E

BMT/10/15
ORIGINAL: English
DATE: November 9, 2006

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS

GENEVA

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

Tenth Session
Seoul, November 21 to 23, 2006

DNA-BASED IDENTIFICATION SYSTEM FOR SOYBEAN

Document prepared by an expert from Argentina

DNA-BASED IDENTIFICATION SYSTEM FOR SOYBEAN

Loray Maria[1]; Vicario, Ana[1]; Boca, Teresa[2]; Hopp, Esteban[3]
[1]Laboratorio de Marcadores Moleculares, Instituto Nacional de Semillas (INASE),
Av Paseo Colón 922, 4° piso, CP 1063, Buenos Aires, Argentina.
[2]Instituto de Clima y Agua, CICVyA-CNIA INTA, CC 25, CP 1712 Castelar, Argentina.
[3]Instituto de Biotecnología, CICVyA-CNIA INTA, CC 25, CP 1712 Castelar, Argentina.

SUMMARY:

1. In the case of soybean, conventional morphological descriptors are still the standard methodology to examine the distinctness of a new variety. It has been proposed that molecular markers and particularly microsatellites might offer a convenient complement to conventional descriptors to genotype varieties. Previous studies carried out by our laboratory have shown that microsatellite patterns of a given variety may vary through time and that it is possible to find additional alleles to those found in a first explorative analysis. The purpose of this work was to develop a reliable identification system for soybean to compare known variety patterns with an unknown sample. In this work we show that repeated individual DNA analysis of a large number of seeds (180 seeds per variety for 3 soybean varieties for up to 8 SSR microsatellites) showed that it is possible to clearly differentiate the predominant major alleles that characterize a given variety from rare alleles that are also present at frequencies below 1%. However, in practical terms, the cost of such an approach makes it too high. Therefore, we tested an alternative approach based on pooling DNA extracts from several individual plants in which predominant alleles may compete with rare alleles in PCR. Sensitivity tests indicated that rare alleles could be detected in pooled DNA samples only at the lowest dilutions respecting the predominant allele. This means that PCR analysis of pooled samples of about, for example, 100 individuals is feasible for obtaining SSR patterns based on the predominant allele, because low frequency alleles have little chance to show up due to their dilution. An alternative is to generate 3 or 4 independent samples of 5-10 plants. Only those alleles that are shared in the 3 or 4 samples should be taken in consideration for genotyping the variety.

Introduction

2. The importance of soybean for Argentina results from the steady increase in production and cultivated area during the last 8 years, in the context of the international seed and grains trade. Forecasts for the 2006/2007 campaign indicate that Argentina will produce about of 41.3 million tons in about half its cultivated area, 15 million hectares.

3. It is of great importance for offices granting breeder's rights to be able to complement and reinforce these rights and to have the possibility to adequately control the seed commerce especially for self-pollinated species where violation of breeders' rights is often detected. As field trials may last two or three years, faster and cheaper assays are needed.

4. Among the diverse data sets used by researchers to analyze genetic diversity in crop plants, DNA-based markers were found to provide a reliable identification of genotypes, with unique DNA profiles useful for the characterization of new varieties. Therefore, molecular markers could be considered as a complement to the traditional system of morphological traits.

5. Regarding this, the Molecular Markers Laboratory at INASE, decided to use soybean, a self-pollinated species, as a model to study for the possible application of DNA-based markers for variety identification in relation to the breeder's rights.

6. Previous analysis developed by our laboratory on soybean focused on a case study of DNA-based-DUS testing. The DNA markers chosen were SSR, as they proved to have the best fit to pedigree data while maintaining an acceptable correlation to morphological-based clustering. Identification was performed using 5 bulked plants per variety analyzed with 30 SSR. Heterogeneity was assessed using 15 individual plants from six commercial varieties, analyzed with 15 SSR. Stability was assessed by studying 7 commercial varieties over a 4-year period with 32 SSR. These analyses showed that the SSR profile per variety varied in accordance with the number of plants analyzed and over the years (Giancola, 2002; Vicario, 2002).

7. The aim of this study is to test, in a large sample, the heterogeneity of Argentine soybean varieties and determine the number of plants to be analyzed in order to obtain a detailed allelic profile feasible to be used for identification when comparing an unknown sample with data from already characterized varieties.

