Crop Breeding and Applied Biotechnology 7:155-164, 2007
Brazilian Society of Plant Breeding. Printed in Brazil

Microsatellite molecular markers in the cultivar identification of Brazilian soybean for human consumption

Alexandre Franco Garcia[1], João Luiz Alberini[2], Maria Imaculada Zucchi[3*], and Anete Pereira De Souza[4]

Received 22 April 2005

Accepted 20 April 2007

ABSTRACT - *The objective of this research was to exploit the variability of 69 microsatellite marker loci in 18 soybean varieties (group I) to select 10 loci for routine cultivar identification. The selected loci were used to evaluate 32 varieties (group II). In group I an average of 3.36 alleles per locus was observed of a total of 232 alleles and an average of 5.1 alleles per locus for the 32 varieties evaluated in group II. The mean distance values were 0.67 and 0.82 for group I and group II, respectively. The 10 loci selected and evaluated in group II were the most effective to discriminate the genotypes and, therefore, recommended for the characterization of Brazilian soybean cultivars for human nutrition.*

Key-words: *Glycine max*, SSR, cultivar protection, germplasm.

INTRODUCTION

Soybean is world-wide the most important oil crop. Subproducts are derived from the grain, mainly flour and oil, besides others used in the food agroindustry and chemical industry. Brazil accounts for about 27% of the worldwide harvested raw product and also ranks second in the world soybean production, with a prospect of becoming the greatest producer in the growing season 2007/2008. The development of the Brazilian agribusiness is intrinsically related to soybean, since in 2004, soybean exports from Brazil reached US$ 10.0 billion (Embrapa 2003, MAPA 2005).

In view of the impact the development of new cultivars has had on the world agriculture, several industrialized countries have implanted protection systems of improved varieties. Brazil has a national legislation that regularizes the protection system of the country's lines and cultivars. The Law of Cultivar Protection requires that cultivars submitted to trademark registration and protection must present traits of distinction, homogeneity and stability (DHS).

To discriminate genotypes, morphologic and biochemical traits must be evaluated. The use of this kind of trait data as identification criteria hinders the registration of new cultivars with identical traits to others registered before. Since the number of new cultivars is on the rise whereas the number of differentiating traits is constant, the clear distinction of

[1] Departamento de Genética, Escola Superior de Agricultura Luiz de Queiroz (ESALQ/USP), Avenida Pádua Dias, 11, C.P. 9, 13.418-000, Piracicaba, SP, Brasil
[2] Naturale, Rua Dr. Jaime Ribeiro Luz, 971, Bairro Santa Mônica, 34.408-188, Uberlândia, MG, Brasil
[3] Instituto Agronômico, Avenida Barão de Itapura, 1481, C.P. 28, 13.012-970, Campinas, SP, Brasil. *E-mail: mizucchi@iac.sp.gov.br
[4] Departamento de Genética e Evolução, Instituto de Biologia, Universidade Estadual de Campinas, Cidade Universitária Zeferino Vaz, Barão Geraldo, 13.083-970, Campinas, SP, Brasil

different cultivars has become more difficult, as for example in the case of related cultivars, for which the set of differentiating traits tends to be the same (Rongwen et al. 1995, Priolli et al. 2002).

A complementary form of increasing the number of differentiators is to exploit the variability of the DNA molecule, by using microsatellite markers. The analysis of the polymorphism DNA sequence allows for a more accurate genetic identification and plant discrimination.

Molecular markers are frequently used in the analysis of soybean germplasm. Akkaya et al. (1992) used 10 loci in 43 soybean genotypes. Abe et al. (2003) used 20 SSR loci in 131 accesses introduced from 14 Asian countries. Morgante and Olivieri (1994) detected similar levels in 7 SSR loci in a group of 61 genotypes. Even in elite cultivars that have a very narrow genetic base after undergoing intense selection pressure, the number of alleles per locus was high. Rongwen et al. (1995) reported 11 to 26 alleles in 7 SSR loci in a group of 96 soybean genotypes, but observed that this number was reduced to 5 to 10 alleles in the same loci when evaluated in 26 cultivars of breeding programs of the northern USA. Doldi et al. (1997) found 2 to 6 alleles per locus in a group of 18 soybean cultivars using 12 microsatellite loci. Diwan and Cregan (1997) observed on average 10.1 alleles per locus in a total of 20 loci studied in 35 genotypes that represented 95% of all alleles of the germplasm cultivated in the north of the United States. In 186 Brazilian cultivars, Priolli et al. (2002) found 4 to 8 alleles in a total of 12 SSR loci studied. Narvel et al. (2000) detected a total of 397 alleles in 79 elite cultivars and PIs using 74 microsatellite loci.

This study had the objective of assessing the variability of microsatellite locus markers, based on commercial varieties and experimental lines, to obtain and select a set of 10 loci with sufficient resolution and discrimination power to be used complementary to variety identification of Brazilian cultivars for human consumption.

MATERIAL AND METHODS

The company Naturalle Agromercantil S/A developed and provided the 32 soybean varieties that were evaluated here. These varieties are used in human nutrition as beverages, tofu and nato. The so-called Tonyu (Japanese term for soy drink), soymilk, soy drink or soy beverage is an aqueous extract of soybean grains, used as refreshment or for the production of other soy-based products. Tofu or soy cheese is coagulated by salts or acids. Nato is soy product fermented by bacteria (*Bacillus natto*). The varieties used for beverage and tofu production have large grains, high protein content and a pleasant taste, whereas the varieties used for nato have small seeds and high sugar contents.

