



BMT/14/10

ORIGINAL: English

DATE: October 21, 2014

**INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS**

Geneva

**WORKING GROUP ON BIOCHEMICAL AND MOLECULAR TECHNIQUES  
AND DNA-PROFILING IN PARTICULAR****Fourteenth Session  
Seoul, Republic of Korea, November 10 to 13, 2014**

## THE USE OF MOLECULAR MARKERS (SNP) FOR MAIZE DUS TESTING

*Document prepared by an expert from France**Disclaimer: this document does not represent UPOV policies or guidance*

## INTRODUCTION: WHY CONSIDER THE USE OF MOLECULAR TECHNIQUES IN MAIZE DUS TESTING?

1. Due to the extensive size of the maize reference collection (over 4000 inbred lines and 4000 hybrids in 2014), *Groupe d'Etude et de contrôle des Variétés Et des Semences* (GEVES) needed to implement tools and procedures to continue to manage its reference collection in an efficient way, keeping in mind the decrease of cost and the improvement of the efficiency of the system. As an example, in 2013, we had 274 candidate inbred lines and 3741 inbred lines in our reference collection, resulting in more than one million pair-wise comparisons to be made in order to assess the distinctness of the candidate lines.
2. The previous system used in France to select the varieties to be grown and compared in the field trials used a combination of differences between varieties observed on morphological and electrophoresis characteristics. As electrophoresis is a technology hardly ever used now by the breeding companies, there is a need to consider updating the process. GEVES has considered the opportunity of molecular markers.
3. Such works were already presented to the UPOV community by experts from France, using SSR markers (see for example BMT/10/14 on maize inbred lines, BMT/12/19 on spring barley). The model "Combining phenotypic and molecular distances in the management of variety collections" received a positive assessment from UPOV (see UPOV/INF/18/1 "Possible use of Molecular Markers in the Examination of Distinctness, Uniformity and Stability (DUS)") and is included in TGP/15/1 "Guidance on the Use of Biochemical and Molecular Markers in the Examination of Distinctness, Uniformity and Stability (DUS)". Although the use of SSR markers was proven efficient and reliable for the management of the reference collection in maize, it was more difficult to implement a system using SSR markers for other applications such as checking the hybrid conformity, which was still tested using electrophoresis.
4. In the recent years, new technologies were developed and became easily available. Among them, SNP markers since they are evenly distributed throughout the genome, highly informative, co-dominant, reproducible and commonly used by most breeder companies were the best option

## MARKERS SELECTION

5. The work started in 2011, two aspects were considered: a) the choice of the platforms and technology and b) in parallel the marker set to be used.

6. A large variety of platforms and chemistry is currently available, they differ widely with respect to the samples quantity that can be analyzed at the same time, the type of equipment needed, and cost design. After the evaluation of several SNP detection platforms of medium throughput, according to their flexibility and their performances (sensitivity, reliability, and reproducibility), the KASP (Kompetitive Allele Specific PCR) chemistry proposed by LGC Genomics was retained. This PCR based technology allows a graphic data visualization straight after amplification by fluorescence reading (Figure 1).

