule getting hard; S6, uniformly dark brown to black sclerified testa, endosperm filling the seed, and very hard green capsule; S7, mature seeds with mosaic colored testa and dry capsule. These stages correspond to approximately 10−15 DAP (S1 and S2), 15−20 DAP (S3), 25−30 DAP (S4), 30−35 DAP (S5), 40−45 DAP (S6), and 50−60 DAP (S7), on the basis of our determinations, and are in good agreement with those of previous studies (*12*).

Nicotiana tabacum L. var. Wisconsin-38 was used for *Agrobacterium* stable transformation experiments. Plants were grown at 25 °C under controlled conditions in growth cabinets with a 16 h light/8 h dark photoperiod and 70% relative humidity. Developing seeds from transgenic plants were classified according to the following stages: S1, growing translucent seeds (< 0.5 mm), small growing capsule (about 4−5 DAP); S2, growing white seeds (< 0.6 mm), intermediate sized capsule (about 6−7 DAP); S3, milky white seeds almost full sized (about 0.6−0.7 mm), full sized capsule (about 8−10 DAP); S4, browning seeds with hardening testa, full sized (about 0.8 mm) (12−16 DAP); S5, brown seeds with hard testa, predominantly green capsule (18−22 DAP); Sm, dark brown mature seeds, fully brown capsule (> 25 DAP).

Database Searching for Genes Encoding 11S Globulin Isoforms from WGS and EST Sequences of *R. communis*. Two cDNA sequences (GenBank accession no. AF262998 and AF262999) annotated as legumin-like protein from developing *Ricinus communis* were initially used as entrance queries to search for identical and homologous clones in the castor bean WGS and EST databases using the tblastn program (http://www.ncbi.nlm.nih.gov/blast/) at the National Center for Biotechnology Information (NCBI). The obtained sequences were iteratively used for further searches, including translated sequences, until we did not get any new translated sequence with above 50% similarity or 30% identical residues to the query. As a result, 11 nonallelic gene sequences were identified that we have termed RcLEG (see **Table 1**). They were further characterized by searching the castor Transcript Assemblies (TA) database of The Institute for Genomic Research (TIGR; http://plantta.tigr.org) and by comparison of TAs to their respective genomic sequences.

Expression of the *RcLEG* genes in the castor plant has been investigated using EST sequences available in the NCBI's dbEST database (http://www.ncbi.nlm.nih.gov/dbEST/). Data from the following cDNA libraries were employed: five non-normalized libraries covering different tissues (leaves, roots, flower buds, and 12−33 and 40−61 DAP seeds; about 50,000 ESTs) from the castor sequencing project (http://castorbean.jcvi.org/) and two normalized cDNA libraries, one from whole developing seeds (*14*) and another from mixed seed stages (about 7−24 DAP) (*15*). Legumin ESTs obtained from these two similar libraries were pooled in our analysis. Finally, a non-normalized cDNA library specifically from the endosperm of 20 DAP seeds (*16*) with 4,720 ESTs was also used.

Cladistic Analysis of Legumin Genes. The amino acid alignment of legumin sequences available in GenBank was achieved using the program Clustal X v.1.7 (http://www.clustal.org/) under default settings and further refined by visual inspection. The alignment output, with a total of 678 positions in the data set, was used to generate a cladogram based on the Minimum Evolution (ME) method, as implemented in the MEGA package v4 (http://www.megasoftware.net/). The bootstrap consensus tree inferred from 1,000 replicates was represented, and branches corresponding to partitions reproduced in less than 50% bootstrap replicates were collapsed. Rooting of the tree was accomplished by using the legumin sequences of gymnosperms *Ginkgo biloba* and *Picea glauca* as outgroups. For sequences selected in **Figure 2**, the alignment was visualized using the Boxshade v. 3.21 software (http://www.ch.embnet.org/software/BOX_form.html).

Semiquantitative RT-PCR Analysis for *RcLEG* Genes. Total RNA was purified from diverse castor tissues using UltraClean Plant RNA Isolation Kit (MoBio, Life Technologies). Oligo dT-directed cDNA was synthesized from 1 to 2 μg of total RNA by employing SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen), following the manufacturer's instructions. Approximately 25−100 ng of the cDNA was amplified in a 40 μL reaction volume using 0.8 units of iProof High-Fidelity DNA Polymerase (BioRad) and a PCR program comprising a denaturation step of 30 s at 98 °C, 20−25 cycles of 10 s at 98 °C, 30 s at 68 °C, 45 s at 72 °C, and a final step of 5 min at 72 °C. Amplification for a putative lysophosphatidate acyltransferase (*RcLPAT*) gene from castor (Acc. No. EU391594) was used as a constitutive control to check for equal template loading. cDNA amounts and the number of cycles (20 cycles for *RcLEG* genes and 25 cycles for *RcLPAT*) were experimentally determined to be nonsaturating, and cDNA loadings for *RcLEG* genes were corrected from the amplifications obtained for the *RcLPAT* gene. Specific oligonucleotide primers and amplification products are summarized in Supporting Information, Table 1. The identity of the PCR fragments was confirmed by direct sequencing using a Perkin-Elmer ABI-310 DNA automated sequencer and BigDye Terminator v3.1 chemistry.

Cloning of *RcLEG* Promoters and Transformation of Tobacco Plants with *RcLEG::GUS* Constructs. The 5′-proximal regulatory regions of *RcLEG1-1* and *RcLEG2-1* genes were cloned by PCR on genomic DNA using specific primer sets (Supporting Information, Table 1) and the Pfx high fidelity polymerase (AccuPrime, Invitrogen). The resulting fragments, about 1.2 Kbp for *RcLEG1-1* and 1.4 Kbp for *RcLEG2-1*, were

directionally cloned into the *Hind*III-*Bam*HI sites of the pCAMBIA-1391Z binary vector (CAMBIA, Australia). The T-DNA region of this plasmid contains a promoter-less GUS gene interrupted by a catalase intron. The resulting constructs were introduced into *A. tumefaciens* LBA4404 cells and then used for tobacco leaf disk transformation, essentially as described by Horsch et al. (*17*), using hygromycin for selection. Plants transformed with the empty vector were used as a negative control.

Transient Expression of *RcLEG::GUS* Constructs by Agroinfiltration (SAAT) of Castor Bean Tissues. For sonication-assisted *Agrobacterium*-mediated transformation (SAAT) of castor tissues, the protocol by Trick and Finer (*18*) was followed with minor modifications, as indicated. *A. tumefaciens* LBA4404 cells containing the two *RcLEG::GUS* constructs, the empty pCAMBIA-1391Z vector (as a negative control), or the pCAMBIA-1305.1 vector (*35S::GUS* fusion) were used at $O.D._{600}$ of 0.2. Castor seeds at different development stages were axenically isolated from freshly collected fruits that were previously sterilized with 50% (v/v) hypochlorite and 0.1% Tween-20, besides whole young leaves. Seeds were excised, either longitudinally into halves or transversely, while the leaves were cut in small squares, using a scalpel.

Explants were sonicated for 15 s in a bath sonicator (Model Ultrasons 512, Selecta; 40 kHz frequency) and incubated on MS-glucose solid medium, supplemented with 100 μM acetosyringone. Tissues were cocultivated for two days at 26 °C under a 16-h light photoperiod, further washed with liquid MS-glucose containing 500 mg/L cefotaxime, and incubated on MS-glucose plates supplemented with 500 mg/L cefotaxime for three additional days, before the determination of GUS activity.

