Cloning, mapping, and analyses of expression of the Em-like gene family in soybean [Glycine max (L). Merr.]

E. S. Calvo  
Iowa State University

E. S. Wurtele  
Iowa State University, mash@iastate.edu

R. C. Shoemaker  
United States Department of Agriculture

Follow this and additional works at: http://lib.dr.iastate.edu/bot_pubs

Part of the Agronomy and Crop Sciences Commons, Botany Commons, Cell and Developmental Biology Commons, and the Plant Breeding and Genetics Commons

Recommended Citation

http://lib.dr.iastate.edu/bot_pubs/67

This Article is brought to you for free and open access by the Botany at Iowa State University Digital Repository. It has been accepted for inclusion in Botany Publication and Papers by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Abstract The entire Em-like Group-1 late embryogenesis abundant (Lea) gene family from soybean was cloned and characterized. The five Group-1 Lea genes (Sle1-5) were divided into two classes based on sequence identity. Sle1-4 were genetically mapped to four different linkage groups. Nucleotide sequencing indicated that Sle1, Sle2, Sle3, and Sle5 encode polypeptides differing primarily by the presence of a repeated 20-amino acid motif. Sle1 and Sle5 were shown by Northern analysis to be expressed in developing embryos weeks earlier than Sle2 and Sle3. Sle4 was shown to be a pseudogene. Maximal levels of mRNA for all functional Sle genes accumulated in maturation-phase seeds, before significant desiccation had occurred, and declined rapidly upon seed imbibition. Desiccation did not induce Sle expression in seeds or vegetative tissue. Sle expression was confined to embryo tissues and Sle mRNA accumulated at similar levels in both the embryo axis and in the cotyledons.

Keywords Em-like gene family • Soybean • Cloning • Characterization • Mapping

Introduction

A considerable amount of information has accumulated on the molecular biology of the events involved in the developmental regulation of expression of several seed-protein genes, particularly those encoding storage proteins (Goldberg et al. 1989; Goldberg et al. 1994). Among these are included the Lea (late embryogenic abundant) genes (Dure 1993). LEA polypeptides can be grouped into at least three structurally distinct groups based on amino acid sequence homology and occurrence of amino acid motifs (Dure et al. 1989). The isolation of homologous genes or cDNAs has been reported from several species (Almoguera and Jordano 1992; Espelund et al. 1992; Litts et al. 1992; Raynal et al. 1989; Ulrich et al. 1990; Williams and Tsang 1991; Wurtele et al. 1993), but none of these reports mentioned genes or cDNAs from legume species.

It has been hypothesized that the LEA proteins are involved in protecting plant cells from the harms of desiccation (Dure et al. 1989). Support for this hypothesis comes mostly from circumstantial evidence, such as abundance in the cell, the time of LEA protein appearance during development, and their hydrophilic nature (Dure 1993). Experimental evidence is limited to biophysical measurements of hydrodynamic properties of the wheat Em protein (McCubbin et al. 1985). Computer modeling of Group-3 proteins has predicted a tertiary structure compatible with the hypothesis that such proteins would be involved in ion sequestration in the desiccated cell (Dure 1993). Whether or not such a structure occurs in vivo remains to be determined. The recent finding of Lea genes in cyanobacteria, an organism more easily studied genetically than higher plants, will certainly contribute to our understanding of
the functions of LEA proteins in late embryo development (Curry and Walker-Simmons 1993).

Much of our information on the regulation of Lea gene expression has been derived from experiments using monocot species (Marcotte et al. 1989; McCarty and Carson 1991; McCarty et al. 1991). However, this information can not necessarily be extrapolated to dicots (Galau et al. 1991). Unlike monocots, where storage-protein gene expression is frequently concentrated in the endosperm, the embryo cotyledons and to some extent the embryo axes are the primary tissue for storage-protein gene expression in seeds of the majority of dicot species (Lopes and Larkins 1993). This is particularly true for legumes such as soybean where the endosperm is very short lived (Chamberlin et al. 1994) and where storage proteins account for 36% or more of the seed dry weight (Wilson 1987). As a consequence, in soybean both the embryo-specific maturation and late embryogenesis programs may take place exclusively in the embryo. Therefore, it is conceivable that embryos of legume species may utilize a somewhat different mechanism from monocots to achieve the temporally different expression of the maturation and late embryogenesis programs.

