Theor Appl Genet (1995) 90: 621–626

A. Kolchinsky · P. M. Gresshoff

A major satellite DNA of soybean is a 92-base pairs tandem repeat

Received: 12 February 1994 / Accepted: 10 June 1994

Abstract We report the cloning, sequencing and analysis of the major repetitive DNA of soybean (*Glycine max*). The repeat, SB92, was cloned as several monomers and trimers produced by digestion with *Xho*I. The deduced consensus sequence of the repeat is 92 base pairs long. Genomic sequences do not fluctuate in length. Their average homology to the consensus sequence is 92%. The consensus of SB92 contains slightly degenerated homologies for several 6-cutters. Therefore, many of them generate a ladder of 92-bp oligomers. The distribution of bands seems to be random, but the occurrence of sites for different 6-cutters varies widely. There is no obvious correlation between the sequences of the neighboring units of SB92 in cloned trimers. Also, there are none of the internal repetitive blocks reported for many satellite DNAs from other species. The SB92 repeat makes up 0.7% of total soybean DNA. This is equivalent to 8×10^4 copies, or 7 megabases. The repeat is organized in giant tandem blocks over 1 Mb in length, and there are fewer blocks than chromosomes. The polymorphism of these blocks is extremely high. The SB92 repeat is present in identical arrangement and number of copies in the ancestral subspecies *Glycine soja*. There are 10 times fewer copies of the repeat in a related species *Vigna unguiculata* (cowpea), and no homologies in several other more distant leguminous plants studied.

Key words Soybean · Satellite DNA · PFGE
Genetic mapping

Introduction

Molecular maps of agriculturally important plants have become a necessary tool in both basic and applied research. Such maps have been successfully developed and utilized for soybean (Keim et al. 1990; Shoemaker et al. 1992; Lark et al. 1993). The integration of molecular markers based on repetitive DNA into these maps is important in several aspects. First, clusters of tandem repetitive DNA show high variability, and therefore provide both markers and fingerprinting opportunity. Second, they are relatively easy targets for in situ hybridization. Third, in some cases they are species specific, variety specific and even chromosome specific. Last, but not least, in many cases they are located in the areas of chromosomes that attract specific attention, namely centromeres and telomeres. This is particularly important in the light of recent data on the predominant localization of plant genes in the telomeric segments of chromosomes (Gill et al. 1993; Moore et al. 1993).

Despite these considerations and the agricultural value of soybean, no tandemly repetitive DNA from this species has been studied (except ribosomal genes, see Doyle 1988; Doyle and Beachy 1985; Kolchinsky and Gresshoff 1992). To fill this gap, we studied a major satellite DNA of soybean. It was cloned and sequenced, and its organization in the genome was studied. This repetitive DNA, SB92, has a simple internal structure with no obvious internal hierarchy. Its gross organization was studied by PFGE of high-molecular-weight DNA. It is found in clusters over 1 Mb in size, and fragments obtained with some restriction enzymes show extremely high variability.

Materials and methods

Materials

The following plant varieties were used: *Glycine max* (L.) Merrill, cv 'Bragg'; *G. max* acc. A81-356022; *G. soja* (Sieb and Zucc.), acc PI468.397; *G. max*, cv 'Peking'; *G. max*, cv 'Enrei'; *G. max*, acc DPS3589; cowpea, *Vigna unguiculata*, cv 'California Blackeye'.

DNA manipulations

Electrophoresis of DNA shown in Fig. 1 was performed in non-denaturing polyacrylamide gels (PAAG) and the gels were stained with silver as described (Bassam et al. 1991). Mono- and trimeric frag-

Communicated by I. Potrkus

A. Kolchinsky (✉) · P. M. Gresshoff
Plant Molecular Genetics and Center for Legume Research, 269, Ellington Bldg, The University of Tennessee, Knoxville, TN 37901-1071, USA