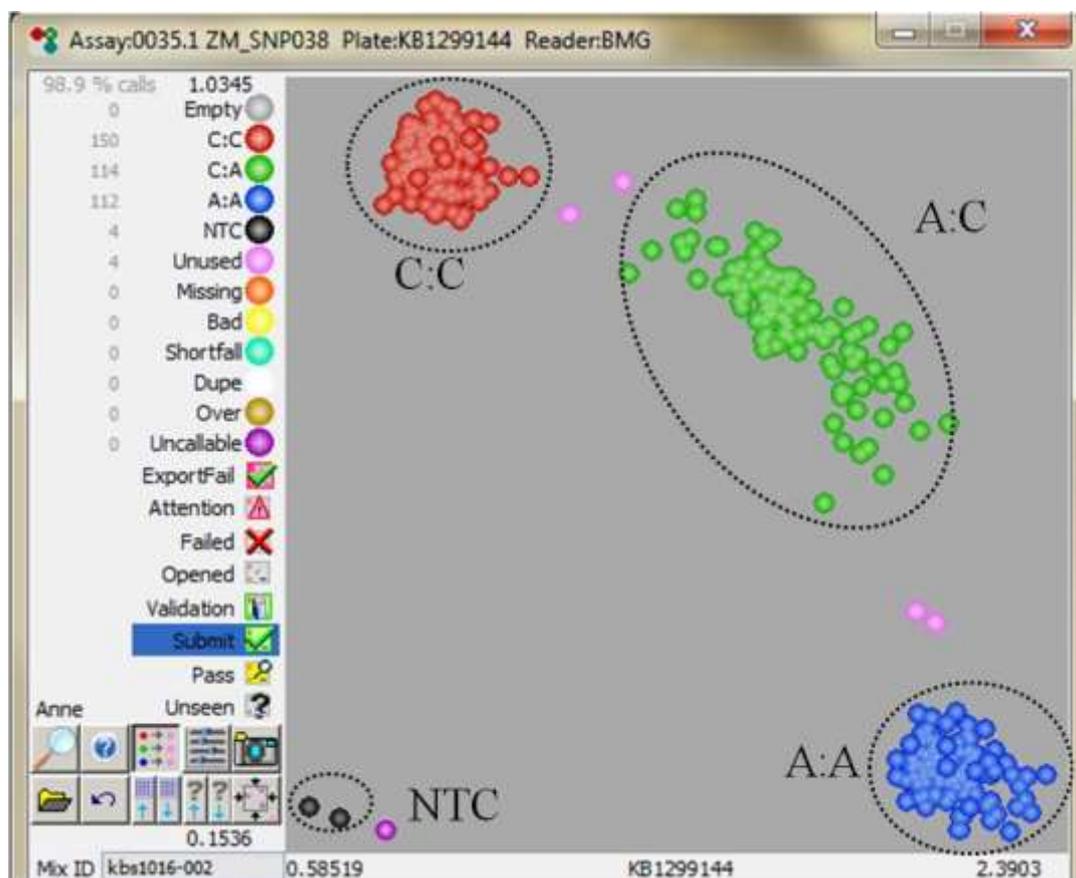


Figure 1: Snapshot displaying the results for one SNP genotyping and 384 samples. Each dot represents a sample. Samples which cluster together will have the same genotypes and will be represented by the same color. Blue and red: homozygous; green: heterozygous; pink: not assigned; purple: no amplification, and black no template control.

7. The choice and selection of maize SNP was a collaborative work based on the Illumina 50k microarray developed by *Union Française des Semenciers* (UFS) and *Institut national de la recherche agronomique* (INRA). From this large amount of markers, two sets of 384 SNP were selected by INRA. The selection of each set was done regarding the quality of the flanked regions, the minor allele frequency score and their distribution on the genome. Combining those two sets and KASP criteria a final set of 384 SNP was determined (Figure 2).



## ASSESSMENTS AND FIRST DATA ANALYSIS

8. SNP primers design and genotyping of more than 4500 inbred lines from GEVES collection and 500 hybrids were carried out by LGC Genomics by the end of 2011.

9. SNP data analysis required a switch from pattern comparison to genetic distance calculation. Genetic distances (the percentage of markers which differ between two lines: sum of the allelic differences on the tested loci) were calculated using R software. On the 384 SNPs tested, 29 were excluded because they showed no amplification. Across all samples, 11 SNPs were monomorphic, and in consequence not included in our marker set.

10. For the 344 SNPs selected, a total of 688 alleles was detected, with each SNP detecting two alleles as expected. Observed heterozygosity ( $H_o$ ), expected heterozygosity ( $H_e$ ) (equivalent to polymorphism information content; PIC), and Minor Allele Frequency (MAF) were computed (Figure 3).