Results and discussion

8. In order to determine a threshold sampling size to analyze heterogeneity, we decided to analyze two different representative samples of 90 individual plants (the original and a repetition) belonging to 3 commonly used commercial varieties, for which we tested up to 8 SSR. A previous study using inbred maize lines, reported a maximum of 48 individual plants to characterize genetic diversity in 7 breeding populations (Warburton, 2002).

9. Global analysis of all 180 plants per variety showed that many markers analyzed (44% of the total) revealed more than one allele and up to 4 different alleles. As expected, in almost every case each variety showed just a major characteristic predominant allele. The only 2 exceptions (microsatellite 173 in variety A and 294 in variety C, see Table 1) are clearly explained by heterogeneity of these two particular varieties for this locus. Thus, frequencies of predominant alleles in the samples were far larger than 90% except in the cases of heterogeneity (larger than 45%). The major allele was sometimes accompanied by one, two or three rare alleles with individual frequencies that were never larger than 6% and lower frequencies at the very limit of detection represented by single cases in heterozygosis (0.3%). Accordingly, for the 3 varieties, the two independent samples reflected the presence of major and rare alleles. As expected, the frequency of the major allele was more or less similar in both samples, while low frequency alleles differed not only in their particular frequencies per locus, but also in their presence due to the fact that statistical errors become larger with these smaller sample sizes. Much larger sample sizes would be necessary to better assess their accurate frequency, which is not the purpose of this work. In spite of the large difference between the relative frequencies of predominant and rare microsatellite alleles, sample sizes should not be too small to avoid the possibility of diagnosing a rare allele as predominant due to sampling errors.

10. In order to reduce costs and due to the necessity to analyze several plants in order to determine the predominant allele (of which are sometimes two, due to heterogeneity), it was decided to pool DNA samples from different plants and perform a single PCR detection.

11. A test on PCR sensitivity was performed to determine the appropriate size of the pool to be analyzed for which all the alleles present could be identified. Five microsatellites (168, 173, 175, 231 and 294) were selected after having been found to be polymorphic in at least one of the three soybean varieties analyzed before.

12. Extracted DNAs were carefully quantified and five different mixes were prepared (1/10, 1/20, 1/30, 1/40 and 1/50). These mixes consisted of DNA containing one target allele (belonging to a variety containing it as single rare allele) diluted in DNA from plants having a solely predominant allele. All DNA mixes were then used as DNA templates in PCR reactions using the corresponding allelic controls.

13. It is known that when more than one target nucleotide sequences compete for limited PCR reactives in a single solution, several factors may influence the final result by selectively amplifying one substrate instead of another. In this case, different results were obtained depending on the locus and variety combinations. Table 2 shows that in 5 out of 7 mixtures of increasingly diluted rare alleles, competitive PCR made them undetectable, compared to undiluted DNA samples. Thus, maximal dilution for which it was possible to get differential amplification of a rare allele, varied independently of the SSR or variety analyzed ranging from no dilution up to two cases of 1/50. These results showed that most rare alleles of bellow 1% frequency (like the ones shown in Table 1) may be out-competed by the predominant substrate in the PCR at the lowest dilutions tested, indicating that a pool of 100 plants will have a very high probability of detecting only predominant alleles (assuming that only one plant of the 100 possesses the rare allele in a heterozygous state). Thus, the identification system for soybean variety protection purposes could be based on bulked DNA in order to obtain a profile for the most frequent allele(s).

14. Alternatively, we propose to generate 3 or 4 independent samples of 5-10 plants from which DNA should be extracted and pooled to perform molecular marker PCR analysis. If more than one allele is detected, only those that are shared in the 3 or 4 samples should be taken into consideration for genotyping the variety.

15. Appearance of low frequency rare alleles due to evolution is nor surprising. Microsatellite mutation rate was estimated to reach frequencies as high as 10^{-5} (Diwan and Cregan, 1997) or 10^{-3} (Heckenberger et al, 2002). Besides, even when the varieties under study have been tested for homogeneity of morphological traits, an important proportion of the background genome may still be segregating. This was confirmed in the present study since two varieties out of three were found to be heterozygous for at least one microatellite locus. Other factors like purity maintenance or cross-pollination may also affect the appearance of new alleles. A way to reduce the influence of the latter factors is to include DNA-based markers during the breeding processes.

