E

BMT/7/13
ORIGINAL: English
DATE: November 12, 2001

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
GENEVA

WORKING GROUP ON BIOCHEMICAL AND MOLECULAR TECHNIQUES AND DNA-PROFILING IN PARTICULAR

Seventh Session
Hanover, Germany, November 21 to 23, 2001

APPLICATION OF MICROSATELLITE MARKERS FOR THE ASSESSMENT OF DISTINCTNESS, UNIFORMITY AND STABILITY (DUS TESTING) OF COMMERCIAL SOYBEAN VARIETIES

prepared by experts from Argentina

Application of microsatellite markers for the assessment of distinctness, uniformity and stability (DUS testing) of commercial soybean varieties

Vicario, A[1]; Giancola, S. Echaide, M[1]; *Hopp, E[2]

[1]Laboratorio de Marcadores Moleculares, Instituto Nacional de Semillas (INASE), Av. Paseo Colón 922, 4° piso, CP 1063, Buenos Aires, Argentina.
[2]Instituto de Biotecnología, CICVyA-CNIA INTA and Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, CC 25, CP 1712 Castelar, Argentina.
[3]Present address: Institut National de la Recherche Agronomique (INRA). Station de Génétique et d'Amélioration des Plantes. Centre de Recherche de Versailles, route de Saint-Cyr, F-78026 Versailles Cedex. France.
Comunicating author: Hopp, H. Esteban. E-mail: ehopp@inta.gov.ar. Telephone: 54 11 4621 1447/1676. Fax: 54 11 4481 2975.

I. Introduction

Although soybean [*Glycine max* (L.) Merr.] is a very old crop (dates from 4500 years ago), it is a relatively new crop for the three main producing countries (USA, Brazil and Argentina). China seems to be the origin of most modern varieties cultivated in the West. Nowadays, Argentine is the third most important world producer of soybean with an estimated production rate of 26.0 million ton/year (2000-2001). Soybean occupies 10.15 million ha (Estimaciones Agrícolas, 2001), representing about 27.5% of the total cultivated land in Argentina. It is cultivated from the Southern part of the Province of Buenos Aires (23° southern latitude) to the most northern provinces (Salta and Jujuy at 38° southern latitude). To cope with this demand, more than 350 different varieties belonging to maturity groups I to IX were registered at the National Seed Institute of Argentina (INASE) during the last 19 years.

INASE follows the international criteria of the Plant Variety Protection (PVP) Office for registration, taking into account phenotypic traits to distinguish one variety from another. However, as the number of varieties being registered continues to increase, it becomes increasingly difficult to achieve this with only phenotypic traits.

The ability of molecular markers (particularly microsatellites) to efficiently distinguish between closely related varieties has been reported (Powell *et al.* 1996; Giancola, 1998; Giancola *et al.*, 2001).

Consideration of their future use in complementing morphological descriptors in variety registration and PVP systems is in progress. Indeed, the PVP Office of the USDA, the Agriculture Marketing Service, now accepts microsatellite allelic profiles as supporting evidence for the uniqueness of a new cultivar (Diwan and Cregan, 1997). However, while the ability of molecular markers for differentiation purposes was widely reported in soybean and in other crop systems, not so much work was published regarding their uniformity and stability, which are as important as distinctness in variety registration systems.

In this work, we report the applicability (pros and cons) of microsatellite markers to assess:

a) distinctness, by assessing the separation and fingerprinting potential of SSR applied to 100 soybean varieties from Argentina;

b) uniformity, by analyzing the relative degree of heterogeneity for the microsatellite alleles in different varieties;

c) stability, by comparing microsatellite patterns over time of a group of varieties in prolonged commercial use.

Part of this work was recently presented (and widely discussed in a technical committee) at the last International Seed Testing Association Congress (Vicario *et al.*, 2001).

II. Materials and Methods

One hundred and three varieties registered at the Registro Nacional de Semillas (Argentine National Seed Register) belonging to the most commonly cultivated varieties (in terms of total acreage) in Argentina during the last 19 years, were selected to represent the diversity of commercially cultivated soybean genotypes. All the seeds used for DUS testing in this work were certified and provided by the original registering seed company. Cultivar Williams (of US origin, but widely used for breeding in Argentina) was used as reference standard to name and compare SSR patterns.

DNA for uniformity analysis was prepared from individual plants. DNA for distinctness and stability analysis was prepared after pooling plant material belonging to at least five grown seedlings. The cultivars used in this study are homozygous lines, but five plants per cultivar were pooled for DNA extraction to avoid the possibility of selecting a single contaminating off-type seed. Gilbert *et al.* (1999) also recommended the use of pools from five plants to assess genetic variability with DNA markers in large plant germplasm collections. DNA was isolated as previously described (Saghai-Maroof *et al.*, 1984; Giancola *et al.*, 2001; Vicario *et al.*, 2001).