Histochemical and Fluorometric GUS Assays. The standard procedures of Jefferson et al. (*19*) were followed. In fluorometric experiments, seeds (a minimum of 50, to minimize variation due to segregation) selected at different developmental stages (see Plant Materials) or leaf explants (3 discs from equivalent leaves) were collected from T_0 transgenic tobacco plants and ground in Eppendorf tubes using a pestle and sand. The fluorescence was recorded using a microtiter plate reader fluorometer (Mod. FLx800, BIO-TEK Instruments Inc.) at 360 and 465 nm as excitation and emission filter wavelengths, respectively. The protein concentration of the extracts was estimated by the Bradford method (*20*). GUS specific activity was expressed either as nmol or pmol of released 4-MU per minute and protein milligram. Protein extracts and GUS activity measurements from castor seeds and leaf explants subjected to agroinfiltration (SAAT) were performed in triplicate (5−8 explants per replica), as described for tobacco.

Histochemical localization of GUS activity was carried out by incubating the tissues in 5-bromo-4-chloro-3-indolyl glucuronide (X-gluc) in 100 mM sodium phosphate buffer at pH 7.0, 0.06% Triton X-100, 5 mM potassium ferrocyanide, and 5 mM potassium ferricyanide. The explants were infiltrated under mild-vacuum for 5 min, and the histochemical reaction was performed in the dark at 37 °C for 18 h. Tobacco seeds were mounted on microscope slides with 50% glycerol and photographed with a Pixe-LINK (Mod. PL-A662) camera coupled to a Nikon microscope (Mod. Eclipse 80i). Castor seeds were photographed on a dark field with a Canon EOS 50D camera and a 60 mm f2.8 macro objective attached to a 31 mm extension tube.

RESULTS AND DISCUSSION

Identification of a Gene Family Encoding 11S Globulins from *Ricinus communis*. We have performed an exhaustive search of the Whole Genome Shotgun (WGS) database generated from the 4X draft of the castor bean genome (http://castorbean.jcvi.org/index.shtml), deposited in GenBank (see Materials and Methods). As a result, a total of 11 nonallelic legumin-like genes have been identified (**Table 1**). All except two putative pseudogenes are supported by annotated transcript assemblies generated from a comprehensive expressed sequence tags (EST) database (> 60,000 sequences), most of them obtained as a part of the mentioned sequencing project. Phylogenetic analysis (see below) allowed the identification of two classes of sequences representing two legumin subfamilies which have been termed *RcLEG1* and *RcLEG2*. The *RcLEG1* group comprises six genes, *RcLEG1-1* to

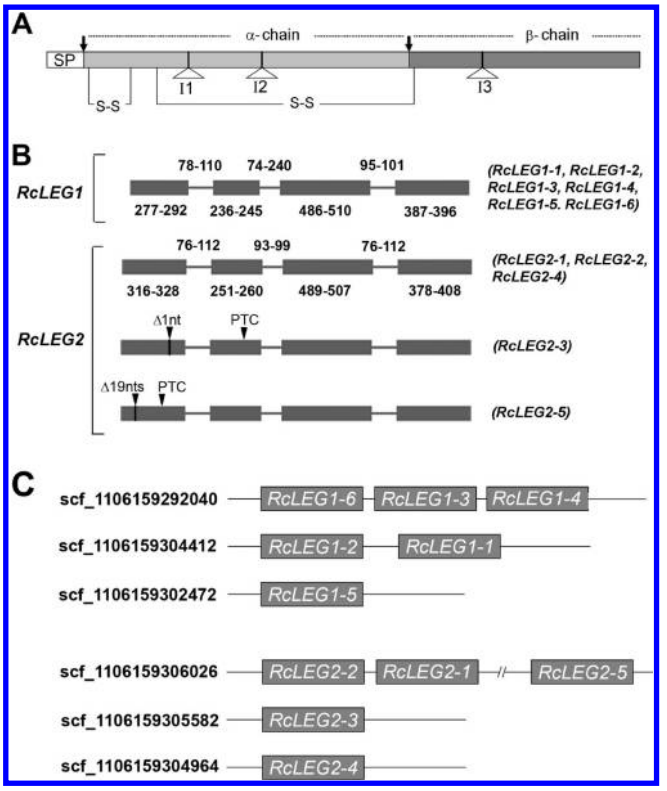

Figure 1. (**A**) Schematic structure of castor legumin proteins depicting the conserved position of introns and the cleavage sites, marked by arrows, involved in post-translational processing. Positions of the signal peptide (SP) and the two resulting polypeptides, α and β chains, are indicated, besides the disulfide bridges participating in the structure of the legumin monomer. (**B**) Structure of legumin genes of the two subfamilies (*RcLEG1* and *RcLEG2*) identified in castor. Lengths in nts for intron and exon regions are shown. Location of the deletions and premature termination codons (PTC) are marked by arrows for the two putative pseudogenes *RcLEG2-3* and *RcLEG2-5*. (**C**) Organization of castor legumin genes into clusters. Relative positions of the genes in the available genomic scaffolds (sfc) are drawn together with their respective GenBank accession numbers.

RcLEG1-6, sharing from 77% to 98% identical nucleotides within their coding regions. Genes *RcLEG1-1* and *RcLEG1-2* are quite similar (98% of identical nts) with only 11 amino acid changes out of 478 residues (3 of them nonconservative) for the protein encoded. The same holds true for *RcLEG1-3* and *RcLEG1-4* (98% nts identity) with 12 changes out of 468 residues (3 nonconservative) in the protein. The *RcLEG2* group is composed of five genes, with similarities ranging from 77% to 95%. However, similarity among members of the two subfamilies is reduced to only 60−68%.

Within the *RcLEG2* subfamily, two sequences, *RcLEG2-3* and *RcLEG2-5*, are not represented by any EST in the database. Further analysis reveals the presence of deletions affecting the first exon in the putative coding regions (**Table 1**), leading to premature termination codons (PTCs). For both genes, the PTCs were located more than 100 nts from the last intron, making them candidates to be down-regulated by the nonsense mediated decay mechanism (*21*). It is therefore likely that they certainly represent pseudogenes.

Gene structures, deduced by comparison of the genomic sequences to their respective TAs, are very similar among castor legumin genes (**Table 1**). They are composed of four exons and three small introns interrupting the coding sequence at equivalent positions of the different genes. This is a fairly conserved structure