The lines used for beverage and tofu production were labeled with TF and the lines used for nato were identified by VG. Fukuyutka, Toromazari and Vinton 81 are standard cultivars for tofu preparation, as well as BR 36, a Brazilian standard cultivar for the same purpose. Conquista and BRS 133 are adapted standard cultivars but not considered specific for nutrition, although Conquista is used for its slightly elevated seed size and protein content (Table 1).

To isolate the genomic DNA of each variety, around 300 mg lyophilized leaf tissue from a 10-plant bulk sample was used, according to the methodology described by Hoisington et al. (1994). The primer pair sequences were obtained from the soybean data base (http://www.soybase.org), preferentially using microsatellites with trinucleotide replications (ATT) distributed in 18 of the 20 linkage groups previously mapped by Cregan et al. (1999). One hundred microsatellites were used, of which 69 were selected to obtain the allele profile. The amplification in a thermocycler MJ-PTC-100 was programmed for 2 minutes at 94 ºC, followed by 33 cycles of 1 minute at 94 ºC, 1 minute at the specific annealing temperature (AT) (from 46 to 55 ºC) of the primer pairs of each microsatellite locus used, 1 minute at 72 ºC and a final of 33 cycles, extension of 10 minutes at 72 ºC. For each amplification reaction we used 50 ng DNA, 0.5 U of Taq DNA polymerase (Invitrogen), 1 x Taq DNA polymerase PCR buffer (20 mM Tris-HCl pH 8.4; 50 mM KCl), 0.8 mM of each primer pair, 1.5 mM $MgCl_2$ and 150 mM dNTP, in a final reaction volume of 25 ml. In a first phase, 5 ml samples of the amplification products were analyzed on 4% Metaphor agarose gel: common agarose (Invitrogen) (1:1), in 0.5 x TBE buffer (Tris/Borato 0.045M and EDTA 0.001M pH 8.3), for 1:30 h at 170 Volts. After staining with 0.5 mg/ml ethidium bromide the fragments were visualized and photographed in a UV transilluminator. Thereafter, 3 ml to 5 ml samples of the amplification products mixed with formamide (1:1) were resolved in denaturing gel containing 6% acrylamide, 7 M urea and 1 x TBE (0.045M Tris/Borato and 0.001M EDTA pH 8.3), during 1:30 h at 75 W. The fragments were visualized by silver-staining as described by Creste et al. (2001).

Table 1. Number and code of the 32 varieties in the evaluations using microsatellite loci

Nr.	Variety code	Protection code[a]	Nr.	Variety code	Protection code[a]
1	FTNagano	NT1	17	VG57-nato	
2	TF99-1648	NT2	18	VG65-nato	NT65
3	Nagano 2		19	TF99-1210	NT18
4	TF99-1239	NT4	20	Fukuyutaka	
5	TF99-1583	NT5	21	Toromazari	
6	TF99-1584	NT6	22	Vinton	
7	TF99-1543		23	TF99-1337	NT16
8	TF99-1496	NT12	24	TF99-1312	NT15
9	TF99-1274		25	TF99-1495	NT11
10	TF99-1214	NT8	26	TF99-1388	
11	TF99-1356		27	TF99-1264	
12	TF99-1253		28	TF99-1548	
13	TF99-1349-B	NT17	29	TF20-65	
14	TF99-1554	NT14	30	Conquista	
15	TF99-1307	NT13	31	BRS133	
16	VG199-nato	NT199	32	BR36	

[a] code used for varieties evaluated with the descriptors for commercial protection

For each one of the polymorphic microsatellite loci, the Polymorphic Information Content (PIC) was calculated according to the formula (Smith et al. 1997) $PIC = 1-\sum_{i=1}^{n} f_i^2$, where f_i is the frequency of the i^{th} allele and n is the number of alleles of the locus.

The genetic distances for 18 varieties were evaluated using 69 marker loci (group I) and the Modified Roger's distance, (MRD) (Wright 1978), which considers the allele frequencies of the codominant markers, as those of the microsatellite type. The distance matrix was constructed using software TFPGA (Miller 1997) and the dendrogram by the UPGMA (Unweighted Pair-Group Method using Arithmetic Averages) grouping criterion with NTSYS (Rohlf 1997). Grouping stability was further tested by the resampling procedure with 10000 bootstrap samples using software DBoot (Coelho 2000).

The 10 microsatellite loci with highest PIC values as well as amplification products of good quality were selected for the identification of this same group plus 14 other genotypes (group II), that is, the 32 genotypes were evaluated considering the 10 selected SSR loci. Then the genetic distances based on the 10 most informative SSR loci were compared with the morphologic descriptors for 16 soybean varieties.

RESULTS AND DISCUSSION

For all varieties, the most common situation was to find a single allele type per locus (homozygous state), although two different types of alleles had been observed in some loci as well (heterozygous state). Since soybean is an autogamous species, it had been expected, as observed, that most of the evaluated loci would be in homozygous state. And since soybean is also a diploid species, each variety was evaluated assuming two alleles per locus, of the same size for the varieties with only one band in the electrophoresis gel (homozygous state) and of different sizes for the varieties with two bands (heterozygous state) (Diwan and Cregan 1997).

There are several possible causes for the appearance of loci in heterozygous state, such as mutation, residual heterozygosity (Diwan and Cregan 1997) or duplication of genome segments (Shoemaker et al. 1996). The use of DNA obtained from plant bulks may further contribute to the increase of loci in heterozygous state, when the bulk genotypes have distinct alleles for a particular locus (Akkaya et al. 1992, Diwan and Cregan 1997). In general however, the presence of these loci does not impair the genotypic identification of the cultivars, since the use of DNA bulks maintains their allele constitution, avoiding that alleles of a single

plant only represent the genotype of the cultivar as a whole (Diwan and Cregan 1997). Nevertheless, different strategies (Akkaya et al. 1992, Smith et al. 1997), including pedigree analyses (Diwan and Cregan 1997), can be used for a more in-depth evaluation of the conditions of the microsatellite loci used here.

The 69 microsatellite loci of the first evaluation of 18 varieties (group I) were assigned to 18 of the 20 linkage groups (LG) (Table 2), with an average 3.63 markers per LG. Based on these loci, 232 alleles were obtained. The number of alleles per locus varied from 2 to 8 (mean alleles per locus 3.36) (Table 3). The 10 microsatellite loci used in the evaluation of the 32 varieties (group II), were distributed in 9 LGs (Table 2). Based on these loci, 51 alleles were detected; the number of alleles per locus varied from of 3 to 7 and the mean of alleles per locus was 5.1 (Table 3).

The total number of alleles of the evaluation in group II was lower than in group I, while the mean of alleles per locus of group II was higher than in group I. The loci with 2 and 3 alleles corresponded to 65 and 10% of the alleles observed for groups I and II, respectively, while the loci with 5 and 6 alleles represented 14 and 50% of the alleles observed for groups I and II, respectively (Table 3). Considering that in the evaluation of the cultivars of group II the 10 microsatellite loci with highest discrimination capacity had been used, it follows that loci with this trait tend to concentrate more alleles in each locus, increasing the mean of alleles per locus and the percentage of loci with more alleles.

The polymorphism of these 10 microsatellite loci was consistent with results of Priolli et al. (2002), Narvel et al. (2000), Doldi et al. (1997) and Akkaya et al. (1992), but lower than stated by Diwan and Cregan (1997), Abe et al. (2003) and Rongwen et al. (1995).

Priolli et al. (2002) used 12 microsatellite loci and detected an average 5.3 alleles per locus, of a total of 64 alleles distributed in loci with a number of alleles varying from 4 to 8. These authors evaluated 186 elite soybean cultivars, developed and released by public and private Brazilian institutions, selected to represent the genotypes cultivated in the country. Considering the large quantity of evaluated cultivars and the representativeness of the genotypes, the authors observed few alleles more than in group II of our study. Besides, the quantity of alleles provided by the 12 selected microsatellite loci was not sufficient to discriminate all evaluated genotypes. This impossibility was attributed to the narrow genetic base of the improvement programs of Brazilian soybean. On the other hand, since the selection criterion of the 12 microsatellite loci was not based on their respective discrimination power, but only on the polymorphisms in acrylamide and replication types, it is possible that other genome regions with greater variability had not been sampled. Rongwen et al. (1995), for example, found 5 to 10 alleles per locus when they evaluated only 26 cultivars of the breeding programs of the northern region of the EUA. Thus, even for genotypes obtained based on a narrow genetic base, some regions could contribute to a clearer discrimination of the genotypes.

Although no detailed information is available on the population from which the soybean genotypes used for human consumption of this study were originated, it is known that this population consists essentially of Brazilian germplasm and that the genotypes derived thereof have a limited genetic base. Moreover, considering that the 186 cultivars were selected to represent the nation-wide cultivated genotypes and that the genotypes of this study represent varieties used in human nutrition only, it is possible to consider these genotypes as a type of specific subgroup of those cultivars, with specific traits of genotypes used for human consumption, such as large grains, high protein content and pleasant taste, for varieties designated for beverage and tofu production and small-seeded varieties with high sugar content, used for nato.

It is therefore expected that the variability within the genotypes designated for human consumption is less, or at the most, equivalent to the variability in the cultivars evaluated by Priolli et al. (2002). On the other hand, one can expect that the discrimination capacity of the 10 microsatellite loci of group II would outmatch the 12 loci used in the 186 cultivars, considering the selection criteria adopted in each case. Taking these two aspects into consideration, even with a high number of cultivars, representative of the genotypes cultivated in the country, some of the genomic variability seems to have remained unexploited, which reduced the values of alleles per locus and average alleles per locus compared to, for example, the study of Rongwen et al. (1995). The values for group II, though quite close to those of Priolli et al. (2002), were obtained from a somewhat lower quantity of genotypes with probable lower variability in the bulks, indicating the efficiency

Table 2. Linkage group, major and minor allele of the locus, number of alleles per locus and PIC of the evaluation of group I. Rows in bold indicate the microsatellite loci used in the evaluation of group II