Thirty-five representative soybean microsatellites containing “ATT” repeated sequence motifs were selected according to their distribution in the genetic map and reported polymorphisms in US soybean. The SSR primers sequence information is available in http://129.186.26.94/SSR.html.

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 *et al.* (2001). 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 manufacturer’s instructions (Promega Biotech, USA).

Each microsatellite allele band was scored as either present (1) or absent (0) across all genotypes to create a binary matrix. The Basic Data Matrix (BDM) was analyzed with the NTSYS 1.8 program using the Jaccard association coefficient (Sneath and Sokal, 1973).

Genetic diversity was measured by evaluating the allele number per locus, similarity values and polymorphic index content (PIC).

PIC was computed as $PIC = 1 - \sum_{i}^{n} p_i^2$ (Anderson *et al.*, 1993). Were p_i is the frequency of the ith allele for that locus. This measure shows the polymorphism revealed for each locus and its average is similar to the Weir (1996) genetic diversity index.

For uniformity testing of 6 Argentine varieties 15 diagnostic SSR were used. The allele number per locus, number of profiles and the number of off-type plants for each SSR were calculated. A total number of off-type plants/locus/variety was also calculated. This value shows the variability within variety for all SSR analyzed. The variability revealed per SSR through varieties was also calculated.

III. Results and Discussion

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

In a previous work, the feasibility of generating efficient differentiation of commercial soybean varieties by fingerprinting them using microsatellite markers was demonstrated (Akkaya *et al.*, 1992; Ronwen *et al.*, 1995; Diwan and Cregan, 1997; Giancola, 1998; Giancola *et al.*, 2001). One hundred commercial varieties in Argentina were surveyed with thirty representative SSR markers, selected on the basis of their specific position in the soybean genetic map and resolution of banding patterns. Between 1 and 2 loci for each linkage group were selected in order to cover up to 20 linkage groups, and provide fairly good genome coverage.

Applying the formula of Brown *et al.* (1996), 5-6 SSR would be sufficient to separate these 100 varieties (the exact calculated value is 5.025 SSR markers). Thus, the number of unique genotypes that can be generated with just ten markers with this average PIC is 9,537 according to the same authors. So, with a rather limited number of loci (30) it was possible to uniquely differentiate and fingerprint each of the varieties.

Interestingly, a quantitative estimation of the so called “minimum genetic distances” calculated on the basis of similarity indexes using morphological descriptors coincided with that calculated with SSR (0.8) and both types of descriptors were shown to complement each other when combined in a common matrix (Giancola, 1998).

In conclusion, the analysis allowed us to separate and obtain a fingerprint or “unique genotypic identity document” for each one of the 100 analyzed varieties. Thus, in soybean, these markers can be effectively applied for germplasm classification and could be used for plant variety protection. This result also fully confirms that microsatellite markers are a valuable tool to accomplish distinctness in cultivars from very diverse origins, levels of genetic diversity and breeding methodology and can be easily used to complement the present PVP system of identification.

Uniformity and stability: analysis of the relative degree of heterogeneity and conservation for the microsatellite alleles in different varieties and landraces

In order to analyze the feasibility of adapting microsatellite markers to the present PVP system, their applicability for uniformity and stability testing must be analyzed as rigorously

as for distinctness. Heterogeneity of the genetic material under analysis can be easily assessed by scoring the number of different alleles for each microsatellite locus after analysis of DNA pooled from different seeds belonging to the same cultivar (see Table 1). Argentine varieties showed heterogeneity in 27 varieties (27% of the total), represented by 51 heterogeneous patterns out of 3,000, for 30 microsatellites.

Table 2 shows some more detailed uniformity results for 6 commercial varieties. All varieties analyzed were shown to be morphologically uniform at the field level. However, SSR analysis revealed differences. According to INASE's regulations, just one seed in 1,000 is allowed to be an 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 (see Table 2).

Table 3 shows the results of the stability testing of seven varieties of prolonged commercial use, during four years, using 32 polymorphic SSR. For most of the markers, there was no significant variation in the haplotype allelic pattern 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 (highlighted in italics and bold, in Table 3). Only 9 out of 32 SSR analyzed presented variation among varieties. All varieties except one (B, SSR 414 and 045) presented an existing new allele for 32 SSR analyzed. These variations could be due to several causes ranging from SSR mutation, to seed mixture and/or to wrong purity maintenance caused by cross-pollination (which is estimated to be 0.5% to 1% for soybean, Carlson and Lerseten, 1987). An unequal cross-over could easily explain the generation of 2 new alleles, one larger and one smaller than the standard one (considered to be the allele found in 1996 seeds). In cases A (SSR 177 and 253), B (SSR 414, 114 and 577), C and F a smaller variant of a putative unequal cross-over is observed, while for A (SSR 534) and B (SSR 045) the larger is observed. These last two cases could also be generated by DNA polymerization slippage errors. The DNA polymerase may have "slipped" one repetition unit due to the high number of repetitions and added nucleotides. If this is the case, the resulting "instability" may be attributed more to a high mutation rate of the marker than to mixing or contamination of seed material. In this context, the high mutation rate does not seem to be locus specific, since the variations observed were scattered through different markers.

