Theor Appl Genet (1997) 95:723–733

N. Diwan · P. B. Cregan

Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean

Received: 24 March 1997 / Accepted: 4 April 1997

Abstract Simple Sequence Repeat (SSR) allele sizing provides a useful tool for genotype identification, pedigree analysis, and for estimating genetic distance between organisms. Soybean [*Glycine max* (L.) Merr.] cultivars are identified for Plant Variety Protection (PVP) purposes by standard pigmentation and morphological traits. However, many commercial soybeans arise from a limited number of elite lines and are often indistinguishable based on these traits. A system based on SSR markers would provide unique DNA profiles of cultivars. Fluorescent labeling of alleles combined with automated sizing with internal size standards in each gel lane was used as an alternative to standard [^{32}P] labeling to assess genetic variability in soybean. Allelic frequencies at 20 SSR loci were determined in 35 soybean genotypes that account for greater than 95% of the alleles in North American soybean cultivars based upon pedigree analysis. An average of 10.1 alleles per locus (range: 5–17), with a mean gene diversity of 0.80 (range: 0.50 to 0.87) were observed at the 20 SSR loci. The 20 loci successfully distinguished modern soybean cultivars that are identical for morphological and pigmentation traits, as well as 7 soybean genotypes reported to be indistinguishable using 17 RFLP probes. Pedigrees of 7 cultivars were studied to estimate stability of SSRs in soybean across generations. Of the 7 pedigrees 6 had one locus in the progeny with an allele(s) that was not present in either parent. These new alleles are most likely the result of mutation. The mutation rate of SSR alleles in soybean was similar to that reported in humans. To avoid difficulty associated with mutation, DNA fingerprint data should be determined from the bulk of 30-50 plants of a cultivar.

Key words DNA fingerprinting · Microsatellites · Plant variety protection · Mutation rate

Communicated by A. L. Kahler

N. Diwan[1] · P. B Cregan (✉)
Soybean and Alfalfa Research Laboratory, USDA-ARS, BARC-West, Bldg. 011, HH-19, Beltsville, MD 20705, USA

Present address:
[1] Plant Genome Center, Department of Plant Genetics, Weizmann Institute, Rehovot, Israel 76100

Introduction

Simple Sequence Repeat (SSR) DNA markers, also known as microsatellites or Short Tandem Repeats (STR), have been successfully used by human geneticists for parentage testing, forensic identification, and medical diagnostics (Edwards et al. 1992; Alford et al. 1994; Hammond et al. 1994). In these studies SSRs were found to provide accurate, reliable, and rapid testing for these purposes. SSRs have also been used in the construction of genetic linkage maps of human (Decker et al. 1992; Hudson et al. 1992; Weissenbach et al. 1992; Wilkie et al. 1992; Gyapay et al. 1994; Murray et al. 1994; Hudson et al. 1995) as well as numerous other mammalian and other animal species such as rat (Serikawa et al. 1992; Goldmuntz et al. 1993; Du et al. 1995), mouse (McAleer et al. 1992; Dietrich et al. 1994), pig (Ellegren et al. 1993; Rohrer et al. 1994; Schook et al. 1994), cattle (Moore et al. 1994) and chicken (Cheng and Crittenden 1994; Crooijmans et al. 1994).

In plants, SSRs have been used for the construction of linkage maps in *Arabidopsis thaliana* (Bell and Ecker, 1994), soybean [*Glycine max* (L.) Merr.] (Akkaya et al. 1995), and maize (Senior et al. 1996). These studies showed that the genotyping of homozygous and heterozygous genotypes using SSRs is readily accomplished and that SSRs are inherited in a Mendelian fashion. SSRs were also used in crop species such as soybean (Morgante and Olivieri 1993; Cregan et al. 1994, Rongwen et al. 1995; Maughan et al. 1995), grape

(*Vitis vinifera*) (Thomas and Scott 1993) and rapeseed (*Brassica napus* L.) (Kresovich et al. 1995) to study the allelic profiles of genotypes for purposes of genotype identification.

SSR markers exhibit high levels of length polymorphism in soybean. Akkaya et al. (1992) reported as many as 8 alleles per SSR locus among a group of 43 soybean and wild soybean (*G. soja* Sieb. & Zucc.) genotypes. Morgante and Olivieri (1993) found similar levels of allelic diversity in soybean. Rongwen et al. (1995) defined allelic diversity at 7 SSR loci in a group of 96 soybean genotypes. All genotypes with the exception of 2 which were 84% identical by descent had unique allelic profiles. Maughan et al. (1995) detected a total of 79 alleles at 5 SSR loci in a sample of 94 accessions of cultivated and wild soybean. The high level of SSR length polymorphism in soybean makes it likely that a particular locus will be polymorphic even in populations resulting from the hybridization of adapted genotypes. This is in contrast to the lower genetic variation revealed by Restriction Fragment Length Polymorphism (RFLP). Keim et al. (1989) used 17 RFLP probes to characterize a group of 58 wild and adapted soybean genotypes. Despite the large number of probes, 7 genotypes had identical RFLP patterns and were therefore indistinguishable.

The Plant Variety Protection (PVP) Act of 1970 was established to help plant breeders "protect" their crop cultivars from commercial exploitation by others. While not having the same exclusivity as a patent, Plant Variety Protection, nonetheless, offers the owner of a plant cultivar legal protection for the exclusive sale of a protected cultivar. The Plant Variety Protection Office (PVPO) of the Agricultural Marketing Service, US Department of Agriculture, is charged with granting PVP certificates for new crop cultivars. A new cultivar may be protected under the PVP Act if it is demonstrated to be distinct from "all other varieties within that species". For purposes of PVP, soybean cultivars are currently distinguished by standard traits such as: flower and pubescence color, leaf shape, growth habit, maturity, and other conventional morphological and disease resistance traits. However, most commercial soybean cultivars arise from hybridization between members of an elite group of genotypes, and the amount of genetic variability among these cultivars is small. As a result, new cultivars are often indistinguishable based on these standard pigmentation and morphological traits. Additionally, as more cultivars have received protection and thereby increased the size of the PVPO database, it has become more difficult to distinguish new cultivars from those in the database. Thus, a system which is based on DNA markers and which would provide unique DNA profiles or 'fingerprints' for cultivars is needed to facilitate the protection of varieties under the PVP Act. SSR markers have a high probability of functioning successfully as a tool for cultivar identification. Combining the informativeness of SSR markers with the precision of automated allele sizing would offer a rapid and standardized system for genotype identification which could be used as an alternative to conventional microsatellite allele sizing using [^{32}P]-labeling.

The automated sizing of fluorescently labeled SSR alleles is a system in which one of the polymerase chain reaction (PCR) primers to an SSR locus is labeled with one of three colored fluorescent tags. The resulting PCR product(s) carrying different colored fluorescent labels are separated on acrylamide gels. A series of size standards labeled with a fourth fluorescent tag are employed in each gel lane to allow exact sizing of the labeled SSR alleles. The report by Kresovich et al. (1995) analyzing SSR allele sizes in rapeseed is the only published study in plants in which an automated system for the detection of fluorescent-labeled PCR products has been used. Reputed advantages of the automated system as compared to the sizing of PCR products on standard sequencing gels are: (1) the ability to visualize and analyze several products in the same lane within the same size range, but with different fluorescent labels; (2) single base resolution over a wide size range from 75–350 bases; (3) automated sizing; (4) automated data output; and (5) elimination of radioactivity. These advantages increase the efficiency and rapidity of the analysis. The system has been used in a number of studies for human identification and found to increase both the efficiency and accuracy of SSR allele size determination (Ziegle et al. 1992; Fregeau and Forney 1993; Holgersson et al. 1994; Scharf et al. 1995).

The objectives of the study described here were: (1) to establish a system to quantify genetic variability in soybean using allelic frequencies at 20 SSR loci in 35 soybean genotypes thought to account for greater than 95% of the alleles in North American soybean cultivars; (2) to determine if this system can be used to distinguish modern cultivars with identical morphological, pigmentation, and other traits commonly used for purposes of cultivar identification; and (3) to study pedigree relationships using SSRs in order to provide a preliminary estimate of the stability of SSR markers in soybean over generations.

Materials and methods

Soybean plant material and DNA isolation

Ancestors of North American soybean cultivars

Based upon pedigree analysis, Gizlice et al. (1994) identified a group of 35 genotypes to represent greater than 95% of the allelic variation present in North American cultivated soybean germplasm. Seeds of each of the 35 ancestors as well as of the cultivar 'Williams' were obtained from the USDA Soybean Germplasm Collection courtesy of Dr. R. Nelson (USDA-ARS, University of Illinois, Urbana). DNA was harvested from a bulk of the leaf tissue of 5–7 plants of each genotype.

Seven genotypes with identical RFLP profiles

Based upon analysis with 17 RFLP probes, Keim et al. (1989) described 7 soybean genotypes with identical profiles. Seeds of these genotypes were obtained courtesy of Dr. R.C. Shoemaker (USDA-ARS, Iowa State University, Ames). DNA was harvested from a composite of 5-7 plants of each genotype.

Modern cultivars with similar morphological, pigmentation and other traits

In collaboration with Dr. J. Strachan, Plant Variety Protection Office (PVPO), USDA, a search was made of the PVPO soybean database to identify groups of cultivars with similar morphological, pigmentation and growth habit characteristics (Table 1). Ten Maturity Group I, 7 Maturity Group II, 10 Maturity Group IV, and 9 Maturity Group VI cultivars were identified, and seeds were obtained from the public or private developer of each (Table 2). DNA was harvested from a composite of a minimum of 30 plants of each cultivar. In all cases DNA was isolated using a standard CTAB (hexadecylatrimethylammonium bromide) procedure (Keim et al. 1988).

Pedigree analysis

In order to study the feasibility of following pedigree relationships with SSRs, plants from seven crosses, including the parents and the resulting cultivar, were examined. The crosses were: (1) 'Forrest' = 'Dyer' × 'Bragg'; (2) 'Bragg' = 'Jackson' × 'D49-2491'; (3) 'D49-2491' = 'S-100' × 'CNS'; (4) 'Lee' = 'S-100 × CNS'; (5) 'Adams' = 'Illini' × 'Dunfield'; (6) 'Clark' = 'Lincoln' × ('Lincoln' × 'Richland'); and (7) 'Amsoy' = 'Adams' × 'Harosoy'. DNA was harvested from a composite of 5–7 plants of each cultivar.

Simple sequence repeat loci

Twenty SSR loci with either di-or tri-nucleotide repeats were selected for this study (Table 3). The SSR loci were chosen based on their high level of length polymorphism as determined on a set of 10 soybean genotypes and because previous examination indicated the production of no spurious PCR products in addition to the predicted products. All SSR primer sequences, except those listed below, were previously published (Akkaya et al. 1995). Except for S45035, which was selected from a GenBank sequence, the complete locus name includes the BARC prefix, which is the acronym for Beltsville Agriculture Research Center. In the text, these loci are referred to by the suffix alone.

S45035	Forward:	5′ tttgtgaacgatagaaatttat 3′
	Reverse:	5′ aggggaaaatttttaaaga 3′
BARC-Satt038	Forward:	5′ aatctttttttctttctattaagtt 3′
	Reverse:	5′ cattgaaatggttttagtca 3′
BARC-Sat_069	Forward:	5′ cgaccagctgaagaaa 3′
	Reverse:	5′ ctgaatacccatcattacttaa 3′
BARC-Satt100	Forward:	5′ acctcattttggcataaa 3′
	Reverse:	5′ ttggaaaacaagtaataataaca 3′
BARC-Satt114	Forward:	5′ gggttatcctccccaata 3′
	Reverse:	5′ atatgggatgataaggtgaaa 3′

The development of the SSR markers, including the selection of SSR-containing sequences from GenBank and SSR-containing clones from a genomic soybean DNA library, and the selection and preparation of PCR primers to the SSR loci have been described elsewhere (Cregan et al. 1994).

Simple sequence repeat allele sizing

Forward primers of the 20 SSR loci were labeled with either blue (FAM), green (TET), or yellow (HEX) fluorescent tags (AB-PEC, Foster City, Calif.) (Ziegle et al. 1992) (Table 3). PCR reaction mixes contained 30 ng of soybean genomic DNA, 1.5 m*M* Mg_{2+}, 0.15 μM of 3′ and 5′ end primers, 200 μM of each nucleotide, 1 × PCR buffer containing 50 m*M* KCl, 10 m*M* TRIS-HCl pH 9.0, 0.1% Triton X-100, 1 unit *Taq* DNA polymerase in a total volume of 10 μl. Cycling consisted of 25 s of denaturation at 94°C, 25 s of annealing at 46°C, and 25 s of extension at 68°C for 32 cycles on a MJ Research model PTC-100 thermocycler (MJ Research, Watertown, Mass.).

Amplification products from different SSR loci carrying the same fluorescent label can be simultaneously analyzed in the same gel lane if allele size ranges do not overlap. Therefore, after cycling, 1.5 μl of each FAM-labeled, 2.0 μl of each TET-labeled, and 4.0 μl of each HEX-labeled PCR product was combined and brought to a total volume of 20 μl. A total of 0.5 μl of fluorescent TAMRA (red)-labeled internal size standard and 0.5 μl of a loading buffer (AB-PEC, Foster City, Calif.) were added to a 4.0-μl aliquot of the combined PCR products. The sample of the combined PCR products was loaded and separated on a ABI Prism 373A DNA sequencer (AB-PEC, Foster City, Calif.). GeneScan 672™ software (AB-PEC, Foster City, Calif.) was used for gel analysis. The Local Southern option in GeneScan 672™ was used for computation of allele size, and Genotyper™ software (AB-PEC, Foster City, Calif.) was applied for accurate visualization of the alleles and for automated data output. All allele sizing was verified using standard [^{32}P]-labeled PCR products separated on sequencing gels as described elsewhere (Cregan et al. 1994).

Statistical analysis

Gene diversity (Weir 1990) was calculated as: $1 - \sum P_{ij}^2$ where P_{ij} is the frequency of the *j*th allele for *i*th locus summed across all alleles in the locus. Soybean is a self-fertilized species and, therefore, in most

Table 1 Morphological and pigmentation traits of commercial soybean cultivars within Maturity Groups I, II, IV, and VI

Trait	Maturity Group I	Maturity Group II	Maturity Group IV	Maturity Group VI
Seedcoat color	Yellow	Yellow	Yellow	Yellow
Hilum color	Yellow	Yellow	Black	Buff
Cotyledon color	Yellow	Yellow	Yellow	Yellow
Leaflet shape	Ovate	Ovate	Ovate	Ovate
Flower color	Purple	Purple	White	White
Pod color	Brown	Brown	Tan	Tan
Pubescence color	Gray	Gray	Brown	Gray
Plant habit	Indeterminate	Indeterminate	Indeterminate	Determinate

Table 2 Soybean cultivars from Maturity Groups I, II, IV and VI that were used in the study, and the public or private developer[a]

Maturity Group I		Maturity Group II		Maturity Group IV		Maturity Group VI	
Cultivar name	Company/Institution	Cultivar name	Company/Institution	Cultivar name	Company/Institution	Cultivar name	Company/Institution
Hardin 91	Iowa State University	HS 2821	Growmark, Inc.	Nile	SIU Carbondale	HSC 623	Helena Chem. Co. Hyperformer Seed
CM 182	Callahan Enterprises	CM 274	Callahan Enterprises	Hartz 4464	Jacob Hartz Seed Co.	Hartz 922	Jacob Hartz Seed Co.
BT 1790	Ziller Seed Farms	A2187	Asgrow Seed Co.	A4715	Asgrow Seed Co.	A6961	Asgrow Seed Co.
3172	Ciba Geigy	Amcor 89	Ohio Res. Develop. Center and USDA	FFR 464	FFR Cooperative	60400	Stratton Seed Co.
DSR-138	Dairyland Seed Co.	J-220	Mycogen	Pioneer 9443	Pioneer Hi-Bred Int.	Hartz 5050	Jacob Hartz Seed Co.
DSR-189	Dairyland Seed Co.	CM 205	Callahan Enterprises	Pioneer 9444	Pioneer Hi-Bred Int.	Hartz 608	Jacob Hartz Seed Co.
B1420	Mycogen Plant Sciences	Pavone	Midwest Oilseeds	DSR-440	Dairyland Seed Co.	Hartz 507	Jacob Hartz Seed Co.
Pioneer 9141	Pioneer Hi-Bred Int.			CX411	DeKalb	Pioneer 9584	Pioneer Hi-Bred Int.
S16-60	Northrup King Co.			Bronson	USDA/Purdue	Pioneer 9692	Pioneer Hi-Bred Int.
B117	Northrup King Co.			Pioneer 9472	Pioneer Hi-Bred Int.		

[a] The authors wish to thank the various organizations and individuals who kindly supplied seeds of the cultivars indicated in this study

cases only 1 allele was detected per locus in each genotype. However, in a few instances 2 alleles per locus were detected. This would be the anticipated result if a genotype were derived from a single plant that was heterozygous for a locus in question. To accommodate these instances each of the genotypes was recorded as having 2 alleles at a locus: identical for genotypes with 1 allele, and different for genotypes that had 2 alleles. A program was written using SAS IML (SAS 1994) to compare each of the 2 alleles at any given locus of all genotypes. Thus, a pair of genotypes could be identical at a locus (possess 1 or 2 alleles of equal size, score = 1); be 50% similar (possess 1 identical and 1 dissimilar allele, score = 0.5); or be dissimilar (possess no allele of similar size, score = 0.0). A simple genetic dissimilarity coefficient 1 − (score/20), was calculated between each pair of 36 genotypes. The 36 genotypes included the 35 ancestral genotypes and cv 'Williams'. This matrix of genetic dissimilarity coefficients was used to determine clusters of genotypes using PROC CLUSTER in SAS (SAS 1994) and unweighted pair-group method using arithmetic average (UPGMA; Method = Average). A dendrogram was drawn by 'Dendro macro'-version 2 (Paul Nicholson, University of Leeds, 1994). Genetic dissimilarity coefficients were also calculated among the 7 soybean genotypes that had identical RFLP profiles (Keim *et al.* 1989) as well as among each group of cultivars within Maturity Groups I, II, IV, and VI described above.

Distinguishing improperly recorded pedigrees, outcrossing and seed mixtures from mutation

As a result of an incorrectly recorded pedigree, seed mixture during harvesting or storage, outcrossing, or mutation, alleles may appear in a soybean genotype that are not present in either recorded parent. To distinguish the first three circumstances from mutation, we chose five sets of 3 cultivars at random from within each of the Maturity Group I, II, IV, and VI cultivars described above. Two members of each set were randomly selected as "parents" and the third as the "progeny" of the first two in order to mimic a situation in which an incorrectly recorded pedigree, outcrossing, or a seed mixture had occurred. The alleles present at the 20 SSR loci in each of the five sets were used to estimate the average proportion of alleles which would be present in a "progeny" that were not present in either "parent". Because of the high genetic similarity of genotypes within a Maturity Group, especially genotypes which are phenotypically indistinguishable, this analysis should provide a conservative estimate of the rate of non-parental alleles that might be anticipated in a genotype whose parents were incorrectly recorded or which had been contaminated by outcrossing or a seed mixture.

Results and discussion

Automated sizing of simple sequence repeat alleles

A gel image from the AB-PEC Prism 373A sequencer is shown in Fig. 1. Fluorescent labeled PCR products from 5 SSR loci were analyzed in each lane. Two of the loci, Sat_038 and Satt038, with $(AT)_n$ and $(ATT)_n$ repeats, respectively, were labeled with FAM (blue), 2 others, Satt114 and Satt030, both with $(ATT)_n$ repeats, were labeled with HEX (yellow); and the last locus, Satt009, also with an $(ATT)_n$ SSR, was labeled with TET (green). Satt009 had the widest allele size range of the 5 loci (163–244 bp) and could not be easily combined with any other TET-labeled locus. Some of the alleles, such as the green-labeled alleles of Satt009 in lanes 1–4, appear double banded. These two bands

Table 3 Allele number, size range, and gene diversity of 20 fluorescent labeled SSR loci in 35 ancestral soybean genotypes that represent 95% of the alleles in North America soybean germplasm, and allele number and gene diversity of the same SSR loci in 36 commercial soybean lines from Maturity Groups I, II, IV, and VI

Locus	Fluorescent label	Linkage group[b]	Ancestral genotypes			Commercial genotypes	
			Number of alleles	Allele size range (bp)	Gene diversity	Number of alleles	Gene diversity
BARC-Sat_022	FAM	D2	16	201–249	0.90	12	0.86
BARC-Sat_038	FAM	O	5	113–133	0.68	4	0.56
GMABAB[b]	FAM	N1	10	145–169	0.87	9	0.80
BARC-Satt001	FAM	K	8	83–122	0.77	6	0.69
BARC-Satt005	FAM	D1	11	141–183	0.87	7	0.73
BARC-Satt038	FAM	G	7	151–187	0.78	4	0.67
BARC-Satt045	FAM	E	9	125–153	0.72	3	0.28
BARC-Satt046	FAM	K	8	177–240	0.77	4	0.69
BARC-Satt070	FAM	B2	8	128–176	0.77	3	0.64
S45035[b]	TET	D1	14	158–214	0.86	6	0.63
BARC-Satt002	TET	D2	7	127–145	0.77	3	0.50
BARC-Satt009	TET	N1	13	163–244	0.82	6	0.76
BARC-Satt012	TET	G	11	112–171	0.81	6	0.73
BARC-Satt042	TET	A1	6	159–177	0.64	5	0.71
BARC-Sat_036	HEX	D1	16	132–185	0.91	9	0.84
BARC-Sat_043	HEX	K	14	135–171	0.87	13	0.87
BARC-Sat_069	HEX	D1	10	152–184	0.83	7	0.78
BARC-Satt030	HEX	F	8	141–171	0.71	5	0.73
BARC-Satt100	HEX	C2	12	114–165	0.89	5	0.65
BARC-Satt114	HEX	F	8	82–121	0.79	5	0.75
Mean			10.1		0.80	6.1	0.69

[b] GMABAB and S45035 were selected from Genbank sequences
[a] Linkage groups reported by Shoemaker and Specht (1995)

resulted from slippage or stuttering commonly observed in SSR-containing amplification products and differ from one another by 3 bp. Likewise, in the dinucleotide locus Sat_038, three bands differing by only 2 bp were occasionally apparent, e.g., Lanes 9 and 17. A few of the genotypes contained 2 alleles at some loci; for example, 'Fiskeby V' in lane 22, with alleles of 98 and 109 bp at the Satt114 locus. The automated sizing of alleles using fluorescent-labeled SSR primers and the AB-PEC Prism 373A Sequencer was found to be an efficient way to rapidly assay allelic diversity in a large number of genotypes. The ability to analyze a number of alleles from different loci in one lane, up to 5 in our study (Fig. 1), significantly reduces the number of gels per study. With careful selection of loci this number can be greatly increased. Additionally, the GENOTYPER™ analysis software was useful in allele visualization and sizing; however, small gel-to-gel variation required additional analysis as described below.

Variation in allele size, as calculated against the in-lane TAMRA-labeled size standard, was found among and within gels. This variation in allele size was similar in magnitude to variation found in other studies, ± 0.2 and ± 0.5 bases, within and between gels, respectively, (Smith 1995; Fregeau and Forney 1993). In order to accurately size alleles, cv 'Williams' was run as a standard for every locus and gel combination, and variation among gels was calculated based on the variation in 'Williams' allele sizes from the different gels. Additionally, allele sizes were converted from the actual fragment length in bases into a relative system. The median allele was given the value of zero; the remaining alleles were given relative values of ± 2, ± 4, ± 6, etc. for di-nucleotide, and ± 3, ± 6, ± 9, etc. for trinucleotide repeats. Relative allele sizes were verified by comparison to radio-labeled allele data (Cregan et al. 1994). This type of relative allele sizing helped in significantly increasing the repeatability and accuracy of allele sizing. Accurate, repeatable allele sizing was more difficult with almost all of the di-nucleotide repeat loci, which produced more stutter bands than the tri-nucleotide loci (Cregan et al. 1994). This stuttering made it difficult to determine the main peak (or band, as compared with [^{32}P]-labeled products) of the allele upon which the size was established. On the basis of our analysis, all the tri-nucleotide loci, with the exception of Satt012, produced clean unambiguous and easily determined allele sizing data.

Simple sequence repeat allelic variation in North America ancestral soybean

Germplasm

The 20 SSR loci showed high levels of length polymorphism among the 35 ancestral genotypes (Table 3). An average of 10.1 alleles per locus was present in the 35

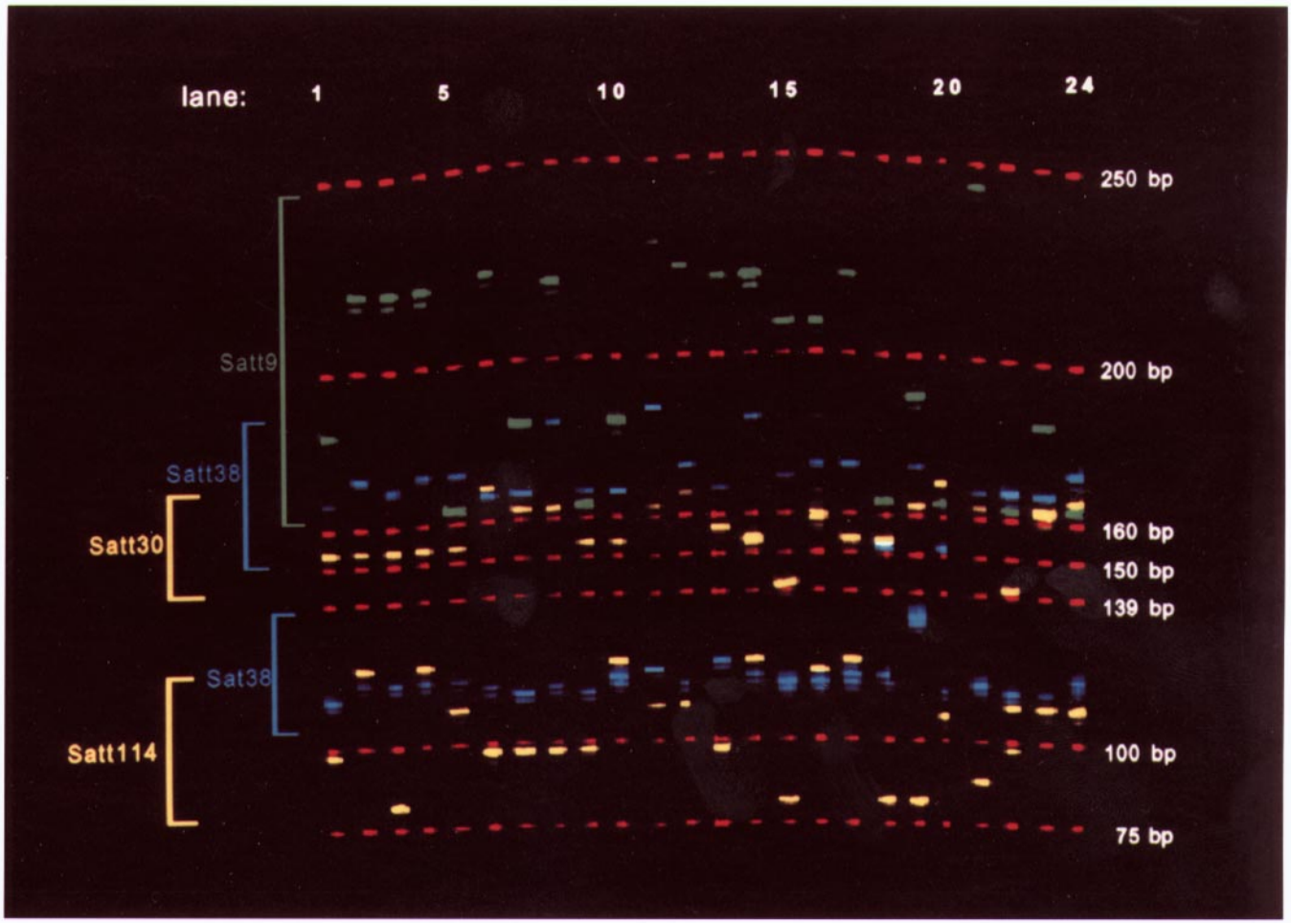

Fig. 1 Gel image of 24 soybean genotypes with 5 fluorescent-labeled SSR loci taken from the ABI Prism 373A DNA sequencer using GeneScan analysis. *Lane 24* is 'Williams' soybean

ancestral genotypes. Loci Sat_022 and Sat_036, both containing $(AT)_n$ repeats, had the highest number of alleles with a total number of 16, while locus Sat_038, also with an $(AT)_n$ core motif, had the least, with only 5 alleles among the 35 genotypes. Gene diversity associated with each of the 20 SSR loci ranged from 0.64 to 0.91, with a mean value of 0.80. This is in contrast to the low level of polymorphism detected by single-locus RFLP probes in soybean. For example, the most informative of 17 RFLP probes used in a group of 58 adapted and wild soybean genotypes had a gene diversity of 0.55 (Keim et al, 1989), while the 16 remaining probes had values of less than 0.50. Only 2 alleles were observed in 15 of the 17 RFLP loci. At the remaining 2 loci, 3 alleles were detected. In subsequent work (Keim *et al.* 1992), 38 ancestral and adapted cultivars were analyzed with 132 RFLP probes. Only 69% of the probes were polymorphic, and more than 2 alleles were detected at only 3 loci. Because of the limited number of RFLP alleles per locus in soybean, the amount of information generated by individual soybean RFLP probes is limited. Without the use of a large number of RFLP probes (50-100) it is quite likely that many modern soybean cultivars would be indistinguishable. These results are in sharp contrast to SSR markers in soybean, which are significantly more informative than RFLPs, and indicate the advantage of SSR markers in soybean genotype identification.

The 20 SSR loci had unique allelic profiles for all but 2 of the 35 ancestral soybean genotypes and cv 'Williams'. The 2 indistinguishable genotypes were 'Illini' and 'AK(Harrow)' (Fig. 2). This is not unexpected since 'Illini' is a selection from 'AK(Harrow)'. Two other genotypes 'Arksoy' and 'Ralsoy', were found to be 90% genetically similar. 'Ralsoy' is a selection from 'Arksoy', therefore, this similarity was also anticipated. Except for these very obvious genetic relationships of a line that was selected from another, or a direct pedigree relationship such as cv 'Williams' and cv 'Lincoln' ('Lincoln' is twice a grandparent of 'Williams'), the analysis was not successful in revealing other genetic relationships among the ancestral genotypes such as relatedness of genotypes within Maturity Group or geographical origin (Fig. 2). This could be attributed to

the relatively small number of loci that were used in the study. In a study by Keim et al. (1992) using 132 RFLP probes, 5 cultivars adapted to the Southern US formed a separate cluster. However, these 5 genotypes also had direct pedigree relationships. Similar to our results, the remaining genotypes in the Keim et al. (1992) study also did not group according to any criteria other than pedigree relationships.

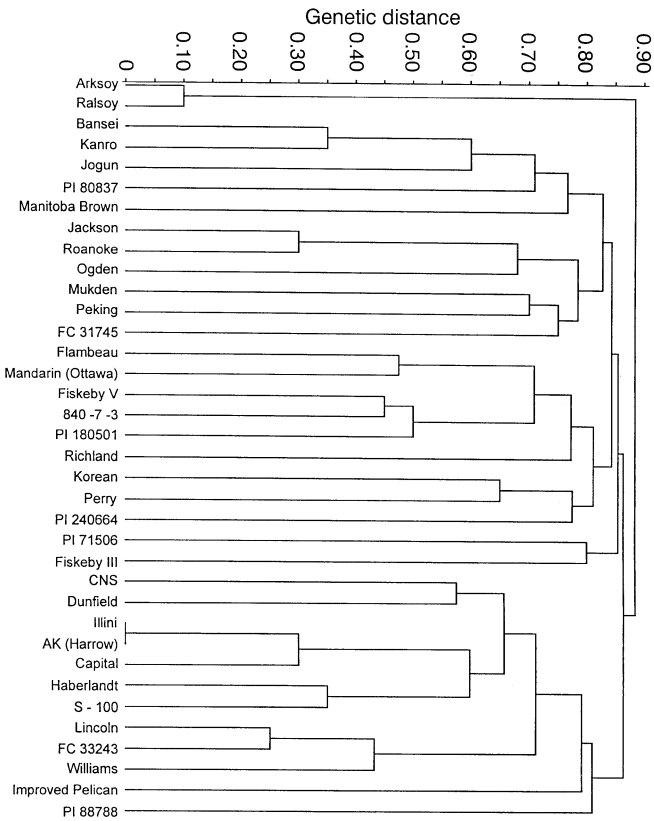

Fig. 2 Genetic distance among 36 soybean genotypes based on allelic profiles of 20 SSR loci. Thirty-five of the genotypes are ancestral soybean genotypes that represent 95% of the alleles in North America soybean germplasm. The 36th genotype is cv 'Williams', the source of the clones from which the library-derived SSR markers were developed

Simple sequence repeat markers to distinguish similar soybean genotypes

Seven genotypes reported to be indistinguishable using 17 RFLP probes (Keim et al. 1989) had unique SSR allelic profiles. The number of alleles observed within each of the 20 SSR loci in these 7 genotypes ranged from 2 to 4, with a mean of 2.95, and the gene diversity values ranged from 0.24 to 0.72, with a mean value of 0.54. Genetic dissimilarity coefficients for the 7 genotypes ranged from 0.4 to 0.8, with a mean of 0.62. Thus, these 7 genotypes, which were identical using 17 RFLP probes, were readily distinguishable based on SSR allelic profile.

Simple sequence repeat markers to distinguish soybean cultivars from maturity groups I, II, IV, and VI which are identical for standard morphological and pigmentation traits

Cultivars within Maturity Groups I, II, IV, and VI (Table 2) were easily distinguished using allelic profiles of the 20 SSR loci. Only in the case of 1 locus, Satt100 in Maturity Group IV, was there no allelic variation among the genotypes within a Maturity Group (Table 4). All other loci had at least 2 and up to 7 alleles (Sat_043 in Group IV) within each Maturity Group. Mean number of alleles ranged from 3.1 to 3.6; gene diversity ranged form 0.00 to 0.81, with mean gene diversity ranging from 0.53 to 0.57 (Table 4). Genetic dissimilarity coefficients were also relatively high, with the mean genetic dissimilarity coefficient ranging from 0.58 to 0.64 (Table 4). The high gene diversity values demonstrate the ability of SSR markers to distinguish even among phenotypically similar elite soybean cultivars. High gene diversity scores suggest that a system based on SSR markers using the tri-nucleotide loci identified earlier in this report as well as 10 or 15 more under development should function for quite some time to distinguish closely related soybean genotypes for purposes of Plant Variety Protection. Indeed, the Plant Variety Protection Office of the USDA Agriculture Marketing Service now accepts SSR allelic profiles as supporting evidence for the uniqueness of a new

Table 4 Means and ranges of number of alleles, gene diversity, and genetic dissimilarity coefficients of 20 fluorescent-labeled SSR loci in commercial soybean lines from Maturity Groups I, II, IV, and VI

Maturity Group	Number of lines	Mean number of alleles	Range of allele number	Mean gene diversity	Range of gene diversity	Mean genetic dissimilarity coefficient	Range of genetic dissimilarity coefficients
I	10	3.6	2–6	0.57	0.32–0.78	0.64	0.35–0.80
II	7	3.1	2–5	0.53	0.24–0.73	0.62	0.40–0.83
IV	10	3.4	1–7	0.54	0.00–0.81	0.60	0.33–0.90
VI	9	3.3	2–5	0.53	0.10–0.74	0.58	0.29–0.79

cultivar (Personal communication, J. Strachan, USDA, AMS, PVPO).

Following pedigree relationships using SSRs

In only one of the seven pedigrees used in this study, 'Forrest' = 'Dyer' × 'Bragg', could all the alleles in the progeny be explained by alleles that are present in the parents. In the remaining six crosses, 1 of the 20 loci had an allele in the progeny that was not present in either parent (Table 5). Of these loci, 2 consisted of $(AT)_n$ repeats, with Sat_043 having unexplained alleles in two of the pedigrees, the remaining 3 loci contained $(ATT)_n$ repeats. There are a number of explanations for the presence of alleles in a progeny which are not present in either recorded parent. One of these is an incorrectly recorded pedigree. A second is the possibility of mixtures having occurred via errors in harvesting or through outcrossing. If a genotype were purified by selection of one or a few plants subsequent to a mixture having occurred, "new" or non-parental alleles would appear in the cultivar. In the case of SSR loci, we believe that none of these possibilities is the likely explanation of the non-parental alleles that occur in six of the seven pedigrees studied. This conclusion was drawn from the study of the four groups of cultivars from Maturity Groups I, II, IV, and VI. Within each group, five pairs of cultivars were randomly selected as "parents" and a third line as a "progeny", and the number of alleles which was present in the "progeny" but not in the "parents" was calculated. This calculation indicated that on the average 8.6, 8.6, 9.2, and 6.6 non-parental alleles would appear in the "progeny" lines of the Maturity Group I, II, IV, and VI cultivars, respectively. These values provide estimates of the number of non-parental alleles that would be anticipated in a progeny as a result of an incorrectly recorded pedigree, outcrossing, or a seed mixture, This is much greater than the number observed in the seven actual pedigrees analyzed in the current study where a mean of 1.00 non-parental alleles was observed per pedigree. Thus, a more likely explanation for the results obtained in the seven pedigrees is that of mutation.

Mutation rates of microsatellite alleles in humans have been reported to be 5.6×10^{-4} (Weber and Wong 1993), 2.36×10^{-5} to 1.86×10^{-4} (Hammond et al. 1994), and 2.3×10^{-5} to 1.59×10^{-4} (Edwards et al. 1992), respectively. Higher mutation rates appeared to be associated with greater numbers of repeats in the core motif (Edwards et al. 1992). Based on the assumption that these mutation rates are similar to those in soybean, an estimate can be made of the number of non-parental alleles anticipated in the progeny of two parents. New cultivars of soybean and many other self-fertilizing crops are generally developed by crossing single plants from two lines followed by selfing and selection (Fig. 3). Typically, a line is derived from a single F_4 or F_5 plant. This line is advanced over a 5- to 7-year period during which time local preliminary, local advanced, and regional field trials are conducted. If a line survives the testing process it will be named and released as a new cultivar 5 or 7 years after its original derivation from a single plant. Cultivar development may include purification, but this step will generally include the bulking of many similar single F_8–F_{11} plant progeny. Thus, the allelic profile of the original F_4 or F_5 plant from which the cultivar was derived should remain intact if one analyzes a bulk DNA sample of plants from the new cultivar. In contrast, if the allelic profile of any single plant of the new cultivar is analyzed, mutations would be anticipated at some predictable rate. Thus, the specifics of the procedure for new cultivar development and the rate of mutation of SSR loci must both be considered in order to obtain an estimate of the rate at which new alleles would be anticipated to arise in a new cultivar derived from a single cross of two existing cultivars.

Following the scheme outlined in Fig. 3, there would be perhaps 7 meiotic cycles in both Parent 1 and Parent

Table 5 Pedigree analysis of seven soybean cultivars using 20 SSR loci: the number of alleles in the progeny cultivar that can be explained by the alleles present in each or neither of the parents

Progeny and ♀ and ♂ parents	Alleles present in the progeny derived from		
	♀ Parent[a]	♂ Parent[a]	Neither
Forrest = Dyer × Bragg	20	20	0
Bragg = Jackson × D49-2491	20	18	2[b]
D49-2491 = S-100 × CNS	24	14	2[c]
Lee = S-100 × CNS	20	18	2[d]
Adams = Illini × Dunfield	20	18	2[e]
Clark = Lincoln × (Lincoln × Richland)	27	11	2[f]
Amsoy = Adams × Harosoy	16	22	2[g]

[a] Because 2 alleles are occasionally present at a locus, a particular parent was considered to have contributed either 0, 1 or 2 alleles to its progeny at each locus (see Materials and methods, statistical analysis)

[b] A mixture of 2 'new' alleles which are present in neither parent at the Sat_043 locus

[c,d,e,f,g] Homozygous for a 'new' allele which is not present in either parent at the Sat_043, Satt042, Satt009, Satt045 and GMABAB locus, respectively

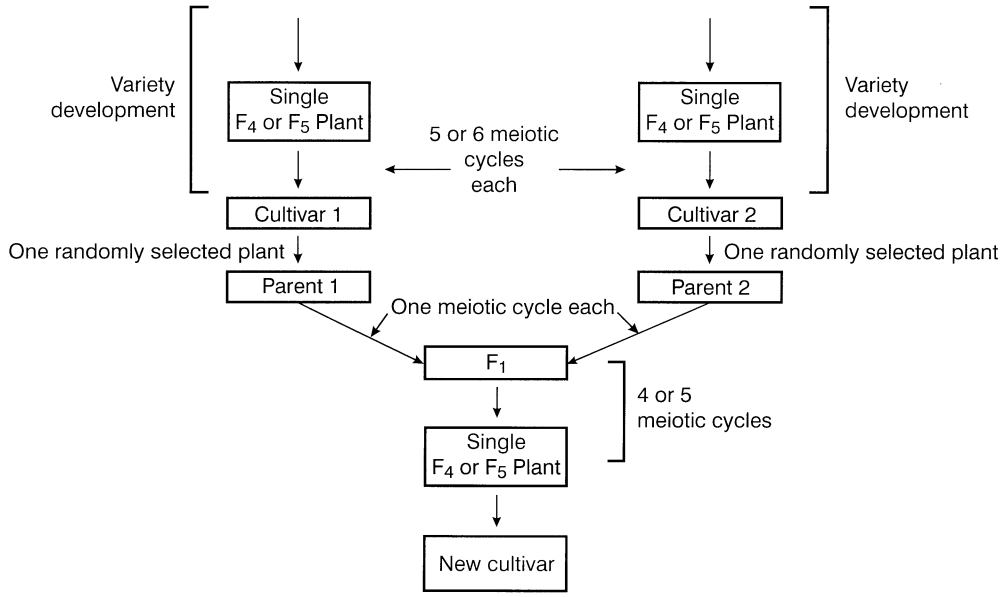

Fig. 3 A diagram of a typical breeding scheme used to produce soybean cultivars

2 between single-plant derivation and the use of a single-plant for crossing. Likewise, in the production of the F_1 and during the selfing generations until the derivation of the new cultivar, an additional 4 or 5 meiotic cycles would occur. Thus, from the single-plant derivation of each of the parents to the single-plant derivation of their progeny, a total of approximately 20 meiotic cycles would ensue during which mutation would occur. However, during the development of a new cultivar selfing will eliminate almost 50% of the mutant alleles that arise. Therefore, the effective number of meioses that would contribute to the occurrence of mutant alleles is only 10. Using the range of SSR mutation rate of 2.3×10^{-5} to 5.6×10^{-4} as reported in humans by Edwards et al. (1992) and Weber and Wong (1993), one would expect to find $10 \times 2.3 \times 10^{-5}$ to $10 \times 5.6 \times 10^{-4}$, or a mutation rate of 2.3×10^{-4} to 5.6×10^{-3} at any given locus. In 20 loci the rate would be $20 \times 2.3 \times 10^{-4}$ to $20 \times 5.6 \times 10^{-3}$ or 4.6×10^{-3} to 0.112. Thus, in the current study one would anticipate a 0.46–11.2% rate of new allele formation or 0.092–2.24 new alleles per pedigree (i.e., per 20 loci). In the seven pedigrees studied here, 7 new alleles or 1.00 per pedigree were observed. While this figure for the number of mutant alleles per pedigree is based upon very limited data, it nonetheless falls well within the range of the anticipated number of new alleles. In our mapping of SSR alleles in soybean our limited observations indicated new allele formation at a rate of 1 per 5,000 or 2×10^{-4}. This rate would produce 0.8 new alleles per pedigree (per 20 loci) using the assumptions of cultivar derivation indicated above. Thus, the observation of mutation rate in mapping populations and that suggested in the analysis of the seven soybean pedigrees appear to be similar.

Mutation should not be an obstacle in the use of SSR markers for genotype identification, mapping or breeding applications. The identification of soybean cultivars for PVP purposes should be determined using bulked DNA of many plants and not that of a single plant. Therefore, mutations in single plants will not alter the SSR allelic constitution of the cultivar. To insure accurate identification of soybean varieties, plant breeders should analyze bulked DNA of 30–50 plants of a cultivar as the standard protocol.

Human geneticists have recently constructed a stepwise mutation model to study genetic distance among related species (Schlotterer and Tauze, 1992; Valdes et al. 1993). The stepwise mutation model argues that alleles at some microsatellite loci evolve by a stepwise mutation process, in which an allele mutates up or down by a small number of repeats. However, infrequent large jumps in repeat number also occur (Di Rienzo et al. 1994). In our soybean pedigree analysis both small and large jumps in repeat number were observed. In the two instances where 'new' alleles were found in the Sat_043 locus (Table 5), both were only 2 or 4 bases different than the parental alleles. In two other instances involving tri-nucleotide repeats, the new allele in 'Clark' at the Satt042 locus was 6 bases different than the allele in either 'Lincoln' or 'Richland', and the same difference of 6 bases was also found in 'Lee' at the Satt042 locus as compared to its parent 'S-100'. In the remaining two cases, larger jumps in repeat number were found, with 'Amsoy' having a new allele which differed from the alleles in 'Adams' and 'Harosoy' by 16 and 18 bases, respectively; and 'Adams' having a new allele at the Satt009 locus that differed by 57 bases from both the 'Illini' and 'Dunfield' alleles. The stepwise mutation theory suggests that alleles that

evolve by stepwise mutations carry information about the amount of time elapsed since two individuals shared a common ancestral allele (Goldstein et al. 1995; Slatkin 1995). Using the stepwise mutation model, investigators have successfully used SSRs in the reconstruction of human phylogeny (Bowocock et al. 1994) and for studying human population structure (Jin and Chakraborty, 1995; Di Rienzo et al. 1994), indicating that mutations at SSR loci may not be a hindrance but rather a tool for evolutionary and systematic studies. It is possible that a similar analysis of SSR allelic variation can be used to define relationships among soybean genotypes with different Asian origins.

This study demonstrates an automated system for unambiguous soybean genotype identification and establishes the ranges of allelic diversity at 20 SSR loci that one would anticipate in North American soybean germplasm. This system provides an alternative to standard [^{32}P] microsatellite allele sizing. Currently, work is underway to establish a standard set of SSR loci for use in soybean plant variety identification. These loci will be selected based upon (1) high level of informativeness (gene diversity), (2) position in separate linkage groups, (3) the production of discrete products with minimal "stutter" bands, and, (4) ability to permit multiplex PCR amplification. The Plant Variety Protection Office of the Agricultural Marketing Service, USDA, will use fingerprinting data derived from this set of loci to guarantee the intellectual property rights of soybean breeders. It is also clear that a range of breeding applications, including QTL analysis, marker-assisted selection, and facilitated backcrossing, are obvious applications of the automated SSR analysis demonstrated here.

Acknowledgments We are grateful to Tyler Jarvik and Dr. J. Strauchan 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 for which the authors are very grateful.

References

Akkaya M S, Bhagwat AA, Cregan PB (1992) Length polymorphism of simple sequence repeat DNA in soybean. Genetics 132:1131–1139

Akkaya MS, Shoemaker RC, Specht JE, Bhagwat AA, Cregan PB (1995) Integration of simple sequence repeat (SSR) DNA markers into a soybean linkage map. Crop Sci 35:1439–1445

Alford RL, Hammond HA, Coto I, Caskey CT (1994) Rapid efficient resolution of parentage by amplification of short tandem repeats. Am J Hum Genet 55:190–195

Bell CJ, Ecker JR (1994) Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. Genomics 19:137–144

Bowocock AM, Ruiz-Linares A, Tomfohrde T, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368:455–457

Cheng HH, Crittenden LB (1994) Microsatellite markers for genetic mapping in the chicken. Poult Sci 73:539–546

Cregan PB, Bhagwat AA, Akkaya MS, Jiang Rongwen (1994) Microsatellite fingerprinting and mapping of soybean. Methods Mol Cell Biol 5:49–61

Crooijmans RPMA, Groenen MAM, Van-Kampen AJA, Van-Poel JJ (1994) New microsatellite markers on the linkage map of the chicken genome. J Hered 85:410–413

Decker RA, Moore J, Ponder B, Weber JL (1992) Linkage mapping of human chromosome 10 microsatellite polymorphisms. Genomics 12:604–606

Dietrich WF, Miller JC, Steen RG, Merchant M, Damron D, Nahf R, Gross A, Joyce DC, Wessel M, Dredge RD, Marquis A, Stein LD, Goodman N, Page DC, Lander ES (1994) A genetic map of the mouse with 4006 simple sequence length polymorphism. Nat Genet 7:220–225

Di Rienzo AA, Peterson C, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple-sequence repeat loci in human populations. Proc Natl Acad Sci USA 19:3166–3170

Du Y, Remmers EF, Zha H, Goldmuntz EA, Mathern P, Crofford LJ, Szpirer J, Szpirer C, Wilder RL (1995) Genetic map of eight microsatellite markers comprising two linkage groups on rat chromosome 6. Cytogen Cell Genet 68:107–111

Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R (1992) Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 12:241–253

Ellegren H, Johansson M, Chowdhary BP, Marklund S, Ruyter D, Marklund L, Brauner-Nielsen P, Edfors-Lilja I (1993) Assignment of 20 microsatellite markers to the porcine linkage map. Genomics 16:431–439

Fregeau, CJ, Fourney RN (1993) DNA typing with fluorescently tagged short tandem repeats: a sensitive and accurate approach to human identification. BioTechniques 5:100–119

Gizlice Z, Carter TE, Burton JW (1994) Genetic base for North American public soybean cultivars released between 1947 and 1988. Crop Sci 34:1143–1151

Goldmuntz EA, Remmers EF, Zha H, Cash JM, Mathern P, Crofford LJ, Wilder RL (1993) Genetic map of 12 polymorphic loci on rat chromosome 1. Genomics 16:761–764

Goldstein DB, Ruiz-Linares A, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distance for use with microsatellite loci. Genetics 139:463–471

Gyapay G, Morissette J, Vignal A, Dib C, Fizames C, Millasseau P, Marc S, Bernardi G, Lathrop M, Weissenbach J (1994) The 1993–94 Genethon human genetic linkage map. Nat Genet 7:246–249

Hammond HA, Jin L, Zhong Y, Caskey CT, Chakraborty R (1994) Evaluation of 13 short tandem repeat loci for use in personal identification applications. Am J Hum Genet 55:175–189

Holgersson S, Karlsson JA, Kihlgren A, Rosen B, Savolainen P, Gyllensten U (1994) Fluorescent-based typing of the two short tandem repeat loci HUMTH01 and HUMACTBP2: reproducibility of size measurements and genetic variation in the Swedish population. Electrophoresis 15:890–895

Hudson TJ, Engelstein M, Lee MK, Ho EC, Rubenfield MJ, Adams CP, Housman DE, Dracopoli NC (1992) Isolation and chromosomal assignment of 100 highly informative human simple sequence repeat polymorphisms. Genomics 13:622–629

Hudson TJ et al. (1995) An STS-based map of the human genome. Science 270:1945–1954

Jin L, Chakraborty R (1995) Population structure, stepwise mutations, heterozygote deficiency and their implication in DNA forensics. Heredity 74:247–285

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

Keim P, Shoemaker RC, Palmer RG (1989) Restriction fragment length polymorphism diversity in soybean. Theor Appl Genet 77:786–792

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

Kresovich S, Szewc-McFadden AK, Bliek SM (1995) Abundance and characterization of simple-sequence repeats (SSRs) isolated from a size-fractionated genomic library of *Brassica napus* L. (rapeseed). Theor Appl Genet 91 : 206–211

Maughan PJ, Saghai Maroof MA, Buss RG (1995) Microsatellite and amplified sequence length polymorphisms in cultivated and wild soybean. Genome 38 : 715–723

McAleer MA, Aitman TJ, Cornall JR, Ghosh S, Hall JRS, Hearne CM, Love JM, Prins JB, Ramachandran S, Rodrigues N, De-Larato NH, Peterson LB, Wicker LS, Todd JA (1992) Linkage analysis of 84 microsatellite markers in intra- and interspecific backcrosses. Mamm Genome 3 : 457–460

Moore SS, Byrne K, Berger KT, Barendse W, McCarthy F, Womack JE, Hetzel DJS (1994) Characterization of 65 bovine microsatellites. Mamm Genome 5 : 84–90

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

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

Rohrer GA, Alexander LJ, Keele JW, Smith TP, Beattie CW (1994) A microsatellite linkage map of the porcine genome. Genetics 136 : 231–245

Rongwen J, Akkaya MS, Lavi U, Cregan PB (1995) The use of microsatellite DNA markers for soybean genotype identification. Theor Appl Genet 90 : 43–48

Scharf SJ, Smith AG, Hansen JA, McFarland C, Erlich HA (1995) Quantitative determination of bone marrow transplant engraftment using fluorescent polymerase chain reaction primers for human identity markers. Blood 85 : 1954–1963

Schlotterer C, Tautz D (1992) Slippage synthesis of simple sequence DNA. Nucleic Acids Res 22 : 285–288

Schook LB, Paszek AA, Louis C, Murtaugh M, Beattie CW, Rohrer GA, Alexander LJ, Wheeler MB (1994) Mapping the swine genome: take home lessons from mouse and man. Anim-Biotech 5 : 129–134

Senior ML, Chin ECL, Lee M, Smith JSC, Stuber CW (1996) Simple sequence repeat markers developed from maize sequences found in the Genbank database: map construction. Crop Sci 36 : 1676–1683

Serikawa T, Kuramoto T, Hilbert P, Mori M, Yamada J, Dubay CJ, Lindpainter K, Ganten D, Guenet JL, Lathrop GM, Beckmann JS (1992) Rat gene mapping using PCR-analyzed microsatellites. Genetics 131 : 701–721

Shoemaker RC, Specht JE (1995) Integration of the soybean molecular and classical genetic linkage groups. Crop Sci 35 : 436–446

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139 : 457–462

Smith RN (1995) Accurate size comparison of short tandem repeat alleles amplified by PCR. BioTechniques 18 : 122–128

Thomas MR, Scott NS (1993) Microsatellite repeats in grapevine reveal DNA polymorphisms when analyzed as sequence-tagged sites (STSs). Theor Appl Genet 86 : 985–990

Valdes AM, Slatkin M, Freimer NB (1993) Allele frequencies at microsatellite loci: the stepwise mutation model revised. Genetics 133 : 737–749

Weber J L, Wong C (1993) Mutation of human short tandem repeats. Hum Mol Genet 2 : 1123–1128

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

Wilkie PJ, Krizman DB, Weber JL (1992) Linkage map of human chromosome 9 microsatellite polymorphisms. Genomics 12 : 607–609

Weir BS (1990) Genetic data analysis methods for discrete genetic data. Sinauer Assoc, Sunderland, Mass.

Ziegle JS, Su Y, Corcoran KP, Nie L, Mayrand PE, Hoff LB (1992) Application of automated DNA sizing technology for genotyping microsatellite loci. Genomics 14 : 1026–1031