The nuclear ribosomal DNA intergenic spacers of wild and cultivated soybean have low variation and cryptic subrepeats

Daniel L. Nickrent and Jennifer A. Patrick

Abstract: The intergenic ribosomal DNA spacers (IGSs) from cultivated soybean (*Glycine max*) and wild soybean (*Glycine soja*) were sequenced and compared with six other legumes. These IGS sequences were 1821 bp in length in *G. soja* and *G. max* cultivars Arksoy, Ransom, and Tokyo, and 1823 bp long in the *G. max* cultivar Columbus. These represent the smallest published plant IGS sequences to date. Two clones from each of the above five cultivars were sequenced and only 22 sites (1.2%) were polymorphic, thereby supporting previous work that showed low genetic variation in cultivated soybean. The amount of variation observed between different clones derived from the same individual was equal to the amount seen between different cultivars. The soybean IGS sequence was aligned with six other published legume sequences and two homologous regions were identified. The first spans positions 706–1017 in the soybean IGS sequence and ends at a putative promoter site that appears conserved among all legumes. The second is located within the 5′ external transcribed spacer, spans positions 1251–1823 in soybean, and includes sequences first identified as subrepeats IV-1 and IV-2 in *Vicia angustifolia*. Sequences homologous to these two subrepeats were identified among all legume species examined and are here designated "cryptic subrepeats" (CS-1, CS-2) given the range in similarity value (79–96% for CS-1 and 60–95% for CS-2). Comparisons of CS-1 and CS-2 sequences within individual species show that divergence (substitutional mutations, insertions, and deletions) is sufficiently high to obscure recognition of the repeat nature of these sequences by routine dot plot analytical methods. The lack of subrepeats in the 5′ half of the soybean IGSs raises questions regarding the role they play in transcription termination or enhancement.

Key words: *Glycine*, Fabaceae, ribosomal DNA, promoter.

Résumé : Les espaceurs intergéniques de l'ADN ribosomique (IGS) ont été séquencés chez le soya cultivé (*Glycine max*) et sauvage (*Glycine soja*) et ces séquences ont été comparées avec celles de six autres légumineuses. Ces séquences IGS mesuraient 1821 pb chez le *G. soja* de même que chez les cultivars Arksoy, Ransom et Tokyo du *G. max* alors qu'elles mesuraient 1823 pb chez le cultivar Columbus du *G. max*. Ces tailles sont les plus petites jamais publiées pour des séquences IGS d'origine végétale. Pour chacun des cinq cultivars, deux clones ont été séquencés et seuls 22 sites (1,2%) se sont avérés polymorphes. La variation observée entre différent clones dérivées du même individu était égale à celle observée entre différents cultivars, une observation qui vient appuyer les travaux antérieurs ayant montré un faible niveau de variation génétique chez le soya cultivé. Les séquences IGS de soya ont été alignées avec celles de six autres séquences publiées chez les légumineuses et ces comparaisons ont fait ressortir deux régions d'homologie. La première s'étend des positions 706 à 1017 de la séquence IGS du soya et se termine à l'endroit d'un possible promoteur qui semble être conservé chez toutes les légumineuses. La seconde région est située à l'intérieur de l'espaceur externe transcrit en 5′, s'étend des positions 1251 à 1823 chez le soya et inclut des séquences appartenant à ce qui a été appelé les sous-unités répétées IV-1 et IV-2 chez le *Vicia angustifolia*. Des séquences homologues à ces sous-unités répétées ont été identifiées chez toutes les espèces de legumineuses examinées et elles sont appellées ici «sous-unités répétées cryptiques» (CS) compte tenu des similitudes observées (79–96% pour la région CS-1 et 60–95% pour la région CS-2). Des comparaisons entre les séquences CS-1 et CS-2 à l'intérieur des diverses espèces montrent que leur divergence (substitutions, insertions et délétions) est suffisamment grande pour rendre difficile à déceler la nature répétée de ces séquences par des méthodes analytiques conventionnelles («dot plot»). L'absence de sous-unités répétées dans la moitié 5′ de l'IGS du soya soulève des interrogations sur le rôle que ces séquences répétées pourraient jouer dans la terminaison ou l'augmentation de la transcription.

Mots clés : *Glycine*, fabacées, ADN ribosomique, promoteur.

[Traduit par la Rédaction]

Corresponding Editor: R.J. Kemble.

Received August 11, 1997. Accepted December 1, 1997.

D.L. Nickrent[1] and J. A. Patrick. Department of Plant Biology, Southern Illinois University, Carbondale, IL 62901–6509, U.S.A.

[1] Author to whom all correspondence should be addressed (e-mail: nickrent@plant.siu.edu).

Introduction

The coding portion of the nuclear ribosomal rDNA cistron (18S, 5.8S, and 26S rDNA) has been extensively characterized at the sequence level for several hundred flowering plants, mainly owing to interest in using these data for phylogenetic analyses. Sequences include those from small-subunit (18S) rDNA (Nickrent and Soltis 1995; Soltis et al. 1997), large-subunit (26S) rDNA (Bult et al. 1995; Kuzoff, et al. 1998), and a large number of internal transcribed spacer (ITS) and 5.8S rDNA sequences (reviewed in Baldwin et al. 1995). In contrast, relatively few sequences exist for the intergenic spacer (IGS) that occurs between the tandem repeats of the coding portions of this cistron. The IGS is the most variable portion of the rDNA cistron and is sometimes referred to as the NTS (nontranscribed spacer), a misnomer, since large portions of it are transcribed. In contrast to the conserved 18S and 26S rDNA, which have phylogenetic utility at deeper divergence levels, IGS sequences often contain sufficient variation to allow examination of genetic relationships between plant varieties, cultivated varieties (cultivars), populations, and individuals (Rogers and Bendich 1987; Hemleben et al. 1988; Jorgansen and Cluster 1988; Schaal and Learn 1988). The IGS varies widely in length among different plant groups, ranging in size from ca. 1 to over 12 kb (Rogers and Bendich 1987). Such length heterogeneity is a result of the presence of at least one, and often many, tandem (direct) or dispersed subrepeat domains. The advent of PCR has precluded the need for cloning, thus greatly facilitating broad-scale sequencing studies of many genes; however, the length of plant IGS sequences and the lack of universally conserved sites among different species within IGSs may explain the relative paucity of these sequences.

The majority of IGS sequences generated to date have been from crop plants representing five major families: Poaceae (Appels and Dvořák 1982; McMullen et al. 1986; Toloczyki and Feix 1986; Barker et al. 1988; Procunier and Kasha 1990; Takaiwa et al. 1990), Brassicaceae (Bennett and Smith 1991; Gruendler et al. 1991; Lakshmikumaran and Negi 1994), Cucurbitaceae (King et al. 1993), Solanaceae (Schmidt-Puchta et al. 1989; Perry and Palukaitis 1990) and Fabaceae. For the legumes, IGS sequence and structural features have been characterized for *Vicia* (Kato et al. 1990; Ueki et al. 1992; Yakura and Nishikawa 1992), *Vigna* (Gernster et al. 1988), *Phaseolus* (Maggini et al. 1992), and *Pisum* (Kato et al. 1990). In this paper, we report the complete IGS sequence of cultivated and wild soybean, *Glycine max* and *Glycine soja*, respectively. Our goals were (*i*) to determine IGS sequences for these species, including several cultivars of *G. max*; (*ii*) to determine whether sufficient variation existed to address intercultivar relationships; and (*iii*) to compare the soybean IGS sequences to those from other related legumes.

Materials and methods

Plant accessions

In this study, *Glycine soja* (United States Department of Agriculture (USDA) plant introduction (PI) number 341–633A, obtained via J. Doyle, Cornell University, Ithaca, N.Y.) and four cultivars of *Glycine max* (Arksoy, Tokyo, Columbus, and Ransom) were used. For the latter, total genomic DNA was extracted from leaves following methods reported in Nickrent (1994). The sequences generated in this study are deposited with GenBank under the numbers AF021095–AF021104 (in which order they are listed in Table 1).

Primers and PCR amplifications

The following primer sequences were used for amplifying soybean IGSs: 3343 forward (5′-GGCCTTGCTGCCACGATCCACTGAG-3′), which occurs at the 3′ end of the large-subunit rDNA, and CNS-1 reverse (5′-CTGAGATTAAGCCTTTGT-3′), which occurs at the 5′ end of the small-subunit rDNA. These priming sites were identified by examining a wide range of angiosperms and other eukaryotes. Amplification reactions were performed in 50-μL volumes containing 10 mM Tris–HCl (pH 9.0), 50 mM KCl, 1.2 mM $MgCl_2$, 0.2 mM of each deoxynucleotide, 125 ng primer, 1.0 U Taq polymerase, and approximately 10 ng genomic DNA. The samples were overlaid with 30 μL of mineral oil. Amplifications were performed using the following parameters: 94°C for 3 min, followed by 35 cycles of 94°C for 1 min, 57°C for 1 min, and 72°C for 2.5 min.

Cloning PCR products

The 1.8 kb PCR amplification product containing the entire IGS region was ligated into the pCR™II vector and chemically transformed into One Shot™ competent cells using the Invitrogen TA Cloning® Kit. Positive colonies were picked and grown overnight in 2 mL of Luria-Bertani broth (Difco Laboratories) containing 50 μg/mL kanamycin. Plasmid DNA was isolated following the minipreparation method, as described by Sambrook et al. (1989). Recombinant plasmids were selected by digestion with *Eco*RI, followed by electrophoresis in 1.0% TBE (89 mM Tris-borate plus 2 mM EDTA, pH 8.0) agarose gels. The presence of a 1.8-kb band in addition to the 3.9-kb vector was considered a positive clone. Recombinant DNA was purified following an ethidium bromide procedure, as described in Stemmer (1991). The recombinant DNA was then linearized with either *Hin*dIII or *Xba*I, two restriction enzymes that do not cut within the IGS region (determined empirically). Recombinant DNA that was cut with *Xba*I was used with forward primers and recombinant DNA that was cut with *Hin*dIII was used with reverse primers for sequencing. A final phenol–chloroform extraction was performed prior to sequencing.

Sequencing

Sequencing reaction methodology was as described in the Sequenase® Version 2.0 DNA Sequencing Kit (United States Biochemical, Cleveland, Ohio). Either the M13 forward (5′-TCCCAGTCACGACGT-3′) or the M13 reverse (5′-AACAGCTATGACCCATG-3′) primer was used to seqence the ends of the 1.8 kb amplified product. Internal primers were then constructed to complete the sequence. These primers were IGS-1 forward (5′-GTGTCTTGCCATGTTATG-3′), IGS-2 reverse (5′-GGTAGGTAGGCAACACA-3′), IGS-3 forward (5′-GACTAAAGCATCACCGGGAA-3′), and IGS-4 reverse (5′-CCCACAATACCAAAGGCAT-3′). The Thermo Sequenase™ Cycle Sequencing Kit (Amersham LifeScience) was used to resolve sequence ambiguities resulting from gel compressions. The ligation-step parameters were 50 cycles of 96°C for 15 s and 55°C for 30 s. The termination-step parameters were 50 cycles of 95°C for 30 s and 72°C for 2 min. Electrophoresis and autoradiography followed standard procedures.

Sequence analysis

The *Glycine* IGS sequences were compared with each other and with other legume sequences available via GenBank: *Phaseolus coccineus* (accession No. X95967), *Vigna radiata* (X17211), *Vicia hirsuta* (X62122), *Vicia angustifolia* (X61082), *Vicia faba* (X16615), and *Pisum sativum* (X70806). Sequence comparisons were made following a multiple sequence alignment conducted manually with

Table 1. Summary of sequence variation in rDNA IGSs among clones of *G. max* and *G. soja*.

	Site																						
Clone	66	87	114	162	213	216	503	557	669	717	770	795	854	1049	1060	1177	1277	1347	1367	1411	1517	1538	1628
Arksoy-1	G	C	A	A	C	—	A	A	A	C	T	T	A	A	T	G	A	A	C	G	G	A	A
Arksoy-4	G	C	C	A	C	—	A	A	A	C	T	T	A	A	T	G	C	G	G	G	G	A	A
Columbus-2	T	T	C	G	T	GT	G	A	G	T	T	T	A	A	T	G	C	G	C	G	G	A	C
Columbus-3	T	T	C	G	T	GT	G	A	G	T	C	T	A	A	T	G	C	G	C	G	G	A	C
Ransom-1	G	C	C	A	C	—	G	A	A	C	T	T	A	T	T	A	C	G	C	C	A	G	C
Ransom-3	G	C	C	A	C	—	G	C	A	C	T	T	A	T	C	A	C	G	C	G	G	A	C
Tokyo-2	G	C	C	A	C	—	A	A	A	C	T	C	A	A	T	G	C	G	C	G	G	A	C
Tokyo-3	G	C	C	A	C	—	A	A	A	C	T	T	A	A	T	G	C	G	C	G	G	A	C
Soja-1 and Soja-2	G	C	C	A	C	—	A	A	A	C	T	T	G	A	T	G	C	G	C	G	G	A	A
Consensus	K	Y	M	R	Y	—	R	M	R	Y	Y	Y	R	W	Y	R	M	R	S	S	R	R	M

SeqApp (D.G. Gilbert 1993)[2]. MACVECTOR version 6.0 was used to generate a dot matrix plot of the soybean IGS sequence and to determine the presence of subrepeats, using the PUSTELL DNA MATRIX option (Pustell and Kafatos 1982), as well as for clustal alignment. The subrepeat sequences were first aligned using CLUSTAL and then adjusted manually, to minimize alignment scores and number of gaps and maximize the number of conserved identical sites. As determined empirically from the data, there was no consistent transition bias, hence substitutional changes were unweighted. These "optimal" alignments were then used to generate a modal sequence to be used for comparison. Each sequence was then compared with the modal sequence, and the numbers of matches, percent matches (proportion per total length), transitions, transversions, and gaps were determined. Transitions and transversions were assessed equal penalty values of –0.5 and gaps were penalized –0.33. In all comparisons, *V. angustifolia* had the highest score (i.e., was most similar to the modal sequence), hence this value was set as 100% and all other scores were scaled proportionally to provide a relative score.

Results and discussion

Sequence variation among the clones and cultivars

Amplification of all soybean genomic DNAs resulted in a single PCR product of approximately 1.8 kb in length. A single clear band, as seen by agarose gel electrophoresis, suggested that no significant length heterogeneity existed within individuals. This result was indirectly confirmed by cloning and sequencing at least two colonies derived from each 1.8 kb PCR product. The IGS sequence of three of the four *G. max* cultivars and *G. soja* was 1821 bp long, whereas the IGS sequence from the cultivar Columbus was 1823 bp long (Fig. 1). The longer sequence was the result of nine instead of eight GT repeats being present at position 216. Only 22 sites among the 1823 nucleotides (1.2%) were polymorphic. A summary of the variable sites by cultivar (Table 1) indicates that all polymorphisms are represented by ditypic sites, with transitions outnumbering transversions ca. 2:1. 'Columbus' was the most variable, showing variation at 9 of the 22 polymorphic sites. The apportionment of variation between different clones of the same cultivar and different cultivars was equal, thus highlight-

[2] D.G. Gilbert. 1993. SeqApp, a biological sequence editor and analysis program for Macintosh computers. Version 1.9a157. Biocomputing Office, Biology Department, Indiana University: ftp://inbio.bio.indiana.edu/molbio/seqapp/

Fig. 1. Consensus of nuclear ribosomal IGS sequences derived from four cultivars of *Glycine max* (Arksoy, Tokyo, Columbus, and Ransom) and *Glycine soja*. The length-variable region composed of eight or nine GT repeats is indicated with dashed underlining. Sequence that is alignable with other legume species is indicated with single underlining, and the putative promoter sequence is indicated with double underlining. Ambiguous nucleotides depict site variation across all sequences (see Table 1): K = G + T; Y = C + T; M = A + C; R = A + G; W = A + T; and S = C + G.

```
CCCTCCCCTC CAAATCACCA ACAAACCCTG TCCCCCCCAA AAAACCAATC 50
AATTCGGTAT CCCTTKCTTC TTCTACTATG TTTGGCYAAA AGAGGCTACC 100
GTTCGTCGGC TCTMTCTCGT TTTGCCTGCC TGGTGCTAAC TAGGTGTGTG 150
CCCGGGCGGA CRGACGACCA AGCCCCTCCT CCCCCTTGTG CCTGCGTGAG 200
TGTGTGTGTG TGYGTGTGCG TGTGACCTTG TGTCTTGCCA TGTTATGACA 250
TGATGATCCA CCACCGAGGC ACAGACTTAG CCAACGGGGC GCAAGCCAAA 300
GTTGGCAAAA TCCTGGCAAC AAAACTGATC ACGGGGGGGA GGAGGGGGCC 350
AACCAATCCG AAACTTAGCA CGGGGCCATA ACCCAGCACC GCCGACCCTA 400
GCCAAGCCAA CAACGGGCAA AAACCACGAA GTTGACCGGG GGGCAAAATT 450
TCAACCAACG GCTTGCCAAG ACCTAGCCAC GGGCCAAAAC CGACCACCGG 500
GCRCATGGAA AAACTTAAGC AATGGCTAAA CCAAGCAGCC CCGCAAACTT 550
GGCATCMGCC ATCATTTGTT TCTTTGCCTC AAAGAAGTTC CCCATAGACT 600
AAAAGCATCA CCGGGAAGGT TTGCAAGGCA AACCTTCACC ACACCTTCCG 650
TCATGGTCCA TGTTCATGRA TTTTTCAAAG TATGCCTTGC CGTTGGTTGT 700
TTGTCACGAA ATTTGGYGGG CACATTGGAT ATATTATTGG TGGATTGTCT 750
GCAAAAAATC ACGCCAAAAY TCATTTTCTA GGTGCCCTTT TTATYTTTTT 800
CGGATTTTAC GCATTTCGCC ACTTTAAAAA TTAATAAAAA AATATTCCCA 850
TGGRAATAAA TTACGAGCCT TGGGGAACTC CGAGATACCA TTTTTCTAAG 900
GGTGCCTGCA AAAAATCTCA TCAAAATTCG ATGTCTAGGT CCCAATTTTG 950
GCCTTATGTC TCCTCGCGAA AAGGTTAATC TACTCCTGTC AATCTGGAAA 1000
TGGCATATTA GCCTATTATA GGGGGGCACC ACACCCCCCC AACGAAAAWT 1050
TTTTTTTTTY TGAAAATGGA TGGGATTCTC ATTTCTAGTG TGGTGTGATG 1100
GCTCATTCCT TCAACAAAGA TGCATTGCAT GCCCAATAGC TTCCGTTTTC 1150
ACATTTAACG GATCCTAGGC GGTGTTRTCC TTGCTGAGGT GTTGTTGGCT 1200
GATTCACGAT TCACTCGACC GGATATGGGT TTTCACCCAA CAAAACTCAA 1250
TTTCGTGCCT AGCGATGCGG ATGAGGMGCC ACTCGATGTT TCCATTTGCC 1300
TGCATGCCTT TGGTATTGTG GGGTGTGGTT CGTTTAGTGG TGTCAARTCT 1350
TGCGAACGTG AGGGTGSATG AGTGGTAATC GGGATGTTTT TGTTGCACGG 1400
CTCCATGCTT GCGCATGCAA CGTGTGACAT GCATCCTTTT CATTGTCACC 1450
TTAGCCTTCG TGGCTTAGGT GCTGACGATT GGTTCCTGTG TTGCCTACCT 1500
ACCAAACGGT ATTCGTRTGG CATGTCAGAC CCATTCCRTC CAGGCCTCTT 1550
TGTTGGGGAC GCTGGAGGAT TTCAAAGTCC CACACGTGAA CCACGTTAGT 1600
CGTCCATCAT AACGGCTTCC TGTGGCAMAC GTGTGGTGAT TGTTCTCTCG 1650
GATGCGGTGC ATGCAGTGAA CGTGGGTGGT CTAGTGCCCC CACATTGTTC 1700
CTTAAAATCG GTTGTCTCGG ATGACGCTTG CCCTGGCGCT TGCTTTGTTC 1750
AACGCGATGC TACTACTCTT AGTTACGTAG CGCTGGCAAT GCTCGTGGCA 1800
GGTGCTTGGT CATTAACGGA TGC 1823
```

ing that as much variation resides within separate IGS cistrons in individual plants as resides among different cultivars. Polymorphisms unique to a particular clone (autapomorphies) were represented at 11 of the 22 polymorphic sites. Only two sites (503 and 1628) were potentially phylogenetically informative (synapomorphic), i.e., differences shared between at least two cultivars.

The low number of substitutions and lack of phylogenetic signal in the existing substitutions clearly indicate that IGS sequences cannot be used to distinguish among the large number of existing cultivars of soybean. Although our sampling was limited to only four *G. max* cultivars, the inclusion of wild soybean (*G. soja*) suggests that the range of variation had been adequately sampled. Doyle and Beachy (1985) observed that among 40 accessions examined, *G. max* and *G. soja* were indistinguishable from one another based upon rDNA repeat length and restriction patterns. The essentially identical IGS sequences for these two taxa reported here, and similarly identical ITS sequences (Nickrent and Doyle 1995), indicate that, with regard to genetic differentiation, *G. soja* and any cultivar of *G. max* are equivalent, thus casting further doubt upon their validity as distinct species. The inability to use IGS sequences to distinguish cultivars of soybean contrasts with other crop species that apparently contain greater genetic variation within the ribosomal loci. IGS sequences have been employed to distinguish cultivars of pumpkin (Torres and Hemleben 1991), potato (Schweizer and Hemleben 1989), radish (Tremousaygue et al. 1988), wheat (May and Appels 1987), and oats (Polanco and Pérez de la Vega 1997) as well as specific genotypes of rice (Cordesse et al. 1992). It has been noted (Delannay et al. 1983) that cultivated soybeans in the U.S. are derived from a very narrow germplasm base. It is estimated that a mere 12 plant introductions collected prior to 1945 contribute approximately 70–80% of the germplasm in northern and southern cultivars (Gizlice et al. 1993; Lorenzen et al. 1995). Molecular approaches that have documented the level of genetic diversity (or lack thereof) among cultivars include isozymes (Cardy and Beversdorf 1984), restriction fragment length polymorphisms (RFLPs) of the nuclear genome (Apuya et al. 1988; Keim et al. 1992), RFLPs of the chloroplast genome (Close et al. 1989), and RFLPs of the mitochondrial genome (Grabau et al. 1992; Hanlon and Grabau 1995). The results of the present study further document the very narrow genetic base from which North American soybeans have been derived.

Features of the soybean IGS

With a length of 1.8 kb, the IGS of soybean is the shortest IGS sequence yet reported for legumes and ranks among the smallest for all angiosperms. Previous analyses (i.e., from restriction fragment length polymorphisms, RFLPs) of nuclear rDNA suggested an IGS length of 1.3 kb for *G. max* and *G. soja*, with 700–1600 cistron copies per haploid genome (Varsanyi-Breiner et al. 1979; Doyle and Beachy 1985); however, our results indicate that this length was an underestimate. Those RFLP analyses also showed that IGS length heterogeneity occurs widely among wild *Glycine* species, for example, from 1.9–3.8 kb in *Glycine tomentella* to a high of 6.1 in *Glycine falcata*, but only single repeat types were observed within individuals. Our finding of length homogeneous IGS sequences further confirms the lack of variation in soybean cultivars. Previous in situ hybridization studies using rDNA probes to soybean nuclei (Skorupska et al. 1989) or metaphase chromosomes (Griffor et al. 1991; Shi et al. 1996) showed that only chromosome 13 is associated with the nucleolar organizing region, thus rDNA genes are present at a single locus. The observed IGS sequence differences seen between clones derived from the same cultivar thus represent polymorphisms present in different repeats within this locus.

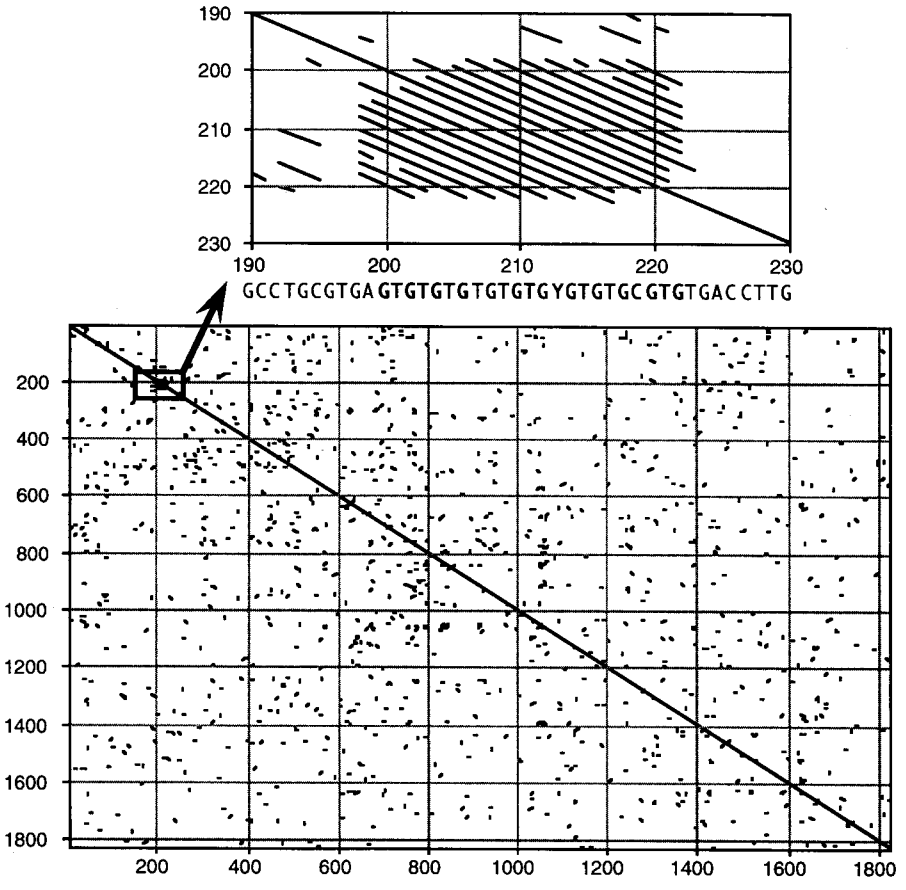
The consensus of all soybean IGS sequences (Fig. 1) was compared with itself, and the resulting PUSTELL MATRIX plot showed no evidence of major subrepeat motifs (Fig. 2). This finding stands in contrast with the organization of other legumes (Gernster et al. 1988; Hemleben et al. 1988; Schiebel et al. 1989; Ueki et al. 1992; Yakura and Nishikawa 1992), which may have as few as one (*Pisum*) to at least five subrepeats (*V. angustifolia* and *V. faba*). Variations on the parameters used in the PUSTELL analysis (e.g., changing the window size, minimum percent scores, etc.) were tried in attempts to detect even weak repeating patterns. Detection of the GT dinucleotide repeats at position 200 highlights the sensitivity of the analysis (Fig. 2) and further confirms the lack of major subrepeats.

Soybean IGS rDNA compared with that of other legumes

With the addition of soybean, IGS sequences currently exist for three members of tribe Phaseoleae (including *Vigna* and *Phaseolus*) and four species representing tribe Vicieae (*Vicia* and *Pisum*). These sequences range in size from 2 to over 4 kbp; however, for the longer sequences, most of the length is accounted for by extensive subrepeat motifs. The pattern of organization is more easily seen when the duplicate copies of tandemly arranged repetitive DNA are removed (Fig. 3). When this is done, the IGS lengths among the seven legume species are more similar. In all legumes, except soybean and pea, IGS organization involves "group A" subrepeat clusters that occur 5′ to the promoter sequence and "group B" subrepeat clusters present 3′ to the promoter. Both the A and B clusters contain from one to three types of subrepeat motifs, depending upon the species. In none of the legumes were motifs from group A interspersed with motifs from group B or vice versa. In *Pisum*, subrepeats occur only 5′ to the promoter. In *Phaseolus* and *V. hirsuta*, IGS organization is comparatively simple, with only two subrepeat types, each being present in multiple tandem arrays. The most complex pattern is seen in *V. angustifolia,* which has five subrepeat types, some of which are tandem and some of which are interspersed (Fig. 3).

To better understand the organization of soybean IGS rDNA, a multiple sequence alignment was conducted with all seven legumes. Few alignments of plant IGS sequences have been published, likely because of the difficulty imposed by their complex structural rearrangements and high substitution rates. An exception is the thorough documentation of IGS sequence variation in *V. faba* and *P. sativum* published by Kato et al. (1990), wherein an alignment of the nonrepetitive sequence shared between the two species was made. Subsequent work has also identified alignable regions in *V. hirsuta* (Yakura and Nishikawa 1992) and *V. angustifolia* (Ueki et al. 1992). Using this information as a starting point, attempts were made to fully align the above species, as well as *V. radiata, P. coccineus,* and *G. max.*

Fig. 2. PUSTELL dot matrix plot of soybean IGS consensus sequence (Fig. 1) to itself. The window size was set to 10, stringency to 65% match, and hash and jump values to 1.0. No major subrepeat motifs were detected. The GT-repeat motif found at positions 200–222 is shown enlarged above.

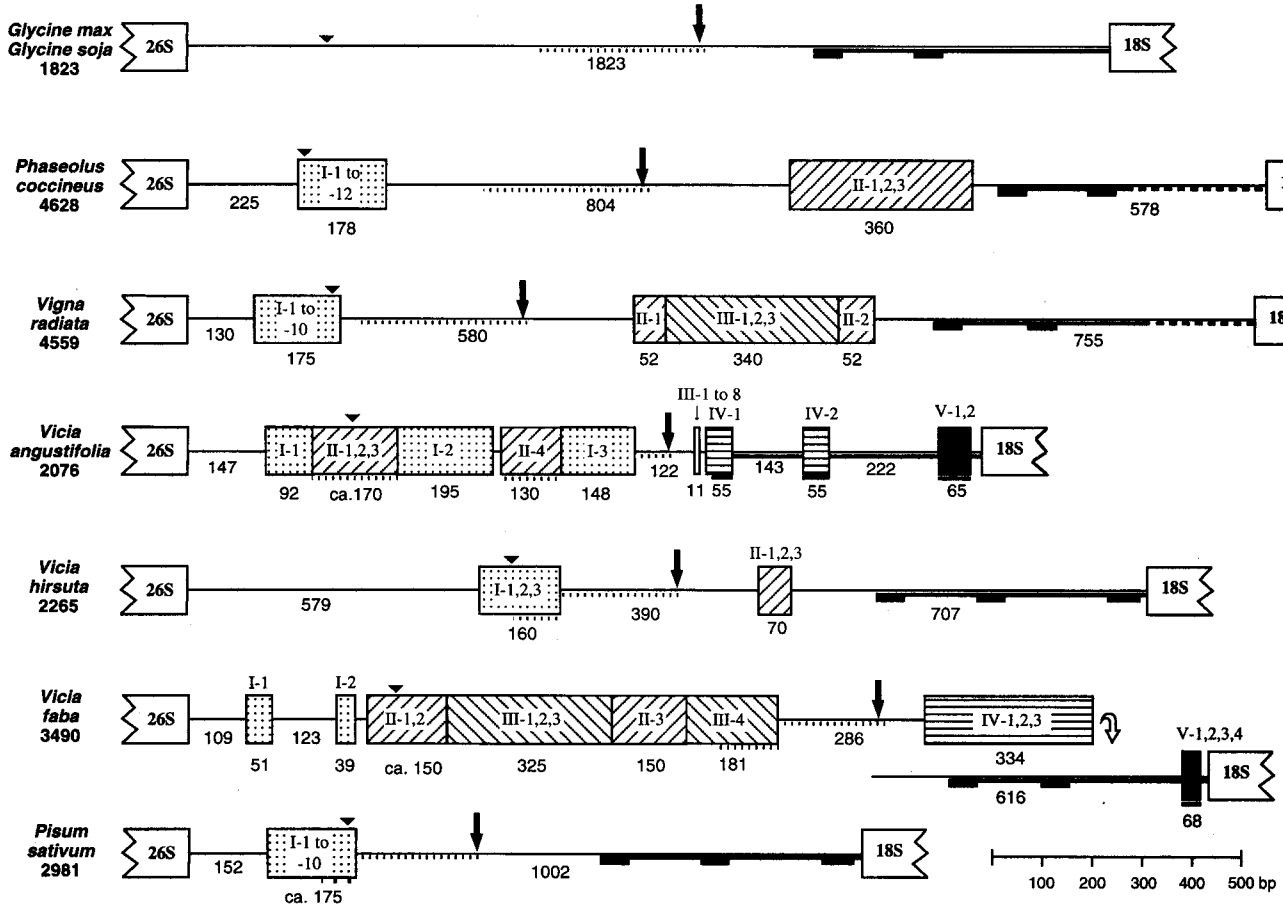


Although all IGS sequences begin with an alignable pyrimidine-rich region just 3′ of the 26S rDNA, sequence variation farther downstream prevents unambiguous alignment of all species. The portions of sequence that can be aligned among all seven species are indicated in Fig. 3. Sequences beginning ca. 300 bp upstream from and including the promoter region (positions 706 to 1026 on *Glycine*; Fig. 1) are readily alignable among all seven species (data not shown; alignment available upon request). Part of this alignable region encompasses nonrepetitive DNA for *Glycine*, *Phaseolus,* and *Vigna,* whereas in others (*Vicia* spp. and *Pisum*), the 5′ portion is found in subrepeat motifs (Fig. 3). This organizational pattern suggests that the rearrangements found in the latter four species have taken place following their divergence from an ancestor shared with the former three species. Conservation of this region among *Vicia* species and *Pisum* was noted by Kato et al. (1990); however, with reference to *Vigna* they state, "there was not a sequence similarity throughout the entire region of the large spacer." Distance analysis of alignable portions of IGS indicate that *Vigna* is genetically more similar to *Phaseolus* and *Glycine* than to *Pisum* or any of the *Vicia* species. Despite this, large portions of the *Vigna* IGS can be aligned with the other legumes (Fig. 3).

As described below, sequence conservation exists at the promoter sequence in all species, yet extensive sequence variation is present just downstream, thereby precluding unambiguous alignment. However, a region farther downstream, corresponding to positions 1251–1823 in *Glycine,* can be aligned for all species except *Phaseolus* and *Vigna* (Fig. 3). For the latter two species, only the 5′ half of this region is alignable. The alignable sequence begins 10 nucleotides from the start of subrepeat IV-1 in *V. angustifolia*. Similarly, sequence corresponding to subrepeat IV-2 in *V. angustifolia* can also be seen in the remaining species. It is worth noting that similarity among these subrepeat regions has not previously been recognized in the other six legume species, owing to substitutional mutations, insertions, and deletions that all result in obscuring sequence similarity. Therefore, we refer to these less obvious duplications as "cryptic subrepeats" (CS-1 and CS-2).

To further examine sequence homology, analyses of alignments of the cryptic subrepeats among all the species, as well as of pairwise alignments within each species, were conducted. Given the selection of available sequences, the modal sequence was influenced more by the presence of three species of *Vicia* and *Pisum* (tribe Vicieae) than by the three other species, which are members of tribe Phaseoleae. Since the genetic distances between all taxa are not equal, lower alignment scores are expected between members of different tribes. For CS-1, percent matches (unpenalized) ranged from 79% for *Phaseolus* to 96% for *V. angustifolia*, with a mean value for

Fig. 3. Diagrammatic representation of IGS sequences among seven legumes where each tandem subrepeat is represented only once. Subrepeat and single-copy spacer lengths (in base pairs) are indicated below the lines. The number of tandem subrepeats is indicated within the boxes. Differences between exact IGS lengths (values below plant names) and lengths calculated from the figure are the result of subrepeat length variation. Promoter sites (indicated by arrows) have been experimentally determined in *Vigna, Vicia angustifolia, Vicia faba*, and *Pisum* (see text). The sites for the other species are based upon sequence comparison (Fig. 5). Dashed lines and solid lines below the IGS sequences indicate regions of similarity that can be aligned. The wider portions of the solid lines indicate cryptic subrepeats that are similar in sequence to subrepeats IV-1 and IV-2 in *V. angustifolia* (Fig. 4). Dashed lines at the 3′ end of this sequence for *Phaseolus* and *Vigna* indicate that these regions are not alignable with other legume species. Putative transcription termination sites are indicated by triangles (experimentally documented only in *Vigna*).



all seven species of 85.8% (Fig. 4). Relative scores (scaled, with penalties imposed) ranged from 72 to 100%, hence CS-1 appears to represent homologous sequence among all seven legume species. The greater sequence divergence for CS-2 is reflected in slightly lower scores among all species. Percent matches ranged from 60% for *Phaseolus* to 95% for *V. angustifolia*, with a mean value for all seven species of 80.1%. These two species also represented the range (35–100%) of relative scores. Members of tribe Vicieae had significantly higher relative scores than members of tribe Phaseoleae. As mentioned above, these relative scores likely reflect the greater evolutionary distance, not necessarily lack of sequence homology. Although the *Pisum* CS-2 sequence matches the modal sequence at 87% of the sites that can be compared, the penalized score is significantly lower (79%), because 12 gaps must be introduced into the alignment. A similar situation occurs with *Phaseolus*, where the number of gaps (11), in combination with greater substitutional mutations, results in a lower score. This suggests that the gap penalty being imposed upon such species may be too severe, resulting in the relative score being a poor indication (underestimation) of sequence homology. In addition to directly comparing subrepeat sequences, we examined the 5′ and 3′ sequences that flank CS-2 and found them to be conserved among all the legumes, thus further supporting the contention that these sequences are homologous.

Comparisons of CS-1 and CS-2 sequences derived from the same species showed that alignment scores were lower than interspecific comparisons of CS-1 only and of CS-2 only (mean = 69%). The lowest percent matches (with and without penalties) were seen in *Pisum* (56%) and *Phaseolus* (58%), whereas the highest score (78%) was that for *Vigna*. The second highest score, 74%, for *V. angustifolia*, required fewer gaps, thus providing a possible explanation of why it was the first species from which this subrepeat was recognized. With a 70% match score, *Glycine* ranks fifth among the seven species, which supports our previous result that demonstrated the lack of major subrepeats. The soybean CS-1/CS-2 alignment resulted in two transitions, 11 transversions, and six gaps. Ap-

Fig. 4. CLUSTAL alignment of the cryptic subrepeats (CS-1 and CS-2) homologous to the IV-1 and IV-2 subrepeats found in *Vicia angustifolia*. Parameters set for both alignments were an open gap penalty of 5, an extended gap penalty of 1, and delayed entry of sequences differing by more than 60%. L, sequence length; M, number of bases matching the modal sequence; % M, percentage of matches based on total length; TN, number of transitions; TV, number of transversions; G, total gaps (sum of nucleotide deletions relative to the modal sequence); S, alignment score (see text); S/L, alignment score divided by length; and RS, relative score scaled by designating the highest value 100%. (A) CS-1. CLUSTAL score 5164 with 18 conserved identity sites. (B) CS-2. CLUSTAL score 5146 with 13 conserved identity sites.

A

```
Modal CS-1 TTTCATGCCTAGCGAKGCA-GAATCGGCGTTGYTTGATGCTTCCATTTG
Phaseolus  TTCCATACCTAGCGATGCG-GACACTGGGTCACTTGATGTTTCCATTTG
Vigna      TT-CATGCCTAGCGATGCTTGGCATTGCGTCACTTGATGTTTC-ATT-G
Glycine    TTTCGTGCCTAGCGATGC--GGATGAGGGCCACTCGATGTTTCCATTTG
Vicia ang  TT-CATGCCTAGCGCGGCA-GAATCGGCGTTGGTTGATGCTTCCATTTG
Vicia hir  TT-CATGCCATGCGA---ACGAATCGGCGTTGTTGGATGCTTCCATTTG
Vicia fab  TTTCATGCCTAG-GCGGCA-GAACCG-C-TTGTTCGATGCTTCCATTTG
Pisum      CTTCATGCCTAGCGCGGAG-GAACCGGCGTTGTTGTATGCTACCATTTG
```

L	M	%M	TN	TV	G	S	S/L	RS
48	38	79	6	4	0	33	69	73
45	36	80	5	4	3	31	69	73
47	38	79	7	2	1	32	68	72
47	45	96	0	2	1	44	94	100
45	41	91	0	3	4	38	85	90
45	42	93	2	1	3	40	89	95
48	40	83	3	5	0	36	75	80

B

```
Modal CS-2 TTRTCAWTTTTCATCGGTTGCCTTGGBGTAATCGGTGTTGGTTCCTTGTTCCTGTTTG
Phaseolus  CC-TCG-TTGGCACCGGTTCCC--CGGGCGAGAGAT-CACGTTC------CCTGTGCT
Vigna      GT-TCGTTTTGCTCCCGTTGCCTT-GGGTGAT-GATGCCCGTTC--AGTCCCTGTGTT
Glycine    TT-TCA-TTGTCACC-TTAGCCTTCGTGGCTTAGGTGCTGACGATTGGTTCCTGTGTT
Vicia ang  TTATCAATTTTCATCGGTC-GCTT-GCGTAATCGGTGTTGGTTTCTTGTTCCTGTTTG
Vicia hir  TTGTCATTTTTCATCGGATGCCTT-GTGCATTTGGTGTTTGTTCGTTGTTCCTGTTTG
Vicia fab  CTATCAATTTTCATCGCTTTGCTT-GCGTAATCG-TGTTGGTTTCTTGTTCCTGTTTG
Pisum      TAGTCA-TTTTCAT--GTT--------GTCATCG-TGTTTGTTATGTGTTCCTGTTTG
```

L	M	%M	TN	TV	G	S	S/L	RS
47	28	60	9	10	11	15	32	35
53	37	70	7	7	5	29	55	60
55	37	67	6	12	3	27	49	54
56	53	95	2	1	2	51	91	100
57	51	89	2	4	1	48	84	92
56	52	93	2	2	2	49	88	97
46	40	87	1	5	12	33	72	79

Fig. 5. Comparison of the putative promoter sequences flanking the RNA pol-I initiation site for seven legumes. Those species for which the transcription initiation site has been determined experimentally are indicated by an asterisk (see text for references). The sequence shown for *Vigna* is the P2 promoter type. The transcription start position (+1) is marked with an arrow. Underlined bases 3′ to the +1 position indicate potential 5′ and 3′ stems of stem-loop structures.

```
                          10        20        30        40        50        60
                          |         ↓ |       |         |         |         |
Glycine max/soja          GCATATTAGCCTATTAT-AGGGGGGCACCACACCCCCCCAACGAAAAWTTTTTTTTTTTYTGAAAAT
Phaseolus coccineus       GCATATTTAGCTATTAT-AGGGGGAGGGTGTTCCTCAATGCATCCAGCTTTGCACCAAGGAAGATT
Vigna radiata *           GCAAAATATCCCCTTAT-AGGGGGGGAGACACTCCTCTCCGCACGCCAGCTTCTGCTCACTTGGCT
Vicia angustifolia *      GCAGTTTAGCCATATAT-ATGGGGGGACAGGCAGGCGCAGGCAGGCAGGCCTGGGCAGCCCGCGCG
Vicia hirsuta             GCTGTTTTGATATATAT-AGGGGGGGGGGGGGATTTTTTGCCTGGCATGGGCGGGGTGCGCGCGGG
Vicia faba *              TCAGTTCAAACATATAT-AGGGGGAGGCCCTTGCCTGGTCCGCCTGGGGCCTTGCCACGCCCAGGC
Pisum sativum *           GCAAATCAAGCTTATAT-AGGGGGAGGCCCTTGACCAGTGCGCCCGTCCCGGGGCCGTCCGAGCGT
```

parently these mutations are sufficient to obscure similarity and thereby prevent detection by PUSTELL analysis. These results also demonstrate that greater sequence divergence has occurred between the cryptic subrepeats within a single IGS cistron than occurs between loci from different species. This may also explain why the IV-1 and IV-2 regions described in *V. angustifolia* have not previously been recognized in the other legume species.

Transcription initiation and termination sites

The transcription initiation site (TIS) for RNA polymerase I (pol I) has been identified using S1 nuclease mapping and primer extension in four legumes, *V. radiata* (Gernster et al. 1988), *V. angustifolia* (Ueki et al. 1992), *V. faba*, and *P. sativum* (Kato et al. 1990), as well as in other crop plants, such as tomato (Perry and Palukaitis 1990), radish (Delcasso-Tremousaygue et al. 1988), maize (McMullen et al. 1986; Toloczyki and Feix 1986), and wheat (Lassner et al. 1987). The rDNA promoter is composed of two domains: a core region that spans positions –40 to +10 and an upstream region that begins at approximately –160 and apparently imparts the species-specificity of pol I (Gernster et al. 1988; Fan et al. 1995). Sequence conservation extends not only to –160 but at least to –300 in the legume species examined in this study (Fig. 1). A promoter sequence of TATA followed by a series of G residues appears conserved not only in plants (Fig. 5) but in animals as well (Coen and Dover 1982; Pikaard and Reeder 1988). The +1 position, determined experimentally for *Vigna*, *V. angustifolia*, and *V. faba*, is an A nucleotide (position 1020 in *Glycine*; Fig. 1) that appears to be conserved among all plant promoters. Among the legumes, a TAT motif occurs just upstream from the +1 site, however, variation is present beyond this point. In this region, the soybean promoter is most similar to that of *Phaseolus*. Multiple TISs have been reported for mung bean (Gernster et al. 1988), *Arabidopsis* (Gruendler et al. 1991), and radish (Delcasso-Tremousaygue et al. 1988), which suggests that different length primary transcripts are produced. Only a single potential promoter region was identified in the soybean IGS, at position 1017 (Fig. 1). Restriction-mapping studies of the soybean rDNA cistron reported by

Jackson and Lark (1982) suggested a putative promoter located between a *BglI* cleavage site and the 18S structural gene. Examination of the soybean IGS sequence (Fig. 1) indicates that only one *BglI* site exists, at position 1454; however, no conserved promoter sequence exists upstream from this site. The assumption that there is a promoter at position 1017 thus indicates that the rRNA precursor molecule is longer than previously suggested by hybridization studies.

By comparison with mammalian systems, Kato et al. (1990) presented a possible secondary structure that flanks the TIS for *V. faba* and *P. sativum*. The 5′ and 3′ stems of this stem-loop structure are indicated in Fig. 5. Attempts to locate corresponding sequence in the other legume species that might engage in pairing was only successful with *V. angustifolia*. For *Glycine*, the only possible stem that can be constructed involves the series of G nucleotides adjacent to the +1 A residue, and the resulting loop would contain only seven nucleotides compared with 16–21 nucleotides present in *V. faba* and *Pisum*. Further work is needed to fully characterize the promoter region in legumes and the possible role of secondary structure in promoter recognition.

In contrast to TISs, in plants, few studies have been made of transcription termination sites (TTS). Nuclease mapping has been conducted on radish (Delcasso-Tremousaygue et al. 1988) and mung bean (Schiebel et al. 1989) to determine the organization of the 3′ external transcribed spacer. For *Vigna*, the TTS (GACTTGGC) occurs within the 175-bp subrepeat and is similar in primary and secondary structure to the T2/T3 termination box (GACTTGC) in *Xenopus* (Schiebel et al. 1989). The presence of inverted repeats within each of the 10 175-bp subrepeats in *Vigna* allows the formation of stem-loop structures (Schiebel et al. 1989). Because the TTS occurs in each subrepeat, transcription termination and initiation sites are brought into closer proximity, thereby allowing "read through enhancement" (Mitchelson and Moss 1987). A search was conducted on the other legume species to determine the presence of the TTS sequence found in *Vigna*. None of the other species had the motif GACTTGGC, however, slight variations were detected in positions that could potentially serve as TTSs. In all *Vicia* species and *Pisum*, the sequence AACTTTGC was found (Fig. 1). In *Phaseolus* and *Glycine* (both members of Phaseoleae), the sequences GACTTGGT and GACTTAGC, respectively, were seen. In all cases except *Glycine*, these potential termination sites were located within subrepeat domains upstream from the promoter (Fig. 1).

If the above proposed sequence is the actual TTS in soybean, its presence within a nonrepetitive spacer region raises questions about the role subrepeats (and structural features formed by them) play in transcription termination or enhancement (Moss 1983; Reeder 1984; Flavell et al. 1986). Unequal crossing over is the most likely mechanism for generating subrepeats (Rogers et al. 1986), thus the lack of subrepeat motifs may account for the small size of the soybean IGS and may indicate low intercistron recombination.

Acknowledgements

This research was supported by a grant from the North Central Soybean Board (U.S.A.). We thank Jeff Doyle for providing a DNA sample of *Glycine soja* and Dr. Lynn Gray for donating an oligonucleotide primer. Two anonymous reviewers provided helpful comments that improved the manuscript.

References

Appels, R., and Dvořák, J. 1982. The wheat ribosomal DNA spacer region: its structure and variation in populations and among species. Theor. Appl. Genet. **63**: 337–348.

Apuya, N., Frazier, B., Keim, P., Roth, E.J., and. Lark, K.G 1988. Restriction length polymorphisms as genetic markers in soybean, *Glycine max* (L.) Merr. Theor. Appl. Genet. **75**: 889–901.

Baldwin, B.G., Sanderson, M.J., Porter, J.M., Wojciechowski, M.F., Campbell, C.S., and Donoghue, M.J. 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. Ann. Mo. Bot. Gard. **82**: 247–277.

Barker, R.F., Harberd, N.P., Jarvis, M.G., and Flavell, R.B. 1988. Structure and evolution of the intergenic region in a ribosomal DNA repeat unit of wheat. J. Mol. Biol. **201**: 1–17.

Bennett, R.I., and Smith, A.G. 1991. The complete nucleotide sequence of the intergenic spacer region of an rDNA operon from *Brassica oleracea* and its comparison with other crucifers. Plant Mol. Biol. **16**: 1095–1098.

Bult, C.J., Sweere, J.A., and Zimmer, E.A. 1995. Cryptic sequence simplicity, nucleotide composition bias, and molecular coevolution in the large subunit of ribosomal DNA in plants: implications for phylogenetic analyses. Ann. Mo. Bot. Gard. **82**: 235–246.

Cardy, B.J., and Beversdorf, W. 1984. Identification of soybean cultivars using isozyme electrophoresis. Seed Sci. Technol. **12**: 943–954.

Close, P.S., Shoemaker, R.C., and Keim, P. 1989. Distribution of restriction site polymorphism within the chloroplast genome of the genus *Glycine*, subgenus *Soja*. Theor. Appl. Genet. **77**: 768–776.

Coen, E.S., and Dover, G.A. 1982. Multiple pol I initiation sequences of rDNA spacers of *Drosophila melanogaster*. Nucleic Acids Res. **10**: 7017–7026.

Cordesse, F., Grellet, F., Reddy, A.S., and Delseny, M. 1992. Genome specificity of rDNA spacer fragments from *Oryza sativa* L. Theor. Appl. Genet. **83**: 864–870.

Delannay, X., Rodgers, D.M., and Palmer, R.G. 1983. Relative genetic contribution among ancestral lines to North American soybean cultivars. Crop Sci. **23**: 944–949.

Delcasso-Tremousaygue, D., Grellet, F., Panabieres, F., and Ananiev, E.D. 1988. Structural and transcriptional characterization of the external spacer of a ribosomal RNA nuclear gene from a higher plant. Eur. J. Biochem. **172**: 767–776.

Doyle, J.J., and Beachy, R.N. 1985. Ribosomal gene variation in soybean (*Glycine*) and its relatives. Theor. Appl. Genet. **70**: 369–376.

Fan, H., Yakura, K., Miyanishi, M., Sugita, M., and Sugiura, M. 1995. In vitro transcription of plant RNA polymerase I-dependent rRNA genes is species-specific. Plant J. **8**: 295–298.

Flavell, R.B., O'Dell, M., Thompson, W.F., Vincentz, M., Sardana, R., and Barker, R.F. 1986. The differential expression of ribosomal RNA genes. Philos. Trans. R. Soc. Lond. B Biol. Sci. **314**: 385–397.

Gernster, J., Schiebel, K., von Waldburg, G., and Hemleben, V. 1988. Complex organization of the length heterogeneous 5′ external spacer of mung bean (*Vigna radiata*) ribosomal DNA. Genome, **30**: 723–733.

Gizlice, Z., Carter, T.E., and Burton, J.W. 1993. Genetic diversity in North American soybean: I. Multivariate analysis of founding stock and relation to coefficient of parentage. Crop. Sci. **33**: 614–620.

Grabau, E.A., Davis, W.H., Phelps, N.D., and Gengenbach, B.G. 1992. Classification of soybean cultivars based upon mitochondrial DNA restriction fragment length polymorphisms. Crop Sci. **32**: 271–274.

Griffor, M.C., Vodkin, L.O., Singh, R.J., and Hymowitz, T. 1991.

Fluorescent in situ hybridization to soybean metaphase chromosomes. Plant Mol. Biol. **17**: 101–109.

Gruendler, P., Unfried, I., Pascher, K., and Schweizer, D. 1991. rDNA intergenic region from *Arabidopsis thaliana*. Structural analysis, intraspecific variation and functional implications. J. Mol. Biol. **221**: 1209–1222.

Hanlon, R., and Grabau, E.A. 1995. Cytoplasmic diversity in old domestic varieties of soybean using two mitochondrial markers. Crop Sci. **35**: 1148–1151.

Hemleben, V., Ganal, M., Gerstner, J., Schiebel, K., and Torres, R.A. 1988. Organization and length heterogeneity of plant ribosomal RNA genes. *In* The architecture of eukaryotic genes. *Edited by* G. Kahl. VHC, Weinheim. pp. 371–383.

Jackson, P.J., and Lark, K.G. 1982. Ribosomal RNA synthesis in soybean suspension cultures growing in different media. Plant Physiol. **69**: 234–239.

Jorgansen, R.A., and Cluster, P.D. 1988. Modes and tempos in the evolution of nuclear ribosomal DNA: new characters for evolutionary studies and new markers for genetic and population studies. Ann. Mo. Bot. Gard. **75**: 1238–1247.

Kato, A., Nakajima, T., Yamashita, J., Yakura, K., and Tanifuji, S. 1990. The structure of the large spacer region of rDNA in *Vicia faba* and *Pisum sativum*. Plant Mol. Biol. **14**: 983–993.

Keim, P., Beavis, W., Schupp, J., and Freestone, R. 1992. Evaluation of soybean RFLP marker diversity in adapted germplasm. Theor. Appl. Genet. **85**: 205–212.

King, K., Torres, R.A., Zentgraf, U., and Hemleben, V. 1993. Molecular evolution of the intergenic spacer in the nuclear ribosomal RNA genes of Cucurbitaceae. J. Mol. Evol. **36**: 144–152.

Kuzaff, R.K., Seveere, J.A., Soltis, D.E., Soltis, P.S., and Zimmer, E.A. 1998. The phylogenetic potential of entire 26S rDNA sequences in plants. Mol. Biol. Evol. **15**: 251–263.

Lakshmikumaran, M., and Negi, M.S. 1994. Structural analysis of two length variants of the rDNA intergenic spacer from *Eruca sativa*. Plant Mol. Biol. **24**: 915–927.

Lassner, M., Anderson, O., and Dvořák, J. 1987. Hypervariation associated with a 12-nucleotide direct repeat and inferences on intergenomic homogenization of ribosomal RNA gene spacers based on the DNA sequence of a clone from the wheat *Nor-D3* locus. Genome, **29**: 770–781.

Lorenzen, L.L., Boutin, S., Young, N., Specht, J.E., and Shoemaker, R.C. 1995. Soybean pedigree analysis using map-based molecular markers: I. Tracking RFLP markers in cultivars. Crop Sci. **35**: 1326–1336.

Maggini, F., Tucci, G., Demartis, A., Gelati, M.T., and Avanzi, S. 1992. Ribosomal RNA genes of *Phaseolus coccineus*. I. Plant Mol. Biol. **18**: 1073–1082.

May, C.E., and Appels, R. 1987. Variability and genetics of spacer DNA sequences between the ribosomal RNA genes of hexaploid wheat (*Triticum aestivum*). Theor. Appl. Genet. **74**: 617–624.

McMullen, M.D., Hunter, B., Phillips, R.L., and Rubenstein, I. 1986. The structure of the maize ribosomal DNA spacer region. Nucleic Acids Res. **14**: 4953–4968.

Mitchelson, K., and Moss, T. 1987. The enhancement of ribosomal transcription by the recycling of RNA polymerase I. Nucleic Acids Res. **15**: 9577–9596.

Moss, T. 1983. A transcriptional function for the repetitive ribosomal spacer in *Xenopus laevis*. Nature (London), **302**: 223–228.

Nickrent, D.L. 1994. From field to film: rapid sequencing methods for field collected plant species. BioTechniques, **16**: 470–475.

Nickrent, D.L., and Doyle, J.J. 1995. A molecular phylogeny of diploid *Glycine* (Fabaceae) based upon nuclear ribosomal ITS sequences. Am. J. Bot. **82** (Suppl.): 153.

Nickrent, D.L., and Soltis, D.E. 1995. A comparison of angiosperm phylogenies based upon complete 18S rDNA and *rbc*L sequences. Ann. Mo. Bot. Gard. **82**: 208–234.

Perry, K.L., and Palukaitis, P. 1990. Transcription of tomato ribosomal DNA and the organization of the intergenic spacer. Mol. Gen. Genet. **221**: 102–112.

Pikaard, C.S., and Reeder, R.H. 1988. Sequence elements essential for function of the *Xenopus laevis* ribosomal DNA enhancers. Mol. Cell. Biol. **8**: 4282–4288.

Polanco, C., and Pérez de la Vega, M. 1997. Intergenic ribosomal spacer variability in hexaploid oat cultivars and landraces. Heredity, **78**: 115–123.

Procunier, J.D., and Kasha, K.J. 1990. The rDNA intergenic spacer region nucleotide sequence of *Hordeum bulbosum* L. Plant Mol. Biol. **15**: 661–663.

Pustell, J.M., and Kafatos, F.C. 1982. A high speed, high capacity homology matrix: zooming through SV40 and polyoma. Nucleic Acids Res. **10**: 4765–4782.

Reeder, R.H. 1984. Enhancers and ribosomal gene spacers. Cell, **38**: 349–351.

Rogers, S.O., and Bendich, A.J. 1987. Ribosomal RNA genes in plants: variability in copy number and in the intergenic spacer. Plant Mol. Biol. **9**: 509–520.

Rogers, S.O., Honda, S., and Bendich, A.J. 1986. Variation in the ribosomal RNA genes among individuals of *Vicia faba*. Plant Mol. Biol. **6**: 339–345.

Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Schaal, B.A., and Learn, G.H.J. 1988. Ribosomal DNA variation with and among plant populations. Ann. Mo. Bot. Gard. **75**: 1207–1216.

Schiebel, K., von Waldburg, G., Gerstner, J., and Hemleben, V. 1989. Termination of transcription of ribosomal RNA genes of mung bean occurs within a 175 bp repetitive element of the spacer region. Mol. Gen. Genet. **218**: 302–307.

Schmidt-Puchta, W., Günther, I., and Sänger, H.L. 1989. Nucleotide sequence of the intergenic spacer (IGS) of the tomato ribosomal DNA. Plant Mol. Biol. **13**: 251–253.

Schweizer, G., and Hemleben, V. 1989. Identification of different cultivars of *Solanum tuberosum* and their somatic hybrids with repetitive gene probes. Vortr. Pflanzenzuecht. **15**: 28–10.

Shi, L., Zhu, T., and Keim, P. 1996. Ribosomal RNA genes in soybean and common bean: chromosomal organization, expression, and evolution. Theor. Appl. Genet. **93**: 136–141.

Shoemaker, R.C., Hatfield, P.M., Palmer, R.G., and Atherly, A.G. 1986. Chloroplast DNA variation in the genus *Glycine* subgenus *Soja*. J. Hered. **77**: 26–30.

Skorupska, H., Albertsen, M.C., Langholz, K.D., and Palmer, R.G. 1989. Detection of ribosomal RNA genes in soybean, *Glycine max* (L.) Merr., by *in situ* hybridization. Genome, **32**: 1091–1095.

Soltis, D.E., Soltis, P.S., Nickrent, D.L., Johnson, L.A., Hahn, W.J., Hoot, S.B., Sweere, J.A., Kuzof, R.K., Kron, K.A., Chase, M.W., Swensen, S.M., Zimmer, E.A., Chaw, S.-M., Gillespie, L.J., Kress, J.W., and Sytsma, K.J. 1997. Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. Ann. Mo. Bot. Gard. **84**: 1–49.

Stemmer, W.P. 1991. A 20 minute ethidium bromide/high salt extraction protocol for plasmid DNA. BioTechniques, **10**: 726.

Takaiwa, F., Kikuchi, S., and Oono, K. 1990. The complete nucleotide sequence of the intergenic spacer between 25S and 17S rDNAs in rice. Plant Mol. Biol. **15**: 933–935.

Toloczyki, C., and Feix, G. 1986. Occurence of 9 homologous repeat units in the external spacer region of a nuclear maize rRNA gene unit. Nucleic Acids Res. **14**: 4969–4986.

Torres, R.A., and Hemleben, V. 1991. Use of ribosomal DNA spacer probes to distinguish cultivars of *Cucurbita pepo* L. and other Cucurbitaceae. Euphytica, **53**: 11–17.

Tremousaygue, D., Grellet, F., Delseny, M., Delourme, R., and Renard, M. 1988. The large spacer of a nuclear ribosomal RNA

gene from radish: organization and use as a probe in rapeseed breeding. Theor. Appl. Genet, **75**: 298–304.

Ueki, M., Uchizawa, E., and Yakura, K. 1992. The nucleotide sequence of the rDNA intergenic spacer region in a wild species of the genus *Vicia, V. angustifolia*. Plant Mol. Biol. **18**: 175–178.

Varsanyi-Breiner, A., Gusella, J.F., Keys, C., and Houseman, D.E. 1979. The organization of a nuclear DNA sequence from a higher plant: molecular cloning and characterization of soybean ribosomal DNA. Gene (Amst.), **7**: 317–334.

Yakura, K., and Nishikawa, K. 1992. The nucleotide sequence of the rDNA spacer region between the 25S and 18S rRNA genes in a species of the genus *Vicia, V. hirsuta*. Plant Mol. Biol. **19**: 537–539.