Czech J. Genet. Plant Breed., 39, 2003 (4): 120–126

Utilisation of SSRs for Characterisation of the Soybean (*Glycine max* (L.) Merr.) Genetic Resources

Martina HUDCOVICOVÁ and Ján KRAIC

Division of Applied Genetics and Breeding, Research Institute of Plant Production, Piešťany, Slovak Republic

Abstract: The SSR profiles of 67 soybean genotypes of various origins have been detected by 188 alleles at the 18 SSR loci. From 4 to 21 alleles were found at each of these loci (average 10.4 per locus) and the gene diversity averaged 0.71. Differentiation of all 67 genotypes each from others has been successful by using of even only 6 of SSR markers (Sat001, Satt005, Satt038, Satt173, Satt177, Satt534) with gene diversity from 0.66 to 0.89. The cumulated probability of obtaining identical soybean SSR profile was 1.11×10^{-9}, which confirms a high potential of SSRs for differentiation of soybean accessions in collections. Clustering of genotypes partially reflects origin and pedigree of analysed soybean accessions.

Keywords: soybean; *Glycine max* (L.) Merr.; microsatellite; genetic diversity

There are developing thousands of breeding lines and hundreds of elite cultivars yearly in the soybean hybridisation programmes over the world. Intensive breeding is engaged with increased genetic uniformity in the frame of species. Generations of new and improved cultivars can be enhanced by new sources of genetic variation, therefore criteria for parental stock selection need to be considered not only by agronomic value, but also from the point of view of their genetic dissimilarity. Parental genotypes with similar agronomic traits but genetically diverse produce highly variable progenies (Cox *et al*. 1985) by heterosis effect (Messmer *et al*. 1993). That is why the evaluation of genetic variation is a very important task not only for population genetics but also for plant breeders.

Soybean is known for a low level of genetic diversity in morphological and RFLP (restriction fragment length polymorphism) markers (Keim *et al*. 1992; Shoemaker *et al*. 1992). Therefore other molecular markers such as RAPDs (random-amplified polymorphic DNA), SSRs (simple sequence repeats), and others have been tested as useful markers for genetic diversity detection in soybean (Akkaya *et al*. 1992). Simple sequence repeats (microsatellites) are sequences of short tandem repeats distributed over the genomes (Hamada *et al*. 1982). The hypervariable number of repeat units makes them an excellent tool for genotype differentiation, pedigree analysis, evaluation of genetic distances among organisms, etc. These markers have been successfully used for the genotyping of many plant species, such as barley (Saghai Maroof *et al*. 1994), tomato (Phillips *et al*. 1994), rapeseed (Kresovich *et al*. 1995), potato (Schneider & Douches 1997) and others. A high level of polymorphism at the SSR loci has been reported also in soybean (Akkaya *et al*. 1992; Cregan *et al*. 1994; Rongwen *et al*. 1995; Diwan & Cregan 1997; Song *et al*. 1999; Brown-Guerida *et al*. 2000; Narvel *et al*. 2000; Meesang *et al*. 2001). All cited authors observed much higher genetic diversity using SSRs than by RFLPs. Diwan and Cregan (1997) by analysing 20 SSR loci were able to distinguish several modern soybean cultivars considered identical on the basis of RFLPs, morphology, and pigmentation traits. Brown-Guerida *et al*. (2000) using SSRs were able to identify related groups of

Supported by the Ministry of Agriculture of the Slovak Republic, Grant No. 27-11.

genotypes and many of the soybean introductions were found to be distinct from the founding stock. They observed higher genetic diversity for three SSR markers than for 46 RAPDs. Prioli *et al*. (2002) only by 12 SSR markers successfully distinguished morphologically similar groups of 186 Brazilian soybean cultivars and variation at the SSR loci agreed with the cultivar pedigree information. Morgante and Olivieri (1993), Maughan *et al*. (1995), Powell *et al*. (1996), Choi *et al*. (1999) observed higher SSR variation in wild soybean (*Glycine soja*) than in cultivated (*Glycine max*).

