(Formerly MENDELIANA)

BAG

Journal of Basic & Applied Genetics

July 2019
Volumen XXX
No. 1
E-ISNN: 1852-6322

Journal of the Argentine Society of Genetics
Revista de la Sociedad Argentina de Genética

www.sag.org.ar/jbag
Buenos Aires, Argentina

DOI: 10.35407/bag.2019.XXX.01

Journal of the
Argentine Society
of Genetics

BAG | Journal of Basic & Applied Genetics

V. XXX - No.1

July 2019

Included in:

Cited by:

Comité Editorial

Imágen de tapa:

Polinizando Traful-biodiversidad
Autora: Bárbara Casali

Contenidos

BAG | Journal of Basic & Applied Genetics

Vol XXX (1): 7-9; July 2019

Journal of the Argentine Society of Genetics

PARTIAL SEQUENCES OF THE GENE THAT CODIFIES FOR THE TRANSCRIPTION FACTOR *VPHSFB1* IN *Vasconcellea pubescens*. FIRST REPORT

SECUENCIAS PARCIALES DEL GEN QUE CODIFICA PARA EL FACTOR DE TRANSCRIPCIÓN *VPHSFB1* EN *Vasconcellea pubescens*. PRIMER REPORTE

Arizala-Quinto E. D [1], Viteri G. [1], Idrovo-Espín F.M. [1,2]

[1] Universidad de Las Américas UDLA. Facultad de Ingeniería y Ciencias Aplicadas. Calle José Queri s/n entre Av. Eloy Alfaro y Granados. Quito-Ecuador

[2] Universidad Central del Ecuador. Facultad de Ciencias Químicas. Francisco Viteri s/n y Gato Sobral. Quito-Ecuador

Corresponding author:
F.M Idrovo-Espín
f.idrovo@udlanet.ec

Cite this article as:
Arizala-Quinto E. D, Viteri G. , Idrovo-Espín F.M. 2019. PARTIAL SEQUENCES OF THE GENE THAT CODIFIES FOR THE TRANSCRIPTION FACTOR *VPHSFB1* IN *Vasconcellea pubescens*. FIRST REPORT. BAG. Journal of Basic and Applied Genetics XXX (1): 7-9.

Received: 12/12/2017
Accepted: 10/08/2018
Author's approval of galley proofs: 05/14/2019

General Editor: Elsa Camadro

DOI: 10.35407/bag.2019.XXX.01.01

ISSN online version: 1852-6233

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

ABSTRACT

Plant heat stress transcription factors (HSFs) are involved in the response to heat. In Arabidopsis thaliana the HSFs genes are completely identified, however there was no information available about these genes in Vasconcellea pubescens (Chamburo) until now. In this preliminary work we describe the VPHSFB1 gene of V. pubescens (gene expression evaluated by RT-PCR and the partial sequence) that was induced by the increment of temperature. From our results, VPHSFB1 could be used as a heat response marker gene in tropical species.

Key words: Caricaceae, gene expression, heat.

RESUMEN

Los factores de transcripción del estrés por calor en plantas (*HSFs*) están involucrados en la respuesta al calor. En *Arabidopsis thaliana* los genes *HSFs* están completamente identificados, sin embargo no había información disponible sobre estos genes en *Vasconcellea pubescens* (Chamburo) hasta ahora. En este trabajo preliminar describimos el gen *VPHSFB1* de *V. pubescens* (expresión génica evaluada por RT-PCR y la secuencia parcial) que fue inducido por el incremento de temperatura. A partir de nuestros resultados, se podría usar a *VPHSFB1* como un gen marcador de respuesta a calor en especies tropicales.

Palabras clave: Caricaceae, expresión génica, calor.

INTRODUCTION

Plant heat stress transcription factors (*HSFs*) are essential components of the signal transduction involved in the expression of genes responsive to this kind of abiotic stress (Nover *et al.*, 2001). In *A. thaliana* 21 members of *HSFs* belonging to three genes classes A, B and C, have been identified (Kotak *et al.*, 2004). Among these, *ATHSFB1* (Class B) is necessary for the expression of heat stress inducible genes (as heat shock protein genes) that are involved in thermotolerance (Ikeda et al., 2011).

Caricaceae is a family composed by six genera, two of them are *Vasconcellea* and *Carica*. The 21 species that belong to genus *Vasconcellea* (collectively known as highland papayas) are distributed in South America, endemically in some countries, as Ecuador (Scheldeman *et al.*, 2011). It has been estimated that *Vasconcellea* diverged from *Carica* 25 Ma ago (Carvahlo and Renner, 2012).

More specifically the exotic species *V. pubescens* has interesting properties and uses, ranging from high levels of antioxidants (Simirgiotis *et al.*, 2009), gastric ulcers treatments (Mello *et al.*, 2008), dermal antitumoral therapy (Dittz *et al.*, 2015) to biofilm production based on Papain against cavities (Torres and Obando, 2016).

In this preliminary work, we report the partial sequence of the *V. pubescens VPHSFB1* gene, a phylogenetic analysis with related sequences and the expression banding pattern of *VPHSFB1* after temperature increase.

MATERIALS AND METHODS

Oligonucleotides for RT-PCR amplification and further sequencing of the amplicons were designed from the *CPHSFB1* gene reported by Tarora *et al.* (2010). Germinated seedlings (75 days old) were subjected to increment of temperature (from 25° C to 33° C or 45° C) for a period of 4 hs; seedlings at 25° C were used as controls. After applying the temperature treatment, RNA was extracted from leaves (PureLink RNA MiniKit, Ambion), then RT-PCR was performed (Superscript III One Step RT-PCR, Invitrogen) and, finally, agarose gel electrophoresis (1% agarose, 45 min, 80 volts) was performed and documented. Amplicons were sequenced twice in UDLA research laboratory (ABI 3130 Genetic Analyzer). Phylogenetic analysis was made in comparison with *HSFs* selected sequences with MEGA7 (Kumar *et al.*, 2016).

RESULTS AND DISCUSSION

Phylogenetic analysis of partial sequences of the VPHSFB1gene

From a PCR product (plants at 25° C) we obtained two partial sequences of *V. pubescens* heat stress transcription factor (Figure 1), hereinafter referred to as *VPHSB1a* (340 bp) and *VPHSB1b* (330 bp).

Despite the fact that the sequences were only fragments of the *VPHSFB1* gene, the phylogenetic tree (Figure 2) exhibited one major clade comprising the *HSF* sequences of *V. pubescens*, *A. thaliana*, *C. papaya* and *Brassica rapa*. Within this clade, a subclade was formed with the Caricaceae members; this was the expected topology since *V. pubescens* and papaya are more related between them than with Arabidopsis. The other sequences in this analysis remained unsolved. Interestingly, the sequences in the Caricaceae subclade seemed to have accumulated changes earlier that the ancestral lineage split between Arabidopsis and *Brassica*. This may have been because *V. pubescens* and papaya are strictly tropical species, as Carvahlo and Renner (2012) have shown in their biogeographic study. Therefore, it is feasible that Caricaceae developed specialized *HSF* genes in order to cope with higher temperatures earlier than Arabidopsis or *Brassica*, which are less adapted to tropical climates.

From the alignment of all sequences (not shown), the highest identity percentages were obtained by comparing *VPHSFB1* with *CPHSFB1*, thus, we conclude that these sequences are orthologs among them.

Expression banding pattern of the VPHSFB1 gene

Although the expression of *VPHSFB1* is constitutive at the assayed temperatures, the intensity of bands obtained by gel electrophoresis (Figure 3) increased at higher temperatures. Previously Tarora *et al.* (2010) characterized the ortholog *CPHSFB1* gene in papaya. In a Northern blot analysis, it was observed that this gene accumulated transcripts differentially after temperature increase (from 24° C to 42° C) and, thus it is responsive to heat stress. This behavior is similar to the observed in our analysis, which revealed the involvement of *VPHSFB1* in the response to temperature increment and, probably, in heat stress.

We conclude that an ortholog *VPHSFB1* gene is present in the genome of *V. pubescens*, which is responsive to temperature increment, and that this gene could be used as a marker for heat stress assays in this tropical species.

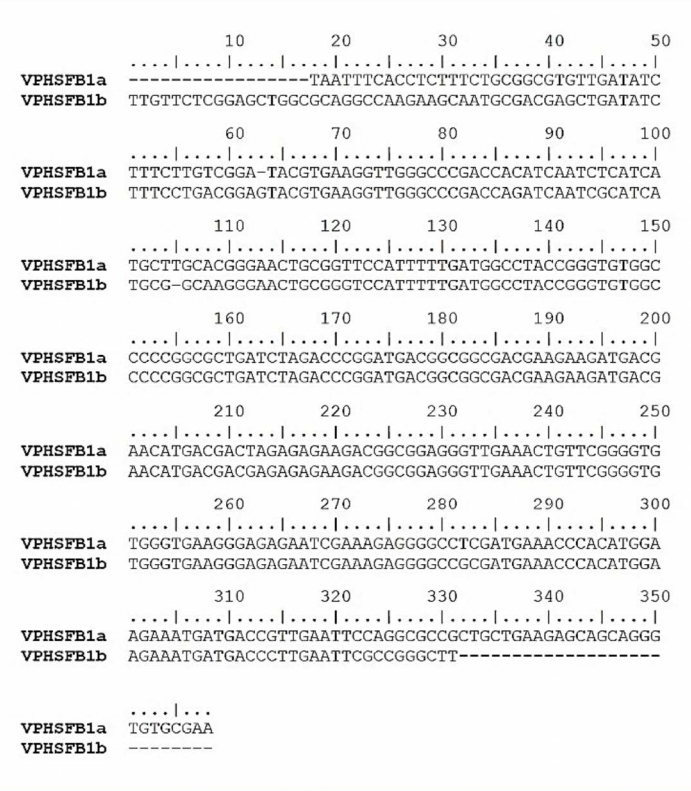
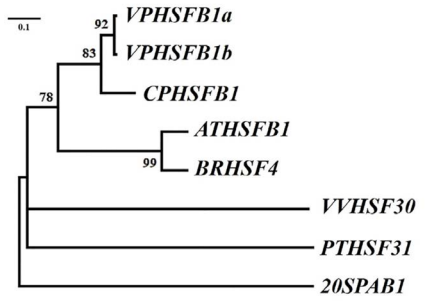


Figure 1. Clustal w alignment of partial sequences of the *VPHSFB1 gene.*



	VPHSFB1a	VPHSFB1b	ATHSFB1
VPHSFB1b	95,19		
ATHSFB1	66,31	66,84	
CPHSFB1	**86,10**	**88,24**	66,84

Figure 2. Maximum Likelihood phylogenetic tree based on a MUSCLE alignment of partial selected sequences *HSFs* genes (*C. papaya* CPHSFB1/AB506766.1, *A. thaliana ATHSFB1/ AT4G36990*, *Brassica rapa BRHSF*/EU186351.1, *Populus trichocarpa PTHSF31*/GI566202080, *Vitis vinifera VVHSF30*/NM001303086.1). The tree was rooted with 20SPAB1 *(AT1G16470.1)* that encodes for the 20S proteasome subunit PAB1 in *A. thaliana* (Iida *et al.*, 2009). The identity percentage of orthologs from *V. pubescens*, *A. thaliana* and *C. papaya* are shown below.

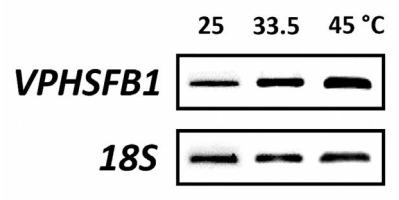


Figure 3. Banding pattern obtained from control plants (25° C) and plants under temperature increase (33.5° C and 45° C). 18S gene expression was used as a positive control. Controls with no template showed any band. The assay was made in triplicates with similar results.

REFERENCES

Carvalho F.A., Renner S.S. (2012) A dated phylogeny of the papaya family (Caricaceae) reveals the crop's closest relatives and the family's biogeographic history. Mol. Phylogenet. Evol. 65: 46-53.

Dittz D., Figueiredo C., Lemos F., Viana C., Andrade S., Souza-Fagundes E., Fujiwara R.T., Salas C.E., Lopes M. (2015) Antiangiogenesis, Loss of Cell Adhesion and Apoptosis Are Involved in the Antitumoral Activity of Proteases from *V. cundinamarcensis* (*C. candamarcensis*) in Murine Melanoma B16F1. Int. J. Mol. Sci. 16: 7027-7044.

Iida K., Fukami-Kobayashi K., Toyoda A., Sakaki Y., Kobayashi M., Seki M., Shinozaki K. (2009) Analysis of multiple occurrences of alternative splicing events in Arabidopsis thaliana using novel sequenced full-length cDNAs. DNA Res. 16: 155-164.

Ikeda M., Mitsuda N., Ohme-Takagi M. (2011). Arabidopsis HsfB1 and HsfB2b act as repressors of the expression of heat-inducible Hsfs but positively regulate the acquired thermotolerance. Plant Physiol. 157(3): 1243-1254.

Kotak S., Port M., Ganguli A., Bicker F., von Koskull-Doring P. (2004) Characterization of C-terminal domains of Arabidopsis heat stress transcription factors (Hsfs) and identification of a new signature combination of plant class A Hsfs with AHA and NES motifs essential for activator function and intracellular localization. Plant J. 39: 98-112.

Kumar S., Stecher G., Tamura K. (2016) MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33: 1870-1974.

Mello V., Gomes M., Lemos F., Delfino J., Andrade S., Lopes M., Salas C. (2008) The gastric ulcer protective and healing role of cysteine proteinases from *Carica candamarcensis*. Phytomedicine 15: 237-244.

Nover L., Bharti K., Döring P., Kumar M.S., Ganguli A., Scharf K.D. (2001) Arabidopsis and the heat stress transcription factor world: how many heat stress transcription factors do we need? Cell Stress Chaperon 6: 177-189.

Scheldeman X., Kyndt T., Coppens d'Eeckenbrugge G., Ming R., Drew R., Droogenbroeck B., Van Damme P., Moore P. (2011). Vasconcellea. In: Kole C. (Ed.) Wild Crop Relatives: Genomic and Breeding Resources. Springer-Verlag, Berlin, pp. 213-249.

Simirgiotis M., Caligari P., Schmeda-Hirschmann G. (2009) Identification of phenolic compounds from the fruits of the mountain papaya *Vasconcellea pubescens* A. DC. grown in Chile by liquid chromatography-UV detection-mass spectrometry. Food Chem. 115: 775-784.

Tarora K., Tamaki M., Shudo A., Urasaki N., Matsumura H., Adaniya S. (2010) Cloning of a heat stress transcription factor, CphsfB1, that is constitutively expressed in radicles and is heat-inducible in the leaves of *Carica papaya*. Plant Cell Tissue Organ Cult. 102: 69-77.

Torres K., Obando G. (2016) Rapid enzimatic disruption of *Enterococcus faecalis* biofilm using *Carica pubescens*: a pilot study. WMCCR 2: 1-4.

ACKNOWLEDGEMENTS

To the Ministry of the Environment of Ecuador (MAE) for the permissions granted (MAE-DNB-CM-2017-0063), to Lien González for her support and valuable comments. This work was supported by UDLA (grant INV/F/PPI/1/0814).

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: 10.35407/bag.2019.XXX.01.02

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.

BAG | Journal of Basic & Applied Genetics

Vol XXX (1): 17-23; July 2019

Journal of the Argentine Society of Genetics

TEST OF INTERACTION IN THE ANALYSIS OF MOLECULAR VARIANCE

PRUEBA DE INTERACCIÓN EN EL ANÁLISIS MOLECULAR DE VARIANZA

Bruno C. [1,2], Videla M.E. [1,2], Balzarini M. [1,2]

[1] Consejo Nacional de Investigaciones Científicas y Técnicas-CONICET

[2] Estadística y Biometría. Facultad de Ciencias Agropecuarias. Universidad Nacional de Córdoba

Corresponding author:
Cecilia Bruno
cebruno@agro.unc.edu.ar

Cite this article as:
Bruno C., Videla M.E., Balzarini M. 2019. TEST OF INTERACTION IN THE ANALYSIS OF MOLECULAR VARIANCE. BAG. Journal of Basic and Applied Genetics XXX (1): 17-23.

Received: 05/04/2019
Accepted: 06/24/2019

General Editor: Elsa Camadro

DOI: 10.35407/bag.2019.XXX.01.03

ISSN online version: 1852-6233

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

ABSTRACT

The genomic diversity, expressed in the differences between molecular haplotypes of a group of individuals, can be divided into components of variability between and within some factor of classification of the individuals. For such variance partitioning, molecular analysis of variance (AMOVA) is used, which is constructed from the multivariate distances between pairs of haplotypes. The classical AMOVA allows the evaluation of the statistical significance of two or more hierarchical factors and consequently there is no interaction test between factors. However, there are situations where the factors that classify individuals are crossed rather than nested, that is, all the levels of a factor are represented in each level of the other one. This paper proposes a statistical test to evaluate the interaction between crossed factors in a Non-Hierarchical AMOVA. The null hypothesis of interaction establishes that the molecular differences between individuals of different levels of a factor are the same for all the levels of the other factor that classifies them. The proposed analysis of interaction in a Non-Hierarchical AMOVA includes: calculation of the distance matrix and partition of it into blocks, subsequent calculation of residuals and analysis of non-parametric variance on the residuals. Its implementation is illustrated in simulated and real scenarios. The results suggest that the proposed interaction test for the Non-Hierarchical AMOVA presents high power.

Key words: genetic variability, non-parametric methods, distances matrix, AMOVA.

RESUMEN

La diversidad genómica, expresada en las diferencias entre haplotipos moleculares de un conjunto de individuos, puede dividirse en componentes de variabilidad entre y dentro de algún factor de clasificación de los individuos. Para tal partición de varianzas, se usa análisis molecular de la varianza (AMOVA), el cual se construye a partir de las distancias multivariadas entre pares de haplotipos. El AMOVA clásico permite evaluar la significancia estadística de dos o más factores jerárquicos y consecuentemente no existe prueba de interacción entre factores. Sin embargo, existen situaciones donde los factores que clasifican a los individuos están cruzados y no anidados, es decir todos los niveles de un factor se encuentran representados en cada nivel del otro factor. Este trabajo propone una prueba estadística para evaluar la interacción entre factores cruzados en un AMOVA No-Jerárquico. La hipótesis nula de interacción establece que las diferencias moleculares entre individuos de distintos niveles de un factor son las mismas para todos los niveles del otro factor que los clasifica. La propuesta de análisis de interacción de factores a partir de distancias en un AMOVA No-Jerárquico comprende: cálculo de la matriz de distancia y partición de la misma en bloques, posterior cálculo de residuos y análisis de varianza no-paramétrico sobre los residuos. Su implementación es ilustrada en escenarios simulados y real. Los resultados sugieren que la prueba de interacción propuesta para el AMOVA No-Jerárquico presenta alta potencia.

Palabras clave: variabilidad genética, métodos no-paramétricos, matrices de distancias. AMOVA.

INTRODUCCIÓN

La estructura genética de poblaciones puede analizarse mediante la comparación de las frecuencias alélicas (Kennington *et al.*, 2003; Hedrick, 2005), mediante métricas de distancias genéticas (Nei, 1973), usando algoritmos de clasificación (Pritchard *et al.*, 2000) y/o con el análisis molecular de la varianza (AMOVA, Excoffier, 1992). La mayoría de los métodos basados en frecuencias alélicas involucran transformaciones no lineales de los datos genéticos y son válidos sólo bajo una serie de supuestos que deben realizarse respecto a los procesos evolutivos subyacentes. Por el contrario, la información sobre divergencia a nivel molecular procesada en el formato de una partición de Análisis de la Varianza (ANAVA) demanda menos supuestos biológicos. Debido a la dimensionalidad de los datos genómicos (naturaleza multivariada), el ANAVA se obtiene a partir de las métricas de distancias entre los pares de haplotipos de ADN. Debido a relaciones entre sumas de cuadrado (SC) y sumas de distancias al cuadrado, la SC asociada con cualquier término de un ANAVA puede ser calculada directamente a partir de las distancias (Gower, 1966; Li, 1976). En el AMOVA se descompone la diversidad genómica (expresada por las diferencias en el total de haplotipos moleculares) como la suma de componentes de variabilidad entre y dentro de grupos de individuos. Estos grupos son conformados por uno o más factores de clasificación usualmente anidados, es decir los niveles de un factor pueden ser distintos para cada nivel del otro factor. Un ejemplo usual de este análisis de variabilidad molecular es el estudio de diferencias entre regiones, entre poblaciones dentro de cada región y dentro de poblaciones. Generalmente, se atribuye una proporción aditiva de la variabilidad total a cada uno de los factores presentes en el diseño del estudio (*i.e.* región, población). La comparación de estos componentes de varianza permite inferir la magnitud de la estructuración genética en el conjunto de todos los haplotipos moleculares bajo estudio.

Los cálculos para estimar componentes de varianza entre y dentro de subgrupos de una estructura jerárquica de factores, se realizan desde la matriz de distancias Euclídeas (al cuadrado) y se usan para contrastar hipótesis sobre variabilidad entre y dentro de grupos (Dyer, 2017). Sin embargo, existen situaciones de estructura de datos, provenientes de estudios experimentales u observacionales, donde los factores que clasifican las muestras se encuentran cruzados en lugar de estar anidados en una estructura jerárquica. Es decir, todos los niveles de un factor se encuentran representados en cada nivel del otro factor. Por ejemplo, en un estudio donde se colecten muestras de haplotipos relacionadas a 3 especies (3 niveles para el factor especie) en cada una de 4 regiones (4 niveles para el factor región), pero los niveles del factor especie son los mismos en cada uno de los niveles del factor región, es necesario evaluar la significancia de la interacción entre especie y región para conocer si las diferencias entre especies dependen de la región desde la que se extrajeron las muestras de haplotipos. Luego, en el ANAVA de perfiles moleculares (AMOVA) No-Jerárquico, es de interés responder si las diferencias moleculares entre los niveles de un factor son las mismos para todos los niveles del otro factor interviniente en la clasificación de los haplotipos. Éste, es el interrogante que se pretende responder con pruebas de interacción entre factores de clasificación no-jerárquicos. Por el contrario, si el muestreo involucra diferentes niveles de un factor en cada uno de los diferentes niveles del otro factor, se usará un AMOVA clásico (AMOVA Jerárquico). En el AMOVA clásico, la prueba de interacción no es factible debido a que no existen muestras moleculares para todas las combinaciones de niveles de los factores intervinientes.

Anderson (2001) propuso un método denominado PERMANOVA, análogo al Análisis de la Varianza Multivariado, particionando las sumas de cuadrado asociadas a sub-matrices de matrices de distancias relacionadas a los factores de clasificación de un conjunto de muestras moleculares. El método puede usarse en diseños con factores de clasificación cruzado, pero sólo en situaciones donde los datos son balanceados, *i.e.*, todos los niveles de los factores tienen la misma cantidad de datos y no falta ningún nivel (Anderson y ter Braack, 2003). La significancia estadística de la interacción entre factores en PERMANOVA es aproximada por permutación.

En este trabajo se propone un método, de inferencia no paramétrica, para evaluar la significancia de la interacción entre factores que es aplicable a estudios moleculares con poblaciones clasificadas a dos vías con estructura cruzada de factores.

MATERIALES Y MÉTODOS

Distancias moleculares

Supongamos que existen n muestras de haplotipos caracterizados con m marcadores moleculares. Si los marcadores se expresan como variables binarias, para cada muestra es posible conformar una observación multivariada (m-dimensional) que lleva valores 1 o 0 para cada uno de los marcadores según el marcador esté presente o ausente en la muestra, respectivamente. El vector booleano m-dimensional denotado como $\mathbf{p'}=[p_1,p_2,..,p_m]$ donde $p_i=1$ con ic1,..., m si el marcador m está presente y $p_i=0$ si está ausente, constituye la observación multivarada a analizar en cada muestra. La diferencia entre dos muestras y_j *vs.* y_k es definida como $\mathbf{p}_j-\mathbf{p}_k$. Se define una métrica de distancia Euclídea entre

las muestras y_j e y_k como $d^2_{jk} = (\mathbf{p}_j - \mathbf{p}_k)' \mathbf{W} (\mathbf{p}_j - \mathbf{p}_k)$ donde **W** es una matriz de pesos que pueden ser diferenciables para los distintos marcadores. Si todos los marcadores se asumen independientes e igualmente informativos entonces **W=I** y la métrica de distancia es igual al número de diferencias entre las dos muestras, *i.e.* complemento a uno del índice de similitud Emparejamiento Simple (Apostol *et al.*, 1993). Una distancia comúnmente usada para análisis de similitud o diferencia molecular es la distancia de Excoffier (Excoffier *et al.*, 1992)

$$D_{Excoffier} = (a + d + c + d)\left(1 - \frac{a+d}{a+d+c+d}\right) \quad [1]$$

donde, *a*, *b*, *c*, y *d* representan las frecuencias de los eventos (1,1), (1,0), (0,1) y (0,0) respectivamente, que surgen al comparar dos individuos para cada alelo. En ésta las disimilitudes o faltas de coincidencia son expresadas como fracción del número total de marcadores que participan en la comparación de cada par de perfiles. Si todos los individuos que se comparan son genotipados con el mismo número de marcadores entonces esta distancia es sólo un múltiplo del complemento a 1 del índice de Emparejamiento Simple (Apostol *et al.*, 1993).

$$D_{ES} = 1 - \left[\frac{a+d}{a+d+c+d}\right] \quad [2]$$

Interacción en Análisis de la Varianza Molecular

A continuación, se propone una prueba estadística desarrollada para evaluar la significancia de la interacción *A*×*B* en un diseño con dos factores A y B cruzados que tiene como *input* a las distancias moleculares entre haplotipos.

La matriz de distancias entre pares de individuos se particiona de manera tal que se identifiquen los siguientes bloques de distancias: 1) Bloque de distancias "Dentro": comprende inter-distancias entre individuos de un mismo grupo, por tanto existen tantos bloques de este tipo como grupos (combinaciones de niveles de los factores) haya; 2) Bloque de distancias "Entre factor *C* dentro de factor *A*", existen tantos bloques como niveles del factor *A*;3) Bloque de distancias "Entre factor *A* dentro de factor *B*" existen tantos bloques como niveles del factor *B* y 4) Bloque de distancias "Entre factor *A* y entre factor *B*", con tantos bloques como (*B*-1) *B* (Fig. 1).

Se obtienen las distancias promedio para cada bloque de distancias (B1 a B4), luego se obtiene el valor absoluto de la diferencia entre una distancia del bloque y la media de las distancias del mismo bloque. Sobre el valor absoluto de la diferencia de cada distancia respecto a su media, se ajusta un análisis de varianza no-paramétrico de Kruskal Wallis (Conover, 1999).

A

Matriz de distancia		Factor A	A_1	A_1	A_1	A_1	A_1	A_1	A_2	A_2	A_2	A_2	A_2	A_2
		Factor B	B_1	B_1	B_1	B_2	B_2	B_2	B_1	B_1	B_1	B_2	B_2	B_2
Factor A	Factor B	Ind	1	2	3	4	5	6	7	8	9	10	11	12
A_1	B_1	1	0											
A_1	B_1	2	1	0										
A_1	B_1	3	1	2	0									
A_1	B_2	4	9	8	8	0								
A_1	B_2	5	9	8	8	2	0							
A_1	B_2	6	10	9	9	1	1	0						
A_2	B_1	7	6	5	5	3	5	4	0					
A_2	B_1	8	6	5	5	3	3	4	2	0				
A_2	B_1	9	7	6	6	2	4	3	1	3	0			
A_2	B_2	10	3	2	4	6	6	7	3	3	4	0		
A_2	B_2	11	3	2	4	6	6	7	3	3	4	2	0	
A_2	B_2	12	4	3	5	5	5	6	2	2	3	1	1	0

B

Bloques		Factor A	A_1	A_1	A_1	A_1	A_1	A_1	A_2	A_2	A_2	A_2	A_2	A_2
		Factor B	B_1	B_1	B_1	B_2	B_2	B_2	B_1	B_1	B_1	B_2	B_2	B_2
Factor A	Factor B	Ind	1	2	3	4	5	6	7	8	9	10	11	12
A_1	B_1	1	B_1			B_2			B_3			B_4		
A_1	B_1	2												
A_1	B_1	3												
A_1	B_2	4	B_2			B_1			B_4			B_3		
A_1	B_2	5												
A_1	B_2	6												
A_2	B_1	7	B_3			B_4			B_1			B_2		
A_2	B_1	8												
A_2	B_1	9												
A_2	B_2	10	B_4			B_3			B_2			B_1		
A_2	B_2	11												
A_2	B_2	12												

C

		Factor A	A_1	A_1	A_1	A_1	A_1	A_1	A_2	A_2	A_2	A_2	A_2	A_2
		Factor B	B_1	B_1	B_1	B_2	B_2	B_2	B_1	B_1	B_1	B_2	B_2	B_2
Factor A	Factor B	Ind	1	2	3	4	5	6	7	8	9	10	11	12
A_1	B_1	1	d_{w1}			$d_{B[A]}$								
A_1	B_1	2												
A_1	B_1	3												
A_1	B_2	4	$d_{B[A]}$			d_{w2}								
A_1	B_2	5												
A_1	B_2	6												
A_2	B_1	7	$d_{A[B]}$			d_{AB}			d_{w1}					
A_2	B_1	8												
A_2	B_1	9												
A_2	B_2	10	d_{AB}			$d_{A[B]}$						d_{w2}		
A_2	B_2	11												
A_2	B_2	12												

Figura 1. Esquema ilustrativo de las distancias entre individuos y los bloques conformados por la partición de la matriz de distancia usados para calcular el Análisis Molecular de la Varianza (AMOVA) No-jerárquico. Panel A distancias entre individuos. Panel B partición de la matriz de distancia en bloques. Panel C distancias promedio para cada bloque conformado por la partición de la matriz de distancia.

Nota: B_1: Bloques de distancia "Dentro", B_2: Bloque de distancias entre factor *B*[*A*], B_3: Bloque de distancia entre *A*[*B*], B_4: Bloque de distancia entre factor *A* y entre factor *B*. Los valores d_{w1} y d_{w2} representan los valores absolutos de las diferencias entre una distancia entre individuos del mismo grupo, conformado por la misma combinación de niveles de ambos factores, respecto a la distancia promedio del mismo bloque. $d_{B[A]}$ es el valor absoluto de la diferencia de cada distancia entre los individuos del factor B dentro del factor A respecto a la distancia promedio del bloque, $d_{A[B]}$ es el valor absoluto de las diferencias entre individuos del factor A dentro del factor B respecto a la distancia promedio del bloque, d_{AB} valor absoluto de la diferencia entre factor A y entre factor B respecto a la distancia promedio del bloque.

Evaluación de la prueba de interacción

El método propuesto se evaluó sobre datos simulados y se ilustró su aplicación sobre un conjunto de datos experimentales. Para la evaluación por simulación de la prueba propuesta para la interacción, se simularon dos situaciones hipotéticas para una matriz $n \times m$ donde n=24 observaciones y m=20 marcadores moleculares codificados como binarios (presencia/ausencia). El diseño de experimentos fue bifactorial con dos niveles cada factor. Las situaciones hipotéticas fueron construidas bajo dos escenarios: (A) Interacción estadísticamente significativa. Esto representa que la hipótesis nula es falsa, es decir que al menos una de las diferencias entre los bloques definidos sobre la matriz de distancias es distinta de cero y (B) No hay interacción entre los factores. En términos estadísticos, este escenario implica que la hipótesis nula es verdadera.

En cada una de las situaciones (A) y (B), se simularon 100 bases de datos a partir de una perturbación introducida a través de una distribución binomial con parámetros n=1 y p=0.20. Sobre cada una de las 100 nuevas matrices de datos binarios creados para cada una de las situaciones, se calcularon tres medidas de distancia entre todos los pares de individuos: la distancia de Excoffier, el complemento a uno del índice de emparejamiento simple y la distancia de Bray-Curtis (Bray y Curtis, 1957).

Se calcularon las distancias promedio para cada bloque de distancias a partir del cual se reagruparon los datos. Sobre el valor absoluto de la diferencia de la distancia entre todos los pares de individuos y la media del bloque al que pertenece según el re-agrupamiento, se ajustó un análisis de la varianza no-paramétrico de Kruskal Wallis. Para estimar el tamaño de muestra se simularon 100 corridas o ajustes del análisis propuesto y bajo hipótesis nula verdadera, se contó la cantidad de veces que la prueba no aceptó la hipótesis es decir donde se rechazó una hipótesis verdadera que suponía la no existencia de interacción. Este valor estimado empíricamente representa el "tamaño de la prueba".

Para estimar la potencia de la prueba, al menos empíricamente, se simularon para la situación (A) diferentes niveles de interacción: (A.a) interacción alta, donde las diferencias entre los individuos de los niveles del factor B para un mismo nivel del factor A eran altas respecto a las diferencias entre los niveles de los factores B con el otro nivel de A donde esas diferencias eran prácticamente nulas, (A.b) interacción media, donde la diferencia entre individuos de los niveles de B que se encuentran bajo un mismo nivel del factor A presentan una diferencia máxima en el 60% de sus loci y (A.c) interacción baja; la diferencia máxima se da solo en un 30% de los loci. Para conocer la potencia de la prueba, se contó en cada situación la cantidad de veces que el valor p del análisis de la varianza no-paramétrico arrojó valores por debajo del nivel de significación (α =0.05).

Los resultados obtenidos fueron comparados con los resultados arrojados por el software PERMANOVA con la distancia de Bray-Curtis y distancia Euclídea (Anderson, 2001).

El conjunto de datos experimentales usado para ilustrar la prueba de interacción involucra muestras de ADN provenientes de un patógeno que habita en distintos cultivos agrícolas. Las muestras fueron recolectadas a partir de 12 plantas infectadas, que se clasificaron según el tipo de hospedero (cultivo de invierno y cultivo de verano), y también según la región. Los dos tipos de hospederos se encontraban en cada región. Se obtuvieron 9 marcadores moleculares polimórficos que fueron codificados como binarios (presencia/ausencia) para cada muestra.

RESULTADOS

En la Tabla 1 se presenta para cada situación simulada (A.a: interacción alta, A.b: interacción media, A.c: interacción baja y B: interacción nula) las tasas de error estimadas para la prueba propuesta en este trabajo (AMOVA No-Jerárquico) y para la prueba de interacción implementada en el software PERMANOVA v.1.6 (Anderson y ter Braak, 2003). Ambas fueron aplicadas teniendo como *input* una matriz de distancias entre individuos de dimensión 24×24 y calculada en base a la métrica de Excoffier.

Los resultados en Tabla 1 sugieren que la prueba de interacción propuesta para el AMOVA No-Jerárquico presenta una alta potencia, complemento a uno de la Tasa de Error de Tipo II, para detectar el efecto de interacción entre los factores, aun cuando este efecto es bajo. PERMANOVA v.1.6 a través de 4999 corridas de permutación de filas, produjo también buena potencia aunque siempre menor que el método propuesto. En escenarios donde la diferenciación entre grupos es mayor, como los casos de alta y media interacción, las Tasas de Error II alcanzada por el método propuesto fueron bajas (0.05 en condiciones de alta interacción y del 0.2 en situaciones de media interacción). La ventaja del método propuesto para evaluar interacción en un AMOVA No-Jerárquico es que no necesita que el diseño del estudio tenga el mismo número de repeticiones para cada nivel de factores como requiere PERMANOVA.

Para implementar la prueba estadística de interacción a través del AMOVA No-Jeráquico propuesto, se construyeron las matrices de distancia entre los perfiles moleculares de los genomas del patógeno genotipado en hospedantes de verano e invierno en dos zonas agrícolas (Fig. 2). El ANOVA No-Jerárquico detectó una interacción estadísticamente significativa entre los niveles de los factores, es decir que el perfil molecular del patógeno es diferente según el cultivo donde se hospede y la región

en la que se encuentre. El valor del estadístico Kruskal-Wallis fue de 4.4262 con un valor-p=0.03539. Para el mismo conjunto de datos, PERMANOVA no encontró una interacción estadísticamente significativa. Los análisis se realizaron en el software R (R Core Team, 2013) con el código que ponemos a disposición del lector en el repositorio GitHub @estadistica-aplicada (https://github.com/estadistica-aplicada). El algoritmo para evaluar la interacción en AMOVA No Jerárquico también fue incorporado al software Info-Gen (Balzarini y Di Rienzo, 2018).

Tabla 1. Tasas de Error Tipo II obtenidas por simulación para la prueba de hipótesis de la interacción, bajo situaciones de alta, media y baja interacción y Tasas de Error Tipo I bajo interacción nula, de los procedimientos AMOVA No-Jerárquico y PERMANOVA con el método de permutación de filas.

Procedimiento	Interacción			
	Alta	Media	Baja	Nula
FIDA	0.05	0.2	0.28	0.097
PERMANOVA	0	0	0.15	0.05

Nota: por definición el Error de Tipo I se calcula bajo hipótesis nula cierta, es decir, interacción nula y el Error de Tipo II se calcula bajo hipótesis nula falsa, *i.e.*, hay interacción.

A

Matriz de distancia		Factor A	A_1	A_1	A_1	A_1	A_1	A_1	A_2	A_2	A_2	A_2	A_2	A_2
		Factor B	B_1	B_1	B_1	B_2	B_2	B_2	B_1	B_1	B_1	B_2	B_2	B_2
Factor A	Factor B	Ind	1	2	3	4	5	6	7	8	9	10	11	12
A_1	B_1	1	**0**											
A_1	B_1	2	**3**	**0**										
A_1	B_1	3	**2**	**3**	**0**									
A_1	B_2	4	6	5	8	**0**								
A_1	B_2	5	4	5	6	**4**	**0**							
A_1	B_2	6	5	6	3	**7**	**7**	**0**						
A_2	B_1	7	2	1	2	6	6	5	**0**					
A_2	B_1	8	2	1	2	6	6	5	**0**	**0**				
A_2	B_1	9	5	6	7	3	1	6	**7**	**7**	**0**			
A_2	B_2	10	6	5	6	2	6	5	6	6	5	**0**		
A_2	B_2	11	6	5	6	4	2	5	6	6	1	**6**	**0**	
A_2	B_2	12	3	6	5	5	3	4	5	5	2	**7**	**3**	**0**

B

Matriz de residuos		Factor A	A_1	A_1	A_1	A_1	A_1	A_1	A_2	A_2	A_2	A_2	A_2	A_2
		Factor B	B_1	B_1	B_1	B_2	B_2	B_2	B_1	B_1	B_1	B_2	B_2	B_2
Factor A	Factor B	Ind	1	2	3	4	5	6	7	8	9	10	11	12
A_1	B_1	1	0											
A_1	B_1	2	0.33	0										
A_1	B_1	3	0.67	0.33	0									
A_1	B_2	4				0								
A_1	B_2	5				2	0							
A_1	B_2	6				1	1	0						
A_2	B_1	7							0					
A_2	B_1	8							4.67	0				
A_2	B_1	9							2.33	2.33	0			
A_2	B_2	10										0		
A_2	B_2	11										0.67	0	
A_2	B_2	12										1.67	2.33	0

Figura 2. Esquema ilustrativo de las distancias entre individuos estimadas desde perfiles multivariados codificados como binarios y los bloques conformados por la partición de la matriz de distancia usados para evaluar interacción en (AMOVA) No-jerárquico. Referencias: El Factor A representa la región del cultivo, con dos niveles (A_1 y A_2) y el Factor B representa si el cultivo hospedero del patógeno es de invierno (B_1) o verano (B_2). Panel A distancias entre individuos. Panel B residuos estimados como la diferencia de la distancia entre un par de individuos dentro del mismo bloque y la distancia promedio del bloque.

DISCUSIÓN

En este trabajo se propone una prueba estadística para analizar la interacción entre factores en un AMOVA No-Jerárquico. La hipótesis nula de interacción establece que las diferencias entre perfiles moleculares agrupados por un factor de clasificación son las mismas para cada nivel del factor de clasificación con el que se ha cruzado. A diferencia del AMOVA de Excoffier (Excoffier *et al.*, 1992) y de PERMANOVA (Anderson, 2001) métodos que permiten evaluar el efecto de factores de clasificación a partir de distancias moleculares, la prueba estadística presentada en nuestro trabajo es específica de interacción y puede implementarse incluso en situaciones de desbalance de datos. Cuando la interacción es estadísticamente significativa, no se recomienda realizar pruebas de efectos principales.

La prueba de interacción que proponemos no demanda la elección de un método de permutación para hallar un pseudo-F que determine la significancia estadística de la interacción de factores, como lo hace PERMANOVA y goza de las propiedades del método no paramétrico de Kruskal-Wallis (Conover, 1999). El algoritmo comprende: cálculo de la matriz de distancia y partición de la misma en bloques de distinto tipos de distancia, posterior cálculo de residuos, análisis de varianza no-paramétrico sobre esos residuos y cálculo del valor-p para determinar la significancia estadística. A través del análisis de varianza no-paramétrico (AMOVA No-Jerárquico) se evalúa la hipótesis de homogeneidad de varianzas dentro de los distintos bloques. El valor promedio de la variable de análisis (rangos de los valores

absolutos de los residuos) será el mismo para todos los bloques sólo cuando la variabilidad dentro de cada uno de los cuatro bloques sea la misma. Si los valores esperados en los bloques que contiene distancias "entre" (B2, B3, B4) no difieren significativamente de los valores esperados en el bloque B1 (que contiene distancias "dentro" de un grupo de individuos o entre individuos de un mismo grupo) entonces no existe evidencia para rechazar la hipótesis nula y suponer que hay interacción entre los factores.

La interacción se produce cuando las diferencias entre los niveles de un factor dentro de un mismo nivel de un segundo factor, son distintas a las diferencias obtenidas para otros niveles del segundo factor. Si esto ocurre (*i.e.* si hay interacción entre factores) los valores absolutos de los residuos provenientes de distancias que involucran perfiles de distintos niveles de uno o ambos factores (Bloques B2, B3 y B4), serán mayores que aquellos provenientes de distancias del bloque B1, donde sólo se estima la variabilidad residual y no la debida a la interacción. Para evitar supuestos distribucionales en la variable de análisis (valor absoluto de los residuos entre distancias observadas y distancias esperada sin interacción) se ajusta una prueba no paramétrica, basada en rango, por lo que el valor de significancia es obtenido como en Kruskal Wallis.

Warton *et al.* (2012) discuten que una propiedad crítica de los datos discretos es que la media y la varianza suelen estar relacionadas. Postulan que si la dispersión o variabilidad entre individuos, se define en función de los cambios que sufre la relación entre la media y la varianza, el efecto de la variabilidad genética puede confundirse con el efecto de la variabilidad subyacente. Para evitar ese efecto confundido, es importante utilizar una métrica de distancia entre individuos de diferente grupo o taxón que contemple apropiadamente la relación entre media y varianza. Dicho de otra manera, una consecuencia de la selección incorrecta de la métrica, es que en grupos de individuos con alta varianza, las diferencias entre grupos serán detectada con baja potencia. Este último resultado es indeseable si el objetivo es la búsqueda de estructura genética (Warton *et al.*, 2012). Por otro lado, Jost (2008) discutió que usar una medida de diferenciación genética como un estimador de divergencia corregido por el sesgo del muestreo, evita cualquier impacto que pueda tener la diversidad dentro de grupo para poder estimar la diferenciación entre grupos de poblaciones (Bird *et al.*, 2014). Si bien la métrica de Excoffier es ampliamente utilizada para estudios de variabilidad genética entre y dentro de grupos en el AMOVA Jerárquico como otros índices de similitud para datos binarios (Bruno *et al.*, 2003) pueden ser transformados a distancias para este tipo de análisis de la varianza molecular .

Nosotros usamos PERMANOVA con la distancia de Bray-Curtis (Anderson, 2001), una técnica ampliamente utilizada para evaluar simultáneamente la respuesta de abundancia en datos multivariados para uno o más factores (Andreson, 2001) para comparar el desempeño del método propuesto. Al igual que Warton *et al.* (2012) los valores-p calculados por PERMANOVA tendieron a ser más pequeños que con la prueba de interacción propuesta (mayor significancia) en conjuntos de datos donde la varianza entre individuos de distinto grupo es alta. Por el contrario, los valores-p que calcula PERMANOVA tienden a ser más grande, es decir, menos significativos, cuando el efecto entre grupos tiene menos variabilidad. Esto indica que la potencia de PERMANOVA para detectar diferencias entre grupos cuando la variabilidad entre ellos es menos variable, es más pequeña que cuando el efecto entre grupos presenta mayor variabilidad (Warton *et al.*, 2012). Con el método no paramétrico propuesto para la prueba de interacción en AMOVA No-Jerárquico, la potencia también disminuye conforme disminuye el efecto de la interacción, pero esta disminución es menor que la que se evidencia con PERMANOVA.

BIBLIOGRAFÍA

Anderson M.J. (2001) A new method for non-parametric multivariate analysis of variance. Austral Ecology 26:32-46.

Anderson M.J., ter Braak T. (2003) Permutation test for multi-factorial analysis of variance. Journal of Statistical Computation and Simulation. 73 (2): 85-113.

Apostol B.L., Black W.C., Miller B.R., Reiter P., Beaty B.J. (1993) Estimation of the number of full sibling families at an oviposition site using RAPD-PCR markers: applications to the mosquito Aedes aegypti. Theoretical and Applied Genetics 86: 991-1000.

Balzarini M.G., Di Rienzo J.A. Info-Gen versión 2018. FCA, Universidad Nacional de Córdoba, Argentina. URL http://www.info-gen.com.ar

Bird C.E, Karl S.A., Smouse P.E., Toonen R.J. (2014). Detectin and measuring genetic differentiation. Cust Issues 19:31-39.

Bray J.R., Curtis J.L. (1957) An ordination of the upland forest communities of southern Wiscosin. Ecol. Monogr. 27: 325-349.

Bruno C, Balzarini M., Di Rienzo J. (2003). Comparación de Medidas de Distancia entre Perfiles RAPD individuales. Journal of Basic & Applied Genetics, 15 (2):69-78. ISSN BAG-1666-0390

Conover W.J. (1999) Practical nonparametric statistics, 3rd edition. New York: John Wiley & Sons.

Dyer R. (2017) Applied Population Genetics. Disponible en https://dyerlab.github.io/applied_population_genetics/index.html#interactive-content

Excoffier L., Smouse P.E., Quattro J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131: 479-491.

Gower J.C. (1966) Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53 (3 and 4): 325-338.

Hedrick P. (2005) Genetic of Populations. Arizona State University. Jones and Bartlett publishers. Third Edition.

Kennington W.J., Gockel J., Partridge L. (2003) Testing for Asymmetrical Gene Flow in a Drosophila melanogaster Body-Size Cline. Genetics 165: 667-673.

Li W.H. (1976) A mixed model of mutation for electrophoretic identity of proteins within and between populations. Genetics 83:423-432

Jost L. (2008). G_{ST} and its relatives do not measure differentiation. Mol Ecol. 17: 4015-4026.

Nei M. (1973) The theory and estimation of genetic distance. In: MORTON, N. (ed.), Genetic structure of Populations University of Hawaii, Honolulu, pp. 45-54.

Pritchard J.K., Stephens M., Donnelly P. (2000) Inference of Population Structure Using Multilocus Genotype Data. Genetics 155: 945-959.

R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Warton D.I., Wright S.T., Wang Y. (2012) Distance-based multivariate analyses confound location and dispersion effect. Methods in Ecology and Evolution 3 :89-101.

BAG | Journal of Basic & Applied Genetics

Vol XXX (1): 25-33; July 2019

Journal of the Argentine Society of Genetics

GENOTYPIC DIVERSITY IN 291 MAIZE LINES FROM CIMMYT AND PHENOTYPIC CHARACTERIZATION IN SOUTHERN CÓRDOBA, ARGENTINA

DIVERSIDAD GENOTÍPICA DE 291 LÍNEAS DE MAÍZ DE CIMMYT Y CARACTERIZACIÓN FENOTÍPICA EN EL SUR DE CÓRDOBA, ARGENTINA

Rossi E.A. [1,2], Ruiz M.[1,2], Di Renzo M.[2], Bonamico N.C.[1,2]

[1] Instituto de Investigaciones Agrobiotecnológicas (UNRC-CONICET)

[2] Facultad de Agronomía y Veterinaria, Universidad Nacional de Río Cuarto

Corresponding author:
Ezequiel Rossi
ezequiel_455@hotmail.com

Cite this article as:
Rossi E.A., Ruiz M., Di Renzo M., Bonamico N.C. 2019. GENOTYPIC DIVERSITY IN 291 MAIZE LINES FROM CIMMYT AND PHENOTYPIC CHARACTERIZATION IN SOUTHERN CORDOBA, ARGENTINA. BAG. Journal of Basic and Applied Genetics XXX (1): 25-33.

Received: 04/26/2018
Revised version received: 06/25/2019
Accepted: 03/15/2019

General Editor: Elsa Camadro

DOI: 10.35407/bag.2019.XXX.01.04

ISSN online version: 1852-6233

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

ABSTRACT

CIMMYT maize inbred lines (CMLs) are freely distributed to breeding programs around the world. Better information on phenotypic and genotypic diversity may provide guidance to breeders on how to use more efficiently the CMLs in their breeding programs. In this study a group of 291 CIMMYT maize inbred lines, was phenotyped by nine agro-morphological traits in south Córdoba, Argentina and genotyped using 18,082 SNPs. Based on the geographic information and the environmental adaptation, 291 CMLs were classified into eight subgroups. Anthesis-silking interval (IAE) was the trait with higher phenotypic diversity. A 40% of maize inbred lines, with IAE less than five days, show a good adaptation to growing conditions in south Córdoba, Argentina. The low phenotypic variation explained by environmental adaptation subgroups indicates that population structure is only a minor factor contributing to phenotypic diversity in this panel. Principal component analysis (ACP) allowed us to obtain phenotypic and genotypic orderings. Generalized procrustes analysis (APG) indicated a 60% consensus between both data type from the total panel of maize lines. In each environmental adaptation subgroup, the APG consensus was higher. This result, which might indicate linkage disequilibrium between SNPs markers and the genes controlling these agro-morphological traits, is promising and could be used as an initial tool in the identification of Quantitative Trait Loci (QTL). Information on genetic diversity, population structure and phenotypic diversity in local environments will help maize breeders to better understand how to use the current CIMMYT maize inbred lines group.

Key words: broad-sense heritability, multivariate analysis, SNPs, agro-morphological traits.

RESUMEN

El maíz (*Zea mays* L.) posee un genoma complejo y una amplia diversidad genética. La información de caracteres fenotípicos y marcadores moleculares en su conjunto provee una mejor descripción e interpretación de la variabilidad genética. El objetivo del trabajo fue estimar la diversidad genética y caracterizar fenotípicamente un panel de 291 líneas de maíz de CIMMYT. Las líneas corresponden a ocho grupos establecidos de acuerdo a su adaptación ambiental y origen geográfico. Éstas se evaluaron fenotípicamente por medio de nueve caracteres agro-morfológicos en tres ambientes del sur de la provincia de Córdoba, Argentina. El intervalo antesis-estigma (IAE) presentó la mayor variabilidad fenotípica. El 40% de las líneas de maíz registraron un IAE menor a cinco días, indicando buena adaptación a los ambientes de evaluación. La estructura poblacional dada por los subgrupos de adaptación ambiental es sólo un factor menor que contribuye a la variabilidad fenotípica del panel estudiado. El análisis de componentes principales (ACP) permitió obtener el ordenamiento fenotípico y el genotípico, mientras que el análisis de procrustes generalizado indicó un consenso del 60% entre ambos ordenamientos para el total de líneas. El consenso entre el ordenamiento obtenido con caracteres agro-morfológicos y con marcadores moleculares indica desequilibrio de ligamiento entre los SNPs y los genes que controlan los caracteres agro-morfológicos. Los resultados muestran una amplia diversidad genética en el germoplasma evaluado, lo que sugiere que esta colección de líneas es un recurso importante para impulsar ganancias genéticas futuras en los programas de mejoramiento de maíz de Argentina.

Palabras clave: heredabilidad en sentido amplio, análisis multivariado, SNPs, caracteres agro-morfológicos.

INTRODUCCIÓN

El Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT) es uno de los centros más importantes de recolección, conservación y utilización de germoplasma de maíz (Yan *et al.*, 2009). Las líneas desarrolladas en el CIMMYT constituyen una fuente pública mundial de germoplasma de maíz, las cuales se distribuyen por medio del Tratado Internacional sobre los Recursos Fitogenéticos para la Alimentación y la Agricultura (ITPGRFA) (Chen *et al.*, 2016). Las líneas de maíz del CIMMYT (CML) se han desarrollado para un amplio número de ambientes en todo el mundo. Cada ambiente presenta diferentes temperaturas durante la estación de crecimiento, diferentes altitudes, así como otros aspectos que definen la adaptación del cultivo. El CIMMYT considera cuatro mega-ambientes principales, y cuatro subprogramas de investigación de maíz que abordan las necesidades de los productores en estas áreas. Los mega-ambientes son (i) tropicales de altitud baja, (ii) subtropical, (iii) altitud media, y (iv) ambientes de altitud alta (Xia *et al.*, 2004., 2005). Según el programa de mejoramiento donde se desarrollaron, estos mega-ambientes contienen líneas de distintos orígenes: México, Tailandia, Kenia, Zimbabue y Colombia (Wu *et al.*, 2016). Comprender las relaciones entre las líneas de diferentes mega-ambientes es de utilidad para incorporar germoplasma exótico en los diferentes programas de mejoramiento. Las líneas de CIMMYT se distribuyen libremente a los programas de mejoramiento de todo el mundo, donde se cruzan con líneas adaptadas localmente. Información sobre el comportamiento de las CML en distintos ambientes permite orientar a los mejoradores sobre cómo utilizar de manera más eficiente las líneas del CIMMYT en los programas de mejoramiento (Xia *et al.*, 2005).

La variabilidad presente en el germoplasma disponible es requisito fundamental para identificar genotipos con caracteres específicos (Dinesh *et al.*, 2016). Tradicionalmente, la diversidad genética se estimó a partir de caracteres agro-morfológicos (Yang *et al.*, 2010), con la limitación dada por la interacción en la expresión de éstos con el ambiente (Mienie y Fourie, 2013; Tiwari *et al.*, 2017). El análisis fenotípico sólo considera parte de la variación del germoplasma. Por ello, el análisis genotípico con marcadores moleculares es el complemento adecuado (Warburton *et al.*, 2002). Actualmente, los marcadores SNPs (polimorfismos de un nucleótido) son ampliamente utilizados en estudios genéticos (Wu *et al.*, 2016).

Los caracteres agro-morfológicos y marcadores moleculares proveen conjuntamente una mejor descripción e interpretación de la variabilidad genética. Esto puede deberse a que la región del genoma que se explora es mayor, así como a la incorporación de caracteres de alta heredabilidad que no interaccionan con el ambiente (Demey, 2008). La heredabilidad en sentido estricto sólo cuantifica los efectos genéticos aditivos, mientras que la heredabilidad en sentido amplio comprende la suma de los efectos aditivos, de dominancia y epistáticos (Nyquist, 1991; Falconer y Mackay, 1996).

Distintas configuraciones se obtienen a partir de caracteres agro-morfológicos y marcadores moleculares para un mismo grupo de genotipos (Balzarini *et al.*, 2008). En esta situación el análisis procrustes generalizado (APG) resulta adecuado como método para analizar la congruencia entre las configuraciones individuales (Bruno y Balzarini, 2010). El APG fue aplicado por Bramardi *et al.* (2005) para investigar relaciones entre genotipos de *Cucumis sativus* L. evaluados con caracteres agronómicos y caracterizados con marcadores RAPD, así como también por Hernández *et al.* (2010) para consensuar el ordenamiento obtenido mediante marcadores moleculares y caracteres agro-morfológicos en maíz.

El presente estudio se realizó con el objetivo de: i) estimar la diversidad genética de un panel de líneas de maíz de CIMMYT mediante marcadores moleculares SNPs; ii) caracterizar fenotípicamente el panel de líneas de maíz del CIMMYT mediante caracteres agro-morfológicos en el sur de Córdoba, Argentina; iii) evaluar los efectos de la estructura de la población sobre el fenotipo; iv) evaluar el consenso entre el ordenamiento fenotípico obtenido mediante caracteres agro-morfológicos y el ordenamiento genotípico obtenido a partir de marcadores moleculares.

MATERIALES Y MÉTODOS

Material vegetal y ensayo de campo

El material vegetal estuvo constituido por 291 líneas de maíz desarrolladas y provistas por el CIMMYT (http://www.cimmyt.org/seed-request). La zona de adaptación o el programa de mejoramiento de origen permitieron diferenciar ocho grupos: tropical de altitud baja (Lowland), subtropical (Subtropical), subtropical de África (Africa MA/ST), América del Sur (South America), altitud alta (Highland), altitud baja de Asia (Asia Lowland), de África (Africa Lowland) y de América Latina (Lowland-LA)(http://hdl.handle.net/11529/10246). En la Tabla 1 se presenta un resumen de la diversidad fenotípica para color y textura de grano en cada uno de estos grupos. En los ciclos agrícolas 2015/2016 y 2016/2017 se realizaron tres ensayos a campo con la finalidad de evaluar fenotípicamente las líneas de maíz. En el primer ciclo de evaluación se estableció un ensayo en el campo experimental de la Facultad de Agronomía y Veterinaria de la Universidad Nacional de Río Cuarto (CAMDOCEX, FAV-UNRC, 33° 06′ S; 64° 17′ O). En el segundo ciclo se realizaron dos ensayos, uno en la

localidad de Río Cuarto (CAMDOCEX, FAV-UNRC) y otro en la localidad de Chaján (33° 33′ S 65° 00′ O). El diseño utilizado fue en bloques aumentados parcialmente repetidos (Williams *et al.*, 2011), con un 25% de los genotipos, elegidos al azar, con tres repeticiones y el resto de los genotipos con sólo una repetición. Cada genotipo se estableció en parcelas de un surco de 3,0 m de largo espaciados a 0,52 m. Las diferentes combinaciones año-localidad constituyeron los tres ambientes de evaluación: Río Cuarto, ciclo agrícola 2015/2016 (E1); Río Cuarto, ciclo agrícola 2016/2017 (E2) y Chaján, ciclo agrícola 2016/2017 (E3).

Tabla 1. Número de genotipos por grupo de adaptación, color de grano (blanco y amarillo) y textura de grano (dentado, flint, semi-dentado y semi-flint) de las 291 líneas de maíz de CIMMYT.

Grupo	Total	Blanco	Amarillo	Dentado	Flint	Semi-dentado	Semi-flint
Africa Lowland	8	1	7	1	7	0	0
Africa MA/ST	38	38	0	3	12	8	15
Asia Lowland	20	0	20	3	11	4	2
Highland	9	6	3	2	1	6	0
Lowland	123	73	50	36	41	30	16
Lowland-LA	9	8	1	0	4	2	3
South America	21	5	16	1	12	2	6
Subtropical	63	60	3	30	7	15	11
Total	291	191	100	76	95	67	53

Evaluación fenotípica

Las líneas de maíz se evaluaron mediante nueve caracteres agro-morfológicos. Los caracteres fueron altura de planta (AP), altura de espiga (AE), largo de la panoja (LP), ángulo de inserción de la primera hoja por encima de la espiga principal (AH1), ángulo de inserción de la segunda hoja por encima de la espiga principal (AH2), número de raquis por panoja (RP), días a floración masculina (DFM), días a floración femenina (DFF). Estos se definieron cuando el 50% de las plantas de cada surco alcanzó cada uno de los estadios. A partir de los DFM y los DFF se estimó el carácter intervalo antesis-estigma (IAE). Los caracteres AP, AE, LP, AH1, AH2, y RP se midieron en cinco plantas para obtener el valor medio de cada surco. Los caracteres DFM, DFF e IAE se midieron sólo en los ambientes E1 y E2.

Caracterización genotípica

La caracterización de las líneas de maíz con marcadores moleculares realizada por Chen *et al.* (2016) es de disponibilidad pública (http://data.cimmyt.org/dvn/dv). La información disponible consta de 18.082 marcadores moleculares SNPs, los cuales fueron utilizados en el presente estudio para el panel total de líneas de maíz y cada uno de los grupos de adaptación.

Análisis estadístico

Fenotípico

La normalidad de los caracteres AP, AE, LP, AH1, AH2, RP, DFM, DFF y del IAE se comprobó con el test de Shapiro-Wilks modificado y luego se realizó un Análisis de la Varianza (ANOVA) a través de ambientes según el siguiente modelo estadístico:

$$y_{ijk} = \mu + G_i + E_j + \beta_{k(j)} + GE_{ij} + \varepsilon_{ijk}$$

donde y_{ijk} y_{ijk} es la observación de cada uno de los caracteres medidos, μ es la media general, G_i G_i es el efecto del i-ésimo genotipo, E_j E_j es el efecto del j-ésimo ambiente, $\beta_{k(j)}$ $\beta_{k(j)}$ es el efecto del k-ésimo bloque dentro de cada ambiente, GE_{ij} GE_{ij} es el efecto de la interacción genotipo-ambiente y ε_{ijk} ε_{ijk} es un término de error aleatorio asociado a la observación.

La aplicación de este modelo permitió obtener las medias de cada carácter a través de ambientes y los efectos de la interacción genotipo-ambiente (GE). Los componentes de varianza de genotipo, ambiente, interacción GE y del error residual se estimaron a

partir de los cuadrados medios obtenidos del ANOVA. Los genotipos que presentaron repeticiones permiten obtener una estimación del error aleatorio, mientras que la estimación de la varianza de la interacción genotipo-ambiente (GE) y la varianza genotípica se realizó con la totalidad de los genotipos (Bernardo, 2002). Los componentes de varianza fueron utilizados para estimar la heredabilidad en sentido amplio sobre la media de cada parcela según Holland *et al.* (2003). La estimación de la proporción de la variación fenotípica explicada por la estructura de la población se realizó mediante una regresión lineal múltiple. Para cada carácter, se utilizaron las medias a través de ambientes como variable respuesta y la matriz de estructura poblacional de acuerdo a la zona de adaptación, como variable predictora.

Además, para cada carácter se evaluó la diversidad fenotípica del panel de líneas de maíz mediante la estimación del índice de Shannon-Weaver (S-W). Para esto, cada uno de los caracteres se dividió en 10 categorías utilizando el valor medio (M) y el desvío estándar (DE). En la categoría 1 se incluyeron los valores fenotípicos menores a M - 2DE, mientras que la categoría 10 incluyó a los valores mayores a M + 2DE. El intervalo entre cada una de las categorías fue de 0,5DE.

Los valores medios a través de ambientes de los nueve caracteres agro-morfológicos se analizaron con un enfoque multivariado, el Análisis de Componentes Principales (ACP), para explorar las relaciones entre caracteres, entre genotipos y entre caracteres y genotipos. El ACP se realizó para el panel de 291 líneas de maíz, así como también para cada uno de los grupos de adaptación. Los análisis se realizaron mediante el programa Infostat (Di Rienzo *et al.*, 2018).

Genotípico

El ordenamiento de las líneas basado en la caracterización molecular se realizó a partir de un Análisis de Componentes Principales (ACP) con los 18.082 SNPs. Las dos primeras componentes principales permitieron realizar el ordenamiento o visualización de los genotipos en un plano. El ACP con los 18.082 SNPs también se realizó para el panel total de líneas y para cada uno de los grupos. Los análisis se realizaron mediante el programa TASSEL (Bradbury *et al.*, 2007).

Análisis procrustes generalizado

El consenso entre el ordenamiento fenotípico y el ordenamiento genotípico se estimó mediante un Análisis Procrustes Generalizado (APG) (Gower, 1975). Este se realizó a partir de las componentes principales obtenidas del ordenamiento fenotípico y genotípico, respectivamente. Las componentes principales fenotípicas seleccionadas fueron las que tenían un auto valor igual o mayor a uno. Mientras que, para las componentes principales genotípicas se seleccionó el número de componentes donde se produce la mayor caída en la variación explicada. El APG se realizó con la totalidad de las líneas y para cada uno de los grupos de adaptación.

RESULTADOS

La evaluación fenotípica del panel de líneas de maíz de CIMMYT se realizó mediante la medición de nueve caracteres agro-morfológicos. Los caracteres se midieron en ambientes del sur de Córdoba, Argentina y se observó un amplio rango de variación para cada uno de ellos (Cuadro 2). De acuerdo al cociente entre los valores máximos y mínimos, el IAE fue el carácter que presentó mayor variabilidad fenotípica, el cual varió entre 1 y 22 días. El número de RP fue el otro carácter que presentó amplia variación con una media de 17,6 (±6,6) con un mínimo de 5 raquis y un máximo de 47 raquis. Mientras que los caracteres AP, DFM y DFF fueron los que mostraron menor variabilidad fenotípica con medias de 172,4 (±19), 90,8 (±9,1) y 96,3 (±10,4), respectivamente. El índice de Shannon-Weaver presentó valores entre 1,92, para el IAE, y 2,09 para AP, en tanto que el resto de los caracteres presentó valores intermedios. Al estimar el índice de Shannon-Weaver se pudo observar en base a las frecuencias de cada categoría que el 40% de las líneas de maíz registraron un IAE menor a 5 días (datos no mostrados).

El ANOVA a través de ambientes permitió observar que la interacción GE fue significativa en todos los caracteres excepto para AH1, AH2 e IAE. La heredabilidad en sentido amplio presentó valores intermedios para la mayoría de los caracteres. El carácter RP presentó el mayor valor de heredabilidad (0,73), y el carácter IAE el menor valor (0,07). El porcentaje de variabilidad fenotípica explicada por los distintos grupos de adaptación ambiental fue bajo para los nueve caracteres. Sin embargo, en los caracteres DFM y AH la presencia de grupos explicó el 16% y el 14% de la variación fenotípica, respectivamente.

El análisis de componentes principales (ACP) redujo la dimensión de la matriz de información fenotípica de los nueve caracteres agro-morfológicos medidos. Este análisis permitió obtener cinco componentes principales (CP) significativas que explicaron el 88% de la variabilidad total. La Figura 1 presenta el gráfico generado a partir de las dos primeras CP del ACP que explican el 47% de la variabilidad. En el gráfico se observa que los distintos grupos de adaptación tienen poca influencia en la variación fenotípica de los nueve caracteres medidos ya que no existe separación entre los genotipos pertenecientes a los distintos grupos.

Tabla 2. Estadística descriptiva, índice de Shannon-Weaver, heredabilidad y porcentaje de variación fenotípica explicada por los grupos de adaptación para los nueve caracteres agro-morfológicos medidos en las 291 líneas de maíz de CIMMYT en los ambientes Río Cuarto, ciclo agrícola 2015/2016 (E1); Río Cuarto, ciclo agrícola 2016/2017 (E2) y Chaján, ciclo agrícola 2016/2017 (E3).

Carácter	Mín.	Máx.	Media ± DE	Ambientes	GE	S-W	H^2	R^2
AP (cm)	127,2	221,1	172,4 ± 19,0	E1, E2, E3	***	2,09	0,50	9
AE (cm)	42,1	125,0	79,6 ± 14,1	E1, E2, E3	***	2,06	0,59	12
LP (cm)	21,9	45,3	32,8 ± 4,6	E1, E2, E3	*	2,06	0,46	4
AH1 (°)	16,7	61,3	26,8 ± 5,8	E1, E2, E3	ns	1,99	0,56	14
AH2 (°)	14,2	58,8	25,5 ± 5,9	E1, E2, E3	ns	1,96	0,59	14
RP (N°)	5,4	47,3	17,6 ± 6,6	E1, E2, E3	**	1,96	0,73	3
DFM (días)	67,0	119,0	90,8 ± 9,1	E1, E2	***	2,07	0,51	16
DFF (días)	69,0	129,7	96,3 ± 10,4	E1, E2	**	1,98	0,48	7
IAE (días)	1,0	22,0	7,6 ± 4,2	E1, E2	ns	1,92	0,07	3

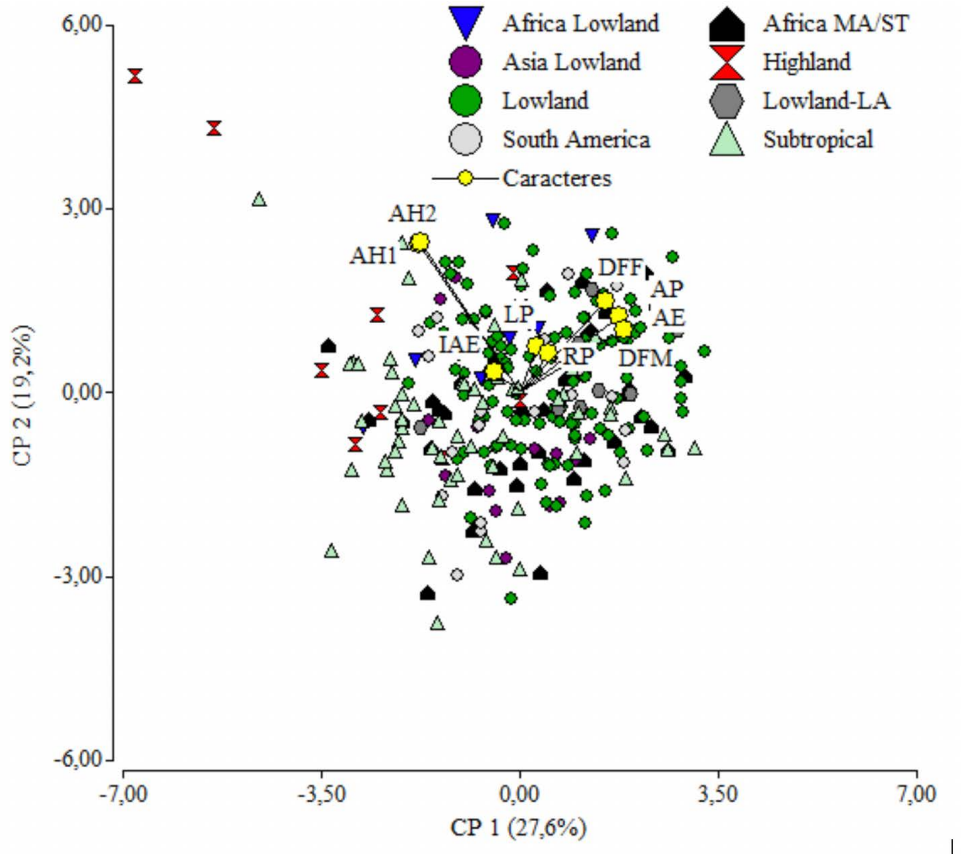


Figura 1. Gráfico biplot a partir de las dos primeras componentes principales obtenidas del análisis de componentes principales con los nueve caracteres agro-morfológicos medidos en las 291 líneas de maíz del CIMMYT evaluadas en tres ambientes del sur de la provincia de Córdoba, Argentina, durante los ciclos agrícolas 2015/2016 y 2016/2017.

AP: altura de planta; AE: altura de espiga; LP: largo de panoja; AH1: ángulo de inserción de la primera hoja por encima de la espiga principal; AH2: ángulo de inserción de la segunda hoja por encima de la espiga principal; RP: número de raquis por panoja; DFM: días a floración masculina; DFF: días a floración femenina; IAE: intervalo antesis-estigma.

Los 18.082 SNPs permitieron estimar el contenido de información polimórfica (PIC) y la diversidad genética presente en el panel de líneas de maíz de CIMMYT. El PIC y la diversidad genética oscilaron entre los distintos grupos. El grupo de altitudes bajas de Latinoamérica (Lowland-LA) registró el menor valor y el grupo Subtropical el mayor valor. El número total de alelos para el panel de líneas fue de 71.996 con una media de 3,98 alelos por locus (Cuadro 3).

El análisis de componentes principales del panel de líneas, a partir de los 18.082 marcadores SNPs, permitió obtener 49 componentes principales que explican el 39% de la variabilidad. El gráfico obtenido con las dos primeras CP se muestra en la Figura 2. En general líneas pertenecientes a un mismo grupo de adaptación ambiental presentan menores distancias. El grupo de líneas tropicales de altitud baja (Lowland) que es el de mayor número de líneas presenta una amplia variación entre sus genotipos, superponiéndose con el resto de los grupos.

Al analizar cada uno de los grupos de adaptación, el ACP realizado con los caracteres agro-morfológicos permitió obtener entre tres y cuatro CP significativas, que explicaron entre 75% y 90% de la variabilidad fenotípica. Mientras que en el ACP realizado a partir de los 18.082 SNPs, entre 4 y 17 CP fueron significativas. Estas explicaron entre 33% y 87% de la variabilidad genotípica presente (Tabla 4).

El porcentaje de consenso para el panel de líneas fue de 60%, mientras que para cada uno de los grupos este valor de consenso fue mayor. Así, en el grupo de líneas correspondientes a altitudes bajas de África (Africa Lowland) se observó el mayor valor de consenso (80%) entre ambos ordenamientos.

Tabla 3. Medidas de resumen relacionadas con la diversidad genética para las 291 líneas de maíz de CIMMYT y para cada uno de los grupos de adaptación, estimadas con 18.082 SNPs.

Grupo	PIC	Diversidad genética	Número medio de alelos por locus	Número total de alelos
Africa Lowland	0,21	0,23	2,10	38.033
Africa MA-ST	0,26	0,29	3,21	58.016
Asia Lowland	0,27	0,30	3,03	54.871
Highland	0,21	0,25	2,00	36.176
Lowland	0,27	0,31	3,86	69.783
Lowland-LA	0,18	0,21	1,80	32.598
South America	0,23	0,26	2,63	47.557
Subtropical	0,29	0,32	3,75	67.853
Total	0,24	0,32	3,98	71.996

Tabla 4. Número de componentes principales significativos y porcentaje de variación, fenotípica y genotípica, explicada por esas componentes. Consenso obtenido mediante el Análisis de Procrustes Generalizado (APG) a partir de información proveniente de caracteres agro-morfológicos y de marcadores moleculares SNPs de las 291 líneas de maíz de CIMMYT.

Grupos	Componentes principales (Fenotipo)	R^2	Componentes principales (Genotipo)	R^2	Consenso (%)
Africa Lowland	3	90	5	86	80
Africa MA-ST	4	81	8	45	68
Asia Lowland	4	83	4	41	74
Highland	3	89	6	87	76
Lowland	4	77	17	35	65
Lowland-LA	3	89	4	77	74
South America	3	75	12	87	73
Subtropical	4	83	10	33	69
Total	5	88	49	39	60

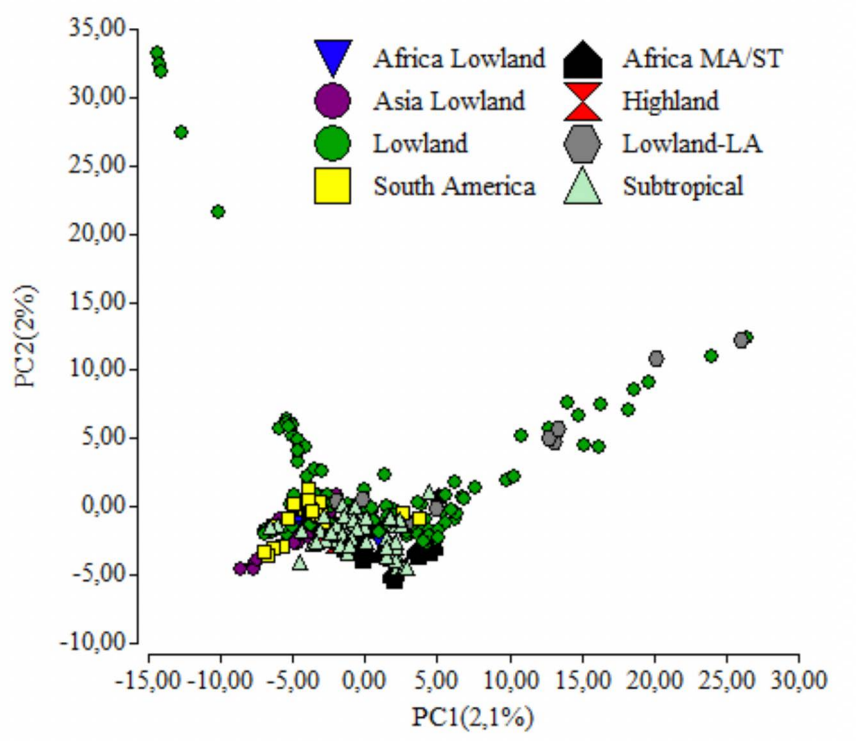


Figura 2. Gráfico de las dos primeras componentes principales obtenidas del análisis de componentes principales a partir de 18.082 SNPs en 291 líneas de maíz del CIMMYT.

DISCUSIÓN

En los programas de mejoramiento genético de maíz a nivel mundial se aplicó selección directa para rendimiento en grano, lo cual llevó a una selección indirecta para caracteres relacionados con el rendimiento como ángulo de hoja e intervalo antesis-estigma. Las hojas erectas permiten una mejor intercepción de la luz ante el incremento de la densidad de plantas, propio del cultivo moderno. Mientras que, intervalos antesis-estigma breves indican que la espiga presentó un ritmo de crecimiento normal sin obstáculos (Duvick, 2004). El grupo de líneas de maíz utilizadas en este estudio mostró estar sometidas a varias generaciones de mejoramiento ya que en general presentan hojas erectas e intervalos antesis-estigma menores a cinco días, lo que indicaría adecuada adaptación a los ambientes de evaluación de este estudio.

La estructura de la población, los ocho grupos en este panel de líneas, explicó en promedio un 9% de la variación fenotípica para los nueve caracteres agro-morfológicos medidos. Esto sugiere que los grupos de adaptación son un factor con una contribución menor a la variación fenotípica en este panel de líneas. Sin embargo, los efectos variaron según el carácter y se observó mayor influencia de la estructura poblacional para los caracteres días a floración masculina y ángulo de inserción de la primera y segunda hoja por encima de la espiga principal.

El conocimiento de la variación fenotípica y genotípica existente entre diversos caracteres agro-morfológicos, así como de la heredabilidad de éstos es importante en los programas de mejoramiento (Beyene *et al.*, 2005). Las líneas de maíz del CIMMYT evaluadas en este estudio presentan amplia variabilidad para los caracteres agro-morfológicos medidos en el sur de la provincia de Córdoba, Argentina. Los valores medios a altos estimados de heredabilidad en sentido amplio en este estudio indican la posibilidad de utilizar las líneas de maíz mejor adaptadas en programas de mejoramiento locales. Éstas pueden incorporarse para el mejoramiento poblacional, para el desarrollo de genotipos híbridos, variedades de polinización abierta o poblaciones de mapeo.

La diversidad genética y el contenido de información polimórfica fueron similares a los informados por Wu *et al.* (2016), quienes caracterizaron la totalidad de líneas de maíz de CIMMYT mediante un amplio set de marcadores moleculares SNPs obtenidos a través de genotipado por secuenciación. Los resultados del presente estudio indican que el panel de líneas caracterizadas fenotípicamente en el sur de Córdoba es una muestra representativa del conjunto de líneas de maíz público de CIMMYT.

La estadística multivariada es ampliamente usada para describir y analizar observaciones multidimensionales obtenidas al compilar información sobre diferentes caracteres en distintos genotipos (Balzarini *et al.*, 2008). Estudios previos informan el uso de análisis de procrustes generalizado para consensuar el ordenamiento obtenido mediante marcadores moleculares y el ordenamiento obtenido mediante caracteres agro-morfológicos. Bramardi *et al.* (2005) aplicaron APG para investigar relaciones entre genotipos de *Cucumis sativus* L. evaluados mediante caracteres agronómicos y caracterizados con marcadores RAPD. Hernández *et al.* (2010) informan un 67% de consenso entre los ordenamientos individuales obtenidos con 14 caracteres fenotípicos y 40 marcadores moleculares de 14 líneas de maíz evaluadas fenotípicamente en Venezuela. En el presente estudio, la matriz de caracteres agro-morfológicos se redujo a cinco componentes principales que explicaron el 88% de la variabilidad total y la matriz de información genotípica se redujo a 49 componentes principales que explicaron el 39% de la variabilidad. El análisis de procrustes generalizado permitió obtener un valor moderado de consenso (60%) entre el ordenamiento de los genotipos evaluados mediante caracteres agro-morfológicos y marcadores moleculares. Este valor de consenso fue mayor al analizar cada grupo por separado, así en el grupo de líneas de baja altitud de África (Africa Lowland) se observó un 80% de consenso. El consenso entre ordenamientos fenotípicos y genotípicos podría ser usado como una herramienta inicial promisoria en la identificación de loci de caracteres cuantitativos (QTL), con la ventaja de poder realizarse para más de un carácter simultáneamente (Demey, 2008). La correlación entre la diversidad agro-morfológica y molecular indicaría la existencia de desequilibrio de ligamiento entre genes que controlan los caracteres agro-morfológicos y los marcadores moleculares (Baranger *et al.*, 2004). En este estudio, el valor de consenso observado entre los ordenamientos individuales, fenotípico y genotípico, indica la presencia de desequilibrio de ligamiento entre loci que controlan los caracteres agro-morfológicos medidos y los marcadores SNPs empleados. Esto pone de manifiesto el potencial de este grupo de líneas de maíz para identificar QTL de caracteres de interés agronómico mediante la implementación de estudios de mapeo por asociación. Además, brinda información sobre la estructura de la población basada en los grupos de adaptación ambiental, la cual debería ser considerada en futuros estudios de mapeo por asociación.

Un panel de mapeo por asociación debe abarcar la máxima diversidad fenotípica y molecular que pueda ser estimada de manera confiable en un ambiente común (Flint-Garcia *et al.*, 2005). El uso de germoplasma con caracteres favorables es el procedimiento más económico y ambientalmente sustentable para lograr incrementos y estabilidad en la producción de maíz (Di Renzo *et al.*, 2002). En la práctica, los programas de mejoramiento de clima templado intercambian germoplasma con programas de mejoramiento subtropical o de mediana altitud (Wu *et al.*, 2016). Los resultados del presente estudio mostraron una amplia diversidad genética en el germoplasma evaluado, el cual en su mayoría corresponde a programas de origen tropical. Esta colección de 291 líneas de maíz de CIMMYT constituye un valioso recurso para impulsar ganancias genéticas futuras en los programas de mejoramiento de maíz en Argentina.

BIBLIOGRAFÍA

Balzarini M., Gonzalez L., Tablada M., Casanoves F., Di Rienzo J., Robledo C. (2008) Infostat. Manual del Usuario, Editorial Brujas, Córdoba, Argentina.

Baranger A., Aubert G., Arnau G., Lainé A.L., Deniot G., Potier J., Weinachter C., Lejeune-Hénