Plant Molecular Biology 7: 451–465, 1986

Two soybean ribulose-1,5-bisphosphate carboxylase small subunit genes share extensive homology even in distant flanking sequences

M.A. Grandbastien[1], S. Berry-Lowe[2], B.W. Shirley & R.B. Meagher*
Department of Genetics, University of Georgia, Athens, GA 30602, U.S.A.
[1]*Present address: Biologie Cellulaire, CNRA Route de Dt. Cyr, 78000 Versailles, France*
[2]*Present address: Carlsberg Laboratory, Department of Physiology, GL. Carlsberg Vej10 DK-2500, Copenhagen Valby, Denmark*

Keywords: alloalleles, DNA sequence, homeologous alleles, light regulation, multigene family, ribulose-1,5-bisphosphate carboxylase small subunit

Summary

Soybean contains a multigene family which encodes the small subunit of ribulose-1,5-bisphosphate carboxylase (RuBPCss). A member of this gene family, SRS4, has been isolated from a soybean genomic DNA library. Its nucleotide sequence has been determined and compared to the sequence of SRS1, a previously characterized RuBPCss gene from soybean. Relevant regulatory sequences such as the PuPuCCAAT boxes, TATA box, the actual start of transcription and poly A addition sites are conserved between the two genes. Using a gene specific synthetic probe to the 3′ flanking region the steady state mRNA levels of SRS4, like SRS1, are shown to be very high in light grown soybean seedlings and low in seedlings grown in darkness. SRS1 and SRS4 are very closely related, the three exons being 96%, 93% and 96.5% homologous in nucleotide sequence. The polypeptide sequences are nearly identical with only one amino acid change in each of the three exons encoding the 178 amino acid precursor polypeptide. The two introns are about 75% homologous and the flanking regions are more than 85% homologous (700 base pairs on the 5′ end and 300 base pairs on the 3′ end). Furthermore, hybridization studies between lambda clones containing the SRS1 and SRS4 genes reveal that a region of strong homology extends at least 4 kb on the 5′ end and about 1.1 kb on the 3′ end. We propose that these two genes may be alloalleles or homeologous alleles. This proposal is consistent with soybean having an allotetraploid origin, and would imply that the divergence of two ancient Papilionoidae species gave rise to these two genes.

Introduction

Ribulose-1,5-bisphosphate carboxylase (EC 4.1.1.39) catalyzes the carboxylation and hydrolytic cleavage of ribulose-1,5-bisphosphate to form two molecules of 3-phosphoglycerate. This soluble enzyme is composed of eight large and eight small subunits (33). In vascular plants and green algae the small subunit of ribulose-1,5-bisphosphate carboxylate (RuBPCss) is encoded by the nuclear genome while the large subunit is encoded in the chloroplast genome (35). The steady state levels of small subunit mRNA and protein appear to be positively controlled by light. Light and phytochrome are known to control the level of transcription in pea (24), *Lemna* (50, 54, 58) and soybean (6). In those plants in which it has been examined, including pea, soybean, petunia, *Lemna* and wheat, the small subunit is encoded by a small multigene family (5, 9, 10, 21, 55). There is little evidence for significant functional differences between the proteins encoded by these multigene family members.

Soybean contains greater than six RuBPCss gene family members. Analysis of a soybean multigene family is further complicated by the fact that soybean may have an allotetraploid origin, i.e., two ancient species of Papillionoideae ($2N \cong 20$) may have fused their genomes to produce the soybean genus, *Glycine* ($2N + 2N' = 40$) (7, 15, 17,19, 25,

*Person to whom offprint requests should be addressed.

32, 53). This fusion would result in the presence of related pairs of alloalleles or homeologous alleles in the soybean genome. One soybean gene, SRS1, has been characterized in detail and shown to be composed of three exons and encodes a mature small subunit polypeptide, which differs in 20 to 30 of its 128 amino acid risidues from the mature sequences found in pea, petunia, *Lemna* and wheat. The levels of small subunit mRNA in light grown seedlings are induced strongly over levels found in seedlings grown in darkness (5). The transcription of SRS1 in isolated nuclei has been shown to be controlled over a 32 fold level by light, phytochrome and darkness (6). In soybean, pea and *Lemna* far red light can be used to block the induction of RuBPCss transcription produced by short exposure to white light (6, 24, 50, 54). Additionally, far red light exposure has been shown to turn off RuBPCss transcription in soybean seedlings grown in continuous light (6).

The regulation of the various gene family members in petunia has been examined and shown to vary greatly in the steady state levels of mRNA produced in the light (21). However, whether the presence of an RuBPCss multigene family represents a fortuitous accident of evolution or results from a selection for differentially regulated small subunit genes is not clear. To further our exploration of the soybean RuBPCss multigene family we have examined a second soybean RuBPCss gene, designated SRS4. We show here that SRS4 is closely related to SRS1 and that the steady state levels of its mRNA are also strongly light regulated. It is possible from the data presented herein that SRS1 and SRS4 are functional light regulated homeologous alleles.

Materials and methods

Isolation and mapping of the lambda clone SRS4 and plasmid subcloning

The RuBPCss clone λSRS4 was isolated from a Charon 4A soybean library as described for λSRS1 (5). Insert fragments from the pea cDNA clone pSS15 (9) and the soybean clone pSRS2.1 containing the 5′ half of the soybean RuBPCss gene (5) were used as probes during the initial screening. The insert of the soybean small subunit clone pSRS0.8 containing the 3′ half of a soybean RuBPCss gene was then used as a probe to confirm the identity of these clones (5).

Phage DNA was purified and physically mapped with restriction endonucleases as described in Nagao *et al* (37, 42). The six EcoRI fragments from SRS1 were cloned into pBR325 plasmid (8) as described for pSRS2.1 and pSRS0.8 (5). The 1.5 kb HindIII and the 4.5 kb EcoRI fragments from λSRS4 containing the polypeptide coding region of the gene were subcloned into plasmid pUC13 in the *E. coli* strain JM101 (56) and plasmid DNA prepared (37).

Characterization of the SRS4 gene

Restriction mapping and fragment purification of plasmid inserts was performed as described by Smith and Birnstiel, (52) with minor modifications as described by Shah *et al.* (49). Fragments were labeled, either by kinasing their 5′ ends or by Klenow reaction with their 3′ ends (49). Labeled DNA fragments were then sequenced according to Maxam and Gilbert (36). The 5′ and 3′ ends of the mRNA was mapped on labeled DNA fragments (23).

Analysis of sequence data

Sequence data was analyzed on a SUN/Intelligenetics Workstation using the DNA sequence analysis program, SEQ. Optimum alignment between SRS1 and SRS4 and homology data were obtained using combinations of the Search subprogram and then Align sub-program. The polypeptide encoding regions of the two genes, SRS1 and SRS4, were analyzed for percent silent and replacement nucleotide substitutions according to Perler *et al.* (44) using a program prepared for us by Ken Rice.

Expression of the SRS4 gene

Gene specific synthetic DNA probes were prepared for SRS4 and a soybean 18S rRNA gene SR1 (22) on an Applied Biosystems Oligonucleotide Synthesizer and purified by acrylamide gel electrophoresis (36). Synthetic probes were labeled by kinasing with [γ-^{32}P]-ATP (34) and used to analyze dot blots of total RNA prepared from plants grown in light and darkness (5, 29, 34).

S1 nuclease mapping

The 5′ end of the SRS4 small subunit mRNA and the 3′ end of the SRS1 small subunit mRNA were mapped by determining the size of the S1 nuclease resistant hybrids formed between the poly (A+) RNA and the template DNA. The DNA fragment used for 5′ S1 analysis of SRS4 was a 450 bp *Sau*3A fragment (Grandbastien *et al.*, unpublished). The fragment used as the DNA template for the 3′ S1 hybridization analysis of SRS1 was the DdeI fragment from pSRS0.8 (5).