The aim of this study was to prove efficiency of SSRs for differentiation of soybeans maintained in the soybean genetic resources collection; to detect variation at the microsatellite loci; and to evaluate distribution of microsatellite alleles over the analysed set of genotypes.

MATERIAL AND METHODS

The seeds of 67 soybean genotypes of different origin, including released cultivars, hybrid lines, landraces, and obsolete cultivars were obtained from the collection of soybean genetic resources from the Gene Bank of the Slovak Republic, Piešťany: Aida, Amsoy 71, Apache, BS 31, Cresir, Crusader, Dunajka, Gadir, Hana, Kalmit, Labrador, Maple Arrow, Polanka, Silvia, Sluna, PY-OT-92-6/1, PY-OT-92-7/2, PY-OT-92-13/1, PY-OT-92-37/2, CMS 2 × Fred, CMS 2 × Adoc, CMS 2 × Canton, Sluna × Crusader, Sluna × Crusader II, Fred, Canton, Adepta, Amsoy × Silvia I, Amsoy × Silvia II, Amsoy × Silvia IV, Dacota, Eko žltá, Fiskeby V, Chabarovskaja, Chmelárova Brněnská, Ishigo Wase, Ishikar Shiro No., Jihomoravská žlutá, Kanagawa Wase, Kina 1, Kobora, Korada, Mutant KG7, Nadneprjanskaja, Nigra, Piava, Polan, Progres, Pulawska, Roudnická černá, Ruská žlutá, T 218H, T 259H, T 266H, T 268H, T 273H, T 274H, T 277H, T 295H, T 310, T 312, T 317, T 54, Terassol, Termition soybean Fiesta, Mutant KG7 II, and Zora.

Genomic DNA was isolated by the method of Dellaporta *et al*. (1993) from young fresh leaves. A sample of each genotype represented bulk DNA collected from 10–15 individual plants. Altogether 18 pairs of specific microsatellite primers (Table 1) (Cregan *et al*. 1999; http://129.186.26.94/SSR.html) were used for amplifications. PCR reactions were carried out in 15 μl volumes and contained 25 ng of DNA, 1 × PCR buffer (50 mmol/l KCl, 10 mmol/l Tris-HCl, pH 8.3), 1.5 mmol/l $MgCl_2$, 0.1 μmol/l of both primers, 0.2 mmol/l each of dNTPs, and 0.6 U of Taq-DNA polymerase. PCR was programmed for initial denaturation 5 min at 94°C, followed by 35 cycles of 1 min at 92°C, 1 min at 50°C, and 1 min at 72°C. Final extension was 10 min at 72°C. Five microliters of the reaction mixture were loaded into 6% denatured polyacrylamide gels. Microsatellites were stained by silver staining method (Bassam *et al*. 1991). Polymorphic DNA segments amplified with each microsatellite primer were considered as different alleles, assigned a letter and each allele was scored as present (1) or absent (0). Based on the frequencies of microsatellite alleles index of diversity (DI) $1 - \sum P_{ij}^2$ (P_{ij} = frequency of the *j*-th allele at the *i*-th locus), the probability of identity (*PI*) $\sum p^4_i + \sum\sum (2p_i p_j)^2$, and polymorphic information context (PIC) $1-(\sum p^2_i) - \sum\sum (2p^2_i p^2_j)$ were calculated (Weir 1990; Paetkau *et al*. 1995; Weber 1990). The unweighted pair group method of cluster analysis using arithmetic means (UPGMA) was used for the grouping of genotypes. A dendrogram was constructed based on Jaccard's similarity coefficient by the statistic software package SPSS 8.0 (SPSS Inc., USA).