Locus	Linkage group	Minor and major fragment (pb)	Nr. of alleles	Allele frequency	PIC
Satt146	F	280-330	8	0.17; 0.11; 0.06; 0.17; 0.17; 0.11; 0.11; 0.11	0.864
Satt308	**M**	**130-175**	**6**	**0.08; 0.22; 0.25; 0.22; 0.17; 0.06**	**0.801**
Satt173	O	200-255	6	0.06; 0.14; 0.25; 0.14; 0.36; 0.06	0.762
Satt329	A2	235-265	6	0.03; 0.22; 0.19; 0.11; 0.39; 0.06	0.745
Satt191	G	200-225	6	0.03; 0.08; 0.08; 0.36; 0.14; 0.31	0.742
Satt154	**D2**	**255-280**	**6**	**0.06; 0.28; 0.42; 0.03; 0.06; 0.17**	**0.715**
Satt152	**N**	**230-260**	**4**	**0.19; 0.25; 0.14; 0.42**	**0.707**
Satt070	**B2**	**145-175**	**4**	**0.25; 0.42; 0.22; 0.11**	**0.702**
Satt294	C1	240-285	4	0.06; 0.36; 0.28; 0.31	0.696
Satt129	**D1a+Q**	**120-130**	**4**	**0.31; 0.22; 0.06; 0.42**	**0.681**
Satt175	**M**	**155-185**	**5**	**0.19; 0.14; 0.11; 0.06; 0.50**	**0.677**
Satt358	O	185-200	4	0.11; 0.19; 0.22; 0.47	0.677
Satt185	**E**	**215-250**	**4**	**0.08; 0.19; 0.47; 0.25**	**0.670**
Satt045	**E**	**135-145**	**4**	**0.19; 0.08; 0.25; 0.47**	**0.670**
Satt263	E	200-245	5	0.03; 0.50; 0.06; 0.17; 0.25	0.656
Satt186	D2	225-250	5	0.53; 0.06; 0.08; 0.22; 0.11	0.650
Satt009	**N**	**160-210**	**3**	**0.39; 0.22; 0.39**	**0.648**
Satt187	**A2**	**240-290**	**3**	**0.36; 0.42; 0.22**	**0.647**
Satt192	H	250-260	4	0.03; 0.33; 0.47; 0.17	0.637
Satt588	K	150-180	5	0.06; 0.56; 0.14; 0.14; 0.11	0.637
Satt408	D1a+Q	185-195	3	0.44; 0.17; 0.39	0.623
Satt357	C2	215-320	3	0.50; 0.22; 0.28	0.623
Satt184	D1a+Q	155-185	3	0.50; 0.17; 0.33	0.611
Satt509	B1	180-250	3	0.14; 0.44; 0.42	0.610
Satt046	K	175-235	3	0.17; 0.31; 0.53	0.600
Satt453	B1	230-255	3	0.44; 0.11; 0.44	0.593
Satt012	G	110-200	4	0.31; 0.03; 0.56; 0.11	0.585
Satt178	K	165-190	3	0.14; 0.56; 0.31	0.579
Satt128	B1	265-280	3	0.50; 0.42; 0.08	0.569
Satt237	N	230-270	4	0.11; 0.61; 0.22; 0.06	0.562
Satt307	C2	150-185	5	0.31; 0.03; 0.03; 0.58; 0.06	0.562
Satt225	A1	95-105	3	0.22; 0.17; 0.61	0.549
Satt042	A1	155-175	4	0.11; 0.64; 0.06; 0.19	0.539
Satt190	C1	180-220	3	0.56; 0.06; 0.39	0.537
Satt102	K	135-150	3	0.56; 0.06; 0.39	0.537
Satt241	O	205-245	3	0.39; 0.56; 0.06	0.537
Satt324	G	200-240	3	0.61; 0.33; 0.06	0.512
Satt072	F	195-200	2	0.53; 0.47	0.498
Satt330	I	110-150	2	0.53; 0.47	0.498
Satt179	D1a+Q	130-190	3	0.06; 0.31; 0.64	0.495
Satt421	A2	230-235	2	0.42; 0.58	0.486
Satt135	D2	170-200	2	0.64; 0.36	0.461
Satt309	G	125-140	3	0.08; 0.69; 0.22	0.461

To be continued...

Locus	Linkage group	Minor and major fragment (pb)	Nr. of alleles	Allele frequency	PIC
Satt020	B2	105-125	3	0.03; 0.31; 0.67	0.461
GMGLPSI2	I	115-120	2	0.36; 0.64	0.461
Satt134	C2	310-320	2	0.67; 0.33	0.444
Satt194	C1	215-245	4	0.72; 0.03; 0.17; 0.08	0.443
Satt376	C2	215-250	3	0.14; 0.72; 0.14	0.440
Satt165	A1	275-280	2	0.31; 0.69	0.424
Satt245	M	195-225	3	0.75; 0.11; 0.14	0.406
Satt172	D1b+W	225-240	3	0.08; 0.17; 0.75	0.403
Satt311	D2	180-195	3	0.17; 0.08; 0.75	0.403
Satt126	B2	120-155	2	0.72; 0.28	0.401
Satt168	B2	225-230	2	0.28; 0.72	0.401
Satt181	H	185-225	3	0.03; 0.75; 0.22	0.387
Satt335	F	150-165	3	0.03; 0.22; 0.75	0.387
Satt227	C2	240-250	3	0.06; 0.17; 0.78	0.364
Satt160	F	250-260	2	0.78; 0.22	0.346
Satt395	F	275-280	2	0.22; 0.78	0.346
Satt314	H	150-235	3	0.81; 0.06; 0.14	0.329
Satt170	C2	125-135	2	0.17; 0.83	0.278
Satt005	A2. D1b+W	155-175	4	0.03; 0.86; 0.06; 0.06	0.252
Satt201	M	270-300	3	0.11; 0.86; 0.03	0.245
Satt384	E	120-150	2	0.89; 0.11	0.198
Satt155	A1	145-155	2	0.89; 0.11	0.198
Satt208	D2	190-195	2	0.08; 0.92	0.153
Sct026	B1	115-120	2	0.06; 0.94	0.105
Satt063	B2	105-145	2	0.94; 0.06	0.105
Satt189	D1b+W	235-245	2	0.97; 0.03	0.054