Conclusions

16. The results show that as more individual plants are observed, more alleles will be found. Thus, even working with self-pollinated varieties, it is possible to find alleles at very low frequencies and, for an unknown sample, they may be erroneously scored as predominant without the use of very large number of individual plants. This will certainly be a problem if fixing a DNA pattern for each variety is the objective for genotyping.

17. In order to obtain a profile based on the most frequent allele(s), the identification system for soybean varieties protection purposes could be carry out on:

1. 100 bulked seeds;

2. the use of 3 or 4 independent bulked seeds of 5-15 plants each.

Either of these procedures will allow an individual identification pattern to be obtained, avoiding the consideration of infrequent alleles that may alter the similarity among samples and protected varieties.

References

- Diwan N, Cregan P. Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. *Theor. Applied Genet*. 1997 95: 723-733.
- Giancola S, Marcucci Poltri S, Lacaze P, Hopp H E. Feasibility of integration of molecular markers and morphological descriptors in a real case study of a plant variety protection system for soybean. *Euphytica* 127: 95–113, 2002.
- Heckenberger M, Bohn M, Ziegle J, Joe L, Hauser J, Hutton M and Melchinger A. Variation of DNA fingerprints among accessions within maize inbred lines and implications for identification of essentially derived varieties. I. Genetic and technical sources of variation in SSR data. *Molecular Breeding* 10: 181–191, 2002.
- Vicario A. Estudio Comparativo de la Diversidad Genética del Germoplasma Comercial de Soja (*Glicine max (*L.) Merr.) y Variedades Tradicionales. Tesis doctoral por la Universidad de Buenos Aires. Octubre de 2002. Facultad de Ciencias Exactas y Naturales – UBA.
- Warburton M, Xianchun X, Crossa J, Franco J, Melchinger A, Frisch M, Bohn M, and Hoisington D. Genetic Characterization of CIMMYT Inbred Maize Lines and Open Pollinated Populations Using Large Scale Fingerprinting Methods. *Crop Sci*. 42:1832–1840 (2002).

Table 1: Microsatellite alleles and allelic frequencies among samples of 90 plants

VARIETY	SSR	Sample 1	Sample 2	TOTAL
A	**231**	1 and 2 (0.988 / 0.012)	1 and 3 (0.983 / 0.017)	1, 2 and 3 (0.986 / 0.014/ 0.014)
	226	1	1	1
	294	1	1	1
	168	1	1	1
	42	1	1	1
	173	1 and 2 (0.461 / 0.539)	1 and 2 (0.551 / 0.449)	1 and 2 (0.506 / 0.494)
	175	1	1	1
	353	1	1	1
B	**173**	1	1	1
	175	1	1	1
	353	1	1	1
	294	1, 2, 3 and 4 (0.011 / 0.005 / 0.139 / 0.844)	1, 2, 3 and 4 (0.011 / 0.017 / 0.051 / 0.92)	1, 2, 3 and 4 (0.011/ 0.011/ 0.095 / 0.882)
	168	1	1, 2 and 3 (0.988 / 0.005 /0.005)	1, 2 and 3 (0.994 / 0.003 / 0.003)
C	**175**	1, 2 and 3 (0.06 / 0.927 / 0.012)	2 (1.00)	1, 2 and 3 (0.029 / 0.964/ 0.006)
	294	1 and 2 (0.566 / 0.434)	1 and 2 (0.517 / 0.483)	1 and 2 (0.530 / 0.470)
	168	1 and 2 (0.975 / 0.025)	1	1 and 2 (0.988 / 0.012)

First column indicates the varieties tested (A, B and C). Second column indicates microsatellite loci analyzed (BARC – ATT repeats). Total data were obtained by combining both samples (i.e. from 180 plants).

Table 2: PCR sensitivity

VARIETY	SSR	AMPLIFICATION (at different dilutions)				
		1/10	1/20	1/30	1/40	1/50
A	**173**	Yes	Yes	Yes	Yes	Yes
A	**231**	Yes	Yes	No	No	No
B	**294**	Yes	No	No	No	No
B	**168**	Yes	No	No	No	No
C	**294**	Yes	Yes	Yes	Yes	Yes
C	**175**	Yes	No	No	No	No
C	**168**	Yes	No	No	No	No

The Table shows the amplification of the rare allele at different dilutions with the predominant allele.

[End of document]