RESULTS AND DISCUSSION

The SSR polymorphism in 67 soybean genotypes showed variation at all 18 analysed microsatellite loci. Altogether 188 alleles were detected at these loci. All genotypes were differentiated from each other. Due to high variation at the SSR loci and high PIC values, the indices of probability had low values (Table 1). Thirty-five accessions analysed by us showed complete genetic homogeneity, i.e. only a single allele (single SSR phenotype) each from 18 analysed loci were detected, whereas the remaining accessions revealed genetic heterogeneity, similar to soybean accessions obtained from farmers and analysed by Meesang *et al*. (2001). Mostly (89%) from 188 detected alleles occurred in low frequency (below 25%) over the analysed set of soybeans. The highest frequency was allele A at the locus Sat168, which occurred in 92.5% of genotypes. Nearly uniform distribution of alleles was detected at the loci Satt173, whereas very low balanced distribution at the locus Sat168 (Figure 1). The number of alleles per locus varied from 21 (locus Sat001 with gene diversity 0.894) to 4 (locus Sat168 with gene diversity 0.141), with an average of 10.4 alleles and average gene diversity 0.71 per locus. It is comparable with results of Diwan and

Table 1. Description and statistical analysis of soybean SSR alleles

SSR locus	Core motif	Forward primer Reverse primer	No. of alleles	DI	PI	PIC
Sat001	$(AT)_{17}$	GCGGATACGACCAAAAATTGTT	21	0.894	0.0185	0.8851
		GCGAACTGCGAAGATACTACCC				
Sat168	$(AT)_{15}$	TGTGGATAAAAGAGCATTCAAAATG	4	0.141	0.7404	0.1377
		GCGATCCTTGTTTATCTCAAAAAAGTGT				
Satt001	$(ATT)_{25}$	AAAGTCTTTAAAAGTGTGTCTTA	10	0.732	0.0633	0.7205
		TTAAAAGAAAAATGCAACAT				
Satt002	$(ATT)_{25}$	TGTGGGTAAAATAGATAAAAAT	8	0.639	0.1270	0.8729
		TCATTTTGAATCGTTGAA				
Satt005	$(ATT)_{19}$	TATCCTAGAGAAGAACTAAAAAA	15	0.862	0.0142	0.8556
		GTCGATTAGGCTTGAAATA				
Satt009	$(ATT)_{14}$	CCAACTTGAAATTACTAGAGAAA	15	0.857	0.0276	0.8447
		CTTACTAGCGTATTAACCCTT				
Satt038	$(ATT)_{17}$	GGGAATCTTTTTTTCTTTCTATTAAGTT	12	0.750	0.0930	0.7177
		GGGCATTGAAATGGTTTTAGTCA				
Satt082	$(ATT)_{13}$	AATTCATTTAGGGAGTTGAT	7	0.728	0.1661	0.6598
		CTAGCCAATGTCATATGACT				
Satt173	$(ATT)_{18}$	TGCGCCATTTATTCTTCA	17	0.887	0.0224	0.8763
		AAGCGAAATCACCTCCTCT				
Satt177	$(ATT)_{16}$	CGTTTCATTCCCATGCCAATA	7	0.662	0.1237	0.6330
		CCCGCATCTTTTTCAACCAC				
Satt242	$(ATT)_{26}$	GCGTTGATCAGGTCGATTTTTATTTGT	14	0.760	0.0490	0.7351
		GCGAGTGCCAACTAACTACTTTTATGA				
Satt244	$(ATT)_{27}$	GCGCCCCATATGTTTAAATTATATGGAG	6	0.670	0.1625	0.6119
		GCGATGGGGATATTTTCTTTATTATCAG				
Satt309	$(ATT)_{13}$	GCGCCTTCAAATTGGCGTCTT	8	0.651	0.2066	0.5874
		GCGCCTTAAATAAAACCCGAAACT				
Satt373	$(ATT)_{21}$	TCCGCGAGATAAATTCGTAAAAT	11	0.760	0.0734	0.7234
		GGCCAGATACCCAAGTTGTACTTGT				
Satt534	$(ATT)_{25}$	CTCCTCCTGCGCAACAACAATA	13	0.854	0.0164	0.8466
		GGGGGATCTAGGCCATGAC				
Satt547	$(ATT)_{18}$	GCGCTATCCGATCCATATGTG	8	0.624	0.2147	0.5540
		TGATTTCGCTAGGTAAAATCA				
GMSC514		TACCTTTCTTGTGAGTCGTA	5	0.538	0.3415	0.4525
		TATTGAGATGGA TATTGTAGATC				
SOYPRP1	$(TAT)_{20}$	CGTGCCAAATTACATCA	7	0.707	0.0681	0.6784
		TGATGGGAACAAGTACATAA				
Average values			**10.4**	**0.71**	**0.1405**	**0.6885**