Table 3. Number of alleles, total of alleles (a) and mean of alleles per locus (b) in the evaluations of group I, group II and group I*

Nr. of alleles	Freq. group I	%	Nr. x freq[c]	Freq. group II	%	Nr. x freq[d]	Freq. group I *	%	Nr. x freq[e]
2	18	26.09	36	0	0	0	0	0	0
3	27	39.13	81	1	10	3	2	20	6
4	13	18.84	52	3	30	12	4	40	16
5	5	7.25	25	1	10	5	2	20	10
6	5	7.25	30	4	40	24	2	20	12
7	0	0	0	1	10	7	0	0	0
8	1	1.45	8	0	0	0	0	0	0
Total	69	100	232[a]	10	100	51[a]	10	100	44[a]
Mean			3.36[b]			5.10[b]			4.40[b]

[c] Nr. of alleles multiplied by the allele frequency of group I (18 varieties evaluated with 69 SSR loci)
[d] Nr. of alleles multiplied by the allele frequency of group II (32 varieties evaluated with 10 SSR loci)
[e] Nr. of alleles multiplied by the allele frequency of group I*(18 varieties evaluated with 10 SSR loci)

of the 10 selected microsatellite loci in the discrimination of these genotypes.

Furthermore, considering the low number of genotypes evaluated compared to Priolli et al. (2002), the polymorphism of the selected loci had possibly not been fully exploited. A comparison of the results

obtained for group II and group I* (Table 3) showed that, even when the 10 selected microsatellite loci are used, the reduction of the number of evaluated genotypes from 32 in group II to 18 in group I*, causes a reduction in the total number of alleles from 51 to 44, respectively, as well as a reduction in average alleles per locus from 5.1 to 4.4, respectively. The variation range of the number of alleles per locus diminished from 3 to 7 for group II and from 3 to 6, for group I*. The loci with 3 and 4 alleles increased from 40% of the alleles observed for group II to 60% for group I* while the loci with 6 and 7 alleles dropped from 50% of the alleles observed to 20%, for groups II and I*, respectively. Therefore, although there is a limit, the evaluation of more genotypes using the 10 microsatellite loci can further enhance the discriminatory power of these loci.

Narvel et al. (2000) evaluated 39 elite soybean cultivars and 40 PIs from different regions of the world, using 74 microsatellite loci (mean of alleles 5.4 per locus; 2 to 11 alleles per locus). The number of alleles detected among the PIs was 30% higher than among the elite cultivars, indicating the loss of variability by the use of elite genotypes. Although the interval of the number of alleles varied from 3 to 7 for group II, the value of the mean number of alleles per locus was consistent with our results. Doldi et al. (1997) evaluated the genetic diversity of 18 soybean genotypes selected for use in breeding programs to increase the protein content of varieties adapted to cultivation conditions of Central Europe. Twelve RAPD primers and twelve microsatellite markers were used for this purpose and 2 to 6 alleles found per locus.

Akkaya et al. (1992) used 10 microsatellite loci to evaluate 43 soybean genotypes and obtained a mean of 7.0 alleles per locus. The evaluated soybean genotypes were derived from 5 groups: 1) ancestor cultivars of the northeastern United States that include genotypes that represent the basis of the American and Canadian germplasm; 2) cultivars of the northeastern USA, which were developed by hybridization; 3) ancestor cultivars of the southeastern USA, including cultivars that represent the basis of the southeastern germplasm; 4) Cultivars of the southeastern USA developed by hybridization and 5) *Glycine soybean* genotypes, wild soybean. Considering that the 43 evaluated soybean genotypes were selected for representing wide genetic diversity, differently from the soybean genotypes for human consumption of this study, the mean of alleles per locus in the two studies was rather alike.

Rongwen et al. (1995) found 11 to 26 alleles per locus in an evaluation of 96 soybean genotypes, and 5 to 10 alleles per locus when the same microsatellite loci were used to investigate 26 cultivars of the North American breeding program, indicating that in spite of the loss of variability in genotypes under strong selection pressure in breeding programs, it is still possible to find variability enough to discriminate genotypes.

In general, more heterogeneous genotypes presented more polymorphism when compared with the soybean genotypes of this study. Abe et al. (2003) used 20 SSR loci in 131 accesses introduced from 14 Asian countries. The loci produced an average of 11.9 alleles per locus, in a range of 8 to 24 alleles. Diwan and Cregan (1997) using 20 SSR loci found a mean of 10.1 alleles per locus when using 35 ancestor soybean genotypes, responsible for 95% of the alleles of the North American soybean germplasm.