Hybridizations were carried out according to Favaloro (23) in a volume of 10 μl covered with ster-

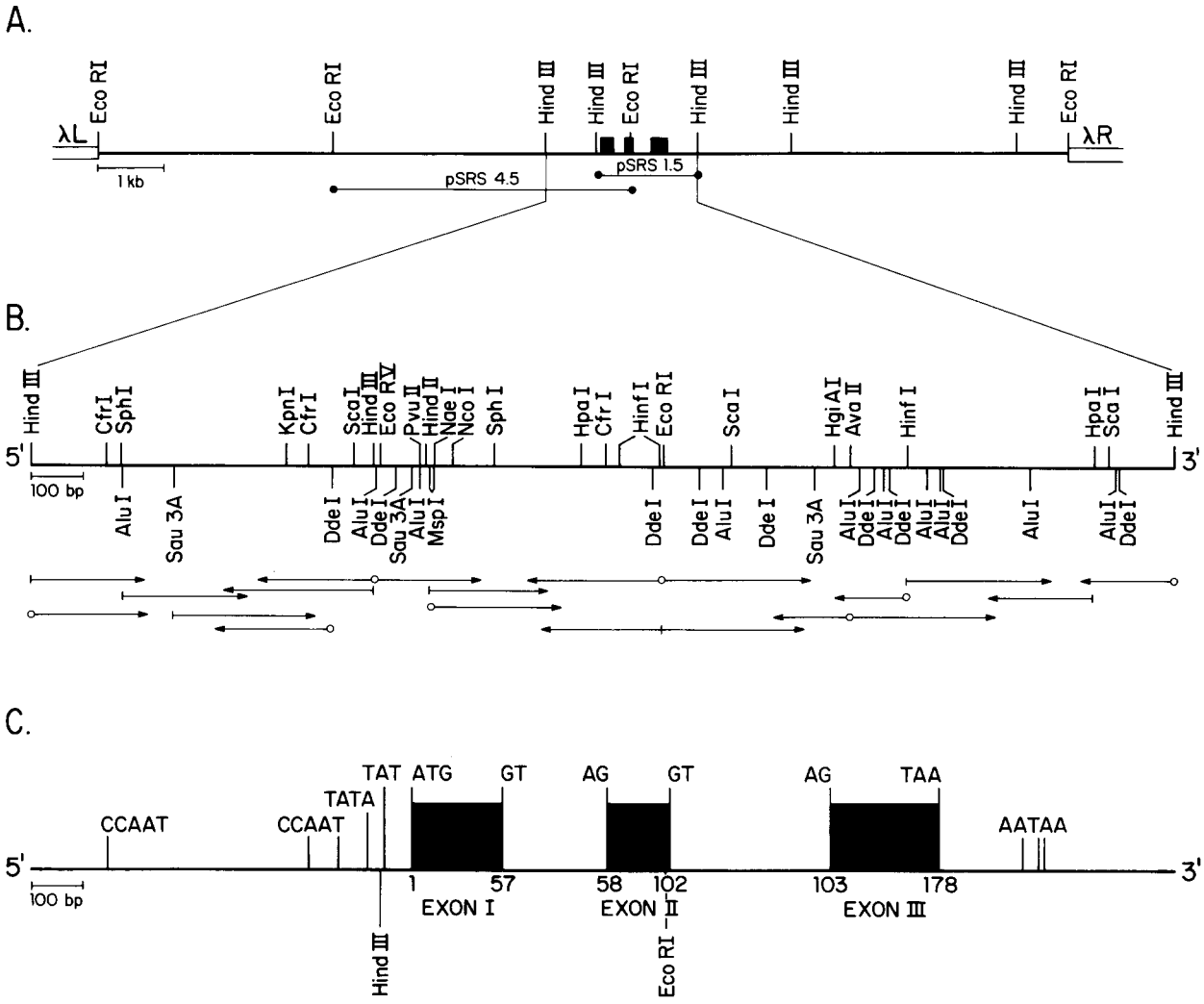


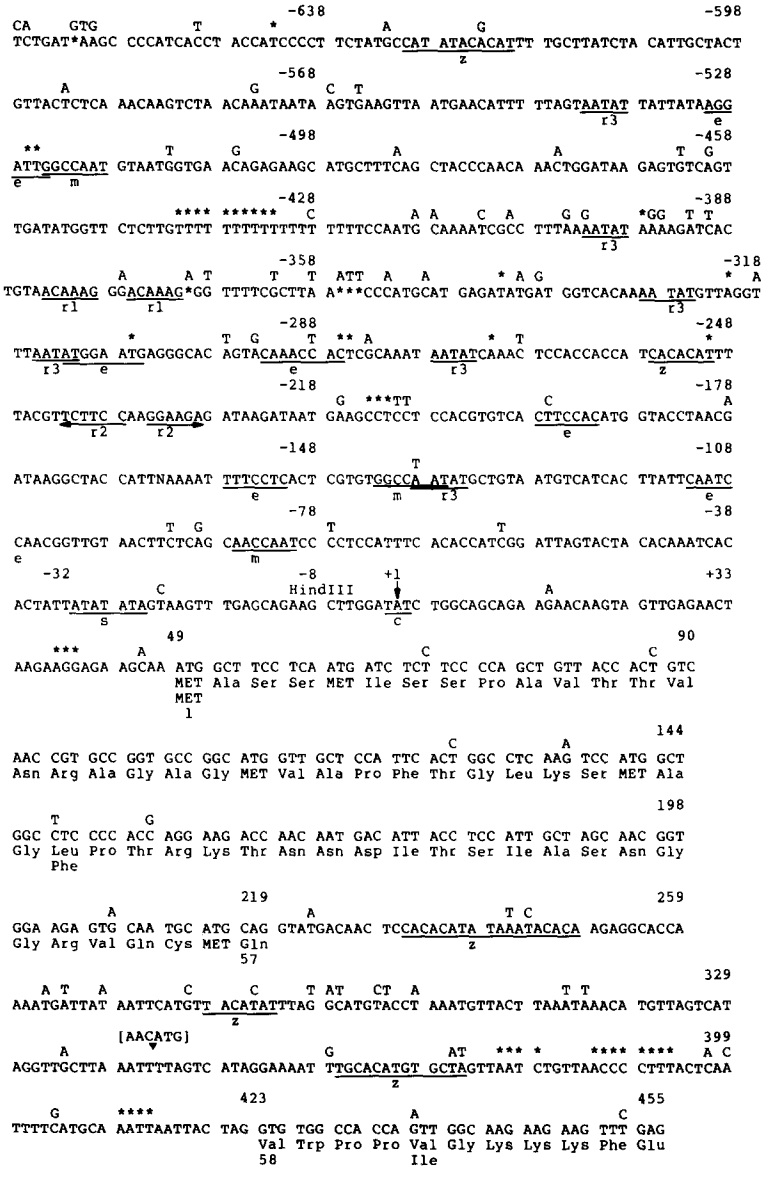
Fig. 1. Physical map of the soybean clone λSRS4 containing the RuBPCase small subunit gene SRS4. Figure 1A shows the EcoRI and HindIII restriction pattern of the 14.5 kb insert of λSRS4. The two fragments subcloned in pUC13, pSRS4.5 and pSRS1.5, that have been used for sequencing are also shown. Figure 1B shows the detailed restriction map of the sequenced region as determined by Smith-Bernstiel restriction mapping and by the sequencing data. The sequencing strategy is represented by the arrows below the map. The 5′-end kinase labeling is noted by a bar and 3′-end Klenow labeling is noted by an open circle. Figure 1C is colinear to the detailed map, and represents the diagram of the gene structure derived from analysis of the sequencing data. Translated portions of exons are schematized by raised blocks. The codon numbers are indicated below each exon. DNA sequences of potential importance are indicated.

ile oil in an eppendorf tube. Five to 10 pmoles of DNA, and 5 μg of poly(A+) RNA, or 25 μg of total RNA was used for the hybridization. The SRS4 5′ end hybridization was carried out at 50 °C. The SRS1 3′ end hybridization was carried out at 46, 48, and 50 °C. All hybridizations were carried out for a minimum of 5 h. The S1 digestion was performed with 50 units of S1 nuclease (Sigma) for 30 min at 37 °C. Resistant hybrids were visualized on a denaturing urea gel after ethanol precipitation.

Results

Isolation and characterization of the SRS4 gene

The soybean RuBPCss gene SRS4 was isolated from a Charon 4A genomic library using a combination of pea and soybean small subunit probes. The restriction pattern of λSRS4 (Fig. 1A) revealed that it had a unique fragment pattern, distinct from the pattern previously described for λSRS1 (5). Southern Blot analysis of these restriction patterns



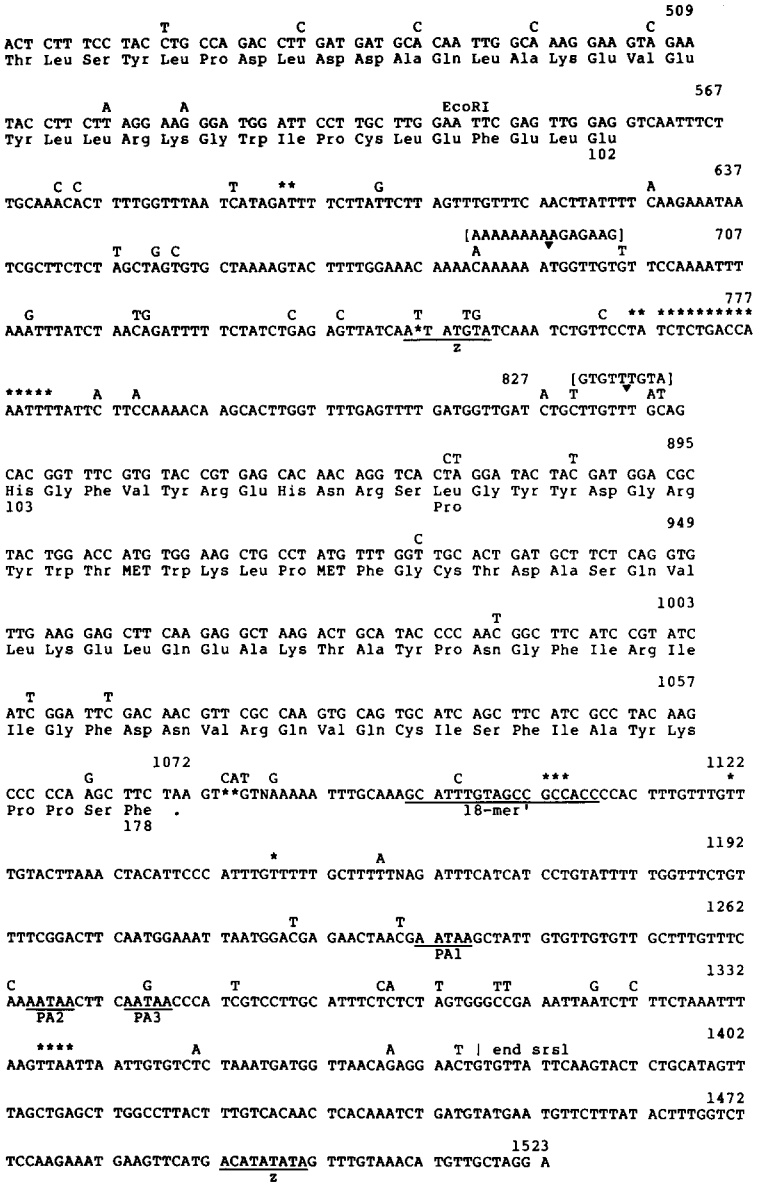


Fig. 2. Complete nucleotide sequence of the soybean RuBPCss gene, SRS4, is presented and compared to SRS1. The sequence of 2190 bp contains flanking sequence, exons, and introns of SRS4 and shows the sequence of the presumed translation product. The nucleotide sequence differences observed in the soybean RuBPCss gene, SRS1, are indicated above the SRS4 sequence. Amino acid differences for SRS1 are indicated below the translation of SRS4. Insertion/deletion differences are indicated by an asterix (*) in the shorter sequence. Three large insertions in the SRS1 sequence relative to SRS4 are shown above both sequences in brackets. Nucleotide numbering is given for SRS4 above the sequences with the start of transcription labeled as +1. Negative numbers are used to identify the 5′ untranscribed flanking sequence. Codon numbers are given below the translation of SRS4 and are the same for both SRS1 and SRS4. Sequences of particular importance are underlined, given letter designations and are discussed in the text.

was then performed at low stringency (6X SSC, 56 °C), using the pea cDNA insert from pSS15 that encodes the entire mature sequence as a probe. It showed the presence of a homologous sequence, located on a 1.5 kb HindIII fragment, and on two contiguous 4.5 kb and 6.4 kb EcoRI fragments (data not shown).

The recombinant pUC13 plasmids carrying the 1.5 kb HindIII fragment and the adjacent upstream 4.5 kb EcoRI fragment were designated pSRS1.5 and pSRS4.5, respectively, and were believed to encode the entire SRS4 gene. The fragments from within these inserts were purified on 4% acrylamide gels. Relevant restriction fragments were sequenced (Fig. 1B) to produce the complete RuBPCss gene sequence of SRS4 shown in Fig. 2. Comparison of the SRS4 gene sequence with other known small subunit sequences and the presumed translation product with known small subunit gene and protein sequences revealed the complete gene structure shown in Fig. 2 and summarized in Fig. 1C.

Analysis of the SRS4 sequence

The SRS4 nucleotide sequence, totaling 2190 bp, was compared with the RuBPCss gene sequence of SRS1 (5) plus additional sequence determined for the 5′ region of SRS1 (Fig. 2). Figure 2 shows a number of mismatches and small deletion/additions between the two sequences.

The comparison of these two nucleotide sequences (Fig. 2) is summarized in Fig. 3. It appears from these data that the two genes are very closely related. The predicted coding sequences of SRS4 consist of three translated regions whose sizes are

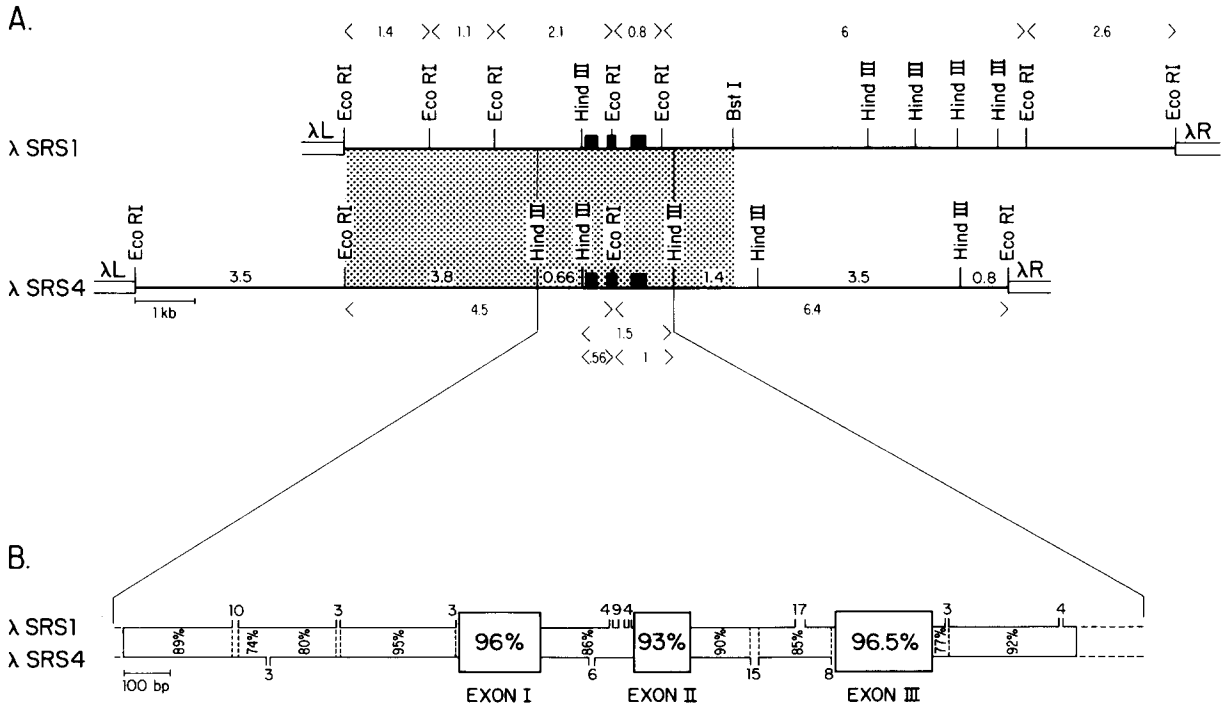


Fig. 3. Summary of homology between SRS1 and SRS4 gene regions. Figure 3A shows the colinearity of the restriction maps for EcoRI, HindIII and BstI sites in SRS1 and SRS4. The grayed area between the two clones represents the extent of homology confirmed by Southern hybridizations. No homology has been detected outside this region. Figure 3B is an expanded map of the two sequenced gene regions showing the homologous zones as calculated from sequencing data. The translated regions are drawn as raised boxes. Deletions larger or equal to 3 bp are indicated by small openings accompanied by the deletion size in base pairs. The percent homology between each sequence has been calculated by the Align program of Bio-Intelligenetics. We have arbitrarily defined each of the homologous regions shown here and given an average percent nucleotide homology. The sizes of a number of relevant restriction fragments is given in kb.

identical to the translated portions of the three exons in SRS1 (171, 135 and 231 bp) that encode a precursor protein of 178 amino acids. The two introns are located at exactly the same positions (after codons 57 and 102). Introns 1 and 2 in SRS4 are 203 and 284 bp long respectively as compared to 193 and 290 bp in SRS1. Fig. 3B shows the percent homology between various portions of the two sequences. The three exons are 96%, 93% and 96.5% homologous in nucleotide sequence, with only one altered amino acid in each exon: codon 34 is a Leu in SRS4 instead of a Phe; codon 62 is Val in SRS4 instead of Ile; and codon 114 is Leu in SRS4 instead of Pro.

Using the quantitative divergence analysis methods of Perler *et al.* (44), we were able to separate the silent nucleotide substitutions from the replacement substitutions and weight them according to their probability of occurrence. By this analysis, SRS1 and SRS4 differ by 12% in silent nucleotide substitutions and 0.9% in replacement nucleotide substitutions relative to the number of potential silent or replacement nucleotide sites. Because of this degree of similarity we have not attempted to correct for multiple hit kinetics. This relatively low predicted level of multiple hits allows a calculation of the number of transitions to transversions to be made. The expected random ratio of transitions to transversions is 1:2. Transition to transversion ratios for the observed nucleotide substitutions between SRS1 and SRS4 was 2.2:1, demonstrating the same bias observed in animal genes (39, 44). In other words, either more transitions are allowed to accumulate or more transitions actually occur than are predicted from random substitution of bases.

The degree of homology between the pairs of introns in both genes is also very high: 85% to 90%, if the gaps and deletions larger than 3 bp are not taken into account. Some insertion-deletion differences between the sequences are moderate (15 and 17 bp in intron 2), but these gaps separate very homologous sequences.

High percentages of sequence homology extend into the flanking sequences of SRS1 and SRS4. A highly homologous zone (95%) extends approximately 320 nucleotides upstream from the start codon (from -270 to +50; Fig. 2). A moderately homologous zone (74%-80%) extends from -270 to -450. At -450 a highly homologous region (90%) starts again.

S1 nuclease analysis

S1 nuclease analysis demonstrates that the 5′ presumptive start of SRS4 transcription lies 48 bp upstream from the ATG codon (Fig. 4). The 5′ terminus of the SRS4 transcript correlates with the previously determined 5′ terminus of the SRS1 transcript located 45 base pairs upstream from the start of SRS1 translation (5). As observed for SRS1, the levels of the SRS4 specific product were higher in RNAs isolated from seedlings grown in the light than in those grown in darkness. The 5′ terminus of the transcript was the same for RNA from light- and dark-grown plants. The 3 nucleotide deletion

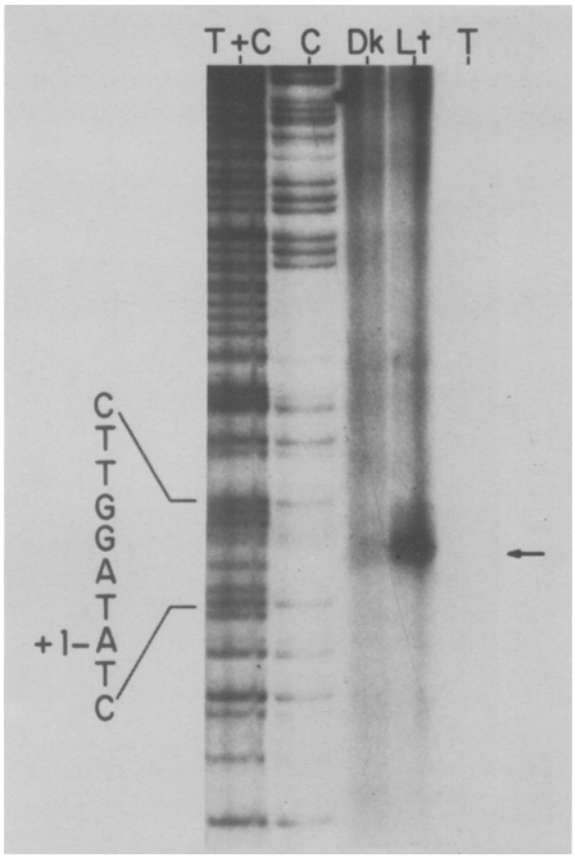


Fig. 4. S1 nuclease mapping of the 5′ end of the SRS4 small subunit gene. A 450 base pair *Sau*3A fragment from the 5′ region of SRS4 was purified, labeled with polynucleotide kinase, subjected to Maxam and Gilbert sequencing reactions, and resolved on a 7% urea-acrylamide gel. Lane T+C and lane C contain the products from two of the four reactions. The S1 treated hybrids formed between the HaeIII-Sau3A fragment and 5 μg poly(A+) RNA from dark-grown plants (DK), 5 μg poly(A+) RNA from light-grown plants (LT), or 10 μg total yeast tRNA (T) are shown.

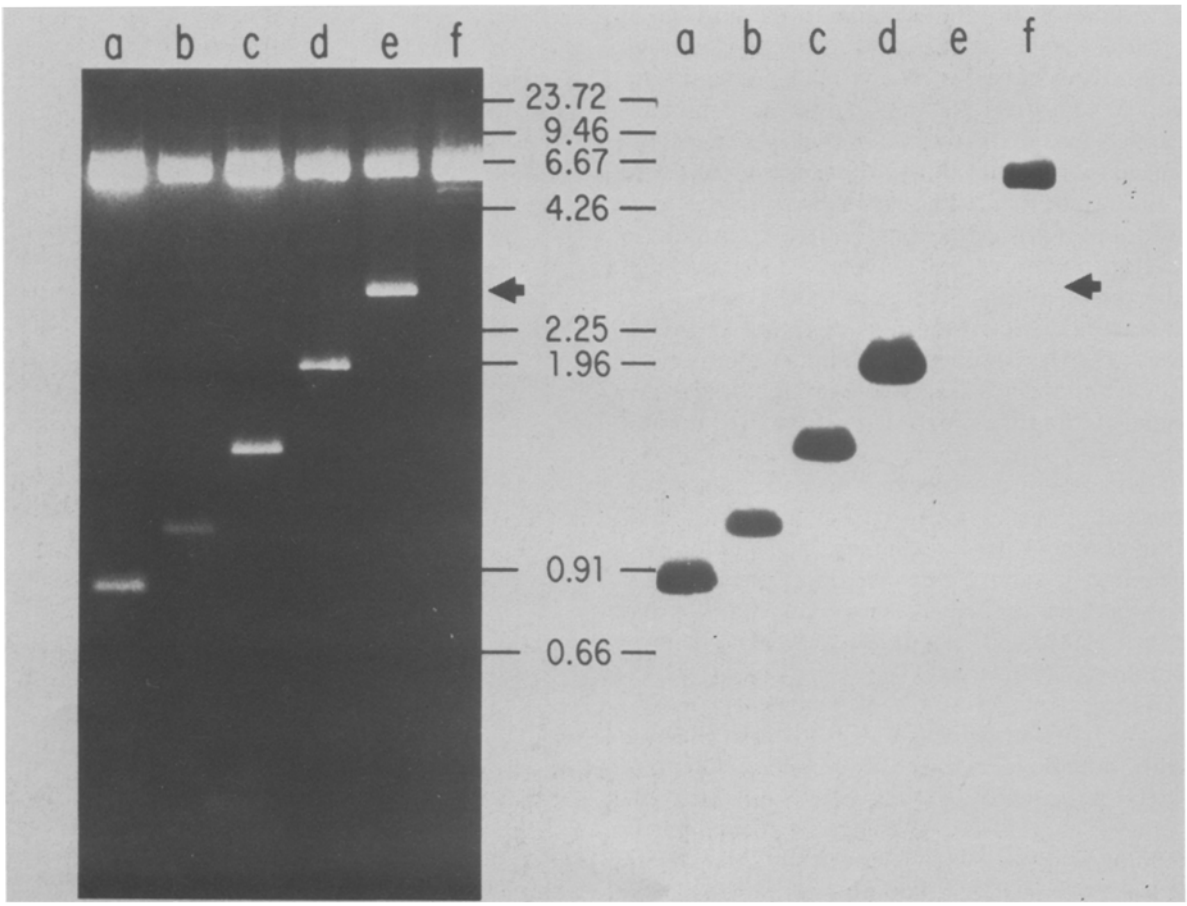
which is present in the SRS1 mRNA leader relative to the presumptive SRS4 mRNA leader should prevent any SRS4 DNA/SRS1 mRNA hybrids from surviving S1 nuclease treatment under the conditions used (22).

Repeated S1 nuclease mapping on the 3′ end of SRS1 gave a smear of protection products in the region containing the two polyadenylation site consensus sequences, AATAA (data not shown). No attempt was made to map the 3′ end of the SRS4 mRNA, which contains two homologous AATAA sequences and a third short distance downstream (see below).

SRS1 and SRS4 genes show extensive sequence homologies even in distant flanking regions

The extremely high level of homology between the SRS1 and SRS4 gene sequences is accompanied by some notable similarities in the restriction maps of the lambda clones (Fig. 3A). In order to determine whether sequence homology extended upstream and downstream from the sequenced regions, λSRS1 and λSRS4 restriction fragments were compared by hybridization experiments at low stringencies. The clones containing these fragments, pSRS1.4, pSRS1.1, pSRS2.1, pSRS0.8, pSRS6 and pSRS2.6, are shown in Fig. 3A.

EcoRI digests of the recombinant plasmids containing the six individual SRS1 EcoRI fragments shown in Fig. 3 were probed with the ^{32}P-labeled λSRS4 clone. The results are shown in Fig. 5A. All of the *Eco*RI insert fragments from SRS1 hybridized with the λSRS4 probe, except the 2.6 kb EcoRI fragment indicated by the arrow. The results of this experiment show that homology can be found along most of the colinearizing maps of λSRS1 and λSRS4. The fact that the 2.6 kb EcoRI



fragment of λSRS1, which maps outside of the colinearizing portion of the two clones, does not hybridize with the probe is consistent with this map. Similarly, the 3.5 kb EcoRI fragment from SRS4, which is located outside of the colinearizing area of the two sequences (see Fig. 3A), did not hybridize when probed with ^{32}P labeled λSRS1 (data not shown).

We confirmed and extended these results by using each of the plasmids containing the six cloned EcoRI fragments from λSRS1 as probes of the EcoRI, HindIII, and EcoRI/HindIII double digests of SRS4. The results are illustrated in Fig. 5B. In order to more precisely determine the extent of homology in the 3′ region, ^{32}P-labeled λSRS4 was used to probe multiple restriction digests of the pSRS6 plasmid containing the 3′ region of SRS1 (Fig. 5C). Although the data are not presented, all the experiments shown in Fig. 5 were also carried out at lower stringency levels (56 °C, 6X SSC). The results were the same as shown in Fig. 5). The results of these experiments are summarized in Fig. 3A. We can conclude that the detectable 5′ flanking sequence homology between the two genes is limited by the size of the λSRS1 clone and that the detectable 3′ flanking sequence homology does not extend far beyond the BstI site in SRS1.

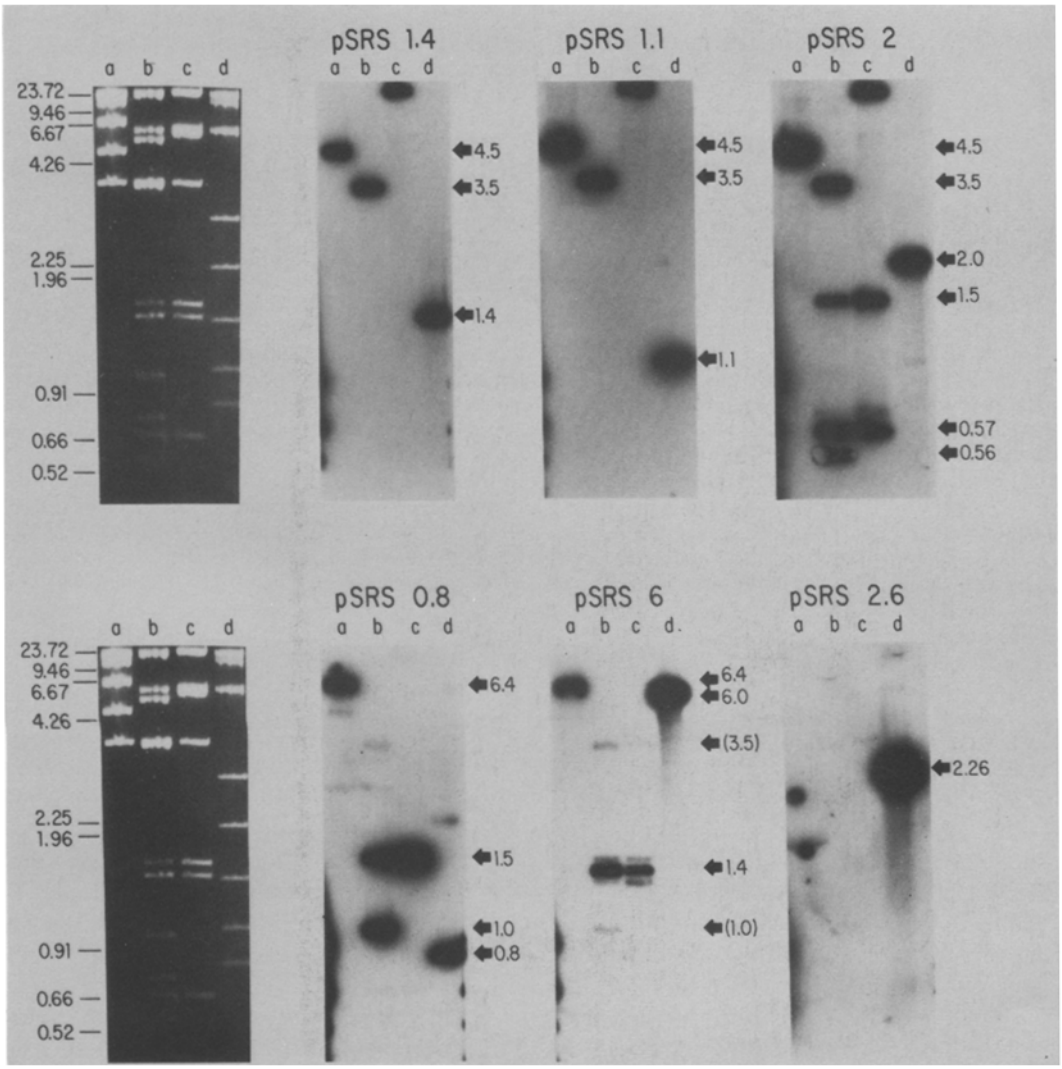


Fig. 5B.

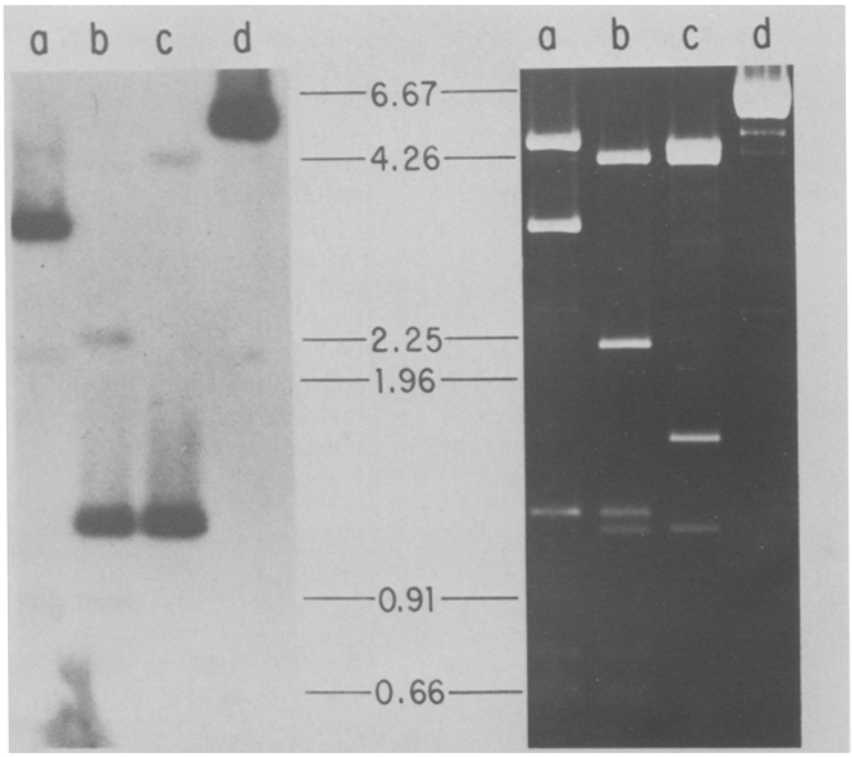


Fig. 5C.

Fig. 5. Extent of homology between SRS1 and SRS4 in distant flanking sequences as determined by Southern hybridization.

