Theor Appl Genet (1988) 75:621–624

5S ribosomal gene variation in the soybean and its progenitor

J. J. Doyle

L. H. Bailey Hortorium, Cornell University, Ithaca NY 14853, USA

Received July 1, 1987; Accepted September 9, 1987
Communicated by J. Mac Key

Summary. The soybean, *Glycine max* and its wild progenitor, *Glycine soja,* have been surveyed for repeat length variation for the nuclearly encoded 5S ribosomal RNA genes. There is little variation among the 33 accessions assayed, with a common repeat length of 345 bases being typical of both taxa. A 334 base size variant was encountered in individuals from two populations of *G. soja* from China. The low level of variability is in marked contrast to the variation observed within and between the species of the perennial subgenus *Glycine.*

Key words: Soybean – *Glycine max* – *Glycine soja* – 5S ribosomal RNA genes

Introduction

The cultivated soybean, *Glycine max* (L.) Merr., is considered a separate taxonomic species from its presumed wild progenitor, *G. soja* Sieb. and Zucc., based largely on morphological characters associated with domestication (Hymowitz and Newell 1980). The two taxa seem to have few, if any, barriers to hybridization, and the morphological extremes are connected by populations sometimes recognized as a third species (*G. gracilis* Skvortz.), but perhaps merely representing the products of crosses between wild and cultivated forms (Broich and Palmer 1980). *G. max* has an extensive distribution worldwide in cultivation, while *G. soja* is native to northern and central China, Japan, Korea, Taiwan, and the western Soviet Union; *G. gracilis* is found only where the two other species come in contact (Hymowitz and Newell 1980; Hymowitz 1970). Geographic distribution and annual habit distinguish these taxa, which are recognized as subgenus *Soja,* from the dozen or so wild perennial species of *Glycine* subgenus *Glycine* that are native almost exclusively to Australia.

The relationships among the taxa of subgenus *Soja* recently have been the focus of studies using variation at the DNA level. Shoemaker et al. (1986) have used variation in chloroplast DNA to demonstrate the existence of several plastome types, some shared among taxa, while Sisson et al. (1978) found diversity among mitochondrial genomes in the subgenus. Our studies of the nuclearly-encoded 18S–25S ribosomal RNA multigene family in both subgenera of *Glycine* (Doyle and Beachy 1985; Doyle 1987) showed that while considerable variation for repeat length and spacer sequence occurs within subgenus *Glycine, G. max* and *G. soja* appear to be identical to one another.

The 5S ribosomal RNA genes constitute a second large nuclearly-encoded multigene family unlinked to the 18S–25S genes. Like the 18S-25S gene family, the 5S genes exist as tandem copies, but are much more numerous (up to 50,000 copies in flax; Goldsbrough et al. 1981), and considerably smaller, with short coding regions and total repeat lengths generally under 500 bp. Size and sequence variation of the intergenic spacer makes 5S genes potentially useful for assaying variability over short evolutionary periods, while 5S rRNAs have been used in phylogenetic studies spanning long evolutionary divergences. In this paper we report 5S rDNA repeat length variation patterns within and between *G. max* and *G. soja.*

Materials and methods

Accessions used in this study are listed in Table 1. All plants were grown from seed obtained from a variety of sources. Voucher specimens are deposited at the Bailey Hortorium (BH).

Table 1. 5S ribosomal gene repeat sizes of accessions used in study. *Abbreviations:* Cv = cultivar; PI = Plant introduction no.; MG = Maturity group; 18S–25S = repeat length of 18S–25S ribosomal RNA gene family, in kilobases; 5S = repeat length of 5S ribosomal gene family, in bases

