E

BMT/11/19
ORIGINAL: English
DATE: August 29, 2008

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
GENEVA

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

Eleventh Session
Madrid, September 16 to 18, 2008

IDENTIFICATION SYSTEM FOR SOYBEAN BASED ON THE MOST FREQUENT SSR ALLELES

Document prepared by an expert from Argentina

Loray, M., Langan, M., Herner, A. y Vicario, A.
(Instituto Nacional de Semillas – INASE. Av. Paseo Colón 922, 4to piso. CP: 1063.
Ciudad de Buenos Aires, Argentina).

BACKGROUND

1. It is of great importance for offices granting plant breeders' 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. In the same way, it is also important for quality departments to develop variety identification systems which also would allow a more reliable seed commerce.

2. DNA-based markers have been found to provide a reliable identification of genotypes, giving unique DNA profiles useful for the characterization of new varieties, and they are also considered as a complement to the traditional system of morphological characteristics.

3. Regarding this, the Molecular Markers Laboratory at the *Instituto Nacional de Semillas* (INASE), decided to use soybean, a self-pollinated species, as a model to study the possible application of DNA-based markers for variety identification in relation to the reinforcement of breeders' rights and also for quality assurance.

4. The DNA markers chosen, for both cases, were Simple Sequence Repeats (SSRs), as they proved to have the best fit to pedigree data while maintaining an acceptable correlation to morphological-based clustering (Giancola, 2002). Also, SSRs are inexpensive, amenable to automation, co-dominant, independent of the environment, highly polymorphic, essentially unlimited and multiallelic, and provide coverage of the entire genome.

5. In a previous study, this group determined the heterogeneity of Argentine soybean varieties and 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. We found that this characterization has to be based on the most frequent allele for each variety in order to avoid mis-classification of samples due to the consideration of infrequent, rare alleles for the variety. The conclusion of that previous study was that the analysis of 100 bulked seeds or, alternatively, 4-5 pools of 5 seeds, will allow the database construction based only on the most frequent allele/s.

6. There are several parameters to be assessed when developing a testing method: limit of detection; limit of quantification; reproducibility; repeatability; specificity; range and lineally; and ruggedness (= robustness).

7. Depending on the method (qualitative or quantitative), it is recommended to assess few or all of them. The case of the SSRs used to characterize soybean varieties, can be considered as a case of qualitative data. So, the parameters to be determined for an intra-laboratory validation are specificity, limit of detection, ruggedness and repeatability.

8. Marker specificity is given by the source of the marker. In this case, all markers used were selected from *http://soybase.org/resources/ssr.php*. As the method will detect a sequence that is expected to be present in all the genome copies, it is possible to say that it

will be far from the limit of detection. But the method validation requires the assessment of the parameters ruggedness and repeatability.

9. The aims of the present study are to test:

a) which system is more efficient for variety identification, i.e. less time and money consuming;
b) the method repeatability, i.e. if the analysis of different samples in the same laboratory, with the same equipment and technicians will lead to the same result; and
c) the method ruggedness, i.e. to test if small deviation of the method parameters will, or will not, change the results.

Materials and methods

10. In order to obtain a SSR profile for soybean based on the most frequent allele(s), 12 commercial soybean varieties were analyzed studying 12 SSRs selected on the basis of their capability to uniquely identify more than 230 soybean varieties from the Argentine reference collection (data obtained in a previous study). Four of these SSRs were also used for Yoon *et al.* (2007) on 96 varieties from the United States Department of Agriculture (USDA) Soybean Germplasm Collection.

11. Two different strategies were studied, both using bulked seeds: 1) a global sample of 100 bulked seeds; and 2) four small samples of 5 seeds each. The most frequent allele/s are those that could be amplified from both 100 bulked seed samples and from all 4 five-seed samples. Alleles present only in 1, 2 or 3 of the five-seed samples are considered as infrequent alleles.

12. DNA extraction, from all different bulked seed samples, was performed on the basis of the Dellaporta method (Dellaporta *et al.*, 1983).

13. PCR products were resolved by standard sequencing electrophoresis on denaturing polyacrylamide gels. Bands were revealed using a sequencing gel silver nitrate staining system. The size of each band was estimated using a molecular size marker and reference alleles loaded in adjacent lines of the gel. The differences in size of the amplified fragment were considered as different alleles.

