BAG | Journal of Basic & Applied Genetics

Vol XXX (1): 11-15; July 2019

Journal of the Argentine Society of Genetics

THE IMPACT OF MOLECULAR GENETICS IN PLANT BREEDING: REALITIES AND PERSPECTIVES

EL IMPACTO DE LA GENÉTICA MOLECULAR EN EL MEJORAMIENTO GENÉTICO VEGETAL: REALIDADES Y PERSPECTIVAS

Zambelli, A.

IIDEAGROS, Facultad de Ciencias Agrarias, Universidad Nacional de Mar Del Plata
Ruta 226 Km 73,5 (7620) Balcarce, Argentina

Corresponding author:
A. Zambelli
andres.zambelli@mdp.edu.ar

Cite this article as:
Zambelli, A. 2019. THE IMPACT OF MOLECULAR GENETICS IN PLANT BREEDING: REALITIES AND PERSPECTIVES. BAG. Journal of Basic and Applied Genetics XXX (1): 11-15.

Received: 03/26/2019
Accepted: 04/05/2019

General Editor: Elsa Camadro

DOI:

ISSN online version: 1852-6233

Available online at
www.sag.org.ar/jbag

ABSTRACT

Even when conventional breeding was effective in achieving a continuous improvement in yield, Molecular Genetics tools applied in plant breeding contributed to maximize genetic gain. Thus, the use of DNA technology applied in agronomic improvement gave rise to Molecular Breeding, discipline which groups the different breeding strategies where genotypic selection, based on DNA markers, are used in combination with or in replacement of phenotypic selection. These strategies can be listed as: marker-assisted selection; marker-assisted backcrossing; marker assisted recurrent selection; and genomic selection. Strong arguments have been made about the potential advantages that Molecular Breeding brings, although little has been devoted to discussing its feasibility in practical applications. The consequence of the lack of a deep analysis when implementing a strategy of Molecular Breeding is its failure, leading to many undesirable outcomes and discouraging breeders from using the technology. The aim of this work is to trigger a debate about the convenience of the use of Molecular Breeding strategies in a breeding program considering the DNA technology of choice, the complexity of the trait of agronomic interest to be improved, the expected accuracy in the selection, and the demanded resources.

Key words: DNA marker, selection, plant improvement.

RESUMEN

El mejoramiento convencional ha sido efectivo para lograr una mejora continua en el rendimiento; sin embargo las herramientas de Genética Molecular aplicadas en el fitomejoramiento han contribuido a maximizar la ganancia genética. Es así que el uso de la tecnología de ADN aplicada en la mejora agronómica dio lugar al Mejoramiento Molecular, disciplina que agrupa las diferentes estrategias en las que la selección genotípica, basada en marcadores de ADN, es utilizada en combinación con, o bien en reemplazo de, la selección fenotípica. Estas estrategias se pueden clasificar como: selección asistida por marcadores; retrocruzamiento asistido por marcadores; selección recurrente asistida por marcadores; y selección genómica. Se han presentado fuertes argumentos sobre las potenciales ventajas que aporta el mejoramiento molecular, aunque poco se ha dedicado a discutir la viabilidad de su aplicación práctica. La consecuencia de la falta de un análisis profundo al implementar una estrategia de este tipo puede ser su fracaso, lo que puede derivar en resultados indeseables, desalentando a los fitomejoradores a usar la tecnología. El objetivo de este trabajo es propiciar un debate sobre la conveniencia del uso práctico de estrategias de mejoramiento molecular teniendo en cuenta la tecnología de ADN elegida, la complejidad del rasgo de interés agronómico que se quiere mejorar, la precisión esperada en la selección y los recursos demandados.

Palabras clave: Marcadores de ADN, fitomejoramiento, selección.

INTRODUCTION

In his review, Dr. Rex Bernardo summarized what his adviser had taught him about plant breeding for complex traits: a breeder created genetic variation by crossing good by good, selected the best progenies in the cross, and synthesized the best progenies into a new and improved cultivar (Dudley and Moll, 1969; Bernardo, 2008). Of course, the reality showed Dr. Bernardo (and everyone, by the way) that the situation, unfortunately, is not so simple.