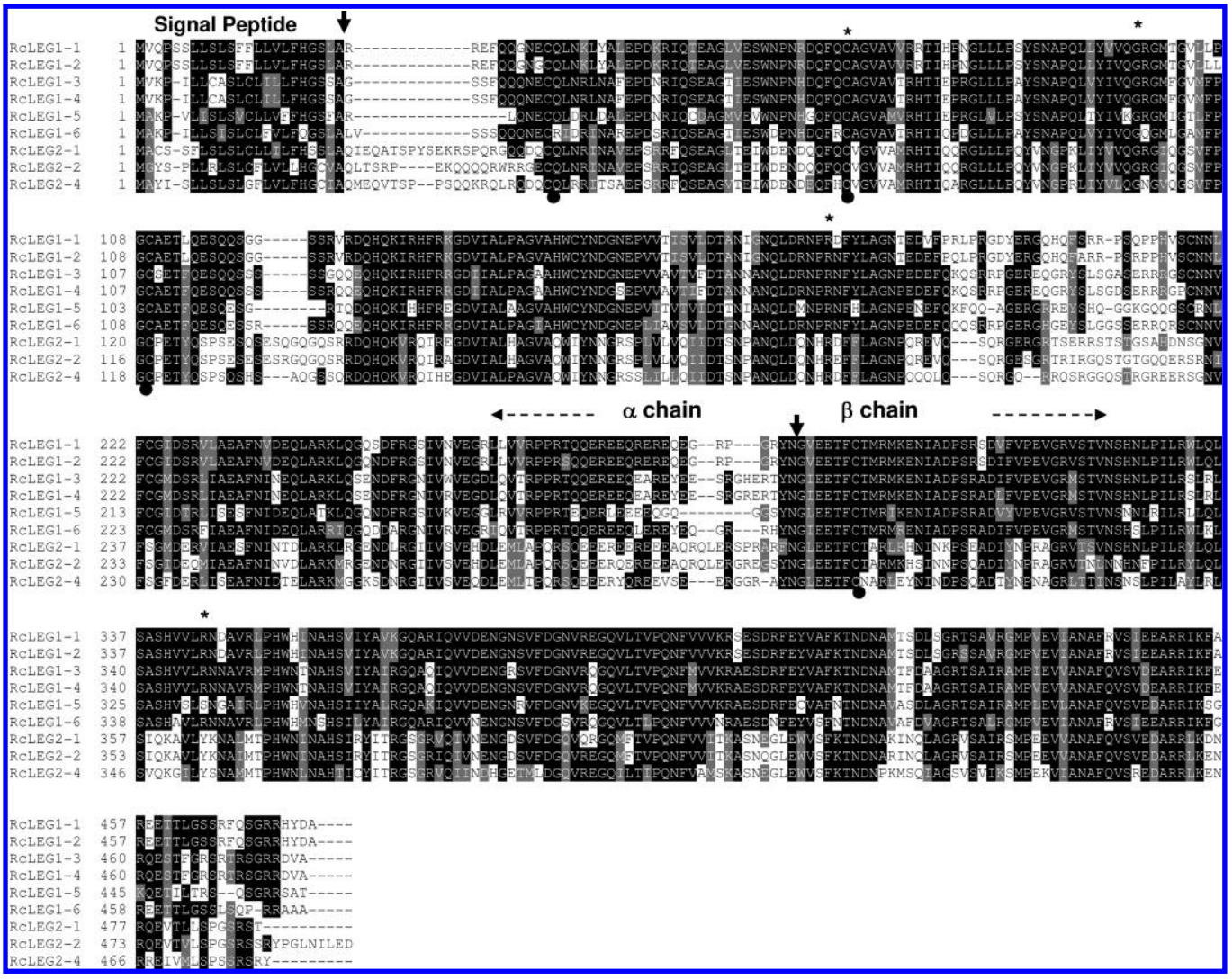

Figure 2. Amino acid sequence alignment of castor legumin isoforms. Identical or similar amino acids are shadowed in black and gray, respectively. Shading is applied when there is agreement for a fraction of sequences above 0.5. Cleavage sites involved in post/cotranslational processing are marked by arrows. Signal peptides have been predicted using the SignalP 3.0 software (http://www.cbs.dtu.dk/services/SignalP/). Asterisks indicate the highly conserved positions of introns. Conserved Cys residues involved in the formation of intrachain and interchain disulfide bridges participating in the structure of the legumin monomer are marked by dots.

in angiosperms, where introns I1 and I2 interrupt the α-polypeptide, while intron I3 interrupts the *β*-polypeptide in the coding regions (**Figure 1A**) (*22*). Just a few legumin genes are divergent in number but not in the position of introns. Thus, introns I1 and I2 have been found in the *cruA* and *HaG3* genes from oil seed rape and sunflower, respectively, while introns I2 and I3 are present in the B subfamily of genes from field bean and in *J*/*K* genes from the pea (*22*).

Analysis of the available genomic scaffolds in GenBank (**Table 1**) indicates that some of the *RcLEG* genes are organized into clusters (**Figure 1C**), a common feature of legumin families (*6*, *7*). Castor legumin genes appear organized into six chromosomal domains, three for each subfamily, that contain up to three genes. These clusters usually group closely related genes, as in the case of *RcLEG1-1* and *RcLEG1-2* (98% identical nts) or *RcLEG1-3*, *RcLEG1-4*, and *RcLEG1-6* (85−98% nts). *RcLEG2-3* is an exception to this rule. This is a putative pseudogene that is very similar to *RcLEG2-1* (95% nts) but is located apart, thus probably representing a recent duplication to a distant place.

Characterization of Castor 11S Legumin Proteins. The primary translation products deduced for *RcLEG* genes are proteins containing 461 to 496 residues (**Table 2**). Sequence alignment and software prediction for the nine legumin products indicate that all are precursor polypeptides containing N-terminal signal sequences (22−23 residues) responsible for ER targeting via a signal recognition particle (SRP) dependent pathway (**Figure 2**). A conservative processing site (Asn-Gly) for splitting each polypeptide into two segments (α and *β* subunits) via proteolytic cleavage is also present as in most other angiosperm legumins (*23*). The predicted sizes for α-subunit isoforms are 28.6−32.6 kDa and 20.2−21.6 for the *β*-subunit (**Table 2**), which are in good agreement to the SDS−PAGE estimated sizes of 29−34 kDa and 20.5−23.5 kDa from the purified castor proteins (*10*). Low isoelectric points (5.4−6.4) are deduced for all of the α-chain isoforms except for the two closely related RcLEG1 and RcLEG2 with pI 8.3 and 8.6, respectively. This is an exception to the reported acidic nature of the α-peptides (pI 4.8−6.2 in castor) and may be indicative of protein phosphorylation. For all *β*-peptide isoforms, high theoretical pI (8.9−10.2) values are calculated, in agreement to estimated pI 7.4−9.4 for the purified proteins (*10*). RcLEG proteins also contain four conservative cysteine residues responsible for an intrachain disulfide bond within the α-subunit

Table 2. Characteristics of the 11S Legumin Proteins from *R. communis*[a]

legumin isoform	GenBank acc. no.	Uniprot cat. no.	amino acid residues[b]	precursor molecular mass (kDa)	molecular mass (α-chain) (kDa)	molecular mass (β-chain) (kDa)	isoelectric point (α/β)
RcLEG1-1	EEF38213	Q9M4Q8	476	53.67	30.14	21.06	**8.27**/8.90
RcLEG1-2	EEF38210	B9SDX6	476	53.65	30.12	21.06	**8.58**/8.90
RcLEG1-3	EEF37813	B9SF36	478	54.01	30.73	20.98	5.50/10.21
RcLEG1-4	EEF37814	B9SF37	478	54.06	30.84	20.93	5.68/9.97
RcLEG1-5	EEF30344	B9T1B8	461	51.16	28.56	20.25	5.41/9.38
RcLEG1-6	EEF37812	B9SF35	475	53.49	30.54	20.62	5.87/9.60
RcLEG2-1	EEF28918	B9T5E7	490	55.26	32.63	20.31	5.80/9.77
RcLEG2-2	EEF28917	B9T5E6	496	56.29	32.37	21.56	6.63/9.60
RcLEG2-4	EEF32199	B9SW16	480	54.06	31.29	20.44	6.36/8.96

[a] Amino acid sequences were deduced from the genomic sequences (WGS) and the available transcript assemblies (see **Table 1**). The number of residues and the molecular mass of the precursor include the signal peptide (SP). [b] Our deduced protein sequences (**Figure 1**) show some differences from those annotated in the GenBank. More specifically, RcLEG1-3 has 30 residues less than the annotated EEF37813 because of the presence of two putative ATG codons with a Kozak context. Comparison to other close legumin proteins (see alignment of **Figure 2**) favors our prediction for the use of the downstream codon. RcLEG2-1 has eight residues less than EEF28918 because of different splicing site predictions of the first intron. Similarly, for RcLEG2-2 alternative predictions in the splicing of the first intron are responsible for the difference in the N-terminal sequence relative to EEF28917. In both cases, our prediction is supported by individual EST sequences and our RT-PCR results. In bold are marked the unusually basic isoelectric points for the α-subunit of the RcLEG1-1 and RcLEG1-2 proteins.

and an interchain disulfide bond linking both subunits in their mature proteins (**Figure 2**).