The number of alleles per locus and relative frequencies can be expressed in a single PIC value, an estimate that expresses the discrimination power of a locus. The PIC value varies from 0, when the locus is monomorphic, to 1, when the locus presents many alleles at similar frequencies (Smith et al. 1997). The PIC values in the evaluation of group I varied from 0.054 to 0.864, when all 69 microsatellite loci were considered and from 0.647 to 0.801 when only the 10 selected microsatellite loci were taken into account (Table 2). The mean PIC values were 0.508 and 0.692, respectively. In the evaluation of group II, the PIC values for the 10 selected microsatellite loci varied from 0.656 to 0.814, with a mean value of 0.747. All PIC values of the 10 selected loci were higher in the evaluation of group II than in group I, with exception of locus Satt 154, which presented the same value in both evaluations. The higher number of evaluated genotypes increased the discrimination power of the selected loci, indicating that the polymorphism of the loci had not been entirely exploited due to sampling.

The information on allele frequency was also used

for a distance matrix to evaluate groups I and II. The values varied from zero to one, where zero indicates identical genotypes and above zero different genotypes. The mean distance values were 0.67 and 0.82 in group I and II, respectively. The mean of the distances in group I was lower since in this evaluation selected microsatellite loci with high PIC values were used, that is, loci with high discrimination power. Most values of group I ranged between 0.6 and 0.8, while in group II the frequency was highest between 0.7 and 1.0. In both evaluations, the most frequent values were similar to those observed by Priolli et al. (2002), who found values varying from 0.4 to 0.9. The most frequent values in the evaluation of group II were closer to 1. However, as mentioned above, the cultivars used by these authors represent the wide genetic variability of genotypes cultivated in Brazil while the genotypes of our evaluations (group I and II) represent only those used in human consumption. Since the genotypes evaluated by Priolli et al. (2002) were genetically less similar than those of group II, the microsatellite loci used here most likely have a higher discrimination power.

In a comparison of the differences between these microsatellite loci with the morphologic descriptors of 16 varieties of group I (Table 1), it was observed that the microsatellite loci detected more differences between the genotypes than the morphologic descriptors used for registration and protection. In most comparisons of pairs of cultivars between 40% and 50% discriminative traits were found; for none of the evaluated traits the differences between the genotypes were higher than 80%. For the microsatellite loci used here, most comparisons revealed over 80% difference between the compared pairs of varieties. Most commonly, the varieties differed by 100%, detected by the 69 SSR markers, which underscores the discriminatory power of this molecular marker type.

The distance matrices of the evaluations of groups I and II were also used to construct dendrograms, which are graphical representations of the genetic distances between cultivars. Two large groupings can be observed in the dendrogram for group I (Figure 1), one with the genotypes 1, 2, 11, 3, 6, 5, 12, 4, 16, 17, and 18 and the other 8, 10, 13, 7, 19, 9, and 15, with 100% consistency in the bootstrap analysis with 10,000 resamplings. Within these groupings, subgroupings can be distinguished. For example, the varieties 16, 17 and 18, also grouped with 100% consistency. These varieties, used for nato, were assigned to the same group, while they were also genetically distant enough to allow a differentiation. The genetic distance between varieties 3 and 6 was small (0.17), that is, the similarity was high, with a consistency of 84% for the node formed in this grouping by resampling. In the dendrogram obtained for group II (Figure 2), two groupings were observed where variety 21 and 29 clustered with the consistency of the node of 84%, calculated with 10,000 bootstrap resamplings. These two cultivars formed a group that differs from the other cultivars with 100% consistency of the node formed between them. The other 30 genotypes form two subgroups - one with cultivars 1, 12, 3, 6, 5, 11, 2, 4, 28, 26, 31, 16, 17, 18, 27, 20, 32 and 30 and a second subgroup with cultivars 7, 10, 19, 9, 15, 22, 14, 25, 8, 13, 23, and 24. A similar pattern was observed in these 30 genotypes as in the group I evaluation, where cultivars 3, 5 and 6 grouped consistently (89%) and of these, 3 and 6 clustered with a consistency of 96%. Varieties 13, 23 and 24 grouped with a node consistency of 97%. Another noteworthy grouping was formed by cultivars 10 and 19 that clustered with a consistency of 91%. The shorter distance values of groupings of the varieties 3, 6 and 5 in group II compared to group I evaluation may be ascribed to sampling, since for genetically nearly identical genotypes more polymorphic loci are necessary to make the differences between them clear (Figures 1 and 2). Thus, the selection of microsatellite loci increased on the one hand the discriminatory power of genotypes and on the other it evidenced the similarity of genetically very similar genotypes. Thus, the short distance value obtained in the grouping of varieties 3 and 6 may indicate genetic identity (Figure 2).

CONCLUSIONS

1. Microsatellite loci are effective in the discrimination of Brazilian soybean varieties for consumption and provide additional information to the morphologic descriptors.

2. Ten informative SSR loci are effective for differentiation and therefore indicated for evaluations of molecular descriptors of soybean cultivars for consumption.