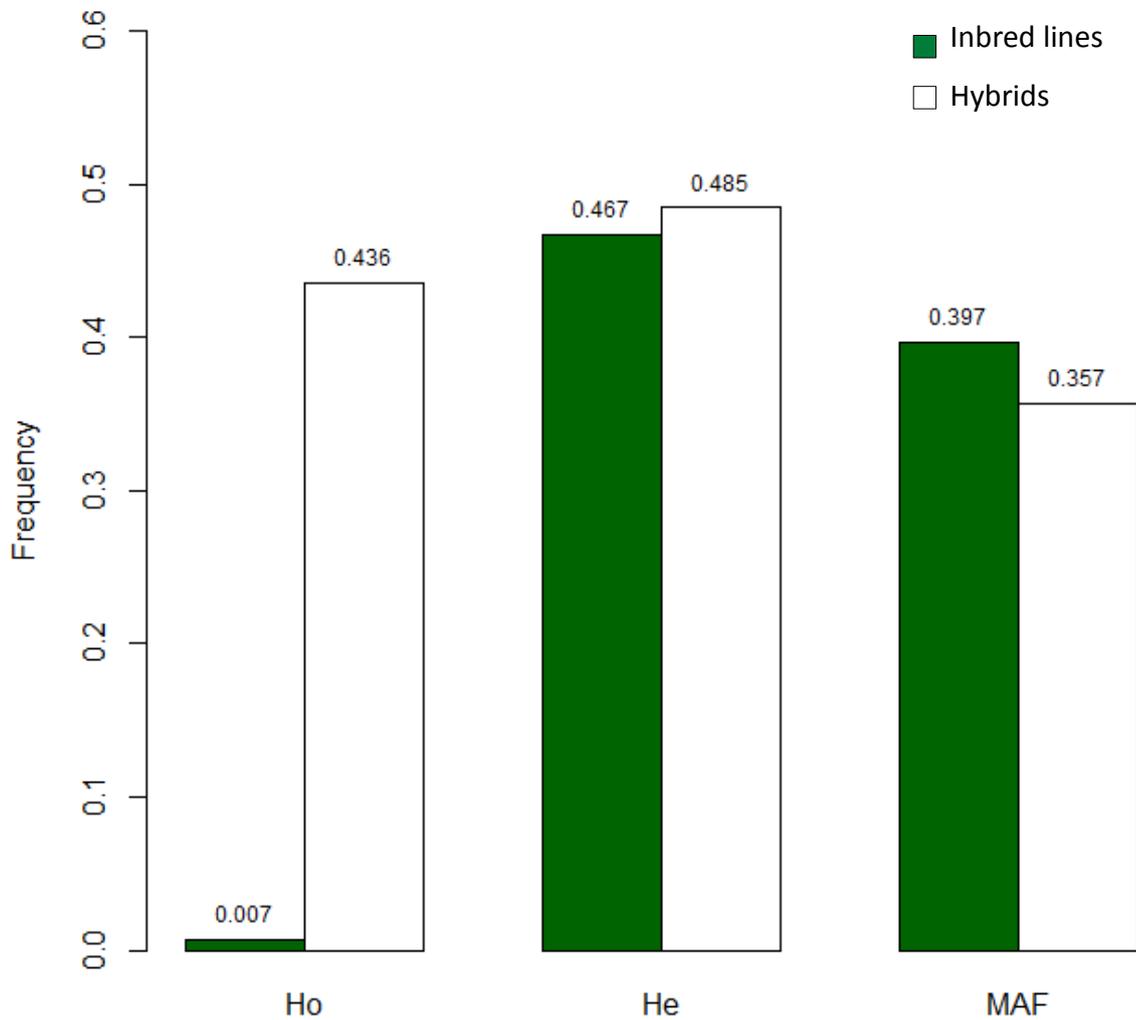


Figure 3: Statistics of the 344 SNPs used for genotyping inbred lines and hybrids from the collection.

11. A subset of 312 SNPs was selected by removing 32 SNPs with high percentage of missing data (> 15%) and ambiguity or irreproducibility in allele calling for known genotypes. Correlation between two genetic distance matrices (Mantel test) calculated from 312 and 344 SNPs was highly significant ( $r= 0.989$ ;  $p<0.001$ ), indicating that no information is lost by removing the 32 SNPs (Figure 4).

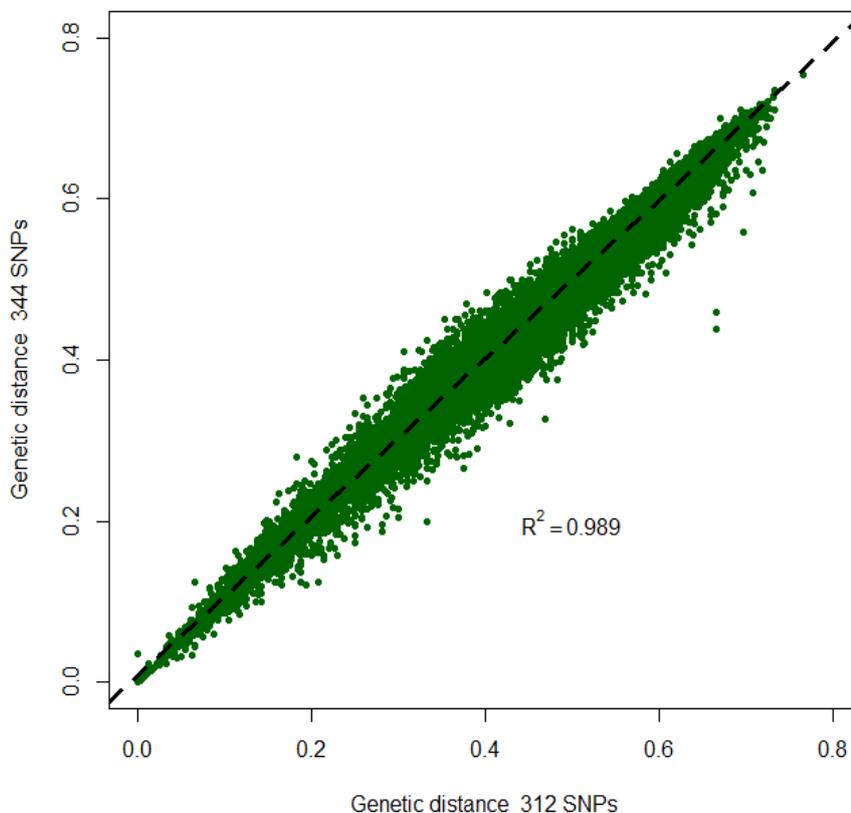


Figure 4: Correlations between genetic distances (from a subset of 1000 samples chosen randomly) calculated from 344 SNPs and a subset of 312 SNPs. Each point represents genetic distance between a pair of samples.

12. This final set of 312 SNPs was used during 2 years (2012-2013) on the new varieties to compare genetic distances and electrophoresis results used for managing field trials and checking the hybrid formulae.

#### THE USE OF MOLECULAR TECHNIQUES FOR THE MANAGEMENT OF THE REFERENCE COLLECTION BY COMBINING MORPHOLOGICAL DISTANCES AND GENETIC DISTANCES

13. From 2011 on, GEVES continued the studies on maize using the complete set of SNP markers, thus replacing the SSR markers. The definition of the “distinctness plus” threshold, which means that the distances between a candidate variety and “distinct plus” varieties are robust enough to take a decision without direct comparison in the growing trial, was a key step of the process of the implementation of the new model combining phenotypic and molecular distances in the management of the maize variety collections. (see document TGP/15/1 “Guidance on the Use of Biochemical and Molecular Markers in the Examination of Distinctness, Uniformity and Stability (DUS)”).

14. We compared two models for the management of the reference collection successively during the years 2011, 2012 and 2013: the model using GAIA combining morphological and isoenzymes data which was the model used as a routine to manage the reference collection of maize inbred lines by GEVES, and the model using GAIA combining morphological distances and genetic distances calculated with the set of SNP markers (as presented in Figure 5). We observed in the field side-by-side comparisons of pairs issued by both models.