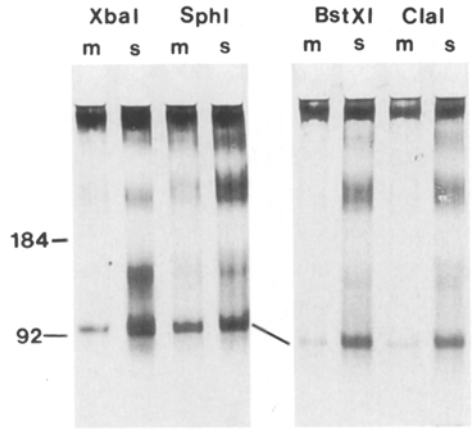

Fig. 1 The 92-bp repeat unit revealed by silver staining of total soybean DNA after electrophoresis in 6% PAAG. The lanes contain approximately 0.5 µg DNA; *m G. max*, *s G. soja*. The restriction enzymes used are indicated *above* the lanes

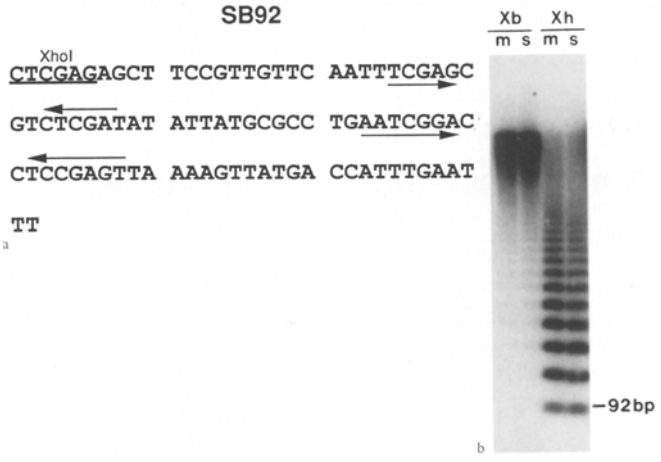

Fig. 2a, b The 92-bp repeats in *G. max* and *G. soja*. genomes. **a** The sequence of SB92 (Accession Z26334); *arrows* indicate palindromes. **b** Southern hybridization of SB92 to *Glycine* sp. DNA. (*m G. max*, *s G. soja*, *Xb XbaI*, *Xh XhoI*.) The size of the monomer is indicated on the *right*

ments of the 92-bp unit generated by *Xho*I digestion were isolated from a larger PAAG and cloned into the *Sal*I site of the pBS(+) plasmid (Stratagene). The inserts were verified by direct polymerase chain reaction (PCR) on colonies (Kolchinsky and Gresshoff, in preparation) and sequenced by a radioactive procedure with a sequencing kit from USB (for monomers) or with a Silver Sequence kit from Promega (for trimers). DNA concentrations were determined using the Hoechst 33258 fluorescent dye and fluorimeter TKO 100 from Hoefer. Southern hybridizations and probe labeling with random primers were performed as described (Sambrook et al. 1989).

Analysis of the DNA sequence and a search in the GenBank were performed by the GCG software package.

High-molecular-weight DNA was isolated in agarose plugs as described (Funke et al. 1993) and separated by PFGE in a CHEF apparatus (BioRad). The telomeric probe was generated as described earlier (Kolchinsky and Gresshoff 1993).

Determination of copy number

Copy number was determined by two methods. In the first method, DNA was digested with *Eco*RI, quantified and run in an agarose gel in parallel with different amounts of the isolated trimer of SB92. The gel was transferred onto Zeta-Probe GT (BioRad) and Southern hybridized to the labeled trimeric fragment. After exposure, the lanes were cut and counted in a scintillation counter. In the second method, measured amounts of soybean DNA and cloned and isolated trimer were added to 0.4 *M* NaOH, boiled in a water bath for 10 min, neutralized with 2 *M* NH_4CH_3COO (pH 4.5) and dot-blotted with proper controls onto a nylon membrane. The membrane was hybridized to the labeled trimer fragment, and the dots were counted. In all cases carrier salmon sperm DNA was added to the samples prior to manipulations. Equivalent amounts of total DNA were loaded. The results obtained by the two methods repeated twice agreed well.

Results

Because of the nature of tandem repeats there are two approaches by which to look for them. The first source is the high-molecular-weight fraction left after the digestion of total DNA with a frequent cutter (so-called relic DNA, see Bedbrook et al. 1980). The second source is the low-molecular-weight fraction generated by a rare cutter. The second approach was chosen in our investigation. The 92-bp repeat was initially found when DNAs from *G. max* and *G. soja* were digested with some 6-cutter restriction enzymes and fractionated on a polyacrylamide gel to resolve fragments under 400 bp. Silver staining of the gels (Bassam et al. 1991) revealed a prominent band about 92 bp long produced by several enzymes (*Bst*XI, *Cla*I, *Sph*I, *Xba*I, *Xho*I) (Fig. 1). To study this product soybean DNA was digested with *Xho*I and run on a preparative polyacrylamide gel. The monomeric 92-bp band and its oligomers were visible after staining with ethidium bromide (not shown). The monomeric and trimeric products were eluted from the gel, cloned and sequenced, and the repeat was designated SB92.

A consensus sequence (Fig. 2A) was derived from ten sequenced repeats (four monomers and two complete trimers). Average homology of the repeats to the consensus was 92%. The search in the GenBank Database did not give any reliable homologies. The dot-plot analysis found two degenerated palindromes at low stringency (8 homologies out of 12-bp-wide window) with their centers at positions 31 and 60 (Fig. 2A). No other internal subrepeats were found. Despite significant divergence of the repeats from the consensus sequence, they never deviated from the 92-bp length.

Southern hybridization of *G. max* and *G. soja* DNA with the cloned repeat showed a simple ladder of fragments generated by *Xho*I and *Xba*I (Fig. 2B). This ladder went to the limit of the resolution of the agarose gel and reached 17-mer without getting diffuse (Fig. 2B). This pattern further proved the uniformity of the length of the repeats. *Xho*I sites were found more frequently than sites for any other 6-cutters tested. The copy number of the repeat in the two genomes looked roughly similar, and it was corroborated by a more accurate estimate (see below).

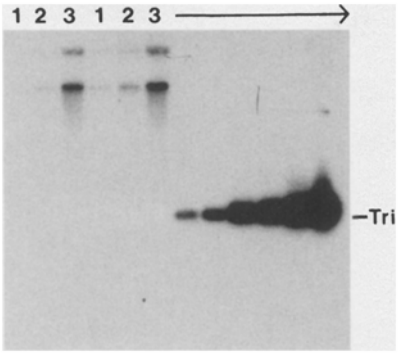

Fig. 3 Determination of copy number of SB92 in soybean DNA. *Lanes 1–3 Eco*RI-digested soybean DNA, 0.1 (*1*), 0.3 (*2*) and 1 (*3*) µg. *Arrow* indicates increasing amounts of the cloned trimer of SB92: 3, 10, 30, 100, 300 and 500 ng

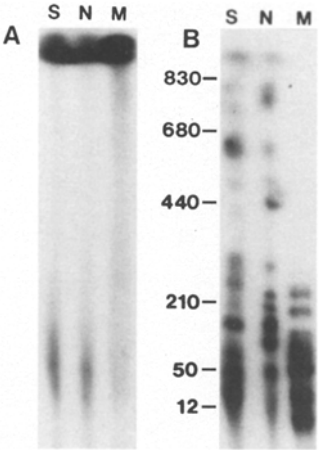

Fig. 6A, B Large blocks of SB92. High-molecular-weight DNA was digested with restriction enzymes *Sfi*I (*S*), *Not*I (*N*) and *Mlu*I (*M*). After PFGE, the membrane was hybridized to SB92 (**panel A**) and then to the telomeric repeat (**B**)

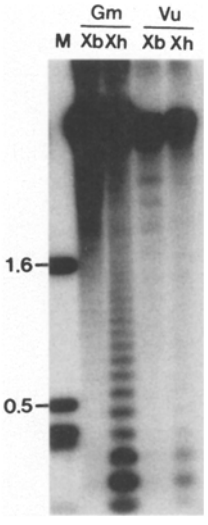

Fig. 4 Hybridization of SB92 to soybean and cowpea DNA. (*M* marker DNA, *Gm* soybean DNA, *Vu* cowpea DNA, *Xb Xba*I, *Xh Xho*I.) Molecular sizes in kb are indicated on the *left*

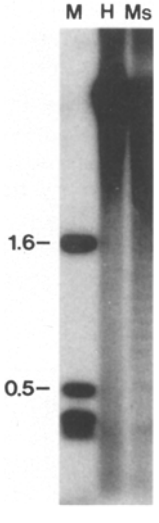

Fig. 5 Infrequent *Msp*I sites in SB92 repetitive clusters are methylated. (*M* marker lane, *H Hpa*II, *Ms Msp*I)

We estimated that SB92 makes up 0.7% of the soybean genome. The size of the haploid soybean genome is 1000 Mb (Gurley et al. 1979), therefore the repeat makes 7 Mb of the total length of DNA, or 8×10^4 copies (Fig. 3). In trying to apply different protocols for the copy number determination, we found it deceptive to use the whole recombinant plasmid as the hybridization probe and for quantitative calibration. The intensity of hybridization is not proportional to the content of the insert in the recombinant plasmid, and this leads to artefacts. Therefore, an isolated trimer of SB92 was used for all hybridizations.

Species-specificity is an important characteristic of tandem repeats. Among the several leguminous plants studied (cowpea, *Lotus japonicus*; alfalfa, *Medicago truncatula*) only cowpea (*Vigna unguiculata*) gave a clear positive signal (Fig. 4). The number of copies was approximately an order of magnitude lower than in soybean, as estimated by counting the lanes from the membrane shown in Fig. 4. Although the repeat length was the same, the organization of the repeats seemed to be more complicated with some bands being more prominent in the 92-bp ladder.

To assess the extent of methylation of SB92, we compared soybean DNA digested with the methylation-sensitive enzyme *Hpa*II and partially methylation-insensitive *Msp*I (Fig. 5). The consensus sequence did not contain the *Msp*I/*Hpa*II site CCGG, which could be generated only by substitutions. The internal C in these rare sites was completely methylated, since there was no ladder in the *Hpa*I lane in Fig. 5. Intensive methylation of satellite DNAs has been reported for virtually all tandem repeats studied (see, for instance, Ganal et al. 1986).

To understand the organization of SB92 repeats, soybean DNA was isolated from protoplasts in agarose blocks, treated with restriction enzymes and fractionated by PFGE. After hybridization with the SB92 probe only very large blocks of repeats were seen (Fig. 6). Their size exceeds 900 kb. The same membrane was stripped of the probe and hybridized to the telomeric repeat. This probe revealed fragments of moderate size only, mostly below 300 kb.