DI = diversity index, PI = probability index, PIC = polymorphic information content

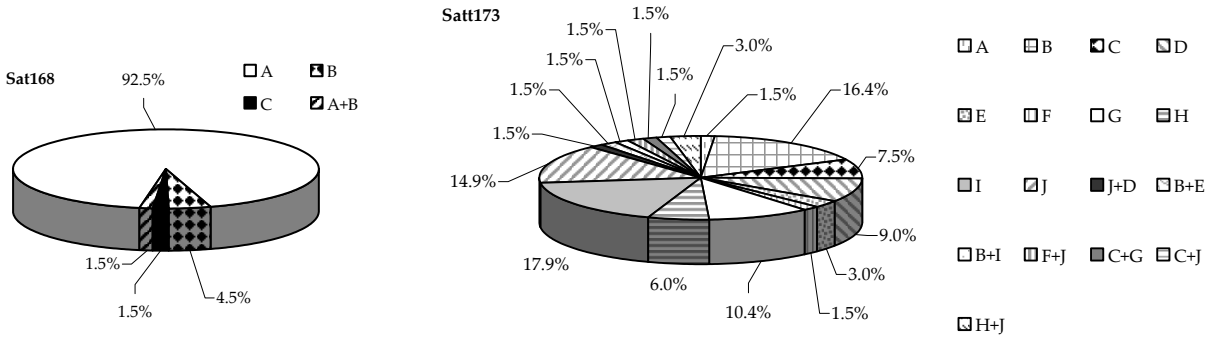

Figure 1. Frequency of SSR alleles (%) at the loci Sat168 and Satt173 in 67 soybean genotypes

Cregan (1997) and Akkaya *et al*. (1992) where average number of alleles was 10.1 and 7, respectively. Lower numbers of alleles and lower values of gene diversity at the SSR loci, in comparison with our results detected Prioli *et al*. (2002). Inspite of these they were able to distinguish 184 SSR patterns in the frame of 186 cultivars using only 12 SSRs. Similar variation found Meesang *et al*. (2001) in 144 soybean accessions using 19 SSR loci and Abe *et al*. (2003) in 131 soybean accessions using 20 SSRs. Rongwen *et al*. (1995) reported 11–26 alleles (average 18.6) at 7 SSR loci in a set of 96 soybeans and the gene diversity for these 7 markers averaged 0.87 for all genotypes. Only 2 closely related genotypes were not distinguished by these 7 markers. Those and our results confirm a very high differentiation capability of SSRs in soybean.

Rongwen *et al*. (1995) suggested that a gene diversity value higher than 0.8 is common for soybean microsatellites and this provides a good basis for DNA profiling of soybean. In our study 12 loci had gene diversity from 0.5 to 0.8 and 5 loci had gene diversity higher than 0.8 (Table 1). Using only these five loci, 2 pairs of genotypes could not be differentiated. BS31 and Mutant KG7/1 can be distinguished only by different alleles at the Satt177 locus. Other pair of soybeans – Fiskeby V and Progress, was distinguished by difference at the Satt038 locus. It means that besides highly informative loci with gene diversity values exceeding 0.8, also loci with lower gene diversity can be very useful for distinguishing soybeans. In our study we found out that a variation at the 6 selected loci (four of them with gene diversity exceeding 0.8 – Sat001, Satt005, Satt173, Satt534, two with lower gene diversity – Satt038, Satt177) was sufficient for differentiation of all of the 67 soybeans. The cumulated probability of identifying genotypes with selected SSR markers was 1.11×10^{-9}. It also confirms a high potential of SSRs in differentiation of soybean genotypes. It also supports that the analyses of variation at additional SSR loci usually enhance differentiation capability of SSRs only slightly, whatever agrees with results of Narvel *et al*. (2000), who in 79 elite soybean cultivars analysed at the 74 SSR loci, obtained lower average gene diversity per locus in comparison to our study. For differentiation of larger or more closely related sets of genotypes 10 to 15 SSR loci should be adequate as recommended by Rongwen *et al*. (1995). For the purposes of genotype characterisation, genetic relationship studies, similarity detection between accessions, and evaluation of their gene diversity, meant that the more SSR markers that were used the better. That is why dendrogram relating 67 soybean genotypes, based on data of all 18 microsatellite loci, expressed distinction of groups with maximum and minimum similarities (Figure 2). There are some reflections of genotypes clustering, geographical origin, and pedigree in the dendrogram. Two Japanese genotypes, Czech landraces and cultivars, American and Canadian cultivars, Polish and Swedish cultivars were grouped together. Hybrids Amsoy × Silvia, CMS 2 based hybrids, Sluna × Crusader were located with their parents Silvia, Adoc, Canton, Fred, and Sluna, respectively. Maple Arrow is nearby Labrador (its pedigree is Mc Call × Mapple Arrow), T312 is nearby T266H (F_3 row of L67-533((Clark × Higan) × SRF 300)), Adepta is nearby Dunajka (Hungarian cultivar of unknown origin × Adepta).