In conclusion, selection of SSR markers for DUS testing must take into account two important 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 mutation rate of certain microsatellite loci may be higher than expected, evidencing differences between genomic backgrounds or instabilities that are not real or just limited to the SSR locus.

Our results clearly show that values of off-type tolerance need to be reexamined if microsatellite markers are implemented to assess both, uniformity and stability. In addition, selection of microsatellite loci must be carefully designed to discard highly hypervariable loci. The results of this work clearly suggest that the analysis of a large number of SSR using representative varieties should be a prerequisite to establish which specific SSR are more stable and thus more appropriate for soybean variety registration purposes.

Table 1: Brief description of the results obtained after analyzing 100 varieties using 30 SSR

The allele number per locus, PIC values, their average and similarity values are shown in this table. The first column shows all SSR used in the analysis (ATT repeats and obtained from "Soybase").

Argentine varieties	Allele number per locus	PIC values
SSR		
2	4	0.41
5	8	0.7
9	5	0.71
168	4	0.56
231	5	0.68
249	4	0.46
534	7	0.72
357	3	0.49
353	3	0.58
324	4	0.67
259	5	0.66
253	4	0.56
577	3	0.62
177	4	0.63
175	4	0.62
45	5	0.54
30	6	0.75
414	6	0.8
173	5	0.62
294	4	0.56
42	5	0.62
172	4	0.5
100	4	0.43
114	4	0.71
226	4	0.65
358	2	0.4
46	4	0.3
70	5	0.56
1	4	0.36
38	3	0.44
Average	4.40 (1.22)	0.57 (0.13)
Similarity values	Max: 0.80 Min: 0.30 Average: 0.262 (0.067)	

Table 2: Uniformity testing of 6 varieties for 15 diagnostic SSR patterns

Representative samples of seeds belonging to six selected Argentine soybean varieties (encoded 01-06 and represented in row 1 by their code name) were screened using 15 SSR of high PIC (identified in column 1). The table shows A: allele number/SSR; B: number of profiles/SSR; C: number of off-type plants (allelic ratio) for each variety. Last row shows the total allelic ratio per variety.

	Varieties																	
	01			**02**			**03**			**04**			**05**			**06**		
SSR	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C
414	2	2	1/15	2	2	1/2	1	1	0/15	-	-	-	-	-	-	-	-	-
231	1	1	0/15	1	1	0/15	1	1	0/15	1	1	0/11	2	2	1/5	1	1	0/11
175	2	2	1/15	2	2	1/2	1	1	0/15	1	1	0/9	1	1	0/5	3	3	4/8
5	1	1	0/10	2	2	1/2	1	1	0/13	-	-	-	-	-	-	-	-	-
324	2	2	1/15	1	1	0/15	1	1	0/15	1	1	0/9	1	1	0/8	1	1	0/8
534	2	2	1/15	1	1	0/15	3	3	2/15	-	-	-	-	-	-	-	-	-
9	1	1	0/14	2	2	1/15	1	1	0/13	-	-	-	-	-	-	-	-	-
177	3	3	2/15	1	1	0/15	1	1	0/15	1	1	0/14	1	1	0/8	2	2	1/15
259	2	3	5/15	2	2	1/2	2	2	1/15	1	1	0/14	1	1	0/8	1	1	0/14
173	1	1	0/15	2	2	3/15	1	1	0/15	2	2	1/8	1	1	0/4	1	1	0/9
226	1	1	1/15	2	2	1/2	2	2	1/13	2	2	1/7	2	2	1/7	3	3	5/10
30	2	2	1/15	2	2	1/2	1	1	0/14	-	-	-	-	-	-	1	1	0/8
42	2	2	5/15	1	1	0/15	1	1	0/15	-	-	-	-	-	-	-	-	-
168	1	1	0/15	2	2	1/2	1	1	0/14	-	-	-	-	-	-	-	-	-
294	1	1	0/15	2	2	1/2	1	1	0/15	1	1	0/9	1	1	0/5	1	1	0/10
Total	**18/219**			**60/225**			**4/217**			**2/81**			**2/50**			**10/93**		

Table 3: Stability testing of seven varieties during four years using 32 diagnostic SSR

Representative samples of seeds belonging to seven selected Argentine soybean varieties of prolonged commercial use (encoded A-G, and represented in the first column by their code name followed by the last two numbers of the year of the collected seed sample) were screened using 32 SSR of high PIC (identified in the first row). Only the 9 SSR that showed variations are included in this table. Specific alleles showing variations in the analyzed period of time (represented by their molecular size in base pairs) are highlighted using bold and italic fonts. The 23 SSR that did not show any variant were: 175, 231, 005, 324, 042, 259, 173, 226, 294, 168, 172, 358, 373, 307, 156, 197, 367, 184, 147, 249, 357, 353, and 100 (see Matherials and Methods for reference number and source). dp: lost data.

