E

BMT-TWA/Soybean/1/2
ORIGINAL: English
DATE: September 11, 2002

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS

GENEVA

AD HOC CROP SUBGROUP ON MOLECULAR TECHNIQUES FOR SOYBEAN

First Session
Rio de Janeiro (Brazil), September 28, 2002

SUMMARY OF THE SSR SOYBEAN RESEARCH FOR DUS TESTING DEVELOPED BY THE MOLECULAR MARKERS LABORATORY AT THE FORMER *INSTITUTO NACIONAL DE SEMILLAS* – INASE – (ARGENTINA)

Document prepared by experts from Argentina

SUMMARY

Use of molecular markers as descriptors for plant variety protection purposes, has been proposed, but few examples of real case applications to national registering systems were reported. This work summarizes the application of microsatellite markers to characterize and differentiate 271 soybean varieties and landraces of commercial use in Argentina, Bolivia and China, examples of large and small size producer countries with limited genetic variability (Argentina and Bolivia) and a large size producer, which is center of origin of this crop (China). Distinctness was assessed, by obtaining a unique genotypic document for varieties using 20-33 microsatellite markers. Uniformity was tested, by analyzing the relative degree of heterogeneity for microsatellite alleles. By a more precise analysis of seven selected varieties, it was shown that tolerance values used for morphological descriptors have to be modified if molecular markers are going to be used for protection purposes. Stability was checked by comparing microsatellite patterns through a four-year period of a group of seven Argentine varieties of prolonged commercial use, with 32 microsatellite markers. Detailed analysis suggests that the observed “instability” may be attributed to high mutation rate of the microsatellite loci, a mixture of seeds, cross-pollination or alleles that have not been detected before.

INTRODUCTION

Due to the importance of soybean in the context of the international seed and grains trade, mainly for Argentina which produces 26 millions of ton a year, the Molecular Markers Lab from ex INASE, decided to use this crop as a model to study the possible application of molecular markers in DUS testing.

In this work, we report the usefulness of microsatellite as a tool to assess:

(a) distinctness, by obtaining a unique genotypic document for varieties from Argentina, Bolivia and China;

(b) uniformity, by analyzing the relative degree of heterogeneity in different varieties and landraces and by analyzing a subset of commercial varieties; and c) stability, by comparing microsatellite patterns through time in a group of Argentine varieties of prolonged commercial use.

MATERIALS AND METHODS

DNA pools belonging to two hundred and seventy one different soybean varieties were screened for microsatellite loci. One hundred and three DNA samples were prepared from Chinese landrace cultivars collected by Professor Gai (Soybean Research Institute, Nanjing Agricultural University. National Center of Soybean Improvement, Ministry of Agriculture); one hundred and fifty six were obtained form the Argentine National Seed Register at ex INASE; twelve Bolivian varieties were obtained form SEMEXA Bolivia; and another one, Williams, was used as a reference.

DNA for distinctness and stability analysis was prepared after pooling plant material belonging to five grown seedlings. DNA for uniformity was prepared from unique plants. All DNA’s were extracted essentially as described by Saghai-Maroof, et al (1984).

Thirty-three representative soybean microsatellites containing "ATT" repeated sequence motifs were selected according to their distribution in the genetic map and reported polymorphisms. PCRs were performed in a 20 µl total volume using a PTC-100 DNA thermocycler (MJ Research Inc., USA), essentially as previously described by Giancola (1998). After PCR, amplification products were resolved by standard sequencing electrophoresis on denaturing polyacrylamide gels following Sambrook, et al (1989) protocols. Bands were revealed using a silver nitrate staining kit according to manufacturers' instructions (Promega Biotech, USA).

Genetic diversity was measured by evaluating the allele number per locus, polymorphic index content (PIC) and similarity values. PIC was computed as $PIC = 1 - \sum_{i}^{n} p_i^2$ (Anderson et al, 1993). Were p_i is the frequency of the i^{th} allele for that locus. This measure shows the polymorphism revealed for each locus.

Each microsatellite allele band was scored as either present (1) or absent (0) across all genotypes to create a binary matrix. Similarity values were calculated analyzing the binary matrix with the Jaccard association coefficient (Sneath and Sokal, 1973) using the NTSyS pc2.0 program.

