

Technical Working Party on Testing Methods and Techniques

TWM/1/15

First Session

Virtual meeting, September 19 to 23, 2022

Original: English

Date: September 14, 2022

VARIETY IDENTIFICATION: SOYBEAN CASE IN ARGENTINA

Document prepared by an expert from Argentina

Disclaimer: this document does not represent UPOV policies or guidance

SUMMARY

In 2016, an Ad Hoc working group on molecular markers in soybean was established by INASE and the private sector, with the aim of selecting SNP markers that in combination with morphological data allows the management of the reference collections (to reduce the size of the field trials). Another important aim was variety identification/verification for control of seed trade. For the selection of SNP markers, the Illumina BARCSOY6KSNP chip was used. This chip contains a selection of 6K SNP markers from the 50K chip, developed by the United States Department of Agriculture (USDA). Two marker sets were selected for variety identification and both were analyzed during this work: 56 previously selected and 32 newly selected in the aim of reducing future costs. Samples consisted on breeder's seeds samples identified as O, certified seeds samples identified as C, grains samples of 19 varieties with replicates identified as A and farmer's samples identified as P. The sample set contains 28 varieties that were represented by breeder's, certified, grains and farmer samples, covering 3 or more generations of successive multiplication of seeds. Genotypic data was obtained by the Genomics and Molecular Markers Laboratory from the Agronomy Faculty, Buenos Aires University using SNPliner from LGC or real time PCR. The coefficient used for obtaining similarity values was the Jaccard association coefficient. The results indicated that the reduced set of 32 markers has less than 1% discrepancies compared to the use of 56 markers (new INASE regulation N°357/22 already in place); for the breeder's seed samples, more than 95% of the samples are correctly verified against the varieties in the data base; regarding thresholds for decision making the data analysis showed that it is possible to establish a 0.8 similarity threshold above which it is possible to be certain about the identity of the variety with 95% of confidence. The annex to this document contains a copy of a presentation on "Variety identification: soybean case in Argentina", prepared by an expert from Argentina, to be made at the first session of the TWM.

[Annex follows]

Variety identification: soybean case in Argentina



Secretaría de Agricultura,
Ganadería y Pesca



Ministerio de Economía
Argentina

Background

2022

In 2016, the **Ad Hoc Working Group** on molecular markers in soybean was established.

This is a working group made up of **breeders and INASE officials** with the aim of selecting SNP markers in soybeans for the management of reference collections to reduce the size of the field trials and for variety identification/verification in the context of trade control in the species.



Secretaría de Agricultura,
Ganadería y Pesca



Ministerio de Economía
Argentina

Objectives

2022

1. To **compare the results obtained by using 56 or 32 SNP markers** for the verification/identification of cultivars over the total number of samples analyzed in order to define the number of SNPs to be used in the context of the trade control.
2. To **evaluate the variation of the similarity values for the same variety through different generations of seed multiplication** using samples sent by the breeders (O), obtained from the trade control and corresponding to the first multiplication (C), sent by the farmers to the collection plants as grains (A) and saved by farmers who are within the ARPOV monitoring system (P).
3. To **establish thresholds of "greater certainty" and "less certainty" in the interpretation of the results**, based on the analysis of the similarity values obtained for the O samples against their corresponding original variety saved in INASE database.



Secretaría de Agricultura,
Ganadería y Pesca



Ministerio de Economía
Argentina

Molecular markers used and analysis

2022

BARCSOY6KSNP chip, which contains a selection of 6K SNP markers from the 50K chip, developed by USDA.

Two marker sets were selected and analyzed during this work: **56 SNP** and **32 SNP**

SNP set were selected by Faculty of Exact and Natural Sciences and the Agronomy Faculty, both of the University of Buenos Aires.

Genotypic Data Base (GDB) software containing all genotypic data provided by breeders when presenting the novel variety before the variety office (**reference genotypic data**).

Data analysis of this work consisted on calculating **Jaccard association coefficient** for the samples against the GDB and drawing conclusions out of the information obtained.



Secretaría de Agricultura,
Ganadería y Pesca



Ministerio de Economía
Argentina

Sample composition and allelic data

2022

Samples consisted on:

- 121 breeder's seeds samples identified as **O**,
- 145 certified seeds samples identified as **C**,
- grains samples of 19 varieties with replicates identified as **A** and
- 36 farmer's samples identified as **P**.
- 28 varieties that were represented by breeder's, certified, grains and farmer's samples, covering 3 or more generations of successive multiplication of seeds.

Genotypic data was obtained by the Genomics and Molecular Markers Laboratory from the Agronomy Faculty of the Buenos Aires University using SNPlane from LGC or real time PCR.



Secretaría de Agricultura,
Ganadería y Pesca



Ministerio de Economía
Argentina

Results

2022

432 samples (O, C, A and P) were genotyped using 56 SNP.

Jaccard similarity values of the samples against the GDB were calculated using the whole set of 56 markers and a subset of 32 SNP.

The table shows the number of times the results obtained using 56 or 32 SNP agreed in identifying a sample.

Agreement between results obtained using 56 or 32 SNP	Number of samples presenting the same results (%)
both sets agree on presenting the declared variety as the most similar one or agree on not identifying the sample (the variety is not present among the 5 most similar ones)	409 samples (94.7%)
both sets agree on presenting the most similar variety among the 5 most similar ones (same or different position)	20 samples (4.6%)
no agreement between both sets of markers	3 samples (0.7%)

This represents a 99.3% coincidence of results obtained with 56 and 32 SNP markers



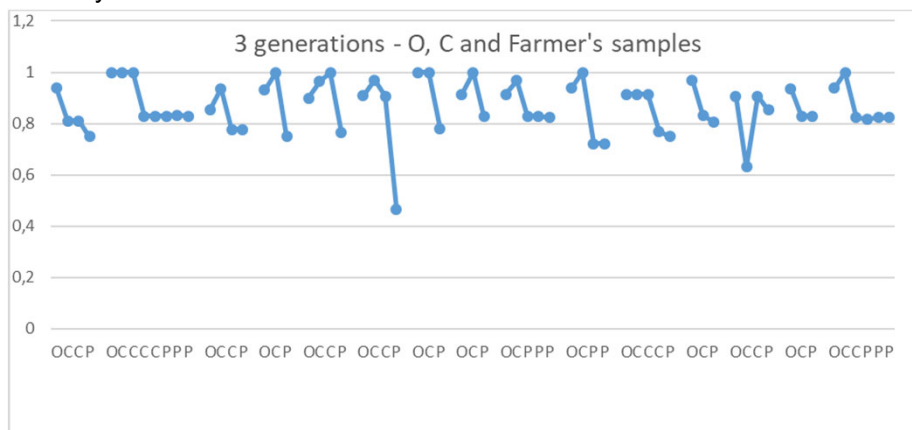
Secretaría de Agricultura,
Ganadería y Pesca



Ministerio de Economía
Argentina

2022

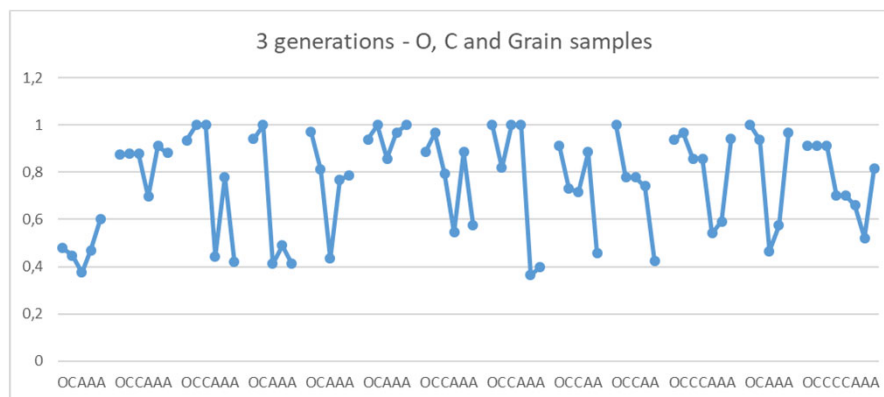
The figure shows **similarity values variation through 3 generations (O, C and P) using 32 SNP**. Y axis shows similarity values and X axis shows sample condition (O C or P). Each data set corresponds to a variety.



2022

The figure shows **similarity values variation through 3 generations (O, C and A) using 32 SNP**.

Y axis shows similarity values and X axis shows sample condition (O C or A). Each data set corresponds to a variety.

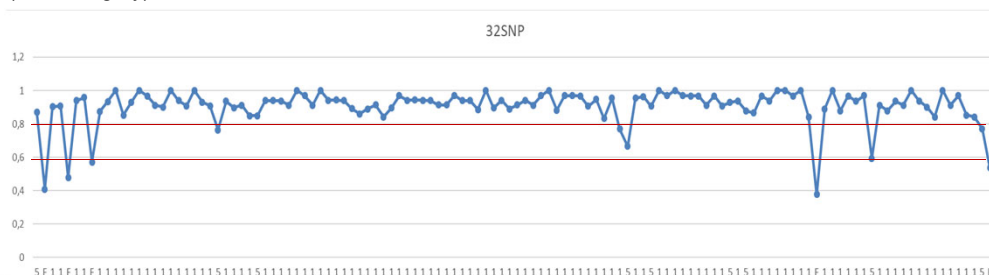


Results

2022

Categories and similarity values for 121 O samples using 32 SNP. The X axis indicates the condition of that variety (1: means that the similarity value obtained for that sample O is the larger one obtained for that variety; 5: means that the variety is among the 5 most similar ones to that sample O; F: means that the similarity value obtained is low and that the variety doesn't fall among the most similar ones to that sample O). There were 108 samples classified as 1, 9 samples classified as 5 (explaining 96.7% of the cases) and 5 samples classified as F (4.13% of the cases; same samples fell in the same category using 56 SNP).

Regarding the **similarity values**, there were 112 samples with similarity values above 0.8 (all 1 and 5 categories) and 5 samples falling between 0.8 and 0.6. This represent 95.7% of the samples O falling above 0.8 threshold. Finally there were 5 samples below 0.6 (all F category).



INASE

Secretaría de Agricultura,
Ganadería y Pesca



Ministerio de Economía
Argentina

Conclusions

2022

It was possible to identify all the samples (O, C, A and P) with both 56 and 32 markers. It was found that there is 99.3% agreement between the results obtained using 56 and 32 SNP markers.

Therefore, **it is possible to reduce the number of markers used for variety identification** from 56 to 32 SNPs. This lead to a new regulation for using SNP markers for soybean in Argentina: INASE regulation N°357/22.

Continue seed multiplication may lead to variation in similarity values of the variety against the database. Also manipulation of grains that will be used as seeds for future sawing may lead to loss of identity or unwanted mixtures.

Regarding the cut-off thresholds for decision making, it is possible to establish a **safe decision limit of up to 0.8** for more than 95% of the cases.

It is possible to use the new set of 32 SNP for variety identification in Argentina.

INASE

Secretaría de Agricultura,
Ganadería y Pesca



Ministerio de Economía
Argentina

This work could be finished thanks to the collaboration of the following colleagues: **2022**

From INASE

Ana Vicario (DEC)
Fernanda Dalmau (DRV)
Mariano Mangieri (DRV)
Mariano Petruzela (DF)
Ignacio Aranciaga (DEC)
Ernesto Loponto (IT)
Ignacio de Pedro (IT)
Maximiliano Burastero (IT)
Martín Aranda (IT)

From the University of Buenos Aires

Marcelo Martí and his team (FCEyN)
Eduardo Pagano and his team (FA)
Marcelo Soria (FA)

From the private sector

Juan Erdmann
Alfredo Paseyro
Edgardo Bresso
Mauro Meier
Gaspar Malone
Julieta Alloatti
Mirta Antongiovanni
Agustina Cereijo
Verónica Vallini
Yanina Moyano
Beatriz Formica
Clarisa Bernardi



Secretaría de Agricultura,
Ganadería y Pesca



Ministerio de Economía
Argentina

2022

Thanks for your attention



Secretaría de Agricultura,
Ganadería y Pesca



Ministerio de Economía
Argentina