Species	Cv	PI no.	Origin	MG	18S –25S	5S
max	Bansei	81031	Japan	II	7.8[a]	345
max	Charlee	71663	China	VII	7.8[a]	345
max	Fuji	81039	Japan	III	7.8[a]	345
max	Harbinsoy	54603-3	China	IV	7.8[a]	345
max	Hidatsa	81038	Japan	00	7.8[a]	345
max	Hongkong	22406	Hongkong	IV	7.8[a]	345
max	Hoosier	30746	China	I	7.8[a]	345
max	Pando	–	Korea	00	7.8[a]	345
max	Poland Yellow	128182	unknown	0	7.8[a]	345
max	Provar	–	China	II	7.8[a]	345
max	Virginia-S	19186D	China	V	7.8[a]	345
soja	–	326582B	USSR	II	7.8[a]	345
soja	–	339732	Korea	IV	7.8[a]	345
soja	–	339871A	Korea	V	7.8[a]	345
soja	–	342619A	USSR	0	7.8[a]	345
soja	–	342621A	USSR	00	7.8[a]	345
soja	–	342622A	USSR	I	7.8[a]	345
soja	–	406684	Japan	III	7.8[a]	345
soja	–	407290	China	II	7.8[a]	345
soja	–	407294	China	II	7.8[a]	345
soja	–	407302	China	V	7.8	345
soja	–	407303	China	VI	7.8	345
soja	–	407304	China	VI	7.8	345
soja	–	407305	China	V	7.8	334
soja	–	407307	China	VI	7.8	334
soja	–	424002	USSR	0	7.8	345
soja	–	440913B	China	II	7.8	345
soja	–	464936	China	VI	7.8	345
soja	–	464935	China	VI	7.8	345
soja	–	464936A	China	VI	7.8	345
soja	–	464937B	China	VI	7.8	345
soja	–	464938	China	V	7.8	345
soja	–	464939B	China	V	7.8	345

[a] Data from Doyle and Beachy 1985

DNA isolation

Total DNA was isolated from greenhouse-grown individual plants using one of two mini-prep procedures. In each case, 0.1–1.3 g fresh weight of fresh leaf tissue was used for isolations. In earlier isolations, the phenol extraction method described previously for *Glycine* (Doyle and Beachy 1985) was used, while in later isolations a more efficient procedure using hexadecyltrimethyl ammonium bromide (ctab) was followed, the details of which are given elsewhere (Doyle and Doyle 1987).

Digestion, electrophoresis, and transfer of DNA

One to five microgram samples of DNA were digested with BamH-I (Bethesda Research Labs of New England Biolabs) according to the specifications of the manufacturer, subjected to gel electrophoresis on 25 cm long 1.5%–2% agarose gels containing 0.89 M tris-borate, 0.089 M boric acid, 0.05 M EDTA, pH 8.0. Following electrophoresis, gels were stained with ethidium bromide and photographed. DNA was transferred to either nitrocellulose or nylon (MSI Nylon 66) using standard procedures (Southern 1975).

DNA hybridization

The 5S ribosomal RNA genes were detected by hybridization to a 5S clone from *Zea mays* (courtesy of Elizabeth Zimmer, Louisiana State University). The total recombinant plasmid was in vitro nick translated (Maniatis et al. 1975) with ^{32}P and hybridized to filters containing *Glycine* DNA. Hybridizations were performed at 42 °C in 5X SSC (1X SSC = 0.15 M NaCl, 0.015 M Na-citrate), 1X Denhardt's solution (0.1% Ficoll, 0.1% polyvinylpyrrolidine 360, 0.1% bovine serum albumin), 0.02 M sodium dodecyl sulfate (SDS), 35% formamide. After overnight hybridization, filters were washed in 2X SSC, 0.4% SDS at 65 °C and exposed to Kodak XAR5 X-ray film with one intensifying screen (Dupont Cronex Lightning Plus) for 1–3 days at –70 °C.

18S–25S rDNA EcoRI profiles for new accessions of *G. soja* were ascertained as previously described (Doyle and Beachy 1985).

Restriction fragment sizes were calculated by comparison of mobilities of known standards run on the same gel. For additional accuracy, some samples were re-run on gels in which Hinf-I digested pBR322 was also electrophoresed; fragments of this plasmid were detected by their homology to the vector bearing the 5S probe and therefore provided labelled standards.

Results and discussion

Thirty-three accessions of *G. max* and *G. soja*, representing a wide range of geographic origins and maturity groups, were tested for variation in 5S repeat length. Digestion with BamH-I followed by electrophoresis, transfer, and detection of 5S ribosomal genes with a heterologous 5S probe yielded the expected ladder-like pattern of monomer, dimer, and multimer repeats (Fig. 1) based on a single repeat size class. *Glycine max* and *G. soja* are typical of diploid members of the genus *Glycine* in having chromosome numbers of $2n = 40$. This number is considered polyploid within the legume tribe Phaseoleae, to which *Glycine* belongs, as most other genera of this tribe are $2n = 22$ (Lackey 1981). This ancient polyploid event, which is hypothesized to have been responsible for increased complexity in soybean multigene families encoding actin (Hightower and Meagher 1985) and leghemoglobin (Lee and Verma 1984), is not observable in either the 18S–25S rDNA family (Doyle and Beachy 1985) nor in the 5S rDNA family. There is no evidence for more than a single major repeat class of 5S rDNA in the genomes of individual plants of *G. max*, *G. soja*, or any other diploid species of the genus, though such within-individual heterogeneity is indeed observed in more recently-derived allopolyploids in *Glycine* subgenus *Glycine* (Fig. 2; Doyle and Brown, in preparation). Assuming that the repeat classes of the two progenitor genomes were initially distinguishable, the loss of heterogeneity could be due to the physical loss of one locus during diploidization or to concerted evolution operating between loci.

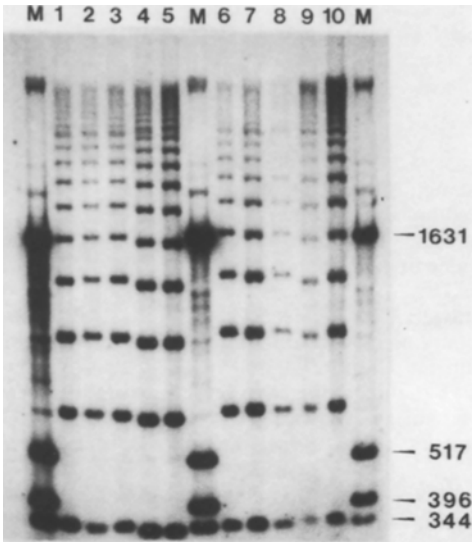

Fig. 1. 5S rDNA variation in *Glycine soja* and *G. max*. Ladder patterns are produced due to the inability of BamH-I to cleave all tandem repeats to completion. The size of the fastest-migrating (*lowest*) fragment in each lane gives the length of a single repeating unit. *M* = size markers, pBR322 digested with HinfI, sizes of fragments given in base pairs; *Lanes:* (*1*) *G. soja* 407303, (*2*) *G. max* cv Pando, (*3*) *G. soja* 407302, (*4*) *G. soja* 407305, (*5*) *G. soja* 407307, (*6*) *G. soja* 464935, (*7*) *G. soja* 464936B, (*8*) *G. max* cv Hidatsa, (*9*) *G. soja* 407304, (*10*) *G. soja* 464937B

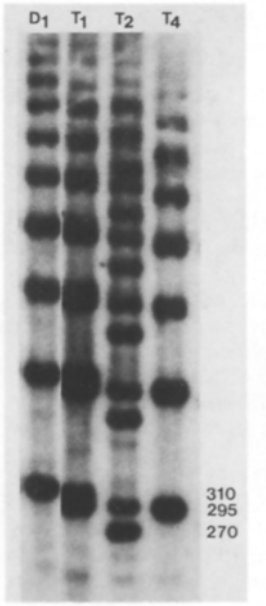

Fig. 2. Within-individual heterogeneity of 5S repeats in polyploids of *Glycine tomentella* (subgenus *Glycine*). T1 is an accession with $2n = 78$, hypothesized to have arisen by hybridization between $2n = 38$ plants such as the accession here *labelled* D1; T1 includes D1 sized repeats as well as a second repeat size which it shares with T2, a $2n = 80$ tetraploid, as well as with T4, also $2n = 80$. T2 also contains a second, smaller repeat class. Diploids of this single species have repeat lengths from 270 to 310 bp

Within the two species of subgenus *Soja* very little variation in the size of the 5S repeat was encountered. With the exception of PIs 407305 and 407307 (Table 1), all *G. soja* accessions had repeat lengths of 345 bp, a size very similar to the 342 bp repeat length reported for another legume, *Lupinus luteus* (Rafalski et al. 1982). The two anomalous *G. soja* accessions were not in any other way unusual, and were found to have 18S–25S rDNA profile identical to other accessions of subgenus *Soja* in both repeat length and EcoRI restriction sites (Table 1). Both accessions are from Shanghai municipality, south of Shanghai City, but come from two different localities (R.L. Bernard, US Soybean Germplasm Collection). It is apparent that polymorphic populations occur in this region, since PI 470304, with a repeat length typical of other *G. soja* accessions, was collected at the same site as 470305. *Glycine max* accessions studied all had repeat lengths identical to the common repeat length found in *G. soja* (345 bp).

The paucity of repeat length variability for the 5S rDNA multigene family within *Glycine* subgenus *Soja* agrees with our findings for the unlinked 18S–25S rDNA family (Doyle and Beachy 1985). As in our previous study, the lack of variability within the cultivated soybean and its wild progenitor is in marked contrast to the higher levels of variation found both between and within the wild perennial species of *Glycine* subgenus *Glycine* (Fig. 2; Doyle and Brown, in preparation). As an example of variation found within a single recognized species of this subgenus, the 5S BamH-I digestion patterns of several *G. tomentella* accessions are shown in Fig. 2. The genomes of two of these plants, both presumed to be allopolyploids, contain 5S repeats derived from three different diploid genome donors, one of which is shared between the two accessions. 5S repeats in this species range in size from 270 to 310 bp, and illustrate both the variability found in single species of the perennial subgenus as well as the additivity of 5S in single allopolyploid plants.

Why there should be such a difference in the variability patterns between the 2 subgenera is not clear. Hu et al. (1985), in their studies of wild *G. soja* populations in China, found no evidence of heterozygosity at a polymorphic amylase isozyme locus in 1,300 plants examined, and concluded that the species is an absolute inbreeder. The species of subgenus *Glycine* produce abundant cleistogamous flowers in addition to chasmogamous flowers, and it appears that much of their seed set could thus be due to selfing. In this respect, then, the two subgenera are quite similar. A potentially significant difference between the wild species of both subgenera could be the existence of the widely distributed cultivated species, *G. max*. The wild perennial taxa of subgenus *Glycine* exist in isolated, often small populations between which gene flow would be predicted to be

low, a situation that could well lead to evolutionary divergence of genes. In contrast, *G. soja* often occurs as a weed in soybean fields, where it is capable of crossing with *G. max*. The cultivated soybean, therefore, could form a bridge through which even the low levels of gene flow predicted for a predominantly inbreeding species could unite wild populations. The processes of concerted evolution then could operate to maintain homogeneity throughout subgenus *Soja*.

The only 5S repeat length variation within subgenus *Soja* is found in the wild species, *Glycine soja*. It is perhaps not unexpected to find that a cultivated species shows a subset of the variability found in its wild progenitor, particularly given the narrow genetic base of cultivated soybean lines grown in the United States (Delannay et al. 1983). The *G. max* accessions studied here were chosen in an effort to sample the native diversity in that species, and were selected to include geographically and physiologically diverse accessions. Yet despite sampling what is presumably a broader genetic base than common U.S. cultived lines (Committee on Genetic Vulnerability of Major Crops 1972), we found no variation within the soybean for 5S repeat length. It is possible, of course, that detailed analysis of 5S sequences would uncover variation; in lupines, Rafalski et al. (1982) report 5S repeats of identical length with differences in base sequence. Nevertheless, it appears that variability in the soybean for this gene family, as for the 18S–25S rDNA, is likely to be low even among otherwise diverse cultivars.

The finding of variation within *G. soja* and uniformity in *G. max* is in contrast to the chloroplast DNA data of Shoemaker et al. (1986), who found that all five *G. soja* accessions studied belonged to only one of five plastome groupings, while *G. max* accessions occurred in three groups. This contrast may, however, be due to the smaller sample size of *G. max* in our study, and the small sample size of *G. soja* in theirs.

It has been suggested that *G. max* and *G. soja* are conspecific (Hymowitz and Newell 1980), based in large part on their interfertility. Similarities between the two taxa are seen at the protein level, in the products of such multigene families as storage proteins (Staswick et al. 1983) and leghemoglobins (Fuchsman and Palmer 1985). Consistent with these findings and with our previous studies of 18S–25S rDNA, our 5S data show little differentiation between the two taxa, and are certainly consistent with the existence of a single species and a widely distributed cultigen.

Acknowledgements. This work was supported by NSF grant BSR-851660, Hatch funds, and by a grant from the Cornell Biotechnology Program, which is sponsored by the New York State Science and Technology Foundation and by a consortium of industries. I am indebted to Roger N. Beachy, in whose laboratory the study was initiated. I am grateful to Reid G. Palmer, R.L. Bernard, and Xavier Delannay for making germplasm available for study. I also wish to thank Jane Leclere Doyle for expert technical and artistic assistance.

References

Broich SL, Palmer RG (1980) A cluster analysis of wild and domesticated soybean phenotypes. Euphytica 29:23–32

Committee on Genetic Vulnerability of Major Crops (1972) Genetic vulnerability of major crops. Natl Acad Sci, Washington DC

Delannay X, Rogers DM, Palmer RG (1983) Relative genetic contributions among ancestral lines to North American soybean cultivars. Crop Sci 23:944–949

Doyle JJ (1987) Variation at the DNA level: uses and potential in legume systematics. In: Stirton CA (ed) Advances in legume systematics, part 3. Royal Botanical Gardens, Kew, pp 1–30

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

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Fuchsman WH, Palmer RG (1985) Conservation of leghemoglobin heterogeneity and structures in cultivated and wild soybean. Can J Bot 63:1951–1956

Goldsbrough PB, Ellis THN, Cullis CA (1981) Organisation of the 5S RNA genes in flax. Nucleic Acids Res 9:5895–5904

Hightower RC, Meagher RB (1985) Divergence and differential expression of soybean actin genes. EMBO J 4:1–8

Hu Z-A, Wang H-X (1985) Genetic structure of natural populations of wild soybean (*Glycine soja*) in Beijing region (China). Acta Bot Sin 27:599–604

Hymowitz TA (1970) On the domestication of the soybean. Econ Bot 24:408–421

Hymowitz TA, Newell CA (1980) Taxonomy, speciation, domestication, dissemination, germplasm resources, and variation in the genus *Glycine*. In: Summerfield RJ, Bunting AH (eds) Advances in legume science. Royal Botanical Gardens, Kew, pp 251–264

Lackey JA (1981) Phaseoleae. In: Polhill RM, Raven PH (eds) Advances in legume systematics, part 1. Royal Botanical Gardens, Kew, pp 301–327

Lee JS, Verma DPS (1984) Structure and chromosomal arrangement of leghemoglobin genes in kidney bean suggest divergence in soybean leghemoglobin gene loci following tetraploidization. EMBO J 3:2745–2752

Maniatis T, Jeffrey A, Klein DG (1975) Nucleotide sequence of the rightward operator of phage lambda. Proc Natl Acad Sci USA 72:1184–1189

Rafalski JA, Wiewiorowski M, Söll D (1982) Organization and nucleotide sequence of nuclear 5S rRNA genes in yellow lupin (*Lupinus luteus*). Nucleic Acids Res 10:7635–7642

Shoemaker RC, Hatfield PM, Palmer RG, Atherly AG (1986) Chloroplast DNA variation in the genus *Glycine* subgenus *Soja*. J Hered 77:26–30

Sisson VA, Brim CA, Levings CS III (1978) Characterization of cytoplasmic diversity in soybeans by restriction endonuclease analysis. Crop Sci 18:991–996

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503–517

Staswick PE, Broué P, Nielsen NC (1983) Glycinin composition of several perennial species related to soybean. Plant Physiol 72:1114–1118