The amino acid compositions of the castor legumins are similar among the two subfamilies (Supporting Information, Table 2) and are in accordance with those of the homologous storage proteins in other plants, with Glx (13.2−17.5%) and Asx (10.5−11.5%) predominating consistently with their N-storage function. Sulfur-containing amino acids (Cys and Met) are similarly represented in the RcLEG1 and RcLEG2 proteins, ranging from 1.2% to 2.2% of Cys and 1.3% to 2.5% of Met (Supporting Information, Table 2). The Met contents of castor legumins are relatively high as compared to those of leguminosae species (0.0% in *Vicia faba* legB) but rather similar to data reported for other dicotyledonous plants such as *Ficus* and somewhat lower than those for *Sesamum* (3.2%) (*24*) and *Perilla* (4.2%) (*25*).

Evolutionary Relationships of Castor Legumins. Evolutionary relationships of castor legumins are summarized in the protein cladogram (**Figure 3**), which includes representatives of gymnosperms and monocotyledonous as well as dicotyledonous angiosperms. According to the phylogenetic tree, two legumin subfamilies can be distinguished, as was already inferred from similarity comparisons. The six RcLEG1 legumins group together with proteins from *Arabidopsis*, *Populus*, *Ficus*, and *Pistacia*, and leguminous species such as *Vicia sativa* and *Glycine max* (86% bootstrap support). The three RcLEG2 legumins group apart from the RcLEG1 group, besides other related proteins of *Populus*, *Ficus*, and *Pistacia* (69% bootstrap). Therefore, it seems that the divergence between the two subfamilies arose rather early in dicotyledonous evolution. The apparent absence of RcLEG2 orthologues in some groups of plants such as *Arabidopsis* and leguminous species suggests that they may have been lost during evolution.

Previous evolutionary data have shown that there is an especially methionine-rich (3−4 mol %) legumin subfamily present in the lower angiosperm clades including the Monocotyledoneae (*26*). This led to the proposal that evolution of legumin genes involved an early gene duplication in angiosperms giving rise to the progenitors of the methionine-rich (MetR) and methionine-poor (MetP) genes described in modern angiosperms (*26*). Our phylogenetic inference suggests that both *RcLEG1* and *RcLEG2* ancestors evolved within the MetP lineage as indicated by the strong support of the clade (90% bootstrap) also including the MetP representative A2 legumin of *Magnolia*. A dedicated database search for castor legumins related to the MetR, using the representative B14 legumin of *Magnolia*, did not render any new member. As in many other modern angiosperms,

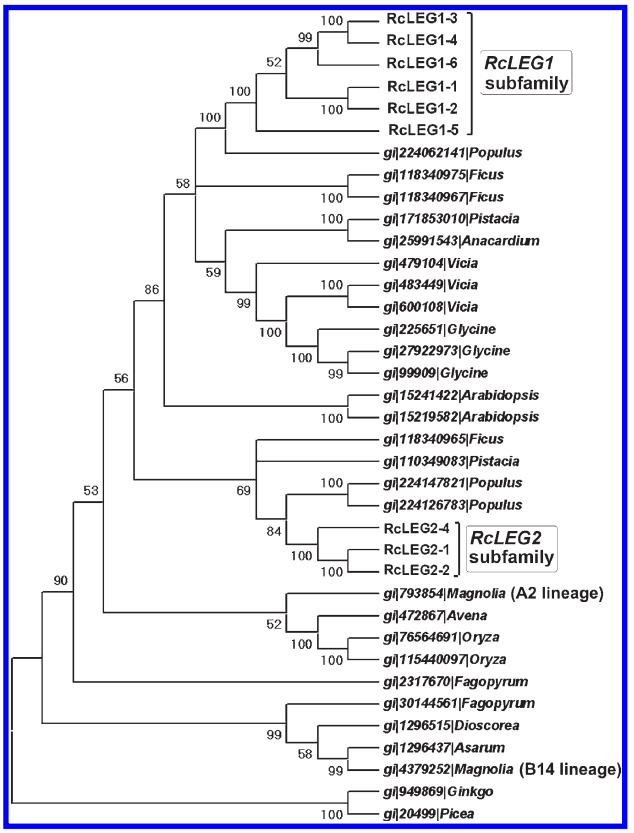

Figure 3. Evolutionary relationships among RcLEG proteins and other plant legumins. Amino acid sequences available in GenBank were aligned, and the resulting matrix was used to infer the evolutionary relationships using the Minimum Evolution method as implemented by the MEGA4 software. The bootstrap consensus tree obtained from 1000 replicates is shown. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. Bootstrap values are provided next to the branches. Placement of the two legumin families, RcLEG1 and RcLEG2, identified in castor are indicated. Rooting of the tree was accomplished by using legumin sequences of the gymnosperms *Ginkgo biloba* and *Picea glauca* as outgroups.

it seems that MetR legumins are missing in castor, in agreement with the view that the B14 orthologue has been lost at some time in angiosperm evolution (*26*).

The presence of 11S subfamilies in all plants that have been studied raises the question of the biological significance, if any, of

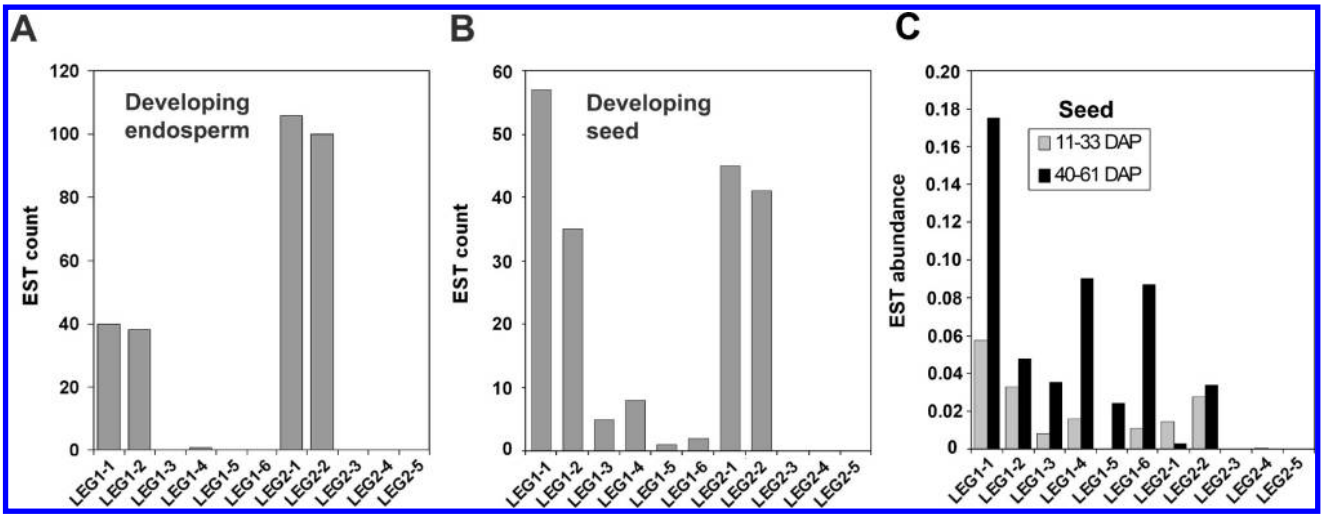

Figure 4. EST counts and abundances for *RcLEG* genes. Data are obtained from (**A**) a non-normalized library from the endosperm of developing seeds, about 20 DAP, with 4,720 ESTs (*16*); (**B**) pooled ESTs of two normalized libraries derived from early developing seeds (*14*, *15*); (**C**) non-normalized libraries generated from 11 to 33 DAP and 40−61 DAP seeds by the TIGR sequencing project (http://castorbean.jcvi.org). EST counts were represented in **A** and **B**, and relative abundances in **C**, calculated as a ratio between EST counts (5′-reads only) and EST library size (7,645 for 11−33 DAP and 3,988 for 40−61 DAP).