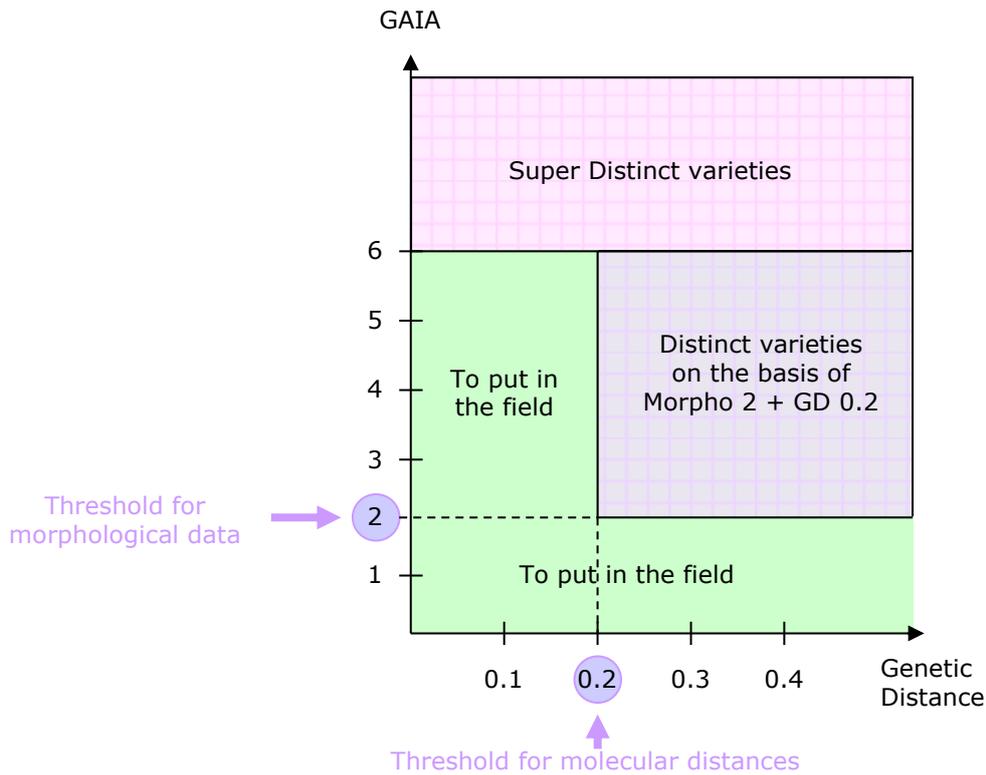


Figure 5: Model combining phenotypic and molecular distances in the management of variety collection. Example on maize inbred lines in France.

15. Evaluations of the link between genetic distance and a global evaluation of distinctness performed by a panel of experts on pairs of varieties were performed in 2011, 2012, and 2013, as described in UPOV document BMT/12/2 "Reports on Developments in UPOV Concerning Biochemical and Molecular Techniques" Annex III "Proposal: "System for combining phenotypic and molecular distances in the management of variety collections"". In 2012 for example, 654 pairs of varieties were grown side by side and tested with the set of SNP. Visual assessment was performed by nine maize DUS crop experts (Figure 6).

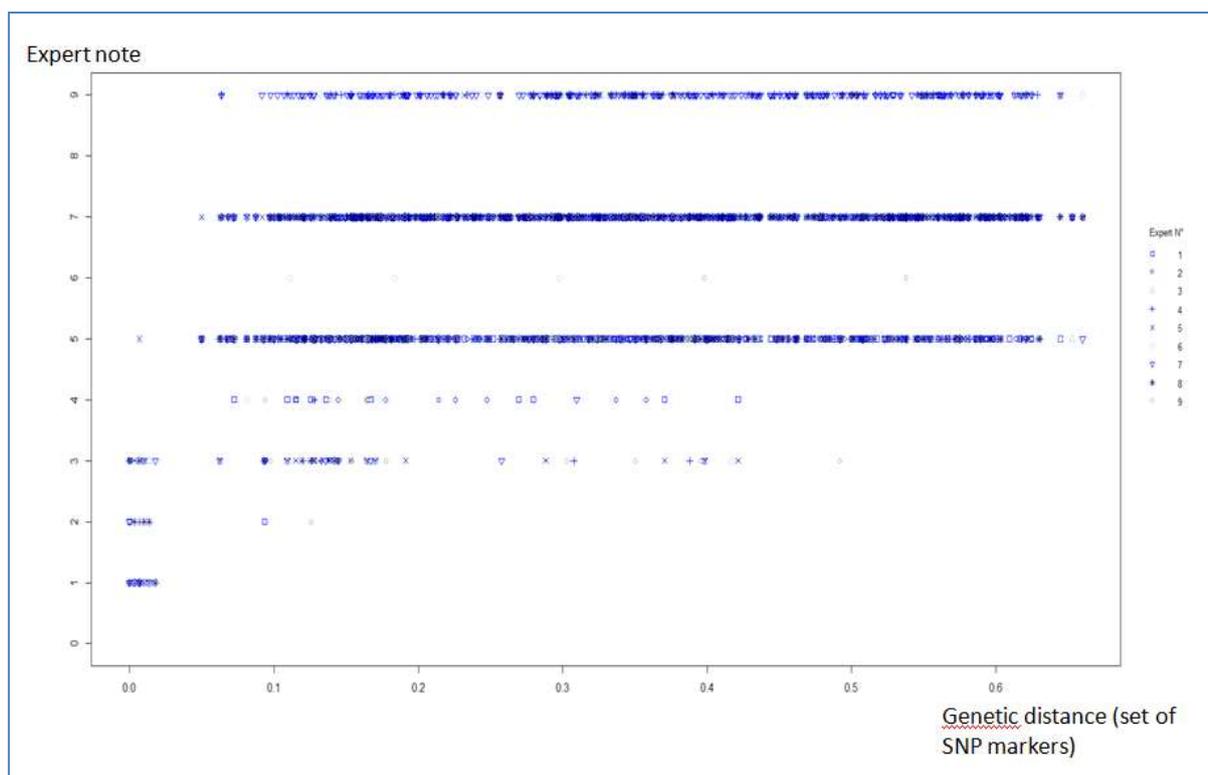


Figure 6: Correlation between the genetic distance and the experts' note on 654 pairs of maize inbred lines observed side-by-side in the field in 2012. The Scale of similarity for the expert note is as follows:

1. the two varieties are similar or very close.
3. the two varieties are distinct but close.
5. the comparison was useful, but the varieties are clearly distinct.
7. the comparison should have been avoided because the varieties are very different.
9. the comparison should have been avoided because the varieties are totally different.