In the classical pedigree breeding method, selecting superior plants bearing traits of higher heritability begins in early generations. However, for traits of low heritability, accurate selection demands the lines to become more homozygous. Commonly, selection of superior plants involves visual assessment for agronomic attributes of interest, as well as laboratory tests for quality or other phenotype feature. When the breeding lines become homozygous (F_5 or further), they can be harvested in bulk and evaluated in replicated field trials. The entire process demands considerable time (depending on the crop, it may range from 5 to 10 years) and money. Even when conventional breeding was effective in achieving a continuous improvement in yield, new technologies were needed to maximize genetic gain. Thus, during the late 1990's, DNA-marker assisted selection offered a promising technology for plant breeding.

The first efforts directed to the design of strategies of plant improvement supported by the use of DNA markers were based on the mapping of quantitative trait *loci* (QTL) in biparental populations. This allowed the development of DNA markers in linkage disequilibrium with them. Its application then focused on recurrent selection schemes to accelerate the pyramiding of QTLs linked to phenotypes of agronomic interest governed by a few genes. Thus, the use of DNA technology applied in agronomic improvement gave rise to Molecular Breeding, discipline which groups the different breeding strategies where genotypic selection, based on DNA markers, are used in combination with, or in replacement of, phenotypic selection. These can be listed as: marker-assisted selection; marker-assisted backcrossing; marker assisted recurrent selection; and genomic selection (Jiang, 2015).

The advantages of using DNA markers to assist selection in plant breeding can be summarized as following:

- It allows selection of traits of interest at early stage of plant growth.
- Unlike the phenotype, the genotype is not affected by environmental conditions.
- It eliminates the need for phenotypic scoring at every breeding generation.
- It provides a uniform and reproducible method for genotype scoring.
- A very small sample of plant, leaf or grain is required for genotyping.
- The release of new cultivars demands a much lower number of breeding generations.

The most widely used technologies are marker-assisted selection (MAS) and marker-assisted backcrossing (MABC). MAS refers to the selection of specific alleles for traits controlled by a few *loci* while MABC is applied to the transfer of a limited number of genes from one genetic background to another, including transgenes.

When setting up a Molecular Breeding (MB) program, different genotyping platforms can be used but the final choice will depend on the requirements of marker density and sample throughput. These platforms range from low-throughput, PCR-based techniques such as the traditional SSRs, to the high-throughput SNP platforms and new sequencing-based methods such as genotyping-by-sequencing (GBS) and amplicon sequencing. Depending on the molecular technology used, DNA markers can be classified into five main types: restriction fragment length polymorphism (RFLP, the first DNA marker available); amplified fragment length polymorphism (AFLP); random amplified polymorphic DNA (RAPD), microsatellites or simple sequence repeats (SSR) and single nucleotide polymorphism (SNP). A comparison of the most conspicuous DNA marker technologies available is summarized in Table 1.

The practical use of MB tools requires very stringent false positive and false negative rates; however,there are a few examples in which some validation of these rates has been conducted. Many studies have investigated the utility of DNA markers in breeding programs; nevertheless, the main criterion that is taken into account at the time of evaluating its usefulness is the genetic linkage of the markers with the QTL, while other issues, such as how reliably the markers classify favorable and unfavorable alleles, are barely analyzed. The consequence of the lack of a deep analysis when implementing a strategy of MB is its failure, leading to many undesirable outcomes and discouraging breeders from using the technology. This determines that in many cases the tool is questioned when in fact what failed was the previous feasibility analysis.

In developing a set of metrics to assess the performance of a candidate DNA marker, it is necessary to break down the features of a marker that impact on its reliability. Thus, Platten *et al.* (2019) proposed to evaluate marker quality based on a measurable quality standard, covering three metric categories: Technical; Biological; and Breeding.

Technical metrics refers to defining the version of the marker (when more than one marker locating close to the QTL is available), call rate, and clarity (that is, how reliable a sample can be classified as allele A, B or heterozygous).