Our study confirmed a high potential of specific microsatellites as excellent molecular markers for soybean genotype identification, differentiation,

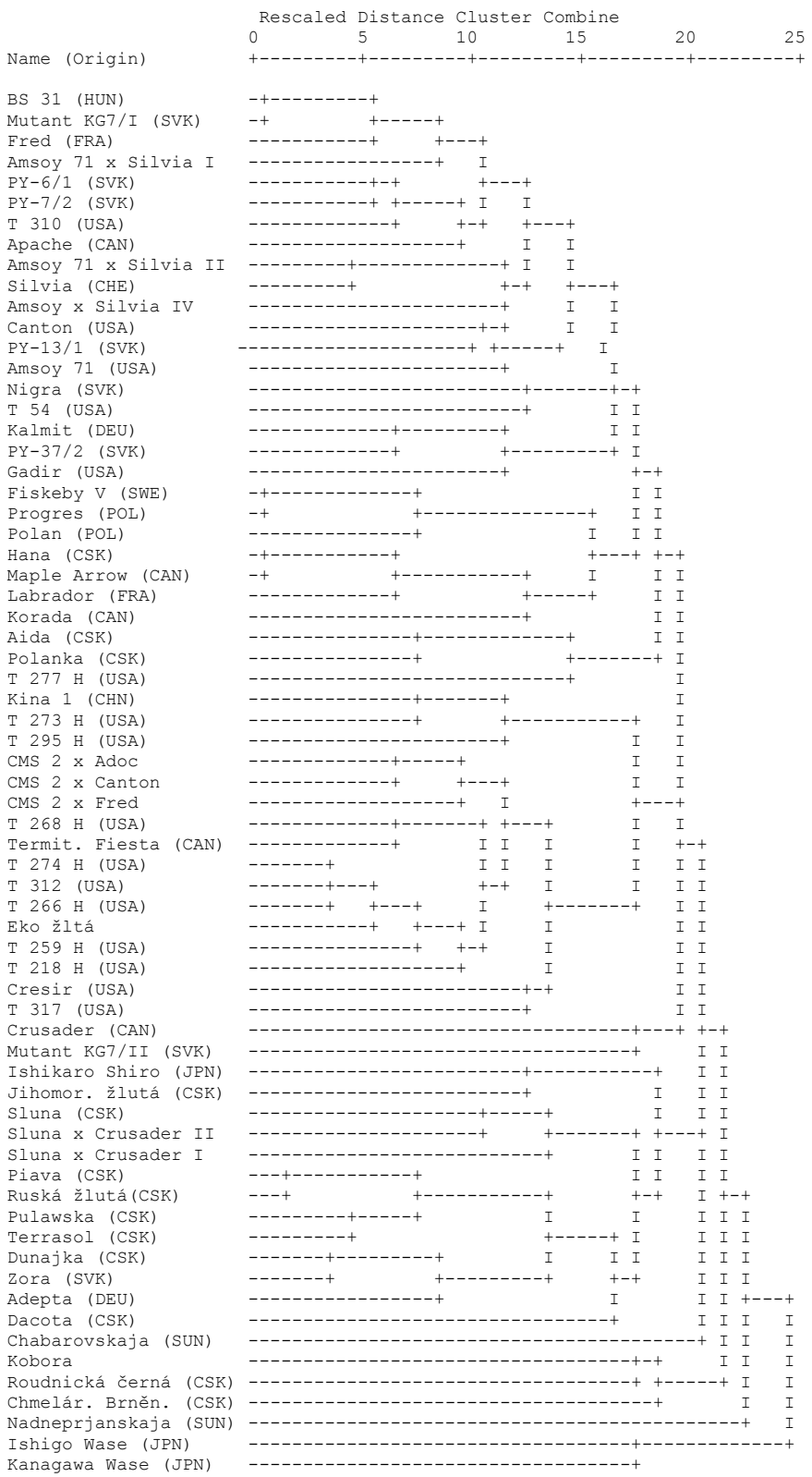

Figure 2. The dendrogram of 67 soybean genotypes differentiated by SSR markers

and evaluation of their genetic variation. Developed DNA profiles of soybeans are usable in soybean genetic resources management in the genebank, especially for differentiation and verification of their identity but also in practice, e.g. in the selection of distant parents to obtain higher genetic variation in progenies, protection of author law, cultivar licences.

Acknowledgements: We would like to thank to Mrs. Jela Klčová for valuable technical assistance, to Mr. František Debre, Ph.D., and Mrs. Katarína Kolenová for providing samples and information about soybean genotypes.

References

Abe J., Xu D.H., Suzuki Y., Kanazawa A. (2003): Soybean germplasm pools in Asia revealed by nuclear SSRs. Theor. Appl. Genet., **106**: 445–453.

Akkaya M.G., Bhawat A., Cregan P.B. (1992): Length polymorphisms of simple sequence repeat DNA in soybean. Genetics, **132**:1131–1139.

Bassam B.J., Caetano-Anolles G., Gresshoff P.M. (1991): Fast and sensitive silver staining of DNA in polyacrylamide gels. Anal. Bioch., **196**: 80–88.

Brown-Guerida G.L., Thompson J.A., Nelson R.L., Warburton M.L. (2000): Evaluation of genetic diversity of soybean introductions and North American ancestors using RAPD and SSR markers. Crop Sci., **40**: 815–823.

Choi I.Y., Kang J.H., Song H.S., Kim N.S. (1999): Genetic diversity measured by simple sequence repeat variations among the wild soybean, *Glycine soja*, collected along the riverside of five major rivers in Korea. Genes Genet. Syst., **74**: 69–177.

Cox T.S., Kiang Y.T., Gorman M.B., Rodgers D.M. (1985): Relationship between coefficient of parentage and genetic similarity indices in soybean. Crop Sci., **25**: 529–532.

Cregan P.B., Bhagwat A.A., Akkaya M.S., Rongwen J. (1994): Microsatellite fingerprinting and mapping of soybean. Meth. Mol. Cell Biol., **5**: 49–61.

Cregan P.B., Jarvik T., Bush A.L., Shoemaker R.C., Lark K.G., Kahler A.L. (1999): An integrated genetic linkage map of the soybean genome. Crop Sci., **39**: 1464–1490.

Dellaporta S. L., Wood J., Hicks J.B. (1993): A plant DNA minipreparation: Version II. Plant Mol. Biol. Rep., **4**: 19–21.

Diwan N., Cregan P.B. (1997): Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. Theor. Appl. Genet., **95**: 723–733.

Hamada H., Petrino M.C., Takugana T. (1982): A novel repeated element with Z-DNA forming potential is widely found in evolutionary diverse eukaryote genomes. Proc. Natl. Acad. Sci. USA, **79**: 6465–6469.

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

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

Maughan P.J., Saghai-Maroof M.A., Buss G.R. (1995): Microsatellite and amplified sequence length polymorphisms in cultivated and wild soybean. Genome, **38**: 715–723.

Meesang N., Ranamukhaarachchi S.L., Petersen M.J., Andersen S.B. (2001): Soybean cultivar identification and genetic purity analysis using microsatellite DNA markers. Seed Sci. Technol., **29**: 637–645.

Messmer M.M., Melchinger A.E., Herrmann R.G., Boppermaier J. (1993): Relationship among early European maize inbreds: II. Comparison of pedigree and RFLP data. Crop Sci., **33**: 944–950.

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

Narvel J.M., Fehr W.R., Chu W.S., Grant D., Shoemaker R.C. (2000): Simple sequence repeat diversity among soybean plant introductions and elite genotypes. Crop Sci., **40**: 1452–1458.

Paetkau D., Calvert W., Stirling I., Strobeck C. (1995): Microsatellite analysis of population structure in Canadian polar bear. Mol. Ecol., **4**: 347–354.

Phillips W.J., Chapman C.G.D., Jack P.I. (1994): Molecular cloning and analysis of one member of a polymorphic family of GACA – hybridising DNA repeats in tomato. Theor. Appl. Genet., **88**: 845–851.

Powell W., Machray G.C., Provan J. (1996): Polymorphism revealed by simple sequence repeats. Trends Plant Sci., **1**: 215–222.

Prioli R.H.G., Mendes C.T. Jr., Arantes N.E., Contel E.P.B. (2002): Characterisation of Brazilian soybean cultivars using microsatellite markers. Genet. Mol. Biol., **25**: 185–193.

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

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

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

Schneider K., Douches D.S. (1997): Assesment of PCR-based simple sequence repeats to fingerprint

North American potato cultivars. Am. Potato J., **74**: 149–160.

Song Q.J., Quigley C.V., Nelson R.L., Carter T.E., Boerma H.R., Strachan J.R. (1999): A selected set of trinucleotide simple sequence repeat markers for soybean cultivar identification. Plant Var. Seeds, **12**: 207–220.

Weber J.L. (1990): Informativeness of human $(dC\text{-}dA)_n \times (dG\text{-}dT)_n$ polymorphism. Genomics, **7**: 524–530.

Weir B.S. (1990): Genetic data analysis. II. Methods for discrete population genetic data. 2nd Ed. Sinauer Associates, Sunderland, Mass.

Received for publication July 14, 2003
Accepted after corrections October 20, 2003

Abstrakt

Hudcovicová M., Kraic J. (2003): **Využitie SSRs pre charakterizáciu genetických zdrojov sóje (***Glycine max* (L.) Merr.**).** Czech J. Genet. Plant Breed., **39**: 120–126.

SSR profily 67 genotypov sóje (*Glycine max* (L.) Merr.) rôzneho pôvodu bolo analyzovaných pomocou 188 alel 18 SSR lokusov. V každom SSR lokuse bolo nájdených 4 až 21 alel (priemerný počet alel na lokus bol 10,4) a priemerná hodnota génovej diverzity bola 0,71. Všetkých 67 genotypov sóje bolo úspešne odlíšených použitím iba 6 vybratých SSR markerov (Sat001, Satt005, Satt038, Satt173, Satt177, Satt534), ktorých hodnoty génovej diverzity sa pohybovali v rozmedzí 0,66–0,89. Kumulovaná pravdepodobnosť nájdenia identického SSR profilu v analyzovanom súbore sóje bola $1{,}11 \times 10^{-9}$, čo potvrdzuje vysoký potenciál SSR markerov v odlišovaní genotypov sóje v kolekciách. Zoskupovanie genotypov v dendrograme čiastočne odzrkadlovalo ich pôvod a rodokmeň.

Kľúčové slová: sója; *Glycine max* (L.) Merr.; mikrosatelity; genetická diverzita

Corresponding author:

Mgr. Martina Hudcovicová, Výskumný ústav rastlinnej výroby, Divízia aplikovanej genetiky a šľachtenia, Bratislavská cesta 122, 92168 Piešťany, Slovenská republika
tel.: + 421 33 772 23 11–12, fax: + 421 33 772 63 06, e-mail: hudcovicova@vurv.sk