14. Repeatability was assessed by analyzing duplicate samples of 21 varieties of high use in Argentina that belong to six different breeding programs. 12 of the 21 varieties had already been used in the previous study. These varieties were characterized using the same 12 SSR.

15. Ruggedness was assessed on a third sample, prepared for each variety, by using different equipment and different reagents than those used for the duplicates. This means a different thermocycler and pipettes set, and a different polymerase quality. Ruggedness was tested by comparing this data with the one obtained for the duplicates.

RESULTS AND DISCUSSION

16. The analysis of microsatellites on 12 varieties showed that the alleles observed for the 100 bulked seed samples were also present in all 4 five-seed samples and they were considered as the most frequent allele(s).

17. However, the alleles that amplified only in 1, 2 or 3 five-seed samples did not amplify in the 100 bulked seed samples and they were considered to be rare or infrequent alleles, and, therefore, were not taken into account to identify the variety.

18. For the 100 bulked seed-samples, varieties showed mostly one allele per marker (89% of the cases). In only 11% of the cases did varieties show two alleles. There were no varieties with three or more alleles per marker.

19. The results observed in this study were, in accordance with previous results obtained from studying individuals plants (documents BMT/10/15 'DNA-Based Identification System for Soybean' and BMT/10/15 Add. 'Addendum to Document BMT/10/15: DNA-Based Identification System for Soybean'). In that study, we found that, in almost every case, each of the three different soybean varieties showed just a major characteristic predominant allele with a far large frequency of 90%, except in the case of heterogeneity (frequencies 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%. It was also demonstrated that most rare alleles of below 1% frequency may be out competed by the predominant substrate in the PCR at the lowest dilution tested (1/50).

20. The strategies compared in the present work, namely, the strategy based on the analysis of 100 bulked seed samples and the strategy based on 4 five-seed samples, were found to be similarly efficient in detecting the predominant major allele(s). Nevertheless, the 4 five-seed samples strategy was found to be less efficient when preparing the samples. Working with the 100 bulked seed sample is less time consuming regarding grinding and DNA extraction, which are also of best quality.

21. In order to confirm that this method was repeatable and robust, three independent samples of 100 bulked seeds were generated from each of 21 soybean varieties. Duplicates for each variety gave the same results, confirming that repeatability is possible for this kind of assay. The third independent sample gave the same results compared with the duplicates confirming the ruggedness of the method.

22. These results confirm that the alleles obtained in this study are definitively the major predominant or most frequent alleles and the candidates to be used to identify these varieties.

23. The results obtained for these two validation parameters are in accordance with a parallel study, carried out by an international forum, which demonstrates that for most cases, high repeatability (and also reproducibility) is possible for soybean (an example of the first step of this study is in Cassarini and Vicario, 2008).

CONCLUSIONS AND FURTHER WORK

24. The development of a robust identification method for soybean has two aims: 1) to reinforce breeder's rights; and 2) to have a method for quality assurance and seed certification.

25. Approximately two hundred soybean varieties of recent commercial introduction will be analyzed using a set of 8 already inter-laboratory validated SSR markers. More SSR markers for soybean are now being validated through an international forum and will be also used in the near future for obtaining unique profiles for Argentine soybean varieties.

26. It is also possible to adapt this method for purity purposes and as complementary information for the National Registration Office. It is now under discussion whether SSR linked to specific genes would be the appropriate markers for those purposes, or other markers such as SNPs would be better. Evidence on that will have to be provided during the next year.

REFERENCES

http://soybase.org/resources/ssr.php

Document BMT/10/15 'DNA-based identification system for soybean'. Tenth session of the Working Group on Biochemical and Molecular Techniques, and DNA Profiling in Particular, held in Seoul from November 21 to 23, 2006.

Dellaporta, S. L., Word, J. and Hicks, J. B. (1983). A plant DNA minipreparation. Plant Molecular Biology Reporter 1: 19-21.

Casarini, E. and Vicario, A.L. (2008). First comparative test on DNA-based methods: final report of the Variety Committee Working Group. Seed Testing International. ISTA News Bulletin 135: 26-28.

[End of document]