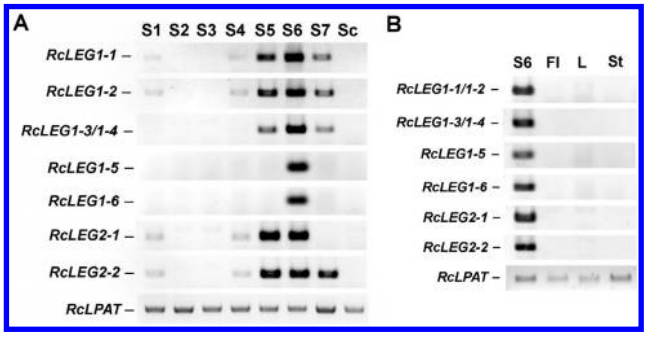

Figure 5. Expression analysis of castor legumin genes along seed development (**A**) and diverse plant organs (**B**) by semiquantitative RT-PCR. (**A**) Analysis of cDNAs from seeds at the S1 to S7 stages (see Materials and Methods), and Sc corresponds to dissected S4 seeds without endosperm. (**B**) Analysis of cDNAs from S6 seeds (S6), flower buds (Fl), leaves (L), and stem (St). Single gene specific primers were used except for the very similar gene pairs RcLEG1-3/RcLEG1-4 (panels **A** and **B**) and RcLEG1-1/RcLEG1-2 (panel **B**), whose transcripts were amplified by the same primers. The identity of the PCR products was confirmed by direct sequencing. A castor putative lysophosphatidate acyltransferase gene (*RcLPAT*) was used as a constitutive control to check for equal template loading. All PCR products were taken at cycle numbers determined to be nonsaturating, 20 cycles for all *RcLEG* genes and 25 cycles for *RcGPAT*.

the subunit heterogeneity. It is not known if all subunits are functionally interchangeable in seed oligomers, but an interesting possibility is that certain subunits in each of the two families could perform different functions, maybe during the assembly of legumin hexamers.

Transcript Analysis of Castor Legumin Genes. Expression of the *RcLEG* genes in the castor plant has been investigated using EST sequences available in the NCBI's dbEST database (http://www.ncbi.nlm.nih.gov/dbEST/). EST databases have been used to obtain gene expression data from the abundance of different messages in unbiased cDNA libraries (*27*). If the EST collection is sufficiently large, it is possible to collect enough data to produce a reasonably accurate estimate of mRNA abundance, at least for highly expressed genes. We have analyzed EST data from the

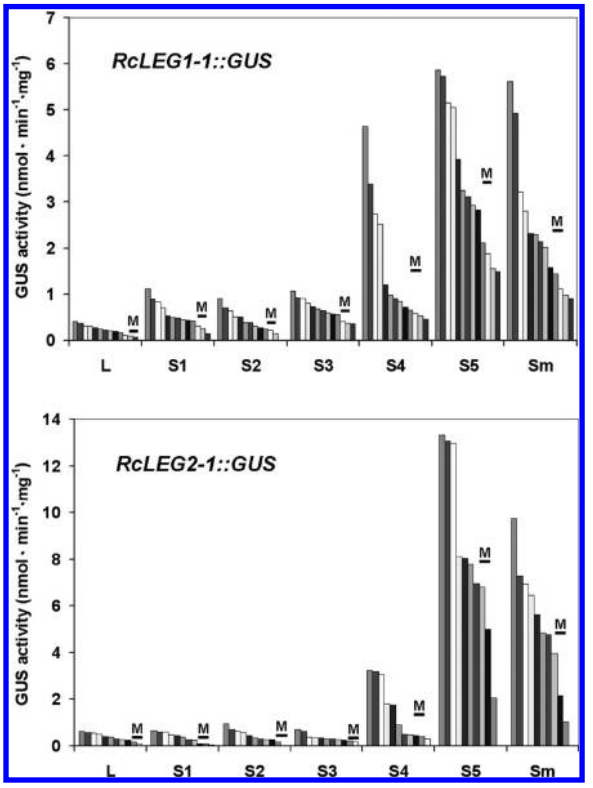

Figure 6. GUS activity in leaves and developing seeds of tobacco transgenic plants containing *RcLEG1-1::GUS* and *RcLEG2-1::GUS* constructs. GUS activity for seeds was determined in protein extracts of leaves (L) and seeds collected at different development stages (S1 to S5 and Sm) defined as described in Materials and Methods. Data are represented in descendent order for 13 (*RcLEG1-1::GUS*) and 11 (*RcLEG2-1::GUS*) selected primary transgenic plants. Mean values are represented by the horizontal lines.

castor sequencing project (http://castorbean.jcvi.org/), which comprises about 50,000 EST generated from non-normalized cDNA libraries for diverse organs, leaf, root, flower bud, and two development ranges (12−33 DAP and 40−61 DAP) of the seed.

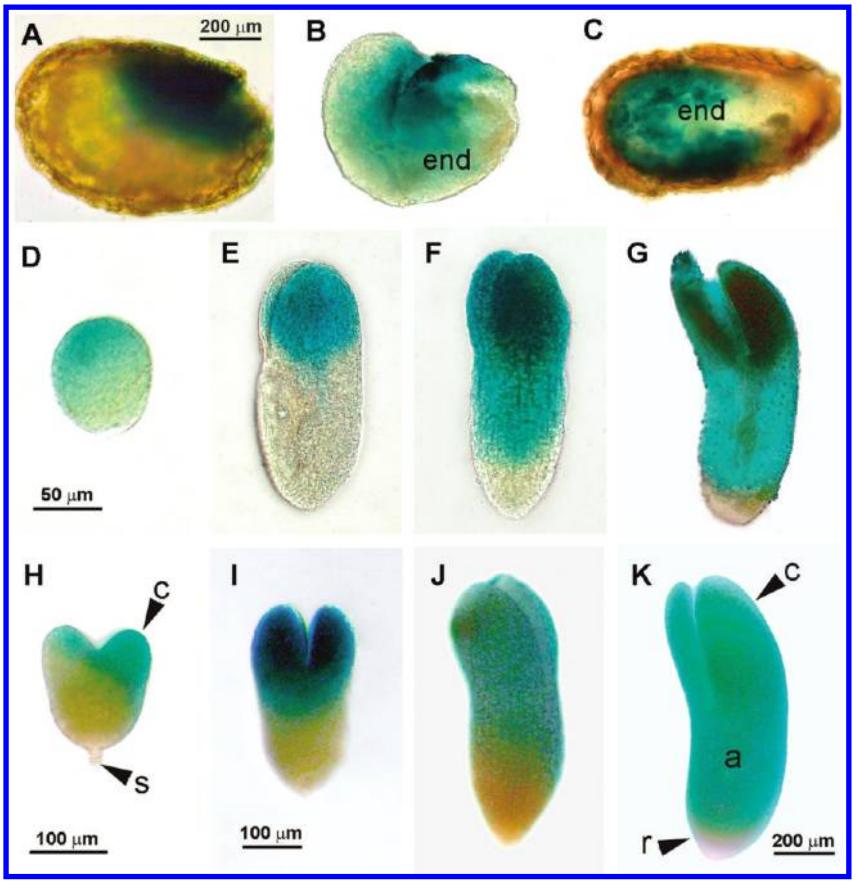

Figure 7. Localization of GUS expression driven by proximal promoter regions of *RcLEG1-1* and *RcLEG2-1* genes in transgenic tobacco seeds. GUS staining of seeds from *RcLEG1-1::GUS* (**A**−**G**) or *RcLEG2-1::GUS* (**H**−**K**) plants is shown at different developmental stages. Bright field microscopy images correspond to (**A**) whole S4 seed; (**B**) dissected S5 endosperm; (**C**) S5 seed section without the embryo; (**D**) globular stage embryo from S4 seed; embryos from S5 seeds (**H**) heart stage; (**E**,**I**) early torpedo stage; (**F**,**J**) late torpedo stage; and (**G**,**K**) cotyledon stage. Abbreviations: end, endosperm; c, cotyledon; s, suspensor; r, radicle; a, seed axis.

Besides this, we used data from three other cDNA libraries. Two of them are derived from whole developing castor seeds of similar developmental stages (*14*, *15*), whose EST have been pooled in our analysis, and another one specifically from developing endosperm (*16*). These data have been considered independently since they correspond to different seed tissues and/or developmental stages, and also, cDNA libraries differ in size and methodologies (see Materials and Methods).

Within the endosperm library, the predominant EST among legumin genes corresponded to the closely related *RcLEG1-1* and *RcLEG1-2* (98% of similarity), besides *RcLEG2-1* and *RcLEG2-2* (**Figure 4A**). This is in agreement with EST counts obtained from the developing seed libraries (**Figure 4B**) and the 12−33 DAP seed library (**Figure 4C**), all of them corresponding to early stages of seed development. Lower EST counts are obtained for *RcLEG1-3*, *RcLEG1-4*, and *RcLEG1-6* and very low counts for *RcLEG1-5* and *RcLEG2-4*. No EST was found in any library for the putative pseudogenes *RcLEG2-3* and *RcLEG2-5*. A comparison of EST abundances among the 12−33 DAP and 40−61 DAP seed libraries (**Figure 4C**) indicates that all *RcLEG1* genes increase their expression during the late development phase. On the contrary, EST abundance for the *RcLEG2-1* gene is strongly reduced in the 40−61 DAP library, while *RcLEG2-2* is similarly represented in both stages. Expression of legumin genes in nonseed organs is almost undetectable, with only few EST in the root library, 1 EST for *RcLEG1-1*, and 2 EST for *RcLEG1-4*, which means a relative abundance 500 times lower than that in the seed.

We also investigated the expression of legumin genes along castor seed development by semiquantitative RT-PCR using specific primer pairs (see Materials and Methods). **Figure 5** summarizes the results obtained for *RcLEG* genes. In all legumin genes, expression is developmentally regulated, with a maximum attained at the S6 stage when the cellular endosperm reaches its full expansion (about 40−45 DAP). However, timing of individual genes clearly differs among them. While the transcripts of genes *RcLEG1-1*, *RcLEG1-2*, *RcLEG2-1*, and *RcLEG2-2* are already detected during early endosperm filling (about 25−30 DAP), with increasing expression in later phases, for others such as *RcLEG1-5* and *RcLEG1-6*, the mRNA is mainly detected during late development (S6 stage). Also, mRNA abundance in late stages differs among genes. Expression of genes such as *RcLEG1-1*, *RcLEG1-2*, and *RcLEG2-2* is maintained up to the S7 stage, when drying of the seed is occurring, while transcripts for *RcLEG2-1*, *RcLEG1-5*, and *RcLEG1-6* are not detected at this stage. Though RT-PCR could not discriminate among the highly similar *RcLEG1-3* and *RcLEG1-4* genes, sequencing of the PCR products showed amplification of the cDNA for the close genes in a similar ratio. These differences in the transcript profiles are in agreement with previous data indicating a differential accumulation pattern of individual legumin proteins along castor seed development (*28*). However, comparison of these results with EST data indicates that the lower EST counts for genes such as *RcLEG1-5* and *RcLEG1-6* are due, at least in part, to the narrower expression window rather than just to a much lower expression level.

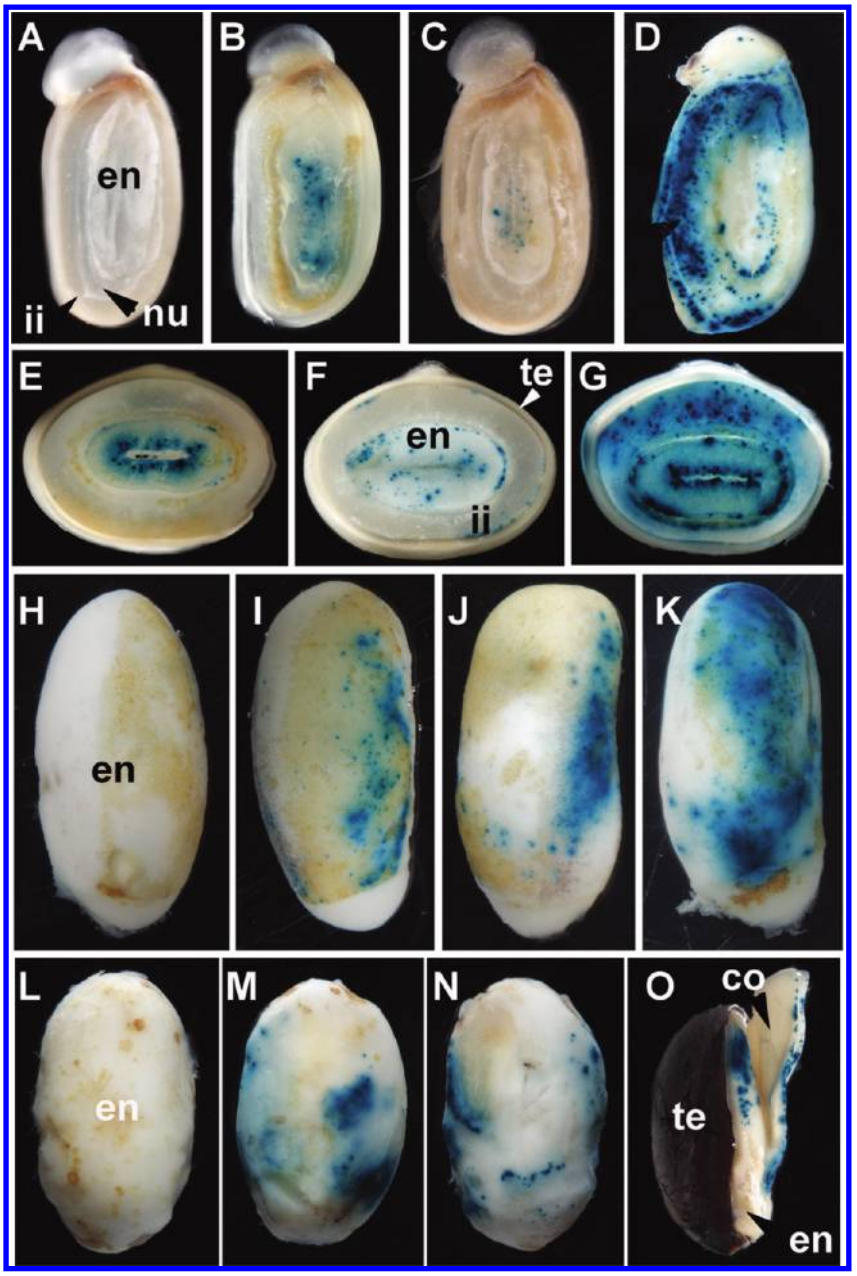

Figure 8. GUS expression directed by *RcLEG1-1* and *RcLEG2-1* promoters in castor seeds. Seeds at different development stages were infiltrated following the SAAT technique with *Agrobacterium* cells containing constructs *RcLEG1-1::GUS* (**B**,**E**,**I**,**M**,**O**), *RcLEG2-1::GUS* (**C**,**F**,**J**,**N**), *35S::GUS* (**D**,**G**,**K**), or the negative control with the empty vector (**A**,**H**,**L**). Dark field photography images correspond to longitudinal sections of S4 seeds (**A**−**D**), transverse (**E**−**G**) sections of S5 seeds, and longitudinal sections of S5 endosperms (**H**−**K**) and S6 endosperms (**L**−**O**). Abbreviations: en, endosperm; ii, inner integument; nu, nucella; co, cotyledon; te, testa.

Very low amplification was obtained for nonendospermic tissues (composed of testa, inner integuments, and nucella) at the S4 stage (Sc in **Figure 5A**), indicating that legumin expression is mainly located in the seed endosperm, at least during this phase. Interestingly, a low but consistent amplification for the major *RcLEG* genes was obtained in the very early S1 stage (10−12 DAP), when the seed is growing and tissues are differentiating, while this is not observed at the S2 and S3 later stages. Transient accumulation of legumin proteins during early seed development has been recorded in tobacco (*29*), where it has been proposed that they could be used as nutritive reserves for the growing embryo.

Expression of *RcLEG* cis-Regulatory Regions in Tobacco and Castor Plants. We have investigated the expression driven by the proximal promoter regions of *RcLEG1-1* and *RcLEG2-1* genes, chosen as representatives for the two legumin subfamilies in castor, by fusion to the *GUS* reporter gene. Appropriate constructs were generated into the binary vector pCAMBIA-1391Z and used for stable transformation of tobacco plants via *Agrobacterium tumefaciens* (see Materials and Methods). A total of 49 primary transgenic lines for *RcLEG1-1* and 37 for *RcLEG2-1* were initially generated, from which 17 and 12 plants, respectively, containing the transgen, were finally selected for detailed analysis. The expression of the reporter gene was monitored in leaves and developing seeds of primary transgenic plants by measuring β-glucuronidase (GUS) activity, both in crude extracts and histochemically.

As shown in **Figure 6**, GUS expression directed by both promoters followed the typical pattern characteristic of seed storage protein accumulation (*7*, *29*). Low activity was obtained

in nonseed tissues, represented by the leaf, relative to that attained in developing seeds. Analysis of seeds collected at different developmental stages, reveal low GUS activity along stages S1 to S3 (equivalent to about 5 to 10 DAP). GUS level increase is first detected in S4 seeds (12−16 DAP) coincident with the period of embryo morphogenesis, reaching a maximum at S5 (18−22 DAP), the maturation stage where most of protein storage deposition occurs. High GUS activity is maintained in the dehydrated mature seed (Sm).

Histochemical analysis of GUS activity in S4 to S5 seeds using the color-yielding substrate X-gluc is shown in **Figure 7**. A similar pattern is obtained for both *RcLEG1-1* and *RcLEG2-1* promoters. GUS staining is first observed in embryos at the globular stage (**Figure 7D**). A gradient of GUS expression along the embryos is seen that is more apparent in successive stages. At the heart stage (**Figure 7E**,**H**,**I**), GUS activity was only present in the cotyledon region. Later on, GUS staining spreads progressively along the embryonic axis, during the torpedo (**Figure 7F**,**J**) and cotyledon (**Figure 7G**,**K**) phases, but remains almost absent in the radicle pole. GUS expression is also observed in the endosperm, initiating in the region surrounding the embryo at the globular stage (**Figure 7A**,**B**) and spreading, later on, to the whole tissue (**Figure 7C**). The pattern described closely resembles that of other legumins whose regulatory regions direct specific and developmentally regulated expression, both in the endosperm and seed embryo, when introduced in tobacco plants (*7*). Heterologous legumin expression also matches the accumulation pattern of the tobacco storage proteins (*29*). That points to the conservation of cis-regulatory elements in the legumin promoters. Analysis of the 5′-upstream regions of the *RcLEG* genes (Supporting Information, Figure 1) reveals the presence of RY boxes which are characteristic and essential motifs present in all legumin genes (*30*). The structure of the six *RcLEG1* promoters is fairly similar, with a conserved CCATGCAAA sequence containing the RY core, located 126−137 nts upstream of ATG. Besides this, a GTGTAAGA sequence resembling the endosperm motif (TGTAAAGT) is conserved, at −150 to −161 nts relative to ATG. A different structure is noticed for the proximal 5′-regulatory sequences of the more diverged *RcLEG2-1* and *RcLEG2-2* genes, which possess two close RY boxes, but placed at different positions among them, in addition to a conserved TTGCCACGTCC motif, related to the CACGTGGC ABA-responsive element (ABRE), located downstream of the RY elements. At present, we do not know if these differences among promoter structures of the two gene subfamilies also involve a differential response to developmental and/or environmental signals, though this is an interesting possibility that deserves further investigation.

The ability of proximal regulatory regions from *RcLEG* genes to direct proper expression into the castor tissues has also been investigated. The *RcLEG::GUS* constructs in the LBA4404 *A. tumefaciens* strain were used for transient expression experiments using the sonication-assisted agroinfiltration technique (SAAT) (*18*) on castor seed explants from different development stages (see Materials and Methods). Control experiments included cells containing either the empty binary vector pCAMBIA-1391Z, as a negative control, or the pCAMBIA-1305.1 plasmid with the *GUS* gene driven by the constitutive *CaMV 35S* promoter. Histochemical GUS detection was performed after two days of coculture with *Agrobacterium*, followed by three days of incubation on MS plates containing acetosyringone. The results, again similar for both *RcLEG* promoters, are summarized in **Figure 8**. GUS staining was observed specifically in the endosperm of seeds at the S4 stage (**Figure 8B**,**C**), corresponding to early endosperm development (about 20−25 DAP). Generalized staining including the inner integuments and nucella tissues around the endosperm is obtained for the *35S* constitutive promoter (**Figure 8D**), while undetectable GUS activity was present in the negative control (**Figure 8A**). A similar pattern was observed for *RcLEG* genes in the S5 stage (30−35 DAP) of middle endosperm enlargement (**Figure 8E**,**F**,**I**,**J**). GUS staining was also detected in the endosperm of S6 seeds (40−45 DAP) when the endosperm already fills the whole seed (**Figure 8M**,**N**) and also in the embryo cotyledons (**Figure 8O**).

In order to test the tissue specificity of the *RcLEG* promoters, fluorometric GUS assays were also performed on developing seeds (S5 stage) and young leaf explants subjected to SAAT (Supporting Information, Figure 2). Detectable GUS activity was present in seeds agroinfiltrated with cells carrying constructs with the two *RcLEG* or the *35S* promoters, while it was very low for the negative control (Supporting Information, Figure 2A), in agreement with the histochemical results. On the contrary, GUS activity above the level of the negative control was only observed in leaves for the construct carrying the constitutive *35S* promoter (Supporting Information, Figure 2B). This supports the notion that the employed proximal regulatory regions of *RcLEG1-1* and *RcLEG2-1* genes direct expression preferentially to the seed tissues, as should be expected from their respective genes. Availability of these nonconstitutive, highly specific seed promoters will be useful in transgenic strategies aimed at the modification of storage components of the castor bean such as oil fatty acids or the knockout of noxious proteins acting as allergens.

ABBREVIATIONS USED

CaMV, cauliflower mosaic virus; CTAB, cetyl trimethyl ammonium bromide; DAP, days after pollination; DTT dithiotreitol; EDTA, ethylenediaminetetraacetic acid; EST, expressed sequence tag; GUS, β-glucuronidase; 4-MU, 4-methyl-umbelliferone; SAAT, sonication assisted *Agrobacterium* transformation; TA, transcript assembly; YEB, yeast extract medium containing beef extract.

ACKNOWLEDGMENT

We are grateful to Leonardo Velasco (Instituto de Agricultura Sostenible, CSIC, Córdoba) for providing us with the IN15 castor seeds. We also thank to Pablo D. Rabinowicz (The Institute for Genetic Resources, TIGR, Castor Bean Project) for kindly sharing useful information about the castor cDNA libraries.