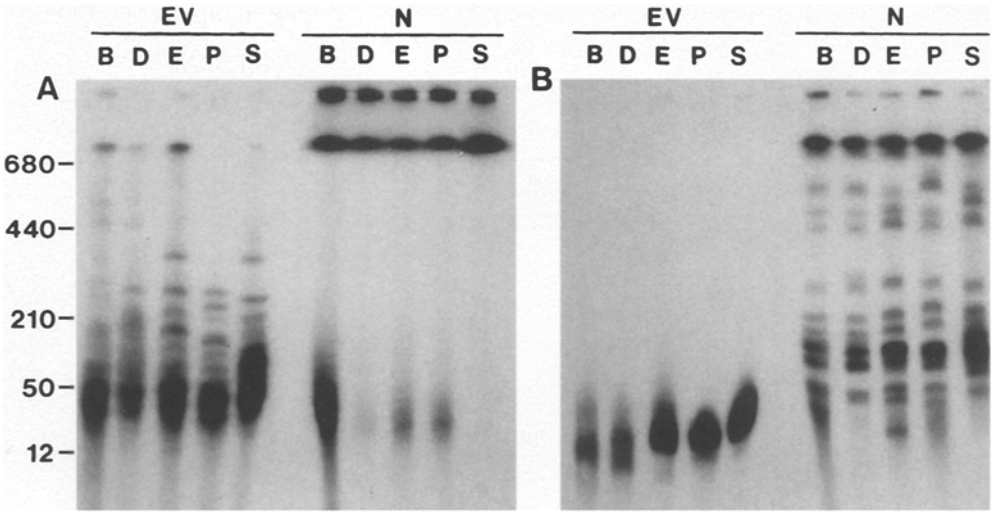

Fig. 7A, B Variability of high-molecular-weight DNA restriction fragments hybridizing to SB92 (**A**) and the telomeric repeat (**B**). (*EV* digestion with *Eco*RV, *N Not*I. *B G. max* cv 'Bragg', *D G. max* cv 'DPS3589', *E G. max* cv 'Enrei', *P G. max* cv 'Peking', *S G. soja*.) The same membrane was hybridized. The size of fragments in kb is indicated on the *left*

Table 1 Variability of repetitive DNA fragments in soybean[a]

Variety	SB92			Telomeric repeat		
	Number of bands	Average F	Average D	Number of bands	Average F	Average D
G. max cv. Bragg	8	0.08	0.92	12	0.76	0.24
G. max cv. DPS	9	0.21	0.89	11	0.72	0.28
G max cv. Enrei	10	0.15	0.85	12	0.67	0.33
G. max cv. Peking	12	0.17	0.83	11	0.66	0.34
G. soja	16	0.15	0.85	10	0.47	0.53
Total average			0.87			0.34

[a] See explanations in the text

The polymorphism of blocks of repeats in related lines and accessions is the most important feature for further genetic mapping. Therefore, high-molecular-weight DNA from four varieties of *G. max* and one accession of *G. soja* were used to assess the variability of the SB92 clusters. After DNA was digested with a 6-cutter, *Eco*RV, and an 8-cutter, *Not*I, restriction enzyme, it was resolved in a PFGE. DNA was Southern-hybridized to the SB92 probe, and then the membrane was stripped and rehybridized to the telomeric repeat for comparison. The blocks of telomeric repeats, as well as subtelomeric repeats, have been shown to be hypervariable in tomato and some other plants (Ganal et al. 1992; Röder et al. 1993). The results of our experiment are shown in Fig. 7 and summarized in Table 1. The proportion F of fragments in common between all pairs of accessions was determined for all samples as F=2nXY/(nX+nY), where nX and nY are the total numbers of fragments for varieties X and Y, and nXY is the number of fragments in common. An average was then calculated for each accession. A corresponding D value (D=1–F) measures the extent of variation. (Calculated according to Nei and Li as quoted by Broun et al. 1992).The calculations showed that in soybean the divergency of SB92 clusters is significantly higher than the divergency of the telomeric repeats.

There is another conclusion that can be drawn from the analysis of the SB92 clusters by PFGE. Rare cutters like *Not*I, *Mlu*I and *Sfi*I (Fig. 6) did not generate any discrete bands below 0.9 Mb. Since the total space of the genome occupied by SB92 is 7 Mb as calculated from its copy number (see above), there could not be more than seven to nine blocks of SB92 per haploid genome. Soybean has 20 chromosomes per haploid genome (Singh and Hymowitz 1988). Therefore, only some of them can contain blocks of SB92.

Finally, there were no *Not*I bands that hybridized simultaneously to the telomeric probe and SB92 probe. Therefore, these repeats are not located within the 1-Mb-long terminal segments of soybean chromosomes.

Discussion

Significant progress has been achieved recently in the genetic studies of soybean. However, to the best of our knowl-

edge, no studies of its repetitive DNA have been published so far. In this article, we report cloning and the study of a 92-bp satellite repeat, SB92.