The minimal number of markers needed to identify a set of genotypes (1) and the number of genotypes that can be identified (2), given a PIC value, were calculated using the following formula from Brown et al, 1996:

$n=\ln X/[\ln(1/(1-D))]$ (1)
$X=[1/(1-D)]^n$ (2)

where X is the number of unique genotypes, D is the genetic diversity over loci, and n is the number of loci analyzed.

Heterogeneity was computed as $H=(1/Nm)*\sum\sum hij$ (Hedrick, 2000), where hij represents the i^{th} variety and the j^{th} locus, N is the number of varieties, and m is the number of loci analyzed. This value, hij, is computed as 1 for heterogeneous varieties and 0 for homogeneous ones.

An allele was considered rare if it only appears once in the group.

RESULTS AND DISCUSSION

Distinctness: generation of a unique genotypic document for varieties and landraces

In previous works, it was demonstrated the feasibility of generating efficient fingerprints of commercial soybean varieties and landraces using microsatellite markers (Giancola, 1998, Vicario, 2000). Two hundred and fifty seven commercial varieties and landraces were surveyed with 20 to 33 representative SSR markers, selected on the basis of their specific position in soybean genetic map and resolution of banding patterns. Between 1 and 2 loci, for each linkage group were selected in order to cover up 20 linkage groups to have good genome coverage.

Table 1 shows all variants analyzed for the 3 groups of varieties tested.

Table 1

	GROUP		
Variable	Chinese	Argentine	Bolivian
Number of varieties analyzed	103	142	12
Number of markers analyzed	33	30	20
Average allele number per locus	11.7	4.6	2.7
Rare alleles	67 (17.3%)	7 (5%)	9 (16.6%)
PIC	0.73	0.58	0.37
Average similarity values	0.137	0.259	0.402
Similarity value range	0 - 0.89	0.05-0.75 (0.9*)	0.20 - 0.75 (1*)
Minimal number of markers	3.56	5.71	5.28
Genotypes that can be identified	10^{18}	10^{11}	$1.2x10^{4}$

*Values for a pair of close related genotypes

For Chinese and Argentine varieties, it was possible to obtain a unique genotypic document. But it was not possible to generate a unique fingerprint for all Bolivian varieties, even when the minimal number of markers theoretically needed to identify all genotypes (5.28) was lower than the number of markers used (20). Only when 28 SSR were used, all Bolivian genotypes could be identified. In the case of Chinese varieties, 5 SSR were enough to identify 103 varieties and in the case of Argentine varieties 27 were needed to distinguish 142.

The number of alleles, average PIC values and similarity values are reflecting the variability within each group. PIC also includes the frequency of each allele, so it better explains how variable or diverse the group is. Genetic similarity estimates are affected by a variety of factors, like the number and distribution of markers in the genome (genome coverage) and the nature of evolutionary mechanisms underlying the variation measured (Powell et al., 1996). The markers used in this work are mapped so they were selected form different linkage groups and rather apart one form the other. They are not strongly linked to any phenotypic trait and are considered to be neutral.

This analysis shows that the number of SSR to be use in a distinguibility test has to be carefully chosen to assure a good genotype differentiation, and to avoid equal or similar fingerprints of close related genotypes.

Uniformity: analysis of the relative degree of heterogeneity in different varieties and landraces

In order to analyze the feasibility of adapting microsatellite markers to the present PVP system, the application to uniformity and stability testing has to be analyzed as rigorously as for distinctness. Heterogeneity of the genetic material was assessed scoring the number of different alleles per locus for each variety. In Table 2, a general uniformity test is described.

Table 2

	CHINESE	ARGENTINE	BOLIVIAN
Heterogeneity	0.092	0.022	0.079
% of heterogeneous varieties	56%	35%	75%
Maximum number of alleles per variety	3	2	2

As expected for landrace material, almost half of the Chinese cultivars showed heterogeneous patterns of bands (i.e. 58 out of the 103 varieties showed more than one microsatellite allele). The totality of the 33 microsatellite markers analyzed showed heterogeneity in at least two varieties (summing up a total of 306 heterogeneous patterns - out of 3399 -) indicating that these landraces are formed by a population of genotypes. Argentine varieties showed heterogeneity in 50 varieties (35% of the total) represented by 91 heterogeneous patterns out of 4118 for 29 microsatellites. Bolivian varieties showed 9 out of 12 heterogeneous varieties (75%), with 19 out of 240 heterogeneous patterns for 20 SSR.