Supporting Information Available: Oligonucleotide primers used in RT-PCR analysis; amino acid composition of castor legumins; sequence alignment and putative cis-elements for 5′-proximal regulatory regions of the *RcLEG* genes; and fluorometric GUS assays of castor tissues after SAAT-agroinfiltration with *RcLEG::GUS* constructs. This material is available free of charge via the Internet at http://pubs.acs.org.

LITERATURE CITED

(1) Weiss, E. A. Castor. In *Oilseed Crops*; Weiss, E. A., Ed.; Blackwell Science: Malden, MA, 2000; pp 13−51.

(2) Conceicao, M. M.; Candeia, R. A.; Silva, F. C.; Bezerra, A. F.; Fernandes, V. J., Jr.; Souza, A. G. Thermo analytical characterization of castor oil biodiesel. *Renewable Sustainable Energy Rev.* **2007**, *11*, 964–975.

(3) Davison, A. G.; Britton, M. G.; Forrester, J. A.; Davies, R. J.; Hughes, D. T. D. Asthma in merchant seamen and laboratory workers caused by allergy to castor beans: Analysis of allergens. *Clin. Allergy* **1983**, *13*, 553–561.

(4) Mills, E. N.; Jenkins, J.; Marigheto, N.; Belton, P. S.; Gunning, A. P.; Morris, V. J. Allergens of the cupin superfamily. *Biochem. Soc. Trans.* **2002**, *30*, 925–929.

(5) Thorpe, S. C.; Kemeny, D. M.; Panzani, R.; Lessof, M. H. Allergy to castor bean 0.1. Its relationship to sensitization to common inhalant allergens (atopy). *J. Allergy Clin. Immunol.* **1988**, *82*, 62–66.
(6) Shewry, P. R.; Napier, J. A.; Tatham, A. S. Seed storage proteins: structures and biosynthesis. *Plant Cell* **1995**, *7*, 945–956.
(7) Thomas, T. L. Gene expression during plant embryogenesis and germination: An overview. *Plant Cell* **1993**, *5*, 1401–1410.
(8) Greenwood, J. S.; Bewley, J. D. Seed development in *Ricinus communis* (castor bean) 0.1. Descriptive morphology. *Can. J. Bot.* **1982**, *60*, 1751–1760.
(9) Youle, R. J.; Huang, A. H. Protein bodies from the endosperm of castor bean: Subfractionation, protein components, lectins, and changes during germination. *Plant Physiol.* **1976**, *58*, 703–709.
(10) Gifford, D. J.; Bewley, J. D. An analysis of the subunit structure of the crystalloid protein complex from castor bean endosperm. *Plant Physiol.* **1983**, *72*, 376–381.
(11) Gifford, D. J.; Bewley, J. D. Synthesis of the crystalloid protein complex in vivo in the endosperm of developing castor bean seeds. *Plant Physiol.* **1984**, *74*, 1006–1009.
(12) Chen, G. Q.; He, X. H.; Liao, L. P.; McKeon, T. A. 2S albumin gene expression in castor plant (*Ricinus communis* L.). *J. Am. Oil Chem. Soc* **2004**, *81*, 867–872.
(13) Tregear, J. W.; Roberts, L. M. The lectin gene family of *Ricinus communis*: cloning of a functional ricin gene and three lectin pseudogenes. *Plant Mol. Biol.* **1992**, *18*, 515–525.
(14) Cahoon, E. B.; Clifton, S.; Pape, D.; Marra, M.; Hillier, L.; Martin, J.; Wylie, T.; Dante, M.; Theising, B.; Bowers, Y.; Gibbons, M.; Ritter, E.; Ronko, I.; Tsagareishvili, R.; DeCarlo, S.; Waterston, R.; Wilson, R. WashU *Ricinus communis* EST Project. Unpublished work, **2003**.
(15) Kroon, J. T.; Kunst, L.; Slabas, A. R.; Smith, M. A. Genomic Resources for Castor Bean (*Ricinus communis*). Generation of a Normalized cDNA Library from Developing Castor Seed of Expressed Sequence Tags. Unpublished work, **2008**.
(16) Lu, C.; Wallis, J. G.; Browse, J. An analysis of expressed sequence tags of developing castor endosperm using a full-length cDNA library. *BMC Plant. Biol.* **2007**, *7*, 42.
(17) Horsch, R. B.; Fry, J. E.; Hoffmann, N. L.; Eichholtz, D.; Rogers, S. G.; Fraley, R. T. A simple and general-method for transferring genes into plants. *Science* **1985**, *227*, 1229–1231.
(18) Trick, H. N.; Finer, J. J. SAAT: sonication-assisted *Agrobacterium*-mediated transformation. *Transgenic Res.* **1997**, *6*, 329–336.
(19) Jefferson, R. A.; Kavanagh, T. A.; Bevan, M. W. GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* **1987**, *6*, 3901–3907.
(20) Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **1976**, *72*, 248–254.
(21) Chang, Y. F.; Imam, J. S.; Wilkinson, M. F. The nonsense-mediated decay RNA surveillance pathway. *Annu. Rev. Biochem.* **2007**, *76*, 51–74.
(22) Hager, K. P.; Muller, B.; Wind, C.; Erbach, S.; Fischer, H. Evolution of legumin genes: loss of an ancestral intron at the beginning of angiosperm diversification. *FEBS Lett.* **1996**, *387*, 94–98.
(23) Dickinson, C. D.; Hussein, E. H.; Nielsen, N. C. Role of post-translational cleavage in glycinin assembly. *Plant Cell* **1989**, *1*, 459–469.
(24) Tai, S. S.; Wu, L. S.; Chen, E. C.; Tzen, J. T. Molecular cloning of 11S globulin and 2S albumin, the two major seed storage proteins in sesame. *J. Agric. Food Chem.* **1999**, *47*, 4932–4938.
(25) Jin, U. H.; Jin, B. R.; Lee, J. W.; Cho, Y. S.; Kwon, O. C.; Kim, Y. K.; Chung, C. H. Characterisation of a methionine-rich storage protein cDNA from perilla (*Perilla frutescens*) seeds. *Aust. J. Plant Physiol.* **2000**, *27*, 701–707.
(26) Fischer, H.; Chen, L.; Wallisch, S. The evolution of angiosperm seed proteins: A methionine-rich legumin subfamily present in lower angiosperm clades. *J. Mol. Evol.* **1996**, *43*, 399–404.
(27) Rafalski, J. A.; Hanafey, M.; Miao, G. H.; Ching, A.; Lee, J. M.; Dolan, M.; Tingey, S. New experimental and computational approaches to the analysis of gene expression. *Acta Biochim. Pol.* **1998**, *45*, 929–934.
(28) Fukasawa, T.; Hara-Nishimura, I.; Nishimura, M. Biosynthesis, intracellular transport and in vitro processing of 11S globulin precursor proteins of developing castor bean endosperm. *Plant Cell Physiol.* **1988**, *29*, 339–345.
(29) Panitz, R.; Manteuffel, R.; Wobus, U. Tobacco embryogenesis: storage-protein-accumulating cells of embryo, suspensor, and endosperm are able to undergo cytokinesis. *Protoplasma* **1999**, *207*, 31–42.
(30) Baumlein, H.; Nagy, I.; Villarroel, R.; Inze, D.; Wobus, U. Cis-analysis of a seed protein gene promoter: the conservative RY repeat CATGCATG within the legumin box is essential for tissue-specific expression of a legumin gene. *Plant J.* **1992**, *2*, 233–239.

Received for review August 24, 2009. Revised manuscript received October 22, 2009. Accepted October 22, 2009. This work was supported by a grant from the Ministerio de Educación y Ciencia (PSE 320100-2006).