The 92-bp repeat is a typical example of satellite DNA. Its organization is simple, and its sequence does not give any hints about its function. While speculations on the origin and function of tandem repeats are abundant in the literature, our results do not support any of these speculations. A hypothesis was put forth that tandem repeats of plants originate from an ancestral tRNA gene (Benslimane et al. 1986), but SB92 does not show any reasonable homology to tRNA genes. Ingham et al. (1993) suggested that tandem repeats of plants consist of a basic 30-bp unit that was in many cases first duplicated or triplicated and then amplified as a block. However, the soybean repeat has neither internal subrepeats nor homologies to any of the plant tandem repeats reviewed in that paper. Recently Vogt (1992) speculated on the possibility that tandem repeats encode specific chromatin structures because of their unusual bending properties through interactions with sequence-specific proteins. He proposed that the $(GNATT)_n$ subrepeat might play an important role in the interaction. The consensus of SB92 contains only two homologies to this sequence, which could be considered random occurrence. Also, no signs of the locus specificity implied by Vogt's hypothesis were found (except for the long-range *Eco*RV-polymorphism). It is worth mentioning, also, that SB92 has no homology to the satellite repeats found in leguminous plants, the 59-bp-long tandem repeat from *Vicia faba* (Kato et al. 1984) and 180-bp-long repeat from alfalfa (Xia and Erickson 1993).

Among the several leguminous plants studied, only the cowpea shows clear homology to SB92. Indeed, according to existing data on molecular mapping the cowpea is the closest species to soybean (Young et al. 1992). Nevertheless, the pattern of divergence of the restriction sites is different, since the cowpea repetitive ladder shows prominent bands. Therefore, some higher-order blocks exist in this genome, as described for many other satellite tandem repeats (for review and interpretation, see Vogt 1992).

In many cases tandem repeats in plants have been found at or near the telomeric ends of chromosomes (for review see Kolchinsky and Gresshoff 1994). In barley and tomato the polymorphism of clusters of these tandem repeats has been successfuly used for the mapping of the ends of linkage groups (Ganal et al. 1992; Röder et al. 1993). We did not find any indications of high-molecular-weight DNA fragments sharing the telomeric sequence and SB92. Therefore, the repeat is not located within approximately 1 Mb from telomeric ends and is probably a centromeric repeat. This may be due to an unusual characteristic of soybean chromosomes: as stated by Singh and Hymowitz (1988), most of them have euchromatic ends, with only 6 out of 40 telomeres being heterochromatic.

The size of the SB92 blocks indicates that they should be located on only a subset of soybean chromosomes. Indeed, *Sfi*I, *Not*I and *Mlu*I produce fragments above 900 kb in size, while telomeric fragments show average-molecular-weight distribution. This experiment cannot be done with frequently cutting restriction enzymes because SB92 would be degenerated and virtually any 4-bp-long site would be found. On the other hand, the rarely cutting enzymes used in this experiment apparently don't have recognition sites inside the repeats and produce unusually large fragments. This makes this satellite DNA an attractive choice for the correlation between linkage groups and chromosomes through in situ hybridization and mapping of SB92 blocks based on their variability.

Both notions of non-telomeric localization and chromosome specificity of the large blocks of SB92 were corroborated by *in situ* hybridization of the same repeat cloned and studied in Paul Keim's laboratory (L. Shi and P. Keim, personal communication). They find approximately six hybridization sites per haploid genome. Therefore, SB92 can be used for chromosome-specific painting of soybean chromosomes for *in situ* studies.

A detailed comparison between the variability of the telomeric repeats and flanking blocks of a satellite repeat in tomato was performed by Broun et al. (1992) The divergence among the telomeric repeats was found to be much higher (D=0.92 and 0.65, correspondingly). In our hands, soybean telomeres show a relatively low extent of polymorphism (0.34), much lower than that in tomato, while the clusters of SB92 in soybean are significantly more polymorphic than the subtelomeric repeats of tomato (D=0.87). The relative stability of long restriction fragments containing soybean telomeres might be determined by their euchromatic nature.

Acknowledgements We thank Fred Allen for soybean lines, Debbie Landau-Ellis for cowpea DNA and Roel Funke for help with the preparation of soybean DNA in agarose plugs. Our gratitude goes also to Paul Keim who shared his unpublished results with us. The study was supported by the Ivan Racheff Endowment, the Tennessee Soybean Promotion Board and the American Soybean Association.

References

Bassam BJ, Caetano-Anollés G, Gresshoff PM (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels. Anal Biochem 196:80–83

Bedbrook IR, Jones I, O'Dell M, Thompson RD, Flavell RB (1980) A molecular description of telomeric heterochromatin in *Secale* species. Cell 19:545–560

Benslimane AA, Dron M, Hartman C, Rode A (1986) Small tandemly repeated DNA sequences of higher plants likely originate from a tRNA ancestor. Nucleic Acids Res 14:8111–8119

Broun P, Ganal MW, Tanksley SD (1992) Telomeric repeats display high levels of heritable polymorphism among closely related plant varieties. Proc Natl Acad Sci USA 89:1354–1357

Doyle JJ (1988) 5S ribosomal gene variation in the soybean and its progenitor. Theor Appl Genet 75:621–624

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

Funke R, Kolchinsky A, Gresshoff PM (1993) Physical mapping of a region in the soybean genome containing duplicated sequences. Plant Mol Biol 22:437–446

Ganal M, Riede I, Hemleben V (1986) Organization and sequence analysis of two related satellite DNAs in cucumber (*Cucumis sativus* L.). J Mol Evol 23:23–30

Ganal MW, Broun P, Tanksley SD (1992) Genetic mapping of tandemly repeated telomeric DNA sequences in tomato. Genomics 14:444–448

Gill KS, Gill BS, Endo TR (1993) A chromosome region-specific mapping strategy reveals gene-rich telomeric ends in wheat. Chromosoma 102:374–381

Gurley WB, Hepburn AG, Key JL (1979) Sequence organization of the soybean genome. Biochim Biophys Acta 561:167–183

Ingham LD, Hanna WW, Baier JW, Hannah LC (1993) Origin of the main class of repetitive DNA within selected *Pennisetum* species. Mol Gen Genet 238:350–356

Kato I, Yakura K, Tanifugi S (1984) Sequence analysis of *Vicia faba* repeated DNA, the *Fok*I element. Nucleic Acids Res 12:6415–6426

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

Kolchinsky A, Gresshoff PM (1992) Nucleotide sequence of the 5S rRNA gene from *Glycine soja*. Plant Mol Biol 19:1045–1047

Kolchinsky A, Gresshoff PM (1993) Direct end-labeling of telomeres. Genome 36:224–229

Kolchinsky A, Gresshoff PM (1994) Plant telomeres as molecular markers. In: Gresshoff PM (ed) Plant genome analysis. CRC Press, Roca Baton, Fla., pp 117–124

Lark KG, Weisemann JM, Matthews BF (1993) A genetic map of soybean (*Glycine max* L.) using an intraspecific cross '*Minsoy*' and '*Noir 1*'. Theor Appl Genet 86:901–906

Moore G, Gale MD, Kurata N, Flavell RB (1993) Molecular analysis of small grain cereal genomes: current status and prospects. Bio/Technology 11:584–589

Röder MS, Lapitan NLV, Sorrels ME, Tanksley SD (1993) Genetic and physical mapping of barley telomeres. Mol Gen Genet 238:294–303

Sambrook J, Fritsch EF, Maniatis T (1989). Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Shoemaker RC, Guffy RD, Lorenzen LL, Specht JE (1992) Molecular genetic map in soybean: map utilization. Crop Sci 32: 1091–1098

Singh RJ, Hymowitz T (1988) The genomic relationship between *Glycine max* (L.) Merr. and *Glycine soja* (Zieb. and Zucc.) as revealed by pachytene chromosome analysis. Theor Appl Genet 76:705–711

Vogt P (1992) Code domains in tandem repetitive DNA sequence structure. Chromosoma 101:585–589

Xia X, Erickson L (1993) An AT-rich satellite DNA sequence, E180, in alfalfa (*Medicago sativa*). Genome 36:427–432

Young ND, Fatokun CA, Danesh D, Menancio-Hautea D (1992) RFLP mapping in cowpea. In: Thottappilly G, Montin LM, Mohan Raj DR, Moore AW (eds) Biotechnology: enhancing research on tropical crops in Africa. CTA/IITA co-publication. IITA, Ibadan, Nigeria, pp 237–246