Previous studies have shown that members of multigene families in soybean pair into more closely related groupings (Grandbastien et al. 1986; Lee and Verma 1984; Nielsen et al. 1989). The analysis of members of a gene family can provide much information on evolutionary processes and genome organization, and gene expression. The primary objective of the research presented here was to clone, sequence, and gene expression. The primary objective of the research presented here was to clone, sequence, and gene expression. The primary objective of the research presented here was to clone, sequence, and gene expression. The primary objective of the research presented here was to clone, sequence, and gene expression. The primary objective of the research presented here was to clone, sequence, and gene expression.

Materials and methods

Plant material

*Glycine max* (L.) Merr. cv ‘Williams 82’ was used as the source of immature seeds obtained from plants grown under standard greenhouse conditions. Flowers were tagged at anthesis, and immature seeds were collected at various days after flowering (DAF). Fifty seeds of approximately the same size were pooled for each time point, their total fresh weight was recorded, and were immediately frozen in liquid nitrogen and stored at −80 °C. Leaf tissue was obtained from 2-week-old plants that were grown in a sandbench in a greenhouse. Seeds were germinated in the dark and stored at −80 °C.

Water-deficit treatments

Seeds containing embryos at 30 DAF were left in excised pods and allowed to dry for 48 h (Rosenberg and Rinnie 1988). Treated embryos (1–2 g fresh weight) were harvested, frozen in liquid nitrogen and stored at −80 °C. For treatment of seedlings, germination papers containing 4-day-old seedlings were removed from the water-containing tray, unfolded, laid over dry trays, and kept in a cabinet for 26 h to ensure a slow dehydration. Tissues of the seedlings were visibly dehydrated as judged by the turgor of the hypocotyl. Seedlings in which the root tip was completely dry were discarded. Roots, hypocotyls, and cotyledons of 15–15 seedlings were harvested and immediately frozen in liquid nitrogen and stored at −80 °C.

For the water-deficit treatment of plantlets, shoots of 5-week-old plants were harvested, weighed, and left on the laboratory bench for about 6.5 h, or until shoots had lost 20% of their initial weight (Plant et al. 1991). Leaves from 12 plants were harvested, pooled, frozen in liquid nitrogen, and stored at −80 °C.

Isolation of cDNA and genomic clones

A lgt11 soybean (cv ‘Enrei’) cDNA library, constructed from mid-maturation developing seeds, was kindly provided by Drs. D. Shibata and R. Whittier (Mitsui Plant Biotechnology Research Institute, Tsukuba, Japan). Bacteriophage manipulation and screening techniques were as described by Sambrook et al. (1989). Nitrocellulose (Schleicher & Schuell) filters were pre-hybridized and hybridized in 6× SS (1× SS is 0.18 M NaCl and 1.9 M sodium citrate, pH 7.0), 1% (w/v) SDS, 25 mM NaHPO₄, pH 6.5, 3× Denhardt’s, and 0.1 mg/ml herming sperm DNA, at 65 °C. Final washes were done in 0.5× SSC, 0.1% SDS, at 56 °C, for 1 h. The probe used for screening the library was the carrot EMB-1 cDNA (Ulrich et al. 1990), labeled with α-[³²P]-dTCTP (Feinberg and Vogelstein 1983). The cDNAs hybridizing to the EMB-1 probe were subcloned into the EcoRI site of pHBluescript KS+ (Stratagene).

Genomic clones for SLE1 and SLE4 were isolated from a λEMBL3 genomic library made from cv ‘Williams’ (Clontech), and those for SLE2 and SLE3 were isolated from a λEMBL3 library made from cv ‘Resnik’ (Clontech). The SLE5 clone was isolated as a 2.5-kb EcoRI fragment from a λgt10 genomic library made after gel-fractionation (2.3–3.5 k b) EcoRI-digested DNA (cv ‘Williams’) in low-melting agarose (FMC). DNA was recovered from the gel using a QIAGEN tip-5 column essentially as described by the manufacturers. Library construction followed procedures described in Sambrook et al. (1989). The three genomic libraries were screened with the soybean cDNA. Screening conditions were as described for the cDNA library except that the final wash was done at 2× SSC, 0.1% SDS, at 50 °C.