Table 1. Comparison of most widely used DNA marker in plants. Adapted from Jiang (2015)

Feature	RFLP	RAPD	AFLP	SSR	SNP
Genomic abundance	High	High	High	Moderate to high	Very high
Genomic coverage	Low copy coding region	Whole genome	Whole genome	Whole genome	Whole genome
Expression / inheritance	Co-dominant	Dominant	Dominant / co-dominant	Co-dominant	Co-dominant
Number of loci	Small ($<10^3$)	Small ($<10^3$)	Moderate (10^3)	High (10^3 - 10^4)	Very high ($>10^5$)
Level of polymorphism	Moderate	High	High	High	High
Type of polymorphism	Single base changes, indels	Single base changes, indels	Single base changes, indels	Changes in length of repeats	Single base changes, indels
Type of probes / primers	Low copy DNA or cDNA clones	10 bp random nucleotides	Specific sequence	Specific sequence	Allele-specific PCR primers
Cloning and / or sequencing	Yes	No	No	Yes	Yes
PCR-based detection	Usually no	Yes	Yes	Yes	Yes
Genotyping throughput	Low	Low	Moderate	Low to moderate	Very high
Amount of DNA required	Large	Small	Moderate	Small	Small
Time demanding	High	Low	Moderate	Low	Low
Ease of automation	Low	Moderate	Moderate	Moderate	High
Development / start-up cost	Moderate to high	Low	Moderate	Moderate to high	High
Cost per analysis	High	Low	Moderate	Moderate to low	Low
Polymorphic loci detected per analysis	1-3	1-5	20-100	1-3	1

Biological metrics imply the characterization of the marker linkage to the QTL of interest and the false positive and false negative rates (FPR and FNR, respectively).

Breeding metrics describe the relative value of applying a marker in a specific breeding program, consisting of three items: breeding program false positive rate (BpFPR); breeding program false negative rate (BpFNR); and marker utility. Thus, BpFPR and BpFNR are equivalent to the FPR and FNR metrics described above but specific to a particular breeding program in which they are assessed. As the breeding pool may be expected to have lower allelic diversity than occurs species-wide, and because selection and genetic drift are modifying patterns of linkage disequilibrium independently across breeding programs, these rates can be quite different from the true FPR and FNR. They will require the characterization of donor and recipient lines, which will involve collecting phenotype data for each program of interest. In other words, it is important to evaluate the marker's reliability for taking breeding decisions in that specific program.

The last Breeding metrics, marker utility, represents the number of cultivars without the desired allele with respect to the number of cultivars with the desired allele at a given QTL in the breeding population (the lower the proportion, the higher the utility).

Platten's proposal provides a systematic and useful set of criteria to establish a superior marker system for a target QTL, allowing the choice of an optimal group of markers when designing an assisted selection strategy (Platten *et al.*, 2019).

It has been found that classical marker-assisted selection (based on the identification of QTL) has worked satisfactorily for simple traits (whose genetic variance is determined by one or a few *loci*). Therefore, the identification and characterization of QTL associated with traits of agronomic importance has been an area that deserved the interest of the scientific community in the last 30 years. A simple exercise gives an account of it: a bibliographic search on the website of the National Library of Agriculture (USDA; https://agricola.nal.usda.gov) covering that period and including as keywords the terms "QTL" and the names of the twelve main crop species in the world, will show a total of 4476 publications, which in many cases documented the discovering of three or even more QTLs. Therefore, being conservative, it would be reasonable to estimate a total of at least 10,000 QTLs published. However, covering all crops, the number of DNA markers effectively applied in breeding selection can be roughly estimated in around 100 (Bernardo, 2008; Collard and Mackill, 2008; St. Clair, 2010; Jian, 2015).

Why this large discrepancy between the number of published QTLs and those that are useful for a marker-assisted selection strategy? Reality indicates that a breeder will replace phenotypic selection by genotypic one only if the QTL on which the DNA marker was designed meets the following requirements: was clearly validated in different environments and genetic backgrounds, and explains a significant proportion of the phenotype variability. Otherwise, the breeder will not use the technology, avoiding the risk of making an inaccurate selection with the consequent loss of useful genetic variability.