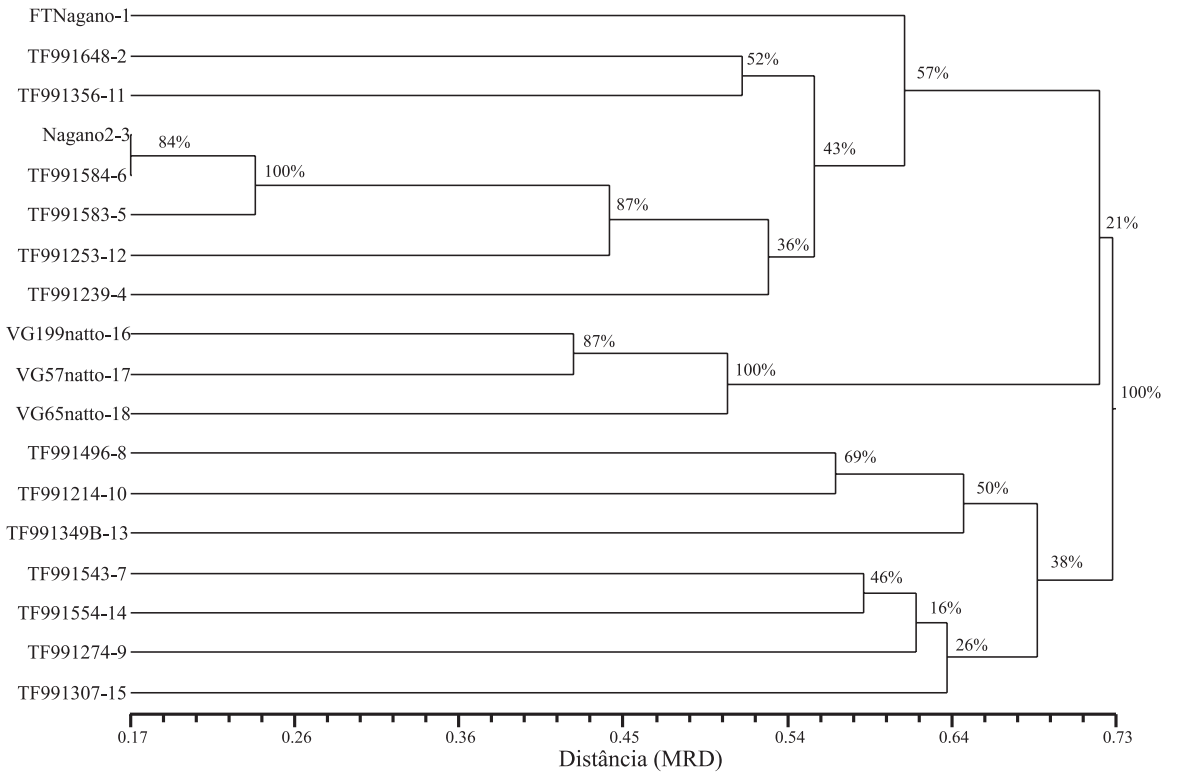

Figure 1. Dendrogram constructed using the modified Roger's distance (MRD) and the UPGMA criterion with data of 69 microsatellites for 18 varieties

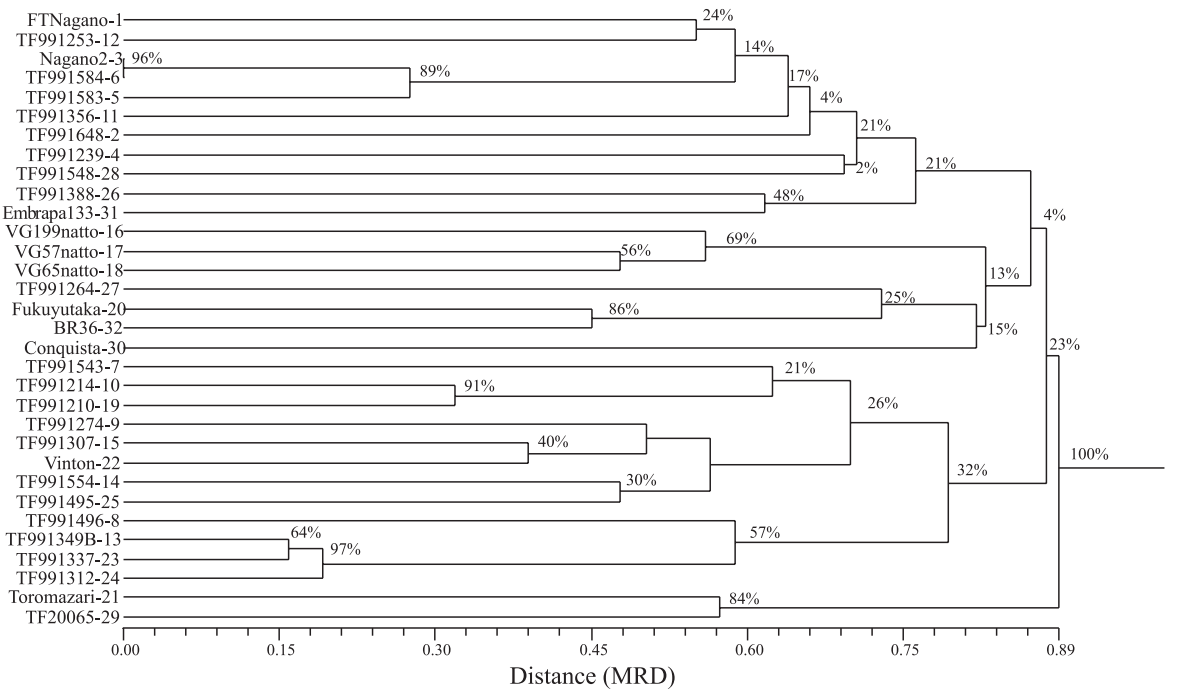

Figure 2. Dendrogram constructed using the modified Roger's distance (MRD) and the UPGMA criterion with data of 10 microsatellites for 32 varieties