16. As shown on Figure 6 in 2012, no parental lines with a genetic distance greater than 0.1 were considered as similar or very close during the expert evaluation. Such evaluations were also performed in 2011 and 2013. They demonstrated that no parental lines with a genetic distance greater than 0.18 were considered as similar or very close during the expert evaluation. To be sure to have a reliable and secure threshold, the threshold of 0.2 for genetic distance was confirmed and adopted, thus implementing a secure system to be sure that the "distinct plus" varieties are robust enough to take a decision without direct comparison in the growing trial.

17. In 2011, 2012 and 2013, the number of pairs issued by the model combining morphological distances and genetic distances was reduced compared with the model using isoenzymes. Thus with the new model (SNP), less varieties would need to be sown and observed in the field (Figure 7).

18. We showed that some pairs were not common to both models; all of them were observed as distinct in the field. Some pairs were common to both models; all of the similar or very close pairs belonged to this group. So, no similar pairs or very close pairs were forgotten in a model or in another. This demonstrates that the confidence of the system based on morphological and biomolecular combination is not altered by switching from enzyme to SNP.

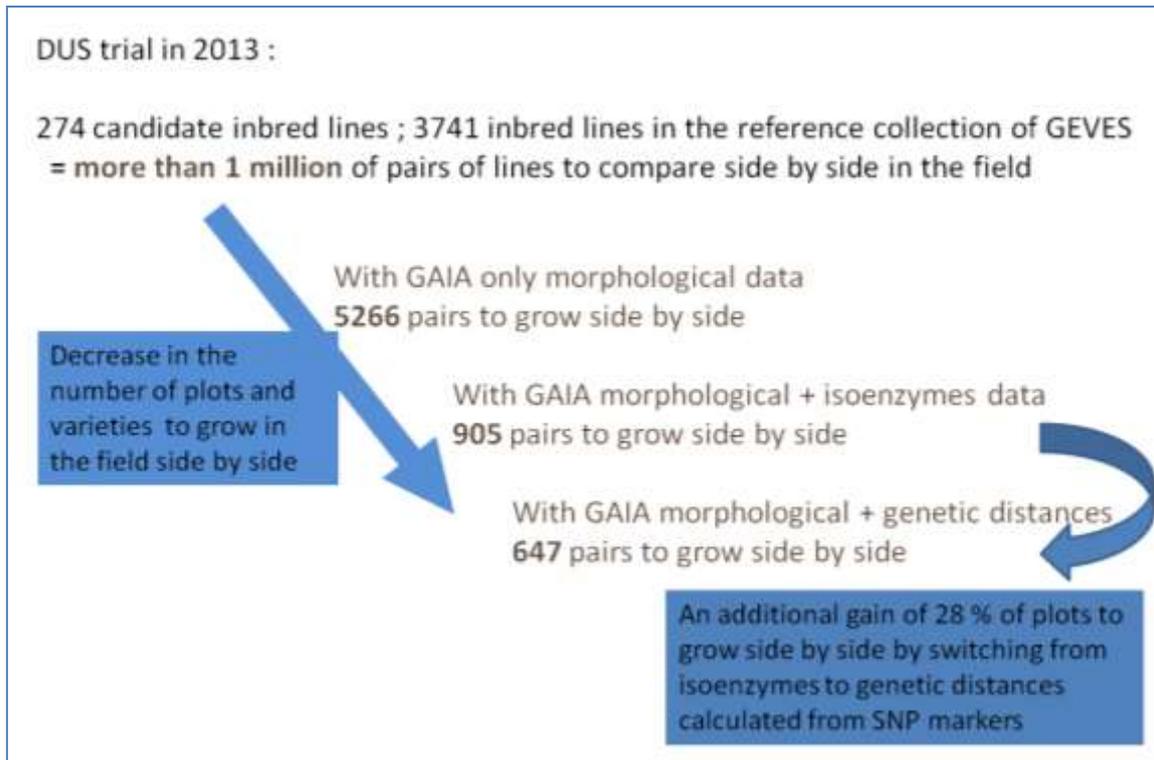


Figure 7: Decrease in the number of pairs of varieties to be observed in the field according to the model used for the management of the reference collection. (Figures from the 2013 DUS trial at GEVES).