The nature of a trait may sometimes suggest that much of the quantitative variation is controlled by many genes with small effects. Even if the effects for a large number of minor QTLs are consistent, pyramiding favorable alleles into a single cultivar becomes increasingly difficult or unfeasible. Examples of such traits are grain yield, quantitative disease resistance and tolerance to abiotic stresses.

To illustrate this situation, suppose the objective is pyramiding four favorable alleles located in four independent QTL. Suppose a cross between two inbred lines, each one carrying two of the QTLs of interest. Which will be the frequency of F_2-offspring carrying the four favorable alleles? Assuming Mendelian inheritance, the expected frequency of homozygotes at each *locus* will be 1/4, therefore the frequency of homozygotes for the four QTLs will be $(1/4)^4 = 1/256$. So, how large should be the F_2 population in order to have a probability of 0.95 to find at least one plant with favorable alleles in homozygous state at all four *loci*? The answer is the population should have 770 recombinant individuals. Even if you can build such population, what will happen with the genetic variability demanded for any breeding program? How big should be the F_2 population if the plan now is to obtain ten individuals with the four homozygous alleles? This simple example clearly shows that a breeding strategy based on pyramiding minor QTLs would be unfeasible.

The arising question is, can DNA markers help in order to develop MB strategies aiming to improve complex traits?

This challenge has led to the development of an alternative MB methodology named genomic selection (GS), genomic selection or genomewide selection (henceforth it will be referred to as GS), emerging as a valuable method for improving complex traits that are controlled by many QTLs with small effects. GS constitutes an approach in which all molecular markers available through the genome are used in order to calculate (predict) breeding values and it was firstly proposed by Meuwissen (2001) to be applied in animal breeding. However, the development of low-cost and high-throughput genotyping platforms has made possible the extension of GS to plant breeding (Rabier *et al.*, 2016; Crossa *et al.*, 2017; Juliana *et al.*, 2017). GS is typically performed among the progeny within a biparental cross between two elite inbreds (breeding population) where phenotypes and genomewide genotypes are investigated in the training population (a subset of the breeding population) to predict significant relationships between phenotypes and genotypes using statistical approaches. Marker effects estimated on the training population will be used to predict the performance of the best candidates in the rest of the breeding population solely based on genomic estimated breeding value (GEBV). Therefore, GS may result in lower costs because the need to evaluate the phenotype performance of the entire breeding population is replaced by a selection based on GEBV. Unlike QTL mapping, GS does not require to identify DNA markers with significant effects for a given trait.

For a better prediction accuracy of GS, a high density genotype is required so that all QTLs (which, as stated above, do not need to be identified) are in linkage disequilibrium with at least one SNP marker (Jiang 2015). The prediction accuracy is expected to increase as the product of heritability (h^2) and size of the training population (N) increases. A low h^2 can be compensated by the use of a large N. It is noteworthy that $N{\cdot}h^2$ determines both the power to detect a QTL and the accuracy of GS. Another important factor to consider is the density of DNA markers, because if it increases, then also accuracy will do. However this positive relationship is not linear, since once having reached a number of 200-500 SNPs, the increase of accuracy is not so evident, becoming unnecessary the increase marker density beyond a few hundred (Hickey *et al.*, 2014; Brandariz and Bernardo, 2019).

Different statistical methods have been developed to predict unobserved individuals in GS. Linear models (e.g., GBLUP) and machine-learning algorithms have been successful in making correct decisions based on genotype data. Also Kernel-based methods, such as Reproducing Kernel Hilbert Spaces (commonly known as RKHS), have extensively delivered good genomic predictions in plants. Several statistical models based on the standard GBLUP that incorporate genotype x environment (G x E) interactions in genomic and pedigree predictions have provided substantial increases in the accuracy of predicting individuals in non-assayed environments helping to exploit positive G x E interactions. Modeling multi-trait multi-environment is essential for improving the prediction accuracy of the performance of newly developed lines in future years. Application of GS in a breeding program should not be focused on predicting all individuals, but rather on classifying individuals into upper, middle, or lower classes, depending on the trait under selection (Crossa *et al.*, 2017).

