Reprinted from *Crop Science*
Vol. 35, No. 5

Integration of Simple Sequence Repeat DNA Markers into a Soybean Linkage Map

Mahinur S. Akkaya, Randy C. Shoemaker, James E. Specht, Arvind A. Bhagwat, and Perry B. Cregan*

ABSTRACT

A total of 40 simple sequence repeat (SSR) or microsatellite DNA markers were mapped in a soybean [*Glycine max* (L.) Merrill] mapping population that consisted of 60 F_2 plants from a cross between near isogenic lines of the cultivars Clark and Harosoy. The first objective of study was to determine the map location of SSR loci in relation to 13 classical loci controlling pigmentation and morphological traits, seven isozyme loci, and a total of 118 RFLP and RAPD markers. The second objective was to determine if the microsatellite loci were randomly distributed in the soybean genome. Linkage analysis with MAPMAKER 3.0b yielded 29 linkage groups with a total map length of 1486 centimorgans (cM). This compares with a map length of 1056 cM if the SSR markers were removed from the data set. Thirty-four of the microsatellite loci were placed in linkage groups. SSR loci were linked to loci controlling nine of the 13 classical traits, and two of seven isozyme loci. Eighteen of the 29 linkage groups contained at least one SSR locus. While this result suggested that the microsatellite loci were randomly distributed throughout the soybean genome, two clusters of five and four SSR loci, spanning 23.4 and 33.6 cM, respectively, were detected. These results indicated a relatively limited amount of clustering of soybean SSR loci, and demonstrated that microsatellite genetic markers should provide an excellent complement to RFLP and RAPD markers for use in soybean molecular biology, genetics, and breeding research. Because SSR markers detect only single genetic loci and are highly polymorphic, they can be extremely informative in pedigree tracing studies, in the analysis of progeny from multiparent matings, in a wide range of mapping applications, and in genotype identification.

GENETIC LINKAGE MAPS serve the plant geneticist in a number of ways from marker assisted selection in plant improvement to map based cloning in molecular genetic research. Thus, in a widely studied and economically important species such as soybean [*Glycine max* (L.) Merrill], a well developed and broadly useful linkage map would be a valuable resource. Ideally, such a map should include many classical markers with discernable phenotypic effects, isozyme markers, as well as a large number of highly informative DNA markers evenly spaced throughout the genome.

Using a mapping population derived from an interspecific *G. max* × *G. soja* Siebold & Zucc. cross, Shoemaker and Olson (1993) developed a molecular genetic linkage map that consisted of about 365 RFLP (Restriction Fragment Length Polymorphism), 11 RAPD (Random Amplified Polymorphic DNA), and only three classical and four isozyme loci. In contrast, the current soybean classical linkage map consists of 63 loci in 19 small linkage groups of a few loci each (Palmer and Hedges, 1993). Because of the almost complete lack of common markers between the molecular and classical maps, map integration has not been possible. The beginning of the integration of the various marker types into a common linkage map was only recently initiated (Shoemaker and Specht, 1995) with a soybean mapping population that resulted from a cross between specially constructed near-isogenic lines of the cultivars Clark and Harosoy. In their recent work, Shoemaker and Specht (1995) attempted to create a linkage map using 13 classical and seven isozyme loci along with 110 RFLP and eight RAPD markers. Their goal was to integrate classical and molecular markers into a common map. These authors used RFLP probes, restriction enzymes, and segregating DNA bands in the Clark × Harosoy mapping population that had previously been mapped in the *G. max* × *G. soja* population (Shoemaker and Olson, 1993). This produced a set of RFLP loci that were directly comparable in the Clark × Harosoy and the *G. max* × *G. soja* mapping populations. With this strategy, some progress was made in associating molecular and classical markers. For example, because of common RFLP loci, two small linkage groups in the Clark × Harosoy map that contained the *L1* and *Dt1* loci were shown to be homologous with Molecular Linkage Group 1 (MLG-1). Because *L1* and *Dt1* were present in Classical Linkage Group 05 (CLG-05) and were linked to isozyme markers *Pgd*1 and *Pgi*1 the position of these isozyme loci in MLG-1 could also be inferred. In like manner, Shoemaker and Specht (1995) were able to assign a number of CLGs to their homologous MLGs.

Two characteristics of RFLP markers in soybean tend to complicate the task of consolidating linkage maps from different mapping populations. First, only rarely have more than two alleles been identified at RFLP loci in soybean. Because these alleles generally have asymmetric frequencies, e.g., $p > 0.9$, $q < 0.1$ (Keim et al., 1989; Keim et al.,1992), the likelihood that any two genotypes will be polymorphic at a particular RFLP locus is relatively low. This is particularly true when the mapping population is created from adapted soybean germplasm (Apuya et al., 1988; Lark et al., 1993). In the case of the work reported by Shoemaker and Specht (1995) only 118 of 365 RFLP markers reported to be

M.S. Akkaya, Dep. of Chemistry, Middle East Technical Univ., Ankara, 06531, Turkey; R.C. Shoemaker, USDA-ARS-FCR, Dep. of Agronomy, Iowa State Univ., Ames, IA 50011; J.E. Specht, Dep. of Agronomy, Univ. Nebraska, Lincoln, NE 68583; A.A. Bhagwat, Agronomy Dept., Univ. of Maryland, College Park, MD 20742; and P.B. Cregan, USDA-ARS, Soybean and Alfalfa Research Lab., Beltsville, MD 20705. Received 4 Feb. 1995. *Corresponding author (pcregan@asrr.arsusda.gov).

Published in Crop Sci. 35:1439–1445 (1995).

Abbreviations: bp, base pairs; CLG , classical linkage group; cM, centimorgan; MLG, moleular linkage group; NIL, near isogenic line; PCR, polymerase chain reaction; RAPD, random amplified polymorphic DNA; RFLP, restriction fragment length polymorphism; SSR, simple sequence repeat.

polymorphic in the *G. max* × *G. soja* population were found to be segregating in the Clark × Harosoy mapping population. Thus, a particular polymorphic fragment that can be mapped in one population may not be segregating in another. A second factor that complicates the use of RFLP in soybean is the multiple DNA banding patterns detected with most probes. This may be the result of the tetraploid origin of soybean (Hymowitz and Singh, 1987). Because one band in a multiple banding pattern may segregate in one population and a different or an additional band in another, one must define an RFLP locus not only by the probe and restriction enzyme being used, but also by the molecular weight of the segregating band(s). Up to nine independent loci have been mapped by specific RFLP probes (Lark et al., 1993; Shoemaker et al., 1992). Thus, when comparing RFLP linkage maps, one must be aware of not only the probe being used, but also the restriction enzyme and the particular fragment being mapped. These various considerations have increased the difficulty in the use of RFLP markers for comparing linkage maps derived from different soybean mapping populations.

One possible solution to this complexity is the use of an alternative DNA marker system. Simple sequence repeat (SSR) or microsatellite DNA markers are now in common use in human and other mammalian genetics (Murray et al., 1994; Dietrich et al., 1994). A SSR is composed of tandemly repeated two to five nucleotide DNA core sequences such as $(CA)_n$, $(ATT)_n$, or $(AGAT)_n$. The DNA sequences flanking microsatellites are generally conserved within individuals of the same species, allowing the selection of polymerase chain reaction (PCR) primers that will amplify the intervening SSR in all genotypes. Variation in the number of tandem repeats, *n*, results in different PCR product lengths (Weber and May, 1989; Litt and Luty, 1989; Tautz, 1989).

Microsatellite markers are being developed in a number of plant species including maize (*Zea mays* L. , Senior and Heun, 1993), brassica (*Brassica* spp., Lagercrantz et al., 1993), wheat (*Triticum aestivum* L., Devos et al., 1994), grape (*Vitis* spp., Thomas and Scott, 1993), barley (*Hordeum vulgaris* L., Saghai-Maroof et al., 1994), rice (*Oryza sativa* L., Wu and Tanksley, 1993; Zhao and Kochert, 1993), sunflower (*Helianthus annuus* L., Brunel, 1994), avocado (*Persea americana* M., Lavi et al., 1995), and *Arabidopsis thaliana* (L.) Heynh. (Bell and Ecker, 1994). In soybean, Akkaya et al. (1992) and Morgante and Olivieri (1993) demonstrated that SSRs were present and exhibited high levels of length polymorphism. More recently reports in soybean (Cregan et al., 1994a; Jiang et al., 1995) have described microsatellite loci with as many as 26 alleles. This high level of allelic diversity makes it likely that a particular locus will be polymorphic in multi-parent populations and even in two-parent populations derived from the hybridization of adapted soybean genotypes. Another virtue of SSR markers is their simplicity. In the development of these markers, care is taken to select PCR primers that produce one amplification product in an inbred soybean genotype (Cregan et al. 1994b). Primer sets producing more than one product are discarded. Thus, the difficulty of the genetic interpretation of multiple banding patterns is eliminated.

The goal of our research group is to develop and map between 250 and 300 SSR loci in soybean. The objective of this report is to summarize the results obtained from the first 40 microsatellite loci developed as a result of this effort, to demonstrate the integration of SSR DNA markers into the linkage map developed from the Clark × Harosoy mapping population reported by Shoemaker and Specht (1995), and to assess the randomness of the distribution of the SSR markers in the resulting linkage map.

MATERIALS AND METHODS

Development of Simple Sequence Repeat Markers

Selection of SSR-containing Sequences from GenBank

Soybean sequences available in GenBank were searched for the presence of all possible mono-, di-, and trinucleotide SSRs. Only those microsatellites with a minimum length of 20 bp (e.g., 10 repeat units in the case of dinucleotide SSRs) were selected.

Selection of Microsatellite-Containing Clones from a Genomic Soybean DNA Library

Soybean DNA (cv. Williams) fragments of 400 to 600 bp in length, obtained by digestion with various restriction enzymes were cloned into pBluescript II KS$^+$ (Stratagene, LaJolla, CA)[1] and used to transform *Escherichia coli* XL1-Blue. The library was screened by colony hybridization (Sambrook et al., 1989). Filters were hybridized to ^{32}P-labeled oligonucleotide probes (oligo-CT, oligo-AT, and oligo-ATT) and exposed to X-ray film. Positive clones identified in the first screening were purified and screened again with the appropriate SSR-containing oligonucleotide probe to verify the presence of the microsatellite. For a more detailed description of these procedures see Cregan et al. (1994b).

The sequence of clones producing positive hybridization signals was determined. Plasmid DNA of sequencing quality was obtained with the QIAwell-8 Plasmid Kit (Qiagen Inc., Chatsworth, CA). Sequencing with an Applied Biosystems International 370A DNA Sequencer (ABI, Foster City, CA) was from both the KS and SK primer sites of pBluescript using the *Taq* Dye Deoxy Terminator Cycle Sequencing Kit (ABI).

Selection and Preparation of PCR Primers to SSR Loci

Primers were selected using DesignerPCR (Research Genetics, Huntsville, AL). In all cases, the thermal melting point (Tm) of primers was 53.5 ± 1.0°C. The Tm was necessarily rather low because of the very low G-C content of many of the target sequences. The selection of nearly identical Tm values for all primers was purposeful; it allowed the use of a standard set of PCR amplification conditions for all loci. Primers were synthesized on a 391 DNA Synthesizer, PCR-MATE (ABI).

[1] Mention of a trademark or proprietary product does not constitute a guarantee or warranty and does not imply its approval over other products that may be suitable.

Soybean Genetic Material and DNA Isolation

A cross was made between near isogenic lines (NIL) of the soybean cultivars Clark and Harosoy to create a mapping population in which a total of 13 classical and seven isozyme loci were segregating (Shoemaker and Specht, 1995). The phenotypes of each of 60 F_2 plants were characterized for each of these traits as well as for 110 RFLP and eight RAPD markers as described by Shoemaker and Specht (1995).

A sample of DNA representative of each F_2 plant was obtained by compositing one leaf from each of about 40 F_3 plants in a given F_2-derived F_3 plant row. DNA was isolated with a standard CTAB (hexadecylatrimethylammonium bromide) procedure (Keim et al., 1988).

Polymerase Chain Reaction Conditions and PCR Product Separation

Reaction mixes contained 30 ng of soybean genomic DNA, 1.5 m*M* Mg^{2+}, 0.15 μ*M* of 3′ and 5′ end primers, 200 μ*M* of each nucleotide, 0.1 μL of 3000 Ci/mmol α-^{32}PdATP (10.0 μCi/μL), 1× PCR Buffer containing 50 m*M* KCl, 10 m*M* Tris-HCl pH 9.0, 0.1% Triton X-100, and 1 unit *Taq* DNA polymerase in a total volume of 10 μL. Cycling consisted of a 25-s denaturation at 94°C, 25-s annealing at 47°C, and 25-s extension at 68°C for 32 cycles on a MJ Research model PTC-100 thermocycler (MJ Research, Inc., Watertown, MA). PCR products (3 μL/lane) were separated on a standard DNA sequencing gel containing 6% polyacrylamide, 8 *M* urea, and 1× TBE, at 60 W constant power for 2 to 3 h. Following drying, gels were exposed to X-ray film.

In a number of instances, when the size difference between DNA amplification products of the two alleles at an SSR locus was greater than 8 to 10 bp in length, a 2% agarose + Synergel (Diversified Biotech, West Roxbury, MA) horizontal gel with ethidium bromide incorporated in the gel was used instead of a sequencing gel. In these instances, the PCR reaction mixes had a total volume of 30 μL, and the concentrations of the reactants and the cycling conditions were identical to those described above, except that no α-^{32}PdATP was used in the reaction mix.

Genetic Mapping

In an initial evaluation, each primer set was tested on the Clark and Harosoy NIL parents of the mapping population. Those primer pairs that produced single PCR products that differed in size in the Clark and Harosoy NIL parents were then used to determine the genotype of each of the 60 F_2 plants in the mapping population.

The total of 40 SSR loci were characterized as described above. Goodness of fit to a 1:2:1 model was tested for each locus using χ^2. The computer program MAPMAKER 3.0b (Lander et al. 1987) was used to order the 40 microsatellite loci with respect to each other and the 110 RFLP, eight RAPD, 13 classical, and seven isozyme loci with a LOD score of 3.0 as the threshold value for the declaration of linkage in the pairwise loci analysis.

Assessment of the Randomness of Distribution of SSR Loci

In order to obtain an estimate of the randomness of distribution of the 40 SSR loci, a predicted distribution was calculated assuming markers would follow a Poisson distribution (Sokal and Rohlf, 1981). It was further assumed that (i) the total length of the soybean linkage map is 3000 cM as estimated by Shoemaker and Olson (1993), and (ii) that any randomly placed marker had an equal chance of falling into one of 60 (3000/50) linkage groups because a maximum distance of 50 cM between loci was used to detect linkage in the MAPMAKER 3.0b analysis. Based on these assumptions, the number of linkage groups expected to contain 0, 1, 2, 3, ... 40 markers was calculated. The observed number of groups in each class, versus the expected, was used to calculate a χ^2 value to test for deviation from randomness. Because the expectation in groups containing from four to 40 markers was less than 1, these groups were pooled with the group containing three markers, thereby producing a χ^2 with three degrees of freedom rather than 40 (Steel and Torrie, 1980).

RESULTS AND DISCUSSION

The 40 soybean SSR loci and the primers used in their amplification are listed in Table 1. The core motif of each locus with a "BARC" prefix is indicated in the name of the locus, i.e., BARC-Sat1 and BARC-Sct28 have $(AT)_n$ and $(CT)_n$ cores, respectively. For purposes of brevity, these loci are subsequently referred to by their suffix alone. Most of the 40 loci have $(AT)_n$ or $(ATT)_n$ core motifs. The exceptions are the two loci with $(CT)_n$, one with $(T)_n$, one with $(A)_n$, and one with $(CTT)_n$ core repeats. The test for goodness-of-fit to a 1:2:1 F_2 segregation ratio generated a significant χ^2 ($P \leq 0.05$) at the Sct28 and Satt22 loci. However, neither locus had a significant lack of fit at the 1% level of probability.

The linkage analysis of the 40 microsatellite, 110 RFLP, eight RAPD, seven pigmentation, six morphological, and seven isozyme loci using MAPMAKER 3.0b produced a genetic map that consisted of 29 linkage groups with a total length of 1486 cM. The previous report of Shoemaker and Specht (1995) in which the same RFLP, RAPD, classical, and isozyme markers were mapped produced a genetic map with 26 linkage groups and a total length of 1056 cM (LOD = 3). Thus, the addition of 40 SSR markers increased the defined map length by almost 40%. However, the *G. max* × *G. soja* map (Shoemaker and Olson, 1993) includes almost 365 RFLP and RAPD loci, with a total recombinational distance of about 3000 cM, in 20 linkage groups. Thus, the current *G. max* × *G. max* map presented here is still in its infancy.

Thirty-four of the 40 SSR loci were placed in 18 different linkage groups (Table 2) while six loci (15%) remained unlinked. A slightly larger percentage, (20.3%) of the 110 RFLP loci were found to be unlinked. Nine linkage groups contained one SSR locus. Six different linkage groups contained two SSR loci, one linkage group included three, and another group contained four. Linkage Group 13 was of interest because it contained a total of nine markers, six of which were SSRs. Microsatellite loci were determined to be linked to loci controlling five of the seven pigmentation traits, four of the six morphological traits and three of the seven isozymes segregating in the Clark × Harosoy mapping population. The addition of the SSR markers was helpful in consolidating what had previously been small linkage groups containing two or three loci in the map reported by Shoemaker and Specht (1995). For example, linkage

Table 1. Forty simple sequence repeat soybean loci mapped in the Clark × Harosoy soybean mapping population, the SSR core motif, and the 5′ end (sense) and 3′ end (anti-sense) polymerase chain reaction primers sequences to each. All primer sequences are given in the 5′ to 3′ direction.

Locus	Core motif	5′ end primer (sense)	3′ end primer (anti-sense)
SoyABAB†	$(AT)_{20}$	caaacataaaaaaggtgaga	aagaacgcacactaatattatt
SoyHSP176†	$(AT)_{15}$	ttttgtttaagttactgtactgt	gctagtcttctacaccttcta
SoyGPATR†	$(CTT)_4CCTCT(CTT)_7$	ggaagaaagtattggtctgt	aggagagagtggagagatta
GMGLPSI2†	$(T)_{19}$	ttcgaagcattcaagg	aaaagacaaaaacatactataaaa
SoyPRP1†	$(ATT)_{19}$	cgtgccaaattacatca	tgatgggaacaagtacataa
SoyLBC†	$(AT)_{27}$	gtggtttcagtgagtgatc	aaggtggagatgaactca
GMRUBP†	$(A)_{20}$	ctggcgtgctaaaagta	ggacagatttgatcaataatt
BARC-Sat1‡	$(AT)_{17}$	gatacgaccaaaaattgtt	aactgcgaagatactaccc
BARC-Sat3‡	$(AT)_{21}$	tgatttttggtgtagaactc	acctcattcgattggttaaac
BARC-Sat9‡	$(TA)_{28}$	cacacgtattgtcttaccac	ctccgagaagcacgta
BARC-Sat20‡	$(AT)_{33}$	aaaggagtaggatgtaagaga	caaaccaccagtgacc
BARC-Sat22‡	$(AT)_{26}$	gccttttctgactgttaa	cagtgactaaaacttactat
BARC-Sat33‡	$(AT)_{16}$	gcagatatttaatcccctcacc	atctcggagagaaagaaatcgc
BARC-Sat36‡	$(TA)_{19}$	actccaagttttttttgttt	ggagttagaggaagagaaca
BARC-Sat38‡	$(AT)_{21}$	cttccaatttgagactctta	gttctttttaacaacactcactt
BARC-Sat39‡	$(AT)_{17}$	caagaataatctaaaggtacactt	agttaaaaaacccacacaac
BARC-Sat42‡	$(AT)_{28}$	catagtaatgctatttgccc	tttttaacagaccactatttt
BARC-Sat43‡	$(AT)_{20}$	ccgtcaatgaatattaaatta	ttacgagtaggaaaataactaaa
BARC-Sat44‡	$(AT)_{21}$	aaaaaatatttatagttacatgtg	ttaccactaagaattaggtctaa
BARC-Sct26‡	$(AG)_{14}$	cgaaacgcaaaatctc	aaaacgtatctgaagtagtgg
BARC-Sct28‡	$(CT)_{14}$	tcgccggtacaaaag	cgaatgaacaaacaaaca
BARC-Satt1‡	$(ATT)_{24}$	aaagtctttaaaagtgtgtctta	ttaaaagaaaaatgcaacat
BARC-Satt2‡	$(ATT)_{16}$	tgtgggtaaaatagataaaaat	tcattttgaatcgttgaa
BARC-Satt6‡	$(ATT)_{10}$	aaaccctttaaattcaaatt	ttaaagcgcataaacaatat
BARC-Satt5‡	$(TAA)_{21}$	tatcctagagaagaactaaaaaa	gtcgattaggcttgaaata
BARC-Satt9‡	$(AAT)_{14}$	ccaacttgaaattactagagaaa	cttactagcgtattaaccctt
BARC-Satt12‡	$(ATT)_{19}$	gcaattagttttaaaatgtttc	agaatagagcctacatataatcata
BARC-Satt14‡	$(ATT)_8$	ttttttctcttcatgtttg	cccgtatagcaaccaa
BARC-Satt20‡	$(ATT)_{16}$	gagaaagaaatgtgttagtgtaa	cttttccttcttattgtttga
BARC-Satt22‡	$(TAA)_{22}$	tgtattttaccttacctttga	aactgccaccaatgac
BARC-Satt30‡	$(TAA)_{18}$	aaaaagtgtaaccaagcc	tcttaaatcttatgttgatgc
BARC-Satt31‡	$(ATT)_{12}$	ttccactttgtatcactttc	tgactgtaaaagaacagataaa
BARC-Satt42‡	$(ATT)_{26}$	gacttaattgcttgctatga	gtggtgcacactcactt
BARC-Satt45‡	$(ATT)_{17}$	tggtttctactttctataattattt	atgcctctccctcct
BARC-Satt46‡	$(ATT)_{21}$	aaaataactaaaatgtcttctca	ttggtcagattattataagattg
BARC-Satt63‡	$(ATT)_{20}$	aaatgattaacaatgtttatgat	acttgcatcagttaataacaa
BARC-Satt70‡	$(ACT)_5(ATT)_{23}$	taaaaattaaaatactagaagacaac	tggcattagaaaatgatatg
BARC-Satt72‡	$(TAA)_9$	ggaaagaatcagcaaaat	cccccacataaataataaa
BARC-Satt77‡	$(ATT)_{13}$	gatctaaagtctgatatttttaacta	aaaaggagaaggaatgc
BARC-Satt82‡	$(ATT)_{10}$	aattcatttagggagttgat	ctagccaatgtcatatgact

† SSR containing soybean loci retrieved from GenBank.

‡ The complete locus name includes the BARC prefix which is the acronym for Beltsville Agricultural Research Center. In the text, these loci are referred to by the suffix alone.

Groups 1 and 16 (Shoemaker and Specht, 1995) were combined into the current Group 1 with a total length of 157.2 cM (Table 2). Likewise, the current Group 6 is composed of markers previously in Groups 6 and 25. RFLP loci such as A199-1 and A605-1 and RAPD locus OPA20 that were unlinked in the original report of Shoemaker and Specht (1995) were placed in linkage Group 13 along with the six SSR markers.

In general, the SSR loci appeared to be scattered across the genome, except, as noted above, for two instances (Table 2, Groups 13 and 18), where clusters of microsatellite loci were detected. The cluster in Group 13 contained the five loci SoyPRP1, Satt46, Satt1, Sat44, and Sat43 that spanned a map distance of 23.4 cM while in Group 18 a cluster of four loci spanned a distance of 33.6 cM. We were concerned, particularly in the case of the cluster in Group 13, that these mapping data suggested a non-random distribution of SSR loci in the soybean genome. In humans, microsatellite markers [mainly $(CA)_n$ repeats] ranged in density from 0.28 markers per megabase on chromosome 21 to 0.75 markers per megabase on chromosome 17 (Weissenbach et al. 1992). Other human geneticists have reported that SSRs are generally distributed randomly throughout the human genome (Decker et al., 1992; Hudson et al., 1992; Wilkie et al., 1992). In mice, Dietrich et al. (1994) indicated that the distribution of $(CA)_n$ repeats was random except in the case of the X chromosome.

The only relevant data for plant SSR markers were provided by Bell and Ecker (1994) who mapped a total of 30 microsatellites in *Arabidopsis*. They found that, based upon their length, two of the five Arabidopsis chromosomes had a smaller than expected number of SSR loci. In addition, two clusters of four microsatellite loci were noted. One spanned a distance of 9.6 cM and the other 5.6 cM. While SSR markers in *Arabidopsis* appeared to deviate somewhat from a random distribution, Bell and Ecker (1994) concluded that a greater number of markers would need to be mapped before a definitive conclusion relative to their random or non-random distribution could be reached.

When the distribution of soybean SSR markers in the current map was compared with the expectation based on a Poisson distribution, a significant deviation from randomness was noted. However, 85% of the accumulated χ^2 was the result of the one cluster of five loci in Linkage Group 13 (Table 2). The expected number of linkage groups with 0, 1, 2, 3, and 4 markers was very

Table 2. Linkage groups containing simple sequence repeat DNA loci in the Clark × Harosoy soybean mapping population detected using MAPMAKER 3.Ob.

Group	Locus	Distance	Group	Locus	Distance	Group	Locus	Distance	Group	Locus	Distance
		cM			cM			cM			cM
1	B219-2	6.3	6	A063-2	4.2	15	OPE142	3.8	23	K002-2	32.9
	Pd1	15.5		L148-1	1.1		Sat39	–		K002-1	37.4
	D2	0.0		A635-1	4.0			3.8		SoyHSP176	13.5
	D1	6.6		K011-3	8.5					K644-4	3.4
	G	28.1		T	31.3	16	Satt30	17.0		K644-1	–
	A702-2	36.7		A538-1	0.0		GMRUBP	20.5			87.2
	Sat36	24.3		P029-1	5.0		A020-2	–			
	Satt77	24.3		Sct28	7.0			37.5	24	SoyABAB	24.4
	A691-1	39.5		A748-1	–					B162-1	12.3
	B214-1	–			61.2	18	A519-1	7.5		Sat33	18.0
		157.2					A593-1	1.3		A280-1	13.5
			7	E2	27.7		A516-1	11.8		Satt9	–
2	A668-1	9.2		Sat38	11.4		Satt63	24.3			68.3
	Sat20	22.3		Pgm1	29.3		Satt20	5.5			
	E014-2	2.0		SoyLBC	–		Satt70	3.8	26	A816-1	18.8
	P1	–			68.5		Sat9	33.1		Satt12	16.2
		33.5					A584-2	–		A121-2	–
			8	Satt6	1.8			87.3			34.9
3	Y9	22.1		Dt1	0.0						
	A661-2	6.8		A461-1	–	21	A664-1	17.8	28	Sat1	8.2
	A069-2	40.7			1.8		B039-2	3.7		Satt82	8.5
	A646-1	0.0					GMGLPSI2	–		K258-2	–
	A454-2	4.2	12	Satt2	14.2			21.5			16.7
	A656-1	10.6		Mdh	14.6						
	pcr2110	0.0		Satt14	–	22	A065-2	6.5	29	Sat22	22.2
	A598-2	3.1			28.8		A083-2	8.9		Satt31	–
	Satt45	6.9					Satt42	0.0			22.2
	R013-1	4.7	13	SoyPRP1	1.8		K647-2	–			
	K274-2	9.6		Satt46	3.0			15.4		Unlinked RFLP Markers: A112-1, A064-1, A404-3, B039-2, A256-3, K002-3, A085-1, A426-3, A586-3, A664-1, A060-1, A135-2, A124-2, L050-3, A685-1, A703-2, B030-2, K401-3, K265-2, R183-2, T092-2	
	A597-1	35.5		Satt1	7.8						
	A136-2	–		Sat44	4.3		Unlinked Classical Markers: I, Pgi, Idh1				
		144.3		OPA20	8.3						
				Sat43	27.4		Unlinked RAPD Markers: OPC-07, UBC-90, pcr2190				
				A199-1	42.0						
				Satt5	24.0					Unlinked SSR Markers: Sct26, Sat42, Sat3, Satt22, Satt72, SoyGPATR	
				A605-1	–						
					118.7						

similar to the observed. Thus, it is reasonable to conclude that with the one exception of the cluster in Linkage Group 13, SSR markers appeared to be dispersed throughout the soybean genome in a random fashion. Clearly, further SSR loci will need to be identified and mapped before the question of the randomness of their distribution can be answered with certainty. However, based upon the 40 loci included in this study, it does not appear likely that clustering of SSR loci in soybean will be a deterrent to the development of a microsatellite-based linkage map covering most of the genome.

The broad distribution of SSR loci in the soybean genome suggests that simple sequence repeats provide a useful source of genetic markers to complement the RFLP and RAPD markers now being used by geneticists and breeders. The high level of polymorphism associated with SSR markers makes it more likely that any given locus will be polymorphic and, thus informative, in any specific biparental population. This was demonstrated by Jiang et al. (1995) who reported that, among a group of 27 soybean cultivars, an average of more than seven alleles was detected at seven different SSR loci. The multiple allelism of SSR loci is also of great advantage in pedigree analysis and in the analysis of multi-parent intermating populations. Furthermore, microsatellite markers avoid the difficulties of genetically dissecting multiple-banding patterns with respect to the multiple genetic loci that often occur with RFLP markers in soybean. Because SSRs are selected as single-locus markers, they map to reproducible positions in the genome without the complications associated with RFLP or RAPD markers. As a result, SSR markers can be used as dependable, comparable, and highly informative points of reference across a wide range of soybean mapping and breeding populations.

One of the drawbacks to SSR markers is their cost of development. However, once loci are defined, the primer sequences are published, and the primers become available commercially at reasonable prices, their cost will not be prohibitive. The detection of SSR alleles using sequencing gels is often viewed as a difficulty. However, while most of the SSR loci reported here were mapped using sequencing gels, a number were mapped using agarose gels. Other plant geneticists use agarose gels to differentiate the amplification products of microsatellite alleles (Bell and Ecker, 1994; Devos et al. 1994; Morgante and Olivieri, 1993; Senior and Heun, 1993). In addition, other techniques for the sizing of SSR alleles are available or under development. These include the non-radioactive sequential detection of alleles at as many as 10 to 20 microsatellite loci in a single gel lane (Vignal et al., 1993), fluorescent tagging and automated sizing of SSR-containing PCR fragments (Ziegle et al., 1992), and capillary electrophoresis using both single (Marino et al., 1994) and multiple (Clark and Mathies, 1993) channel systems for the high-speed separation and sizing

of DNA fragments. In the near future, these and a number of other technologies will offer a range of alternatives for the detection and accurate sizing of SSR alleles. The combination of numerous highly informative markers that can be efficiently and relatively inexpensively detected should promote the broader use and acceptance of DNA markers in soybean genetics and plant breeding research programs.

ACKNOWLEDGMENTS

The authors wish to thank Mr. Edward Fickus and Mr. Charles Quigley for their excellent technical assistance during the course of the work reported here. We are grateful to Drs. K. G. Lark and D. Lohnes for their careful reading and helpful suggestions during the preparation of this manuscript. This research was supported in part by funds from the USDA, Cooperative State Research Service, National Research Initiative, Competitive Grants Program, Grant No. 9300874 to PBC, JES, and RCS for which the authors are very grateful. We also thank the USDA, ARS Postdoctoral Research Associate Program for support of MSA.

REFERENCES

Akkaya, M.S., A.A. Bhagwat, and P.B. Cregan. 1992. Length polymorphism of simple sequence repeat DNA in soybean. Genetics 132:1131–1139.

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

Bell, C.J., and J.R. Ecker. 1994. Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. Genomics 19:137–144.

Brunel, D. 1994. A microsatellite marker in *Helianthus annuus* L. 1994. Plant Mol. Biol. 24:397–400.

Clark, S.M., and R.A. Mathies. 1993. High-speed parallel separation of DNA restriction fragments using capillary array electrophoresis. Anal. Biochem. 215:163–170.

Cregan, P.B., M.S., Akkaya, A.A. Bhagwat, U. Lavi, Jiang Rongwen. 1994a. Length polymorphisms of simple sequence repeat (SSR) DNA as molecular markers in plants. p. 43–49. *In* P.M. Gresshoff (ed.) Plant genome analysis. CRC Press. Boca Raton, FL.

Cregan, P.B., A.A., Bhagwat, M.S. Akkaya, and Jiang Rongwen. 1994b. Microsatellite fingerprinting and mapping of soybean. Methods. Mol. Cell. Biol. 5:49–61.

Decker, R.A., J. Moore, B. Ponder, and J.L. Weber. 1992. Linkage mapping of human chromosome 10 microsatellite polymorphisms. Genomics 12:604–606.

Devos, K.M., G.J. Bryan, A.J. Collins, P. Stephenson, and M.D. Gale. 1995. Application of two microsatellite sequences in wheat storage proteins as molcular markers. Theor. Appl. Genet. 90: 247–252.

Dietrich, W.F., J.C. Miller, R.G. Steen, M. Merchant, D. Damron, R. Nahf, A. Gross, D.C. Joyce, M. Wessel, R.D. Dredge, A. Marquis, L.D. Stein, N. Goodman, D.C. Page, and E.S. Lander. 1994. A genetic map of the mouse with 4006 simple sequence length polymorphism. Nature Genet. 7:220–225.

Hymowitz, T., and R.J. Singh. 1987. Taxomony and speciation. p. 23–48. *In* R.J. Wilcox (ed.) Soybeans: Improvement, production, and uses. 2nd ed. Agron. Monogr. 16. ASA, CSSA, SSSA, Madison, WI.

Hudson, T.J., M. Engelstein, M.K. Lee, E.C. Ho, M.J. Rubenfield, C.P. Adams, D.E. Housman, and N.C. Dracopoli. 1992. Isolation and chromosomal assignment of 100 highly informative human simple sequence repeat polymorphisms. Genomics 13:622–629.

Jiang, Rongwen, M.S. Akkaya, A.A. Bhagwat, U. Lavi, and P.B. Cregan. 1995. The use of microsatellite DNA markers for soybean genotype identification. Theor. Appl. Genet. 90:43–48.

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

Keim, P., T. Olson, and R. Shoemaker. 1988. A rapid protocol for isolating soybean DNA. Soy. Genet. Newsl. 15:150–152.

Keim, P., R.C. Shoemaker, and R.G. Palmer. 1989. Restriction fragment length polymorphism diversity in soybean. Theor. Appl. Genet. 77:786–792.

Lagercrantz, U., H. Ellegren, and L. Andersson. 1993. The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Res. 21:1111–1115.

Lander, E.S., P. Green, J. Abrahamson, A. Barlow, M.J. Daly, S.E. Lincoln, and L. Newburg. 1987. Mapmaker: an interactive computer package for constructing genetic linkage maps of experimental and natural populations. Genomics 1:174–181.

Lark, K.G., J.M. Weisemann, B.F. Matthews, R. Palmer, K. Chase, and T. Macalma. 1993. A genetic map of soybean (*Glycine max* L.) using an introspecific cross of two cultivars: 'Minsoy' and 'Noir 1'. Theor. Appl. Genet. 86:901–906.

Lavi, U., M.S. Akkaya, A.A. Bhagwat, E. Lahav, and P.B. Cregan. 1995. Generation of simple sequence repeat DNA markers in avocado (*Persea americana* M.). Euphytica 80:171–177.

Litt, M., and J.A. Luty. 1989. A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am. J. Hum. Genet. 44: 397–401.

Marino, M.A, L.A.Turni, S.A. DelRio, P.E. Williams, and P.B. Cregan. 1994. The analysis of simple sequence repeat DNA in soybean by capillary gel electrophoresis. Appl. Theor. Electrophoresis (In press).

Morgante, M., and A.M. Olivieri. 1993. PCR-amplified microsatellites as markers in plant genetics. Plant J. 3:175–182.

Murray, J.C., K.H. Buetow, J.L. Weber, S. Ludwigsen, T. Scherpbier-Heddema, F. Manion, J. Quillen, V.C. Sheffield, S. Sunden, G.M. Duyk, J. Weissenbach, G. Gyapay, C. Dib, J. Morrissette, G.M. Lathrop, A. Vignal, R. White, N. Matsunami, S. Gerken, R. Melis, H. Albertsen, R. Plaetke, S. Odelberg, D. Ward, J. Dausset, D. Cohen, and H. Cann. 1994. A comprehensive human linkage map with centimorgan density. Science 265:2049–2054.

Palmer, R.G., and B.R. Hedges. 1993. Linkage map of soybean (*Glycine max* L. Merr.). p. 6.139–6.148. *In* S.J. O'Brien (ed.) Genetic maps: Locus maps of complex genomes. Cold Spring Harbor Laboratory Press, New York.

Saghai-Maroof, M.A., R.M. Biyashev, G.P. Yang, Q. Zhang, and R.W. Allard. 1994. Extraordinarily polymorphic microsatellite DNA in barley: Species diversity, chromosomal locations, and population dynamics. Proc. Natl. Acad. Sci. USA 91:5466–5470.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular cloning: A laboratory manual. Second Ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Senior, M.L., and M. Heun. 1993. Mapping maize microsatellites and polymerase chain reaction confirmation of targeted repeats using a CT primer. Genome 36:884–889.

Shoemaker, R.C., R.D. Guffy, L.L. Lorenzen, and J.E. Specht. 1992. Molecular genetic mapping of soybean: Map utilization. Crop Sci. 32:1091–1098.

Shoemaker, R.C., and T.C. Olson. 1993. Molecular linkage map of soybean (*Glycine max* L. Merr.). p. 6.131–6.138. *In* S.J. O'Brien (ed.) Genetic Maps: Locus maps of complex genomes. Cold Spring Harbor Laboratory Press, New York.

Shoemaker, R.C., and J.E. Specht. 1995. Integration of the soybean molecular and classical genetic linkage groups. Crop Sci. 35:436–446.

Sokal, R.R., and F.J. Rohlf. 1981. Biometry. Freeman, San Francisco, CA.

Steel, R.G.D., and J.H. Torrie. 1980. Principles and procedures of statistics. McGraw-Hill, New York.

Tautz, D. 1989. Hypervariability of simple sequences as a general source of polymorphic DNA markers. Nucleic Acids Res. 17: 6463–6471.

Thomas, M.R., and N.S. Scott. 1993. Microsatellite repeats in grapevine reveal DNA polymorphisms when analysed as sequence-tagged sites (STSs). Theor. Appl. Genet. 86:985–990.

Vignal, A., G. Gyapay, J. Hazan, S. Nguyen, C. Dupraz, N. Cheron, N. Becuwe, M. Tranchant, J. Weissenbach. 1993. Nonradioactive multiplex procedure for genotyping of microsatellite markers. p.

211–221. *In* K.W. Adolph (ed): Methods in molecular genetics–Gene and chromosome analysis. Part A. Academic Press, New York.

Weber, J.K., and P.E. May. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Human Genet. 44:388–397.

Weissenbach, J., G. Gyapay, C. Dib, A. Vignal, J. Morissette, P. Millasseau, G. Vaysseix, and M. Lathrop. 1992. A second-generation linkage map of the human genome. Nature 359:794–801.

Wilkie, P.J., D.B. Krizman, and J.L. Weber. 1992. Linkage map of human chromosome 9 microsatellite polymorphisms. Genomics 12:607–609.

Wu, Kun-sheng, and S.D. Tanksley. 1993. Abundance, polymorphism and genetic mapping of microsatellite in rice. Mol. Gen. Genet. 241:225–235.

Zhao, Xingping, and G. Kochert. 1993. Phylogenetic distribution and genetic mapping of a $(GGC)_n$ microsatellite from rice (*Oryza sativa* L.). Plant Mol. Biol. 21:607–614.

Ziegle, J.S., Y. Su, K.P. Corcoran, L. Nie, P.E. Mayrand, L.B. Hoff, L.J. McBride, M.N. Kronick, and S.R. Diehl. 1992. Application of automated DNA sizing technology for genotyping microsatellite loci. Genomics 14:1026–1031.