19. The combination of phenotypic and molecular distances as defined above for maize inbred lines offers the possibility to achieve a significant reduction in the workload in the field (around 25% reduction in the number of pairs to be compared in the field based on the trials made in 2011, 2012 and 2013). A threshold of 0.2 for genetic distance and a GAIA distance of 6 proved to be appropriate and validated by a panel of maize crop experts. Discussions took place with the breeders and with official bodies, so that the model could be validated and integrated in the French national DUS protocol. The full maize reference collection is now described with the set of SNP markers. Electrophoresis characteristics are not routinely assessed anymore. The model combining genetic and GAIA distances is now routinely used for the management of the reference collection in the frame of DUS test in France.

#### THE USE OF MOLECULAR TECHNIQUES FOR THE RENEWAL OF REFERENCE MATERIAL

20. For the DUS test of maize, it is necessary to maintain a living reference collection. Seeds are stored relatively easily in a cold chamber and when needed, the material can be renewed asking a new sample to the maintainer and then the sample must be compared in the field to check identity and uniformity.

21. However, for about one case out of ten, the old reference material does not germinate, or we don't have enough seeds left. In such cases, we can have problems to ensure that the identity of the newly received seed sample is identical to that of the reference seed sample. In practice, in such cases, the new seed sample is described in the field and then its description is compared with the description of the reference seed sample in the database. But small differences between the description in the database and the description issued in other climatic conditions (year effect) can mislead to the fact that the new seed sample is not a good representation of the variety. Using molecular techniques can be helpful for such cases at least.

22. In 2013, genetic distances between the old seed sample (reference) and the new seed sample received from the maintainer were calculated, for the same maize parental inbred line. The complete set of SNPs was used for the calculation. 309 pairs of seed samples were genotyped in the lab, and they were also observed in the field in side-by-side comparisons between the reference seed sample and the renewed seed sample. The mean observed genetic distance was 0.002. 224 pairs out of 309 pairs showed a genetic distance of 0 (on more than 150 common molecular markers SNPs). More than 90% of the pairs showed a genetic distance smaller than 0.005 (on more than 150 common molecular markers SNPs). All of those pairs were morphologically conform in the field. One pair showed a genetic distance of 0.32 in the lab, and did not conform in the field (Figure 8).

### GD between reference seedlot and new seedlot

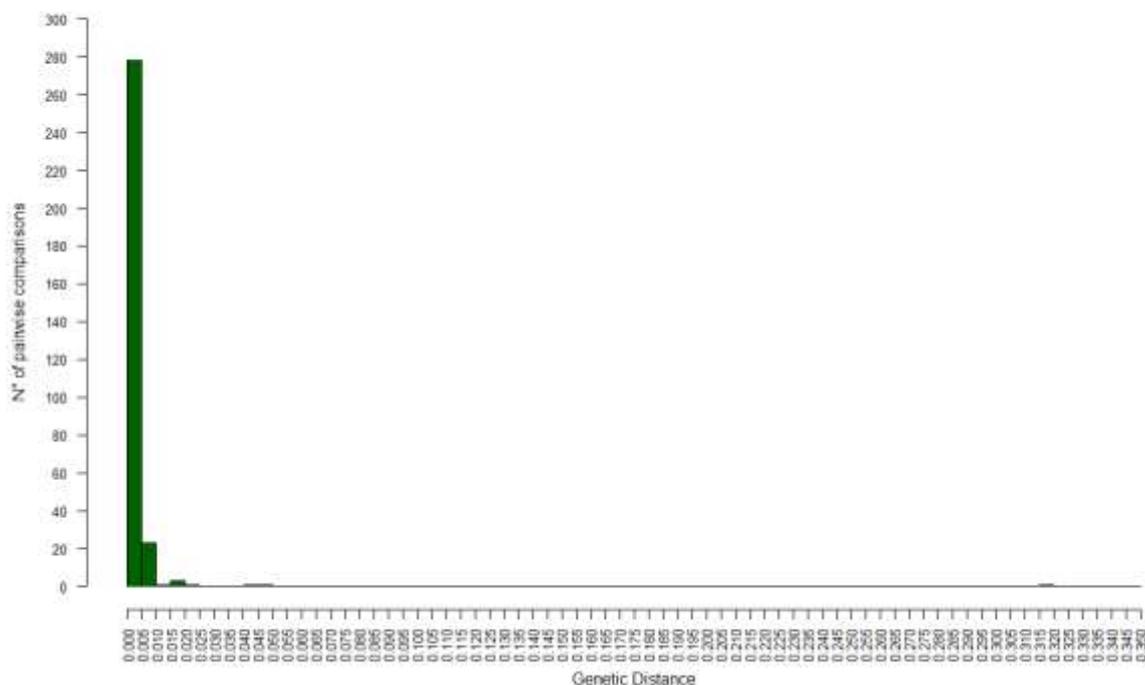


Figure 8: Distribution of pairwise genetic distances between reference seed lot and new seed lot in 2013.

23. It was officially decided at the national level to genotype the seed samples for the renewal of reference material, on a routine basis, using the complete set of markers. Comparisons in the field between the reference sample and the new sample are still performed. The morphological identity is still the key to decide whether the new seed sample can be accepted as reference sample. However, if the reference sample does not germinate or if there are no seed available anymore, a genetic distance of 0 is considered sufficient to accept the new seed sample as reference sample. The new seed sample is observed in the field, so that its uniformity and its germination capacity are checked. If the genetic distance between the reference sample and the new sample is higher than 0.2, then the new seed lot is refused and there is no need to sow it in the field; a new sample is immediately ordered from the maintainer.

24. According to our results, it could also be considered to go further into the decision process for the renewal of maize inbred lines and to propose to stop the direct comparison in the field for pairs with a genetic distance of 0. In this case, no more pairwise comparisons are needed and only the new seed lot will be observed in the field to check the homogeneity.

### THE USE OF MOLECULAR TECHNIQUES FOR CHECKING THE CONFORMITY OF HYBRID FORMULA

25. In France, the DUS test of a hybrid is performed on the basis of the DUS test of the parental lines and the formula. For such a DUS system, checking the conformity of the hybrid formula is an essential part of the DUS test. Until 2014, the conformity of the hybrid formula was performed by GEVES using a set of 6 isoenzymes (14 loci) from the list on the annex of the UPOV Test Guidelines for Maize (document TG/2/7), on all hybrid applications (for the purpose of national listing or PBR).

26. Studies were carried out in 2012 and 2013 to compare the results of conformity of the hybrid formula using either isoenzymes or molecular data. Newly-applied hybrids and their parental lines were genotyped using the complete set of SNP markers. For hybrid conformity with molecular data, the genetic distances between the observed hybrid and the expected hybrid (calculated from the genetic profiles of the parental lines and the formula) were calculated.

27. A first objective was to establish a rule and a threshold for hybrid conformity with molecular data. Our studies showed that a threshold of 0.02 was appropriate. Rules were established after discussions with maize crop experts. As a general principle, the rules previously defined using electrophoresis were translated to a system using genetic distances calculated from SNP data.

## CONCLUSION

28. From the two set of SNP provided by UFS and INRA we have established an optimized set of SNP, with a confirmed and recognized quality for maize genotypes analysis applied to: (a) the management of the reference collection, as well as two other applications; (b) renewal of reference material; and (c) conformity of hybrid formula. The many tests performed at GEVES during the last 3 years produced a lot of supporting data for adopting the SNP technical approach. New rules were adopted at the national official level for DUS testing and this analytic scheme was accepted by the *Community Plant Variety Office of the European Union* (CPVO).

29. GEVES has the equipment and the associated processes settled in its laboratory. Since the beginning of this year, SNP technology is in routine use for maize genotypes analysis and electrophoresis is not used anymore as routine characteristics.

30. In addition to the improvement of our analytic process the consequences of using SNP on maize are: (a) the development of novel tasks for the laboratory part (new skills are needed for SNP data analysis) and; (b) the reduction of field workload for the management of reference collection.

31. Other species are worked out with SNP molecular markers at GEVES to take advantage of this technology for the examination of DUS.

Clarisse MATON, Muriel THOMASSET, Anne BERNOLE, Arnaud REMAY, Bernard AIZAC and René MATHIS  
*Groupe d'Etude et de contrôle des Variétés Et des Semences* (GEVES), France

[End of document]