Fig. 5A: Hybridization of λSRS4 to EcoRI fragment of λSRS1. EcoRI digest of recombinant pBR325 plasmids carrying each EcoRI fragment of λSRS1 were electrophoresed in an 0.8% agarose gel, imprinted onto nitrocellulose and probed with ^{32}P-labeled λSRS4. The ethidium bromide stained gel is shown on the left, the autoradiograph of the nitrocellulose filter imprint is shown on the right. Lane a, pSRS0.8: lane b, pSRS1.1: lane c, pSRS1.4: lane d, pSRS2.1: lane e, pSRS2.6: lane f, pSRS6. The 6.0 kb EcoRI fragment contained in pSRS6 is not resolved from the linearized 5.9 kb pBR325 plasmid in which it is cloned. The filter was hybridized and washed in 3xSCC at 56 °C. Sizes of the HindIII lambda DNA and AluI pBR322 DNA standards are indicated in the margin. The arrow indicates the 2.6 kb λSRS1 fragment that does not hybridize with λSRS4.

Fig. 5B: EcoRI and HindIII single and double digests of λSRS1, probed with each of the six recombinant plasmid carrying EcoRI fragments of λSRS4, (see Figure 3). In each case, the entire plasmid has been ^{32}P labeled. Lane A: EcoRI digests of λSRS4. lane b: EcoRI/HindIII double digest of λSRS4. lane c: HindIII single digest λSRS4. lane d: Control EcoRI digest of λSRS1. Six identical agarose gels were imprinted to nitrocellulose and the filters were hybridized and washed at medium stringency (3 × SSC, 56 °C). Sizes of the HindIII lambda and AluI pBR322 DNA standards are indicated in the margin. One of the ethidium bromide stained 0.8% agarose gels is shown at the left of each row for comparison.

Fig. 5C: Analysis of the 3′ boundary of homology. pSRS6 digested with a) EcoRI/HindIII; b) EcoRI/BstI/HindIII; c) EcoRI/BstI; d) EcoRI; imprinted to nitrocellulose and probed with λSRS4. Hybridization at 3 × SSC, 56 °C.

Steady state levels of SRS4 mRNA

RNA from soybean seedlings grown in light and darkness were analyzed for SRS4 gene specific transcripts. We needed to distinguish between transcripts produces from the closely related SRS4 and SRS1 gene sequences. Consequently, a gene specific 18-mer was synthesized which was complementary to a region in the SRS4 3′ transcribed-non-translated mRNA flanking sequence, which varies by four nucleotides from the SRS1 sequence (labeled 18-mer, Fig. 2). Figure 6B illustrates the specificity of the SRS4-3′ 18-mer probe. The oligomer does not hybridize significantly to SRS1 even under low stringency conditions. From previous genomic hybridization studies using SRS1 and SRS4 gene fragments as probes, we know that there are no other RuBPCss gene sequences closely related to SRS1 and SRS4 in the soybean genome suggesting that this is a gene specific oligomer (4). The possibility that another distantly related gene contains a sequence homologous to one of these probes has not

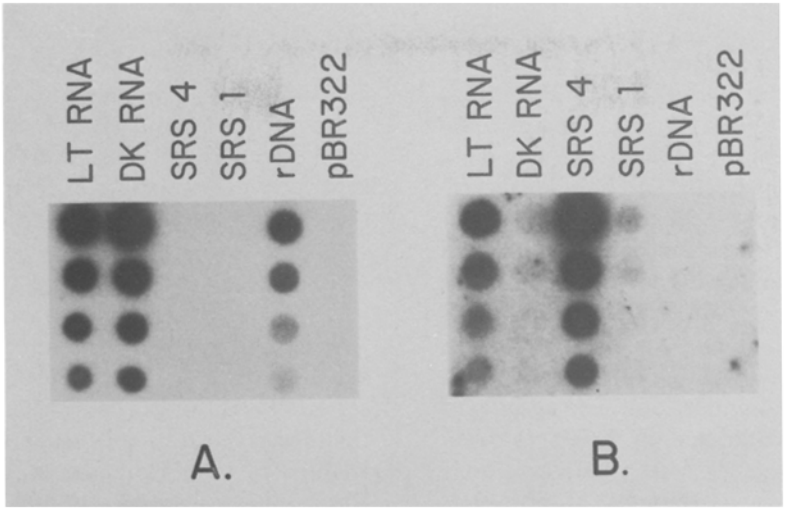


Fig. 6. Light induction of the level of mRNA from SRS4 as measured by hybridization to a SRS4 gene specific 18-mer (6B). Control blot 6A is prepared identically to 6B but is probed with an 18S rRNA specific 40-mer. Total RNA from light- (LT) and dark- (DK) grown 10 day seedling leaves (.02 μg first RNA dots, A; 10 μg first RNA dots, B) SRS4 (pSRS1.5), SRS1 (pSRS0.8), 18S rDNA (pSR1.1) and pBR322 plasmid DNAs (0.01 pmoles first dot) were applied in a two fold dilution series to a Biodyne A filter by suction blotting (DNA) or spotting (RNA). ^{32}P-kinase 18-mers were hybridized for 18 h at 60 °C in 0.25% dry milk/2 × SSC/0.1% SDS, then washed 2 × 10 min at 23 °C, then 1 × 1 min at 60 °C in 1 × SSC/0.1% SDS.

been ruled out. It can be seen in Fig. 6B that the SRS4-3′ 18-mer hybridizes to approximately 10 times more RNA in samples taken in the light than in RNA samples taken in darkness. A universal 18S rRNA specific deoxyoligonucleotide complementary to nucleotides 1764–1803 from a soybean 18S rDNA gene (22) was used to probe an identical dot blot (Fig. 6A). The probe is gene-specific. The data in Fig. 6A demonstrate that the same amount of rRNA is present in the light- and dark-grown total RNA samples and that equal amounts of total RNA from both samples were loaded and bound to the filter.

Discussion

Soybean RuBPC small subunit gene structure: The organization of the three exons encoding the SRS4 precursor protein is identical to most other small subunit genes in dicots (5, 12, 21). The first exon of SRS4 encodes the entire transit peptide and the first two amino acids of the mature protein. Only one amino acid change occurs between the previously determined soybean small subunit transit sequence of SRS1 and the SRS4 transit sequences (5). The lack of conservation for most transit sequences in contrast to the high degree of overall conservation of these two RuBPCss sequences suggests that SRS1 and SRS4 share a recent common ancestral gene (38).

Origin of two soybean RuBPCss genes

The overall sequence homology between SRS1 and SRS4 is similar to that found for a pair of related pea genes and a pair of closely related petunia genes (12, 21). Unlike the petunia and pea genes, however, these two soybean genes have not been shown to be closely linked in the genome (18, 45). The lambda clones λSRS1 and λSRS4 contain 4 kb and 7.5 kb of 5′ flanking sequence and 9 kb and 6 kb of 3′ flanking sequence, respectively. There is no evidence of sequence overlap within these limits. An independent lambda clone containing the SRS1 gene, designated λSRS3, contains another 7 kb of 3′ flanking sequence which does not overlap the λSRS4 clone (5).

While it is likely that the linked and closely related pairs of pea and petunia genes are recently derived by gene duplication, the two soybean genes may have a very different origin. There is substantial evidence that soybean arose from the fusion of two diverged Papilionoideae species with a 2N chromosome number of 20. Thus the 40 chromosomes of soybean may be designated as 2N + 2N′

= 40. Genetic evidence supports this hypothesis by demonstrating that there are related alleles which do not segregate in soybean (15, 17). These potential homeologous alleles or alloalleles may have come from two independent Papilionoideae species. Recently, molecular evidence supporting this hypothesis has come from examining a number of different soybean gene families. The glycinin gene family apparently has a pair of independent genes which share distant flanking sequences and may be homeologous alleles (R.B. Goldberg and R.G. Fischer, personal communication). The actin gene family has been shown to contain three very divergent classes of actin genes with a pair of closely related genes in each class (29). We propose here that the relationship between SRS1 and SRS4 is that of homeologous alleles. SRS1 and SRS4 each would have to have come from a different Papilionoideae subspecies during the formation of the soybean genes, *Glycine*.

A more definitive proof that SRS1 and SRS4 have an alloallelic relationship could come from *in situ* hybridizations to soybean chromosomes. Hybridization to SRS1 or SRS4 gene specific probes should identify two chromosomes each, whereas hybridization to both probes simultaneously should identify four different chromosomes.

If the proposed alloallelic relationship for SRS1 and SRS4 is correct, then the degree of divergence of these two genes should reflect the time of divergence of the two ancient Papilionoideae species from a common ancestor. The formation of *Glycine* would have brought these two genes together again. The nucleotide divergence of 10–20% which was found in the 5′ and 3′ flanking region and in the introns agrees well with the 10–12% silent nucleotide substitution observed in the exons. In animal species the rate of silent or unselected nucleotide substitution ranges from 1% per 0.4 to 1.0 million years (MY) (39, 44). Using these data, we can estimate that these two genes diverged 10 to 30 MY ago. This gene divergence time would then reflect the divergence time for the two Papilionoideae species and indicate an ancient origin for soybean. Considering that the Papilionoideae have been dated back 50–60 MY (16, 41), this large predicted divergence time is plausible. It should be noted that there is not sufficient data in the literature on recently diverged plant genes to make an estimate of silent nucleotide substitution rates in plants. Considering the extensive sequencing data accumulating for plant genes, these values should be forthcoming and revised estimates can be applied to our initial calculations for the divergence time for these two soybean small subunit genes and the two hypothesized Papilionoideae species.

Common sequence elements

The presumptive 5′ promotor sequences in SRS4 are shown in Fig. 1C, and are similar to what is observed in SRS1 (Fig. 2, a,b,c). The start of transcription lies at the *A* (+1) of the *TAT* sequence located 48 bp upstream from the initiation codon ATG. The homologous TAT site of SRS1, located 45 bp upstream from the ATG codon (Fig. 2, c), was shown by S1 protection studies to contain the start of transcription (5). This TAT sequence fits the consensus CAT box sequence, PyPuPy, known for most animal genes (20). SRS4 has the TATATATA sequence identical to that found in SRS1 and closely related to the eukaryotic Goldberg Hogness concensus sequence, TATAA, thought to specify the start site of transcription (ref. 1, 46; Fig. 2, s). The first A of this sequence is -32 bp upstream from the transcription start site in SRS4, whereas it is only 31 bp upstream in SRS4. This spacing is typical of that found in animal genes (11). Both SRS4 and SRS1 (Fig. 2, m) contain the identical AACCAAT sequence at -86 bp upstream from the transcription start site. This sequence is homologous to the PuPuCCCAAT sequence found in many animal gene transcription units and is thought to modulate levels of transcription (2, 20, 26). SRS4 contains an additional GGCCAAT at -141 and both genes contain another repeat of this sequence at -524.

The 3′ end of the SRS4 sequence contains 3 presumptive polydenylation sites (3) with the sequence AATAA (Fig. 1C and Fig. 2; PA). The first two are located at positions similar to the two AATAA sequences found in SRS1 (+159 and +192 bp from the stop codon TAA of SRS4, see PA_1, and PA_2, at 1232 and 1265 in Fig. 2). A third potential site is seen in SRS4, at +301 from the TAA stop codon (PA in Fig. 3). The presence of these multiple polyadenylation sites may have caused the smear of S1 products observed above for SRS1.

These two small subunit genes each contain a relatively large number of alternating purine/pyrimidine tracts in their 5′ flanking se-

quences (Fig. 2, z). Tracts of alternating purines and pyrimidines are found in all eukaryotes examined (27) and have been shown to enhance transcription of several eukaryotic genes and to play an important role in the structure of viral enhancers (42, 28). Both genes contain 5′ flanking sequences homologous to an independent animal viral enhancer consensus sequence $GTGG^{AAA}_{TTT}G$ or its inverse complement (56; Fig. 2, e). This sequence has also been found in pea, wheat and petunia RuBPCss genes (14, 18) and in a light inducible pea chlorophyll a/b binding protein gene (51).

Three sequence elements with potential interest in regulation of these genes exist in the 5′ flank of both genes. The sequence ACAAAG is present as a direct repeat (-378, Fig. 2, r_1). The sequence at -236, *TCTTCC* AA *GGAAGA*, contains a six nucleotide inverted repeat (Fig. 2, r_2). The sequence AATAT is repeated six times in SRS4 and four times in SRS1 (Fig. 2, r_3). Some of these sequences are found in the 5′ flanking regions of other light regulated genes (10, 14, 18, 31, 51).

Expression and regulation of the soybean gene, SRS4

A gene specific probe has been used to demonstrate that SRS4 makes a substantial amount of mRNA in light grown seedlings and low but detectable levels in soybean seedlings germinated and grown in darkness. There is an approximately 10-fold increase in mRNA levels in light grown tissue. We have already shown that transcription of SRS1 is strongly regulated by light and exhibits the classic phytochrome response (5, 6). In separate experiments to be described elsewhere (Berry-Lowe and Meagher, unpublished data), we have shown that the SRS4 gene is also light regulated at the level of transcription. Furthermore, SRS1 and SRS4 together account for at least 80% of the total RuBPCss mRNA in light grown soybean plants (5). As discussed above, the classical regulatory sequences such as the CCAAT, TATATATA, and TAT, AATAA and several interesting repetitive sequences are conserved between (Fig. 2) SRS1 and SRS4. It might have been expected that if SRS1 and SRS4 are homeologous alleles one of them would have lost function with time. A detailed analysis of the expression of these two genes and their potential regulatory sequences in transgenic plants should begin to clarify their functional significance in soybean.

Acknowledgements

We would like to thank Greg Schmidt, Carol Condit, Vance Baird, Glen Galau, Tad Schurr and Mike McLean for their support, suggestions and careful reading of the text, and Colleen McElfresh for preparing the manuscript. We are grateful to Ken Rice for writing the Div program used to calculate silent and replacement nucleotide differences and transition to transversion ratios. This work is in partial fulfillment of a Ph.D. degree in genetics for S.B-L. and B.W.S.

This work was supported in part by grants from the United States Department of Energy and Department of Agriculture-SEA. S.B-L. and B.W.S. were supported by a predoctoral training grant from the National Institutes of Health.

References

1. Benoist C, Chambon P: Deletions covering the putative promoter region of early mRNAs of simian virus 40 do not abolish T-antigen expression. Proc Natl Acad Sci USA 77:3865 – 9, 1980.
2. Benoist C, Chambon P: In vivo sequence requirements of the SV40 early promoter region. Nature 279:336, 1981.
3. Berget, SM: Are U4 small nuclear ribonuclear proteins involved in polyadenylation? Letters to Nature 309:179 – 182, 1984.
4. Berry-Lowe SL: The isolation and characterization of a ribulose-1,5-bisphosphate carboxylase small subunit gene in soybean. Dissertation, University of Georgia, 1985.
5. Berry-Lowe SL, Mc Knight TD, Shah DM, Meagher RB: The nucleotide sequence, expression and evolution of one member of a multigene family encoding the small subunit of ribulose-1,5-bisphosphate carboxylase in soybean. J Mol Appl Genet 1:483 – 498, 1982.
6. Berry-Lowe SL, Meagher RB: Transcriptional regulation of a gene encoding the small subunit of ribulose-1,5-bisphosphate carboxylase in soybean tissue is linked to the phytochrome response. Mol Cell Biol 5:1910 – 1917, 1985.
7. Beversdorf WD, Bingham ET: Male-sterility as a source of haploids and polypoids of *Glycine max*. Can J Genet Cytol 19:283 – 287, 1977.
8. Bolivar F: Construction and characterization of new cloning vehicles, III. Derivatives of plasmid pBR322 carrying unique EcoRI sites for selection of EcoRI generated recombinant molecules. Gene 4:121 – 36, 1978.

9. Broglie R, Bellemaire G, Bartlett SG, Chua N-H, Cashmore AR: Cloned DNA sequences complementary to mRNA encoding precursors to the small subunit of ribulose-1,5-bisphosphate carboxylase and a chlorophyll a/b binding polypeptide. Proc Natl Acad Sci USA 78:7304 – 8, 1981.
10. Broglie R, Coruzzi G, Lamppa G, Keith B, Chua N-H: Structural analysis of nuclear gene coding for the precursor to the small subunit of wheat ribulose-1,5-bisphosphate carboxylase. Biotechnol 1:55 – 61, 1983.
11. Breathnach R, Chambon P: Organization and expression of eukaryotic split genes coding for proteins. Ann Rev Biochem 50:349 – 384.
12. Cashmore T: Nuclear genes encoding the small subunit of ribulose bisphosphate carboxylase. In: Kosuge T, Meridith C, Hollander A (eds) Genetic Engineering of Plants. An Agricultural Perspective. Plenum Press, New York, 1983, pp 29 – 38.
13. Coruzzi G, Broglie R, Cashmore A, Chua N-H: Nucleotide sequences of two pea cDNA clones encoding the small subunit of ribulose-1,5-bisphosphate carboxylase and the major chlorophyll a/b-binding thylakoid polypeptide. J Biol Chem 258:1399 – 1402, 1983.
14. Coruzzi G, Broglie R, Edwards C, Chua N-H: Tissue-specific and light-regulated expression of pea nuclear gene encoding the small subunit of ribulose-1,5-bisphosphate carboxylase. EMBO Journal 3:1671 – 1679, 1984.
15. Crane CF, Beversdorf WD, Bingham ET: Chromosome pairing and associations at meiosis in haploid soybean (*Glycine max*). Can J Genet Cytol 24:293 – 300, 1982.
16. Crepet WL, Taylor DW: The diversification of the Leguminosae: First fossil evidence of the Mimosoidene and Papilionoideae. Science 228:1087 – 1089, 1985.
17. Cutter GL, Bingham ET: Effect of soybean male-sterile gene ms on organization and function of the female gametophyte. Crop Science 17:760 – 764, 1977.
18. Dean C, van den Elzen P, Tamaki S, Dunsmuir P, Bedbrook J: Differential expression of the eight genes of the petunia ribulose bisphosphate carboxylase small subunit multi-gene family. EMBO J 4:3055 – 3061, 1985.
19. Delannay X, Rodgers DM, Palmer RG: Relative genetic contributions among ancestral lines to North American soybean cultivars. Crop Science 23:944 – 949, 1983.
20. Dierks P, van Ooyen A, Mantei N, Weissmann C: DNA sequences preceding the rabbit B-globin gene are required for formation in mouse L cells of β-globin RNA with the correct 5′ terminus. Proc Natl Acad Sci USA 78:1411 – 1415, 1981.
21. Dunsmuir P, Smith S, Bedbrook J: A number of different nuclear genes for the small subunit of RuBPCase are transcribed in petunia. Nuc Acids Res 11:4177 – 4185, 1983.
22. Eckenrode VK, Arnold J, Meagher RB: Comparison of the nucleotide sequence of soybean 18S rRNA with the sequences of other small-subunit rRNAs. J Mol Evol 21:259 – 269, 1985.
23. Favaloro J, Freisman R, Kamen R: Transcription maps of polyoma virusspecific RNA: Analysis by two dimensional nuclease S1 gel mapping. In: Grossman L, Moldave K (eds) Methods in Enzymology. Academic Press, New York, 1980, Vol. 65, pp 718 – 749.
24. Gallagher TF, Ellis RF: Light stimulated transcription of genes for two chloroplast polypeptides in isolated pea leaf nuclei. EMBO J 1:1493 – 1498.
25. Gottlieb LD: Conservation and duplication of isozymes in plants. Science 216:373 – 380, 1982.
26. Grosveld GC, deBoer E, Shewmaker CK, Flavell RA: DNA sequences necessary for transcription of the rabbit B-globin *in vivo*. Nature 295:120 – 126, 1982.
27. Hamada H, Petrino MG, Kakunaga T: A novel element with Z-DNA forming potential is widely found in evolutionarily diverse eukaryotic genomes. Biochem. 79:6465 – 6469, 1982.
28. Hamada H, Seidman M, Howard BJ, Gorman CM: Enhanced gene expression by the poly(dT-dG) -poly(dC-dA) sequence. Cell 12:2622 – 2630, 1984.
29. Hightower RC, Meagher RB: Divergence and differential expression of soybean actin genes. EMBO Journal 4:1 – 8, 1985.
30. Hentschel C, Irminger J-C, Bucher P, Birnstiel ML: Sea urchia histone mRNA termini are located in gene region downstream from putative regulatory sequences. Nature 285:147-151, 1980.
31. Herrera-Estrella L, van den Broeck G, Maenhaut R, van Montagu M, Schell J, Timko M, Cashmore A: Light inducible and chloroplast-associated expression of a chimaeric gene introduced into *Nicotiana tabacum* using a Ti plasmid vector. Nature 310:115 – 120, 1984.
32. Hymowitz T: On the domestication of the soybean. Econ Bot 24:40B SB107.E1G1, 1970.
33. Jenson RG, Bahr JT: Ribulose-1,5-bisphosphate carboxylase-oxygenase. Ann Rev Plant Physiol 28:379 – 400, 1977.
34. Johnson DA, Gautsch JW, Sportsman JR, Elder JH: Improved technique utilizing nonfat dry milk for analysis of proteins and nucleic acids transferred to nitrocellulose. Gene Anal Techn. 1:3 – 8, 1984.
35. Kawashima N, Wildman SG: Studies on fraction 1 protein IV. Mode of inheritance of primary structure in relation to whether chloroplast or nuclear DNA contains the code for a choroplast protein. Biochim Biophys Acta 262:42 – 9, 1972.
36. Maxam A, Gilbert W: Sequencing end-labeled DNA with base-specific chemical cleavages. In: Grossman L, Moldave K (eds) Methods in Enzymology, Academic Press, New York, 1980, Vol. 65, pp 499 – 560.
37. Meagher RB, Shepherd RJ, Boyer HW: The structure of cauliflower mosaic virus 1. A restriction endonuclease map of cauliflower mosaic virus DNA. Virology 80:362 – 75, 1977.
38. Mishkind ML, Wessler SR, Schmidt GW: Functional determinants in transit sequences: Import and partial maturation by vascular plant chloroplasts of ribulose-1,5-bisphosphate carboxylase small subunit of *Chamydomonas*. J Cell Biol 100:226 – 234, 1985.
39. Miyata T, Hayashida H: Recent divergence from a common ancestor of human IFN-alpha genes. Nature 295:165 – 6, 1982.
40. Morelli G, Nagy F, Fraley RT, Rogers SG, Chua N-H: A short conserved sequence is involved in the light-inducibility of a gene encoding ribulose-1,5-bisphosphate

carboxylase small subunit of pea. Nature 315:200 – 204, 1985.
41. Muller J: Fossil-pollen records of extant angiosperms. Botanical Rev 47:1 – 142, 1981.
42. Nagao RT, Shah DM, Eckenrode VK, Meagher RB: Multigene family of actin-related sequences isolated from a soybean genomic library. DNA 1:1 – 9, 1981.
43. Nordheim A, Rich A: Negatively supercoiled simian virus 40 DNA contains Z-DNA segments within transcriptional enhancer elements. Nature 303:674, 1983.
44. Perler F, Efstratiadis A, Lomedico P, Gilbert W, Kolodner R, Dogson J: The evolution of genes: The chicken preproinsulin gene. Cell 20:555 – 66, 1980.
45. Polans NO, Weeden NF, Thompson WF: Inheritance organization and mapping of rbcS and cab multigene families in pea. Proc Natl Acad Sci USA 82:5083 – 5087, 1985.
46. Proudfoot NJ: Eukaryotic promoters? Nature 279:376, 1979.
47. Sasaki Y, Tomoda Y, Kamikubo T: Light regulates the gene expression of ribulose-1,5-bisphosphate carboxylase at the levels of transcription and gene dosage in greening pea leaves. FEBS Lett 173:31 – 35, 1984.
48. Shah DM, Hightower RC, Meagher RB: Complete nucleotide sequence of a soybean actin gene. Proc Natl Acad Sci USA 79:1022 – 6, 1982.
49. Shah DM, Hightower RC, Meagher RB: Genes encoding actin in higher plants: Intron positions are highly conserved but the coding sequences are not. J Mol Appl Gen 2:111 – 126, 1983.
50. Silverthorne J, Tobin EM: Demonstration of transcriptional regulation of specific genes by phytochrome action. Proc Natl Acad Sci USA 81:1112 – 1116, 1984.
51. Simpson J, Timko MP, Cashmore AR, Schell J, Van Montagu M, Herrera-Estrella L: Light-inducible and tissue-specific expression of a chimaeric gene under control of the 5′-flanking sequence of a pea chlorophyll a/b-binding protein gene. EMBO J 4:2723 – 2729, 1985.
52. Smith HO, Birnstiel ML: A simple method for DNA restriction site mapping. Nucl Acids Res 3:2387 – 98, 1976.
53. Sorrells ME, Bingham ET: Reproductive behavior of soybean haploids carrying the ms allele. Can J Genet Cytol 21:449 – 455, 1979.
54. Stiekema WJ, Wimpee CF, Silverthorne J, Tobin EM: Phytochrome control of the expression of two nuclear genes encoding chloroplast proteins in *Lemna gibba* L.G-3. Plant Physiol 72:717 – 724, 1983.
55. Stiekema WJ, Wimpee CF, Tobin EM: Nucleotide sequence encoding the precursor of the small subunit of ribulose-1,5-bisphosphate carboxylase from *Lemna gibba* L.G-3. Nuc Acids Res 11:8051 – 8061, 1983.
56. Viera J, Messing J: The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19:259 – 268, 1982.
57. Weiher H, Konig M, Gruss P: Multiple point mutations affecting the Simian virus 40 enhancer. Science 219:626 – 631, 1983.
58. Wimpee CS: PhD thesis. Organization and expression of light-regulated genes in *Lemna gibba* L. G-3. Univ of Calif, 1984.

Received 2 April 1986; in revised form 31 July 1986; accepted 6 August 1986.