GS is a promising breeding approach that, if used efficiently, provides the opportunity to increase the genetic gain per unit of time and cost. That is why GS is being adopted in plant improvement programs in several crops of commercial importance. However, while there are some efforts focused on the optimal distribution of resources, such as size of training population, marker density and structure of breeding population and their effect on the accuracy and cost of the selection model, more research is needed to cover these issues.

In the case of breeders who work in large seed companies, it is not necessary to convince them on the advantages provided by the application of MB strategies. However, this is not so simple for breeders leading successful improvement programs developed in small companies. Despite they may have heard or even know about the potential advantages that the use of MB strategies offers, the combination of lack of information on how to setup a marker-based approach together with the scarcity of economic resources, move them away from the practical application of DNA markers.

It is important to emphasize that it is not strictly necessary that a MB strategy must be complex and

sophisticated. In many cases, the only use of proved DNA markers for the selection of a simple trait or its use in the recovery of the recurrent genetic background in a backcross, provide an enormous advantage in increasing the genetic gain per time unit. Considering these simple applications as the starting point, their recurrent use can gradually increase the level of complexity, which may lead to GS. Of course, the initial investment demanded for setting up a Molecular Genetics facility is in most cases far away from small companies or public breeding programs. An alternative to solve this limitation could be the development of public-private consortia aimed to establish Molecular Genetics laboratories financed by the partners, which will also be the users. Encouraging thinking about such initiatives may be easier than one believes, and its concretion may allow more breeders to use marker assisted selection technologies, which will ultimately result in delivery of high-yielding crops, contributing to satisfy a growing global food demand.

REFERENCES

Bernardo R. (2008) Molecular Markers and Selection for Complex Traits in Plants: Learning from the Last 20 Years. Crop Sci. 48: 1649–1664.

Brandariz S.P., Bernardo R. (2019) Small ad hoc versus large general training populations for genomewide selection in maize biparental crosses. Theor. Appl. Genet. 132: 347-353.

Collard B.C., Mackill D.J. (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos. Trans. R. Soc. Lond. B Biol. Sci. 363: 557-572.

Crossa J., Pérez-Rodríguez P., Cuevas J., Montesinos-López O., Jarquín D., de los Campos G., Burgueño J., González-Camacho J., Pérez-Elizalde S., Beyene Y., Dreisigacker S., Singh R., Zhang X., Gowda M., Roorkiwal M., Rukoski J., Varshney R.K. (2017) Genomic selection in plant breeding: methods, models, and perspectives. Trends Plant Sci. 22: 961–975.

Dudley J.W., Moll R.H. (1969) Interpretation and use of estimates of heritability and genetic variances in plant breeding. Crop Sci. 9: 257–262.

Hickey J.M., Dreisigacker S., Crossa J., Hearne S., Babu R., Prasanna B.M., Grondona M., Zambelli A., Windhausen V.S., Mathews K., Gorjanc G. (2014) Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. Crop Sci. 54: 1476–1488

Jiang G.L. (2015) Molecular marker-assisted breeding: a plant breeder's review. In: Al-Khayri J., Jain S., Johnson D. (eds) Advances in plant breeding strategies: Breeding, Biotechnology and Molecular Tools. Springer, Cham.

Juliana P., Singh R.P., Poland J., Mondal S., Crossa J., Montesinos-López O.A., Dreisigacker S., Pérez-Rodríguez P., Huerta-Espino J., Crespo-Herrera L., Govindan V. (2018) Prospects and challenges of applied genomic selection—A new paradigm in breeding for grain yield in bread wheat. Plant Genome 11: 180017. doi: 10.3835/plantgenome2018.03.0017

Meuwissen T.H.E., Hayes B.J., Goddard M.E. (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819–1829

Platten J.D., Cobb J.N., Zantua R.E. (2019) Criteria for evaluating molecular markers: Comprehensive quality metrics to improve marker assisted selection. PLoS ONE 14: e0210529.

Rabier C-E., Barre P., Asp T., Charmet G., Mangin B. (2016) On the accuracy of genomic selection. PLoS ONE 11: e0156086.

St. Clair D.A. (2010) Quantitative disease resistance and quantitative resistance loci in breeding. Annu. Rev. Phytopathol. 48: 247-268.