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 XhoI. 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 x 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-
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.

Fig. 1

Fig. 2a

Fig. 2b

ments of the 92-bp unit generated by XhoI digestion were isolated from a larger PAAG and cloned into the SalI 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 EcoRI, 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 NH₄Cl·COO (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 (BstXI, ClaI, SphI, XbaI, XhoI) (Fig. 1). To study this product soybean DNA was digested with XhoI 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 XhoI and XbaI (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. XhoI 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).