Marcadores Moleculares microssatélites na caracterização de cultivares brasileiros de soja para alimentação humana

RESUMO - *O objetivo deste trabalho foi explorar a variabilidade de 69 locos marcadores do tipo microssatélites em 18 variedades de soja (grupo I) para selecionar 10 locos que pudessem ser utilizados rotineiramente na caracterização de cultivares. Os locos selecionados foram utilizados para avaliar 32 variedades (grupo II). No grupo I obteve-se, em média, 3,36 alelos por loco, de um total de 232 alelos. Foram obtidos, em média, 5,1 alelos por loco para o grupo II, de um total de 51 alelos para as 32 variedades avaliadas. A média dos valores de distâncias foram 0,67 e 0,82, para grupo I e grupo II, respectivamente. Os 10 locos selecionados e avaliados no grupo II foram mais efetivos na discriminação dos materiais e, assim, recomendados para a caracterização dessas cultivares de soja brasileiras para alimentação humana.*

Palavras-chave: *Glycine max*, SSR, proteção de cultivares, germoplasma.

REFERENCES

Abe J, Xu DH, Suzuki Y, Kanazawa A and Shimamoto Y (2003). Soybean germplasm pools in Asia revealed by nuclear SSRs**. Theoretical and Applied Genetics 106**: 445-453.

Akkaya MS, Bhagwat AA and Cregan PB (1992) Length polymorphisms of simple sequence repeat DNA in soybean. **Genetics 132**: 1131-1139.

Coelho ASG (2000) **Dboot**: Avaliação dos erros associados a estimativas de distâncias/similaridades genéticas através do procedimento de bootstrap com número variável de marcadores (software). Laboratório de Genética Vegetal. Goiânia: UFG, Instituto de Ciências Biológicas.

Cregan PB, Jarvik T, Bush AL, Shoemaker RC, Lark KG and Kahler AL (1999) An integrated genetic linkage map of the soybean genome. **Crop Science 39**: 1464-1490.

Creste S, Tulmann Neto A and Figueira A (2001) Detection of single sequence repeat polymorphisms in denaturing polyacrilamide sequencing gels by silver staining. **Plant Molecular Biology Reporter 19**: 299-306.

Diwan N and Cregan PB (1997) Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. **Theoretical and Applied Genetics 95**: 723-733.

Doldi ML, Vollmann J and Lelley T. (1997) Genetic diversity in soybean as determined by RAPD and microsatellite analysis, **Plant Breeding 116**: 331-335.

Embrapa (2003) **Tecnologias de produção de soja – região central do Brasil 2004**. Embrapa, Epamig e Fundação Triângulo, Londrina, 237p.

Hoisington D, Khairallah M and González-De-León D (1994) **Laboratory Protocols: CIMMYT applied molecular genetics laboratory**, 2nd ed., CIMMYT, Mexico, p.51.

Mapa (2005). Ministério da Agricultura, Pecuária e Abastecimento. Available at http://www.agricultura.gov.br/pls/portal/docs/PAGE/MAPA/BALANCA_COMERCIAL/NOTA%20DEZEMBRO%202004.PDF. Assessed in 26 August 2005.

Miller MP (1997) **TFPGA – tools for population genetic analyses.** v.1.3., Available at http://www.marksgeneticsoftware.net/tfpga.htm. Assessed in 01 October 2007.

Morgante M and Olivieri AM (1994) Genetic mapping and variability of seven soybean simple sequence repeat loci. **Genome 37**: 763-769.

Narvel JM, Fehr WR, Chu WS, Grant D and Shoemaker RC (2000) Simple sequence repeat diversity among soybean plant introductions and elite genotypes. **Crop Science 40**: 1452-1458.

Priolli RHG, Mendes Jr CT, Arantes NE and Contel EPB (2002) Characterization of Brazilian soybean cultivars using microsatellite markers. **Genetics and Molecular Biology 25**: 185-193.

Rohlf FJ (1997) **NTSYSpc – numerical taxonomy and multivariate analysis system v.2.1**. (CD-ROM) Exerter Publication, Setauket.

Rongwen J, Akkaya MS, Lavi U and Cregan PB (1995) The use of microsatellite DNA markers for soybean genotype identification. **Theoretical and Applied Genetics 19**: 43-48.

Shoemaker RC, Polzin K, Labate J, Specht J, Brummer EC, Olson T, Young N, Concibido V, Wilcox J, Tamulonis JP, Kochert G and Boermaf HR (1996) Genome Duplication in Soybean ***(Glycine* subgenus *soja).* Genetics 144:** 329-338.

Smith JSC, Chin ECL, Shu H, Smith OS, Wall SJ, Senior ML, Mitchell SE, Kresovich S and Ziegle J (1997) An evaluation of the utility of SSR loci as molecular markers in maize *(Zea mays* L.): comparisons with data from RFLPs and pedigree. **Theoretical and Applied Genetics 95**: 163-173.

Wright S (1978) **Evolution and genetics of population Variability within and among natural populations**. vol. 4, University of Chicago Press, Chicago.