DNA sequencing and DNA comparisons

Automated sequencing was performed on both single-and double-strand DNA templates in an ABI sequencing apparatus at the Iowa State University Nucleic Acid Facility. Manual sequencing was performed according to the dyeoxynucleotide chain termination method (Sanger et al. 1977) using a Sequenase 2.0 kit from USB. All sequence analyses were performed with the Mac DNA/asis sequence analysis software (Hitachi) using default settings in the Higgins-Sharp mode. The 3’-UTR regions used for sequence comparisons comprise a 520-bp fragment immediately downstream of the stop codon of each clone. For sequence comparison of the coding regions we omitted the DNA sequence corresponding to the second 20-amino acid hydrophilic motif from clones SLE1 and SLE5. Intron comparisons were made with the entire intron sequences.
Restriction fragment length polymorphism (RELP) mapping

The map locations of four Sle clones were determined by RELP analysis of the F2 segregating population as described by Keim et al. (1990) using the MAPMAKER software (Lander et al. 1987). Placement of gene locations was established with a minimum LOD score of 3 using the Kosambi mapping function. The identity of each locus was assigned based on fragment sizes predicted from the restriction map obtained for individual IEMBL3 clones.

Isolation of DNA and RNA, and hybridization analyses

Soybean leaf DNA was isolated following procedures described by Keim et al. (1990). Total RNA was isolated from different tissues and at different developmental stages by using the mini-prep procedure described by Wadsworth et al. (1988), with an extraction buffer containing guanidine thiocyanate as a chaotropic agent. A chloroform extraction was included immediately after the phenol:chloroform step in order to better remove lipids from the seed tissues.

DNA was digested with different restriction enzymes, fractionated by electrophoresis in 0.8% (w/v) agarose gels, and transferred to nylon membranes (Biotrace, Gelman Sciences) according to standard procedures (Sambrook et al. 1989). Pre-hybridizations and hybridizations were performed as described. Final washes were done in 0.5 x SSC, 0.1% (v/v) SDS, at 65 ºC for 40 min.

Hybridization analyses of RNA blots were performed according to Sambrook et al. (1989). RNA was fractionated in 1% (w/v) agarose gels containing 6% (v/v) formaldehyde. All RNA samples were denatured in the presence of 1 mg ethidium bromide. Nylon membranes were pre-hybridized in a solution containing 50% (v/v) formamide, 5 x SSC, 2 x Denhardt’s, 40 mM NaHPO4, pH 6.5, 10 mM EDTA, and 0.2 mg/ml herring sperm DNA. After 12 h of pre-hybridization, a [32P]-labeled probe was added, and hybridization continued for 15–18 h. Final washes were carried out under conditions given in the figures. The soybean actin cDNA probe, pSAC-7 (Dr. R. Meagher, University of Georgia, Athens), was used as an internal control on the RNA blots. All RNA blots were performed twice with replicated samples from two independent extractions. DNA probes were labeled by random-priming (Feinberg and Vogelstein 1983). Sle1 and Sle5 3'-UTR gene-specific probes comprise the fragments spanning from positions 379 to 687 bp and 1201 to 1740 bp, on the Sle1 cDNA and Sle5 genomic clones, respectively. Oligonucleotide gene-specific probes were hybridized in the absence of Denhardt’s solution at Tm -10 ºC. The oligonucleotides used were: 5'-GGTCTTTGTTTCTGATTCTGTTT-3' and 5'-GACCTAGCTGCTACCTATACCAT-3' for Sle2 and Sle3, respectively. Oligonucleotides were end-labeled using a polynucleotide kinase (Sambrook et al. 1989).

Reverse transcription and polymerase chain reactions (PCR)

One microgram of total RNA, prepared as described above from 87 DAF seeds, was reverse-transcribed using a first-strand cDNA kit (Pharmacia), essentially as described by the manufacturers. One hundredth of the reaction (or 50 ng of soybean genomic DNA in the control reactions) was used as template in a 100 ml PCR [200 mM dNTP, 0.2 mM of each primer, 1 x PCR buffer (Promega), 2 mM MgCl2, 0.1 U Taq polymerase (Promega)] using the combination of primers described in the text. The sequences of the primers used were as follows: P0 (5'-CTTGAGGCTGCTAAGACAT CTGTGCTG); P1/4 (5'-ACCAGACTGTCCATGTTGCTGAG-3). A total of 40 PCR cycles were performed using the following conditions: 45 s at 93 °C; denaturing step; 2 min at 70 °C, annealing step; and 2 min at 72 °C, extension step.