	SSR								
Variety	**414**	**9**	**534**	**30**	**177**	**45**	**114**	**253**	**577**
A-96	295	dp	***260***	164	***120***	140	109	***147+ 135***	119
A-97	295	dp	***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	***dp***	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
D-96	295	214	266	161	108	134	121	147	115
D-97	295	214	266	161	108	134	121	147	115
D-98	295	214	266	161	108	134	121	147	Dp
D-99	295	214	266	161	108	134	121	147	115
E-96	295	214	260	152	113	140	97	135	115
E-97	295	214	260	152	113	140	97	135	115
E-98	295	214	260	152	113	140	97	135	115
E-99	295	214	260	152	113	140	97	135	115
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
G-96	dp	181	266	152	120	134	82	147	119
G-97	301	181	266	152	120	134	dp	147	119
G-98	301	181	266	152	120	134	82	147	119
G-99	301	181	266	152	120	134	82	147	119

IV. Acknowledgements

The authors thank Dr. Perry Cregan (USDA-ARS) who kindly provided primers sequence information before it was actually published. This project was partially supported by a grant PID 024 from ANPCyT, Secretaría de Ciencia y Técnica, Argentina and is part of an INTA-INASE agreement. Lic. A. Vicario holds a fellowship from the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, Argentina) and Dr. H. E. Hopp is a career member of the Consejo de Investigaciones Científicas de la Provincia de Buenos Aires (CIC).

V. References

- Akkaya M, Bhagwat A, Cregan P (1992). Length polimorphisms of simple sequence repeats DNA in soybean. Genetics 132: 1131-1139.
- 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.
- Barrett B. A., Kidwell K. K. (1998). AFLP-based genetic diversity assessment among wheat cultivars from the Pacific Northwest. Crop Sci. 38: 1261-1271.
- 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.
- Giancola S., Marcucci Poltri S., Lacaze P., Hopp H. E. (2001). Feasibility of integration of molecular markers and morphological descriptors in a real case study of a plant variety protection system for soybean. *Euphytica* 2001 (in press).
- Gilbert J., Lewis R. V., Wilkinson M. J., Caligari P. D. S. (1999). Developing an appropriate strategy to assess genetic variability in plant germplasm collections. Theor. Appl. Genet. 98: 1125-1131.
- Goldstein D. B., Linares A. R., Cavalli-Sforza L. L. and Feldman M. W., (1995). An Evaluation of Genetic Distances for Use With Microsatellites Loci. Genetics 139: 463-471.
- González, H. R. (1999). INASE: National Seed Institute, Yearbook, pp 84, Estilos Gráficos S.A. Publisher, Buenos Aires, Argentina.
- Morgante M., Rafalski A., Biddle P., Tingey S. and Olivieri A. M., (1994). Genetic Mapping and Variability of Seven Soybean Simple Sequence Repeats Loci. Genome V: 37.

- 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.
- Rohlf F. J. NTSYS-pc (1989) .Numerical Taxonomy and Multivariate Analysis System. Vs 2.0 Exeter Publ.LTD, Setauket,NY .
- 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.
- Sambrook J., Fritsch E. F., Maniatis T. (1989). Molecular cloning. A laboratory manual. 2° edition. Cols Spring Harbor. Laboratory press, New York.
- 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., Giancola S. M., Echaide M., Hopp H. E. (2001). Application of microsatellites for the assessment of distinctness, uniformity and stability (DUS testing) of commercial soybean varieties and landraces. 26th International Seed Testing Association Congress. Oral presentation. Angers, France. June 2001.
- UPOV/TG/80/3, 1983. Guidelines for the Conduct of Test for Distinctness, Homogeneity and Stability, Soybean (*Glycine max* (L) Merrill).
- USDA, (2000) http://www.fas.usda.gov/oilseeds/circular/1997/97-08/aug97opd2.htm
- Vallone S. D. and Giorda L. M., (1997). Enfermedades de la Soja en Argentina [Soybean diseases in Argentina], pp 72, Ed. INTA Centro Regional Córdoba, Córdoba, Argentina.

[End of document]