For a more detailed homogeneity test, 7 Argentine varieties were analyzed using 15 not linked high PIC value SRR. Table 3 shows a more detailed uniformity results summery.

Table 3

Varieties	01	02	03	04	05	06	07
Off type plants	14/223	60/224	4/217	2/81	2/50	10/93	18/219
		VARIETIES					
Similarity values		02		03			
Range		1 - 0.409		1 - 0.66			
Average		0.803		0.935			

All varieties analyzed showed to be morphologically uniform at the field level. However, SSR analysis revealed differences. According to ex INASE's regulations, just only one seed in 1000 is allowed to be off type (for basic seeds and for morphological descriptors), but this value rises to 4-5 for first or second multiplication seeds (Resolución N°214/99, INASE). All allelic difference ratios were higher than these values when analyzed with SSR markers. The less variable variety is the one encoded as 03. Even for this case, it showed an allelic ratio that is 3.7 times higher than the stipulated for phenotypic descriptors.

When variability of specific loci was analyzed, SSR 226 was the most variable (16 plants off type from 81) and SSR 231 was the less variable (1 plant off type out of 87 analyzed).

Thus, these results clearly show that values of off type tolerance have to be reestablished if microsatellite markers are implemented to assess uniformity and that the selection of microsatellite loci has to be carefully designed.

In 1998 Giancola analyzed 100 commercial varieties and proposed a maximum similarity value of 0.8 to distinguish close varieties using morphological and molecular (SSR) descriptors. Our study showed that in average similarity values for a given variety is around 0.8. Therefore, a value close to 0.8 could be considered as a distinguibility/homogeneity

threshold. In the present uniformity test, variety 02, which was the most heterogeneous, gave an average similarity value of 0.803, and variety 03, the most homogeneous gave a value of 0.935. Though, genotypes with similarity values larger than 0.8 (and up to 1) could be considered the same variety; genotypes with similarity values up to 0.8, could be considered as different varieties.

Stability: SSR patterns change through time

To study the SSR patterns change through time, 7 commercial varieties were surveyed with 32 SSR during 4 years. Table 4 shows a summery of the results of this stability test. Changing patterns are detached in italic and bold.

Table 4

SSR	414	9	534	30	177	45	114	253	577
Variety									
A-96	295	214	*260*	164	*120*	140	109	*147+ 135*	119
A-97	295	214	*260*	164	*120*	140	109	*147*	119
A-98	295	214	*260*	164	*120+ 113*	140	109	*147*	119
A-99	295	214	*260+ 266*	164	*113*	140	109	*147*	119
B-96	*301*	163	260	164	113	*134*	*109*	147	*119*
B-97	*301*	163	260	164	113	*134*	*109*	147	*119*
B-98	*301*	163	260	164	113	*134*	*109*	147	*119*
B-99	*301+ 259*	163	260	164	113	*134+ 143*	*109+ 97*	147	*119+ 115*
C-96	301	*214*	260	161	113	134	109	153	119
C-97	301	*214+ 163*	260	161	113	134	109	153	119
C-98	301	*214+ 163*	260	161	113	134	109	153	119
C-99	301	*163*	260	161	113	134	109	153	119
F-96	301	163	260	*164*	113	134	109	147	119
F-97	301	dp	260	*dp*	113	dp	109	147	119
F-98	301	136	260	*152*	113	134	109	147	119

For most of the markers, there was no significant variation in the allelic haplotype as a whole. In spite of this, 4 out of 7 varieties showed "new" alleles in 1, 3 or 4 of the SSR loci analyzed. Variations in SSR patterns could be due to several causes like seed mix and/or wrong purity maintenance, SSR mutation (slippage) or cross-pollination (which is estimated to be 0.03% to 2.5% for soybean (Carlson and Lerseten, 1987, Sediyama et al, 1999). Another possibility is that, as each sample are formed by a pool of five plants, the "new" allele found could be one already present in the variety and not detected before in any of the years studied.

CONCLUSIONS

- The analysis allowed us to obtain a fingerprint (unique genotypic identity document) for almost all analyzed varieties.
- A theoretical number of 4-6 SSR are sufficient for the characterization of soybean genotypes. Though, the experience shows that very related genotypes would need more than 20 to efficiently distinguish all genotypes.
- This analysis shows that the number of SSR to be use in a distinguibility test has to be carefully chosen to assure a good genotype differentiation, and to avoid equal or similar fingerprints of close related genotypes.
- The similarity values found using SSR could allow us to think in a possible threshold (of 0.8 or a value close to it), above which a variety would be consider uniform.
- If SSR or other DNA markers are going to be used for DUS testing, the nowadays off type plants number allowed would need to be re-established.
- Selection of markers for uniformity and stability testing has to take into account two issues. First, differences in neutral markers may not be as significant for application purposes as trait characteristics that assure the farmer homogeneous agronomic properties of the seeds they purchase. Second, the dynamics of pattern changing of certain microsatellite loci may be higher than it is for morphological descriptors, suggesting differences in genomic backgrounds or instabilities that are not real.
- These results show that these markers could be effectively applied for germplasm classifications and its intellectual protection. The analysis of a larger number of SSR and representative varieties will let us establish which and how many SSR are more adequate for variety registration.

REFERENCES

- Anderson J. A., Churchill G. A., Autrique J. E., Tanksley S. D., Sorrells M. E. (1993). Optimising parental selection for genetic linkage maps. Genome 36:181-186.
- Brown S. M., Hopkins M. S., Mitchell S. E., Senior M. L., Wang T. Y., Duncan R. R., Gonzalez-Candelas F., Kresovich S. (1996). Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum [*Sorghum bicolor* (L.) Moench]. Theor. Appl. Genet. 93: 190-198.
- Carlson J. B. and Lersten N. R., (1987). Soybeans: Improvement, Production and Uses. Chapter 3: Vegetative Morphology. J. R. Wilcox, editor. 865p.
- Cregan P., (1999), http://129.186.26.94/SSR.html.
- Diwan N. and Cregan P., (1997). Automated Sizing of Fluorescent-labeled Simple Sequence Repeat (SSR) Markers to Assay Genetic Variation in Soybean. TAG: 95: 723-733.

- Giancola S. M., 1998. Caracterización y diferenciación de germoplasma de soja (*Glycyne max L. Merr*.) mediante marcadores moleculares. [Characterization and differentiation of soybean germplasm using molecular markers] Tesis Doctoral. Facultad de Ciencias Exactas y Naturales, UBA, [Ph.D. Thesis, Faculty of Sciences, University of Buenos Aires], Buenos Aires.
- Hedrick P W (2000). Genetics of populations. Jones and Bartlett publishers, Sudbury, Massachusetts. Segunda edición, 2000. Arizona State University.
- Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tiggey S, Rafalsky A. (1996). The comparision of RFLP, RAPD, AFLP and SSR (microsatellites). Markers for germoplasm analysis. Mol. Breed. 2: 225-238.
- Resolución N°214/99, INASE.
- Saghai-Maroof M. A, Soliman K. M, Jorgensen R. A, Allard R. W. (1981). Ribosomal DNA spacer length polymorphism in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc. Natl. Acad. Sci. USA 81: 8014-8018.
- Sambrook J, Fritsch E. F Maniatis, T (1989). Molecular cloning. A laboratory manual. 2° Ed. Cold Spring Harbor Laboratory Press, New York.
- Sediyama T, Teixeira R, Reis, M. S (1999). Melhoramiento da soja. In: Melhoramiento de espécies cultivadas. Aluzío Borém Editor. P. 487-533.
- Sneath P. H. A. and Sokal R. R., (1973). Numerical Taxomony. The Principles and Practice of Numerical Classification. Freeman. San Francisco, CA, XV, 573p.
- Vicario, A. L, Langan M, Echaide, M y Hopp, E (2000). Quantification and Comparison of Genetic Diversity of Chinese and Argentine Soybeans (*Glycine Max* L. Merr). Plant and Animal Genome Conference VIII. January, 2000. San Diego, United States.

[End of document]