Results

Screening of the cDNA library

Over 500,000 recombinant bacteriophage from the soybean cDNA library were screened by hybridization with the radioactively labeled EMB-1 cDNA of carrot (Ulrich et al. 1990). This screen resulted in the selection of eight soybean cDNA clones. Restriction mapping and partial DNA sequencing of the soybean clones suggested that all of these clones contained identical cDNAs. For this reason, only one clone, Sle1 (soybean late embryogenic), was examined in detail.

The cDNA clone Sle1 was 676 bp in length, excluding the poly-A tail, and contained a translational open reading frame that codes for a polypeptide of 112 amino acids. The deduced Sle1 polypeptide showed the characteristics of a Group-1 LEA protein (Dure 1993). Sequence comparison with all previously described Group-1 LEA proteins revealed that Sle1 had on average 75% amino acid identity and 85% similarity to its homologs. The Sle1 polypeptide also contained a duplicated, highly hydrophilic, 20-amino-acid motif that had been previously found in the barley, cotton, and Arabidopsis homologs (Espelund et al. 1992; Gaubier et al. 1993).

Genetic mapping of the Sle gene family

DNAs isolated from G. max (A81-356022) and G. soja (PI 468.916) were subjected to restriction endonuclease digestion and analyzed by hybridization to the Sle1 cDNA probe. The Sle1 cDNA hybridized to several (3–6) genomic fragments for all of the restriction enzymes tested (Fig. 1). This observation is in agreement with other studies indicating that most Group-1 Lea genes appear to be members of multi-gene families (Espelund et al. 1992; Futers et al. 1990; Gaubier et al. 1993; Litts et al. 1987).

In order to identify all of the members of the Sle gene family in soybean, we screened two independent IEMBL3 genomic libraries under low-stringency conditions (final wash in 2 x SSC, 0.1% SDS, 50 °C). Extensive restriction mapping of several clones enabled us to identify four different non-overlapping genomic clones containing Sle-hybridizing fragments. A detailed restriction map of these four clones is presented in Fig. 2. The clones were designated Sle1-4, with Sle1 being the clone that most resembles the Sle1 cDNA restriction map (this was later confirmed by sequence analyses). These four clones explained all but 1 of the six HindIII and 5 EcoRI fragments observed in Fig. 1. Therefore, we also screened an EcoRI size-fractionated library constructed in lgt10 vector. This allowed us to isolate a fifth clone, Sle5, as a 2.5 kb EcoRI fragment (Fig. 2).
Figure 1 shows that the Sle-hybridizing fragments are polymorphic between the two soybean species tested. RFLP mapping of the two Dral and the EcoRI polymorphisms displayed in Fig. 1 revealed three independent Sle loci mapping to linkage groups D1, G, and N of the USDA-ARS soybean RFLP linkage map (Fig. 3). Based on the restriction map of the clones and the absence of polymorphisms among the cv ‘Resnik’, cv ‘Williams’, and A81-356022 G. max genotypes (data not shown), we assigned a specific clone to each one of the mapped polymorphisms. The fourth IEMBL3 clone, Sle4, was also mapped to the RFLP linkage map by mapping a low-copy TaqI fragment contiguous to the Sle hybridizing region (Fig. 2). All four Sle genes mapped to independent linkage groups. We were unable to unambiguously map the fifth clone, Sle5, but preliminary data indicated that this locus may map to a fifth linkage group (unpublished results).

Sequence analysis of genomic clones

Results from DNA sequence comparison among the five genes are presented in Table 1. Sequence identity throughout the coding regions was very high among all clones. However, the clones were divided into two classes based on the sequence identity of the coding regions. Sle2 and Sle3 were more related to each other (87.2% identity) than to the remaining clones (average of 72.9% identity). Sle1, Sle4, and Sle5 comprised the second class of clones and showed higher homology to each other (average of 88.4% identity) than to either clone Sle2 or Sle3. Comparison of the deduced polypeptide sequences of each Sle locus (partial sequence for Sle5 polypeptide) showed that they differed primarily in that Sle1 and Sle5 had a repeated 20-amino acid hydrophilic motif. The Sle4 locus encoded a polypeptide containing a premature stop codon and therefore was likely to be a pseudogene.

A single intron of variable length was found in the same position in all five clones (Fig. 4) as was found for other Sle homologs (Galau et al. 1992; Gaubier et al. 1993; Litts et al. 1991, 1992; Wurtele et al. 1993). Sle5 has an unusually large intron; over 1,700 bp (Fig. 4; data not shown). We were unable to determine the complete sequence of the Sle5 intron because the 2.5-kb fragment comprising the Sle5 clone terminates at its 5′ end within the intron. Intron lengths for clones Sle1, 2, 3 and 4 were 250, 509, 811 and 249 bp, respectively (Fig. 4). Intron lengths from previously sequenced Sle
Table 1 DNA sequence identity (%) among the members of the *Sle* gene family

<table>
<thead>
<tr>
<th></th>
<th>Coding region</th>
<th>Intron</th>
<th>3'-UTR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sle2</td>
<td>Sle3</td>
<td>Sle4</td>
</tr>
<tr>
<td><em>Sle</em>1</td>
<td>71.1</td>
<td>77.5</td>
<td>88.7</td>
</tr>
<tr>
<td><em>Sle</em>2</td>
<td>87.2</td>
<td>72.8</td>
<td>67.7</td>
</tr>
<tr>
<td><em>Sle</em>3</td>
<td>76.8</td>
<td>70.8</td>
<td>89.4</td>
</tr>
<tr>
<td><em>Sle</em>4</td>
<td>89.4</td>
<td>42.0</td>
<td>89.4</td>
</tr>
</tbody>
</table>

*Fig. 3* Locations of four *Sle* loci in the soybean USDA-RFLP linkage map. *Numbers to the left* of each linkage group represents distances between markers, in centiMorgans.

homologs in six different plant species have ranged from 71 bp in wheat (Litts et al. 1991) to 185 bp in *Arabidopsis* (Gaubier et al. 1993). Sequence identity of *Sle*1, *Sle*4, and *Sle*5 introns averaged 52.7% (Table 1). *Sle*4 and *Sle*5 introns showed the highest identity (65%), differing almost exclusively by the presence of three insertions in the longer *Sle*5 intron. *Sle*2 and *Sle*3 introns showed 44% identity. Sequence identity across introns from the two classes was found to be very low (31%) and thereby also corroborated the existence of two different classes.
Fig. 4 See page 963 for legend
DNA sequence of the five Sle genes. Intron sequences are shown in lowercase. Sequences of the deduced polypeptides are shown underneath each DNA sequence. The 20-amino acid hydrophilic motif is underlined. The start and the end of the Sle1 cDNA is shown with an arrowhead above the Sle1 sequence.

Sequences downstream of the translation stop site of each gene (Table 1) had a high overall sequence identity within each class, averaging 57.7% for Sle2 and Sle3, and 54.1% for the three remaining genes. When the comparison was made across the two classes, this value dropped to 24.8%. As observed for the introns, Sle4 and Sle5 3'-UTR were more identical to each other than to Sle1 3'-UTR.

Sle mRNA accumulation in embryos

Sle mRNA accumulation was examined by Northern analyses of total RNA extracted from vegetative tissues and developing seeds of soybean using the Sle1 cDNA as a probe under low-stringency conditions (final wash in 2 × SSC, 0.1% SDS at 50 °C). This probe was known to cross-hybridize with all other Sle mRNA under these conditions (data not shown). No hybridization was detected in the RNA from leaf, stem, or root tissues (data not shown). A single band corresponding to mRNA of approximately 850 bp was detected in seed tissues (Fig. 5). To better localize the accumulation of Sle mRNA in the seed, we dissected mature dry seeds and analyzed total RNA isolated from the embryonic axis and from the cotyledons by RNA hybridization analysis. Fig. 5 shows that approximately equal amounts of Sle mRNA accumulated in the two embryonic tissues.

The steady-state levels of Sle mRNA decreased rapidly after imbibition (Fig. 6), as do other Group-1 Lea genes (Raynal et al. 1989; Williams and Tsang 1991). By 32 h post-imbibition, Sle mRNA could not be detected in the seedling. In contrast, actin mRNA accumulated during germination. Hybridization with the Sle1 cDNA probe showed that Sle mRNA could be detected at relatively low levels in 25 DAF seeds which have already entered the early-cotyledonary stage of seed development (Fig. 7). The Sle mRNA steady-state level remained constant throughout seed development until a noticeable increase occurred around 80 DAF, at a time when seeds had reached maximal size and were starting to turn...
yellow in the region surrounding the embryo-axis. At 87 DAF maximum levels of Sle mRNA were detected. At 105 DAF, seeds were fully matured and dried, and levels of Sle mRNA were still high. Sle mRNA reached its maximum levels before any appreciable loss of fresh weight had occurred in the seed. In contrast, expression of the β-conglycinin storage protein peaked much earlier than Sle and was steadily decreasing by the time Sle expression started to increase. The results depicted on Fig. 7 shows that although all four genes were expressed, their temporal pattern of expression differed considerably, with Sle1 and Sle5 being expressed much earlier in seed development than Sle2 and Sle3.

Sequence analysis indicated that Sle4 is a pseudogene. To confirm that Sle4 mRNA was not present we performed rare transcript (RT)-PCR (Frohman et al. 1988) analysis on developing seeds using a set of primers which would specifically amplify mRNA derived from Sle1 (as an internal control of the RT-PCR reactions) and Sle4 loci (Fig. 8A). Since these two clones differed by the presence of a 20-amino acid repeat, we predicted that their RT-PCR products would differ by 60 bp (199 and 139 bp, respectively; Fig. 8A). The results showed that only one band could be seen after PCR amplification (Fig. 8B, lane 4). The band size corresponded to the expected amplification product from a Sle1-derived mRNA (199 bp). That the set of primers and reaction conditions chosen were adequate to amplify the Sle4 gene was shown by the control reactions performed with genomic DNA as template (Fig. 8B, lanes 1 and 3), where the bands corresponding to amplification of the two genes were present. Therefore, the absence of the expected Sle4 mRNA band in the RT-PCR reaction was strong evidence that Sle4 mRNA did not accumulate at appreciable levels in the cell.

Sle mRNA accumulation in embryos and vegetative tissue in response to water deficit

Seed desiccation has been suggested as the requirement for completion of seed maturation (Finkelstein et al. 1987), and thus Lea gene expression may be associated with this physiological change. However, desiccation of the embryos in the pod, as described by Rosenberg and
Fig. 8A, B RT-PCR analysis of expression of Sle1 and Sle4 loci.

A Schematic representation of the Sle1 and Sle4 coding regions (blank boxes) and introns (shaded boxes). The arrows denote the location of binding sites for the primers (P0 and P1/4) used to amplify both loci. The 20-amino acid hydrophilic motif is represented by a hatched box. Predicted size of PCR products are indicated; numbers in parentheses are the expected size of fragments after splicing of the intron from each locus.

B Ethidium bromide-stained (1.3%) agarose gel showing the results after the PCR amplification using either soybean genomic DNA (lane 1 from cv ‘Williams’; lane 3 from cv ‘Williams 82’) or first-strand cDNA from soybean (cv ‘Williams’) developing seed (lane 4) as template for the reactions. The 123-bp ladder appears in lane 2.

Rinne (1988), resulted in a reduction of Sle mRNA accumulation in 30 DAF seeds (Fig. 9A, lane 4). Water deficit did not induce Sle mRNA accumulation to detectable levels in leaves of 5-week-old plants (Fig. 9B). Furthermore, cotyledons, hypocotyls, and roots of 4-day-old seedlings did not accumulate Sle mRNA to detectable levels when subjected to water deficit (data not shown).

Discussion

The gene family presented here represents the largest Em-like gene family described in plants and the first among legume species. On the basis of sequence identity, the five members of the soybean Sle gene family were grouped into two classes. It has been suggested that soybean is an ancient allotetraploid, and putative homeologous genes have been molecularly characterized for several gene families in soybean (Grandbastien et al. 1986; Lee and Verma 1984; Nielsen et al. 1989). The separation of the Sle gene family into two distinct groups based on sequence similarity and temporal expression pattern supports the hypothesis of a tetraploid origin of the soybean. However, since the two ancestral genomes are not known, we cannot unambiguously test this hypothesis.

Sequence and RT-PCR analyses showed that Sle4 was a pseudogene and Sle4 mRNA did not accumulate in the cell. We do not know whether Sle4 is being transcribed. Frameshift mutations that lead to the introduction of premature stop codons have been held responsible for the absence of mRNA accumulation in soybean (Jofuku and Goldberg 1989) and Phaseolus vulgaris (Voelker et al. 1990). Therefore, it is possible that the nonsense point mutation in Sle4 is preventing Sle4 mRNA accumulation in soybean seeds.

The high levels of nucleotide sequence conservation observed among the introns of the Sle4 pseudogene and its closest homologs (Sle1 and Sle5) suggests that the point mutation that originated the premature stop codon may be a relatively recent event. Thus, a functional allele of Sle4 may still be present elsewhere in the soybean germplasm.

The proteins encoded by the Sle gene family differ mostly by the repetition of a 20-amino acid internal hydrophilic motif. A similar situation occurs with the
homologous barley B19 (Espelund et al. 1992), cotton D19 (Galau et al. 1992), and Arabidopsis AtEm (Gaubier et al. 1993) proteins, with some proteins having as many as four of the 20-amino acid repeats. It has been proposed for the B19 proteins of barley that their variable number of hydrophilic repeats, their differential response to ABA and osmotic stress, and their different mRNA abundance impart different functions in the cell (Espelund et al. 1992). Our data show that in soybean the two classes of expressed Sle genes have different temporal patterns of expression, with the genes encoding proteins that contain two copies of the hydrophilic motif (Sle1 and Sle5) being expressed at much earlier stages of seed development than the genes containing a single copy of the motif (Sle2 and Sle3).

Dessication of immature zygotic embryos in the pods induces the expression of many genes normally expressed in mature soybean embryos, including some Lea genes (Hsing et al. 1992; Rosenberg and Rinne 1988). Under similar conditions of water deficit, we were not able to induce Sle mRNA accumulation. This lack of induction of Sle mRNA accumulation in soybean zygotic embryos by water deficit contrasts with the situation observed in cotton (Hughes and Galau 1991), wheat (Berge et al. 1989), barley (Espelund et al. 1992), and maize (Butler and Cuming, 1993), in which desiccation and osmotic stress induce Group-1 Lea mRNA accumulation.

In vegetative organs, the expression of many of the Lea genes coding for Group 2, -3 and -4 Lea proteins is induced by water deficit. This, together with the accumulation of Lea mRNAs in seeds just prior to desiccation and structural analysis of the encoded proteins, has led to the suggestion that Lea proteins may protect the mature plant under conditions of water deficit (Dure 1993). However, despite similarities in the timing of expression, each group of Lea genes encodes a type of protein of distinct structure and potentially diverse function. The failure of vegetative organs to respond to water deficit by Sle mRNA accumulation indicates that the Group-1 Lea proteins in soybean may not function in situ to protect the vegetative plant against water deficit.

The identification of two distinct groups of sle genes supports the hypothesis of the tetraploid origin of the soybean genome. These groups differ in sequence divergence and expression patterns. Cloning and initial characterization of a soybean gene family that has been intensively studied in a broad range of species may provide a means by which we can study the effect of genome diploidization on expression of duplicated genes.

Acknowledgements The authors would like to thank Drs. D. Shibata and R. F. Whitter for providing us with a cDNA library, Dr. R. Meagher for the soybean actin cDNA, pSAC-7, and Dr. K. Polzin for assistance in assembling the IEMBL3 restriction maps. E.S.C. is grateful for his financial support provided by the Brazilian National Council for Scientific Development (CNPq).

References


Keim P, Diers BW, Olson TC, Shoemaker RC (1990) RFLP mapping in soybean: an association between marker loci and variation in quantitative traits. Genetics 126: 735–742


Litts JC, Colwell GW, Chakerian RL, Quatrano RS (1991) Sequence analysis of a functional member of the Em gene family from wheat. DNA Sequence-J DNA Sequencing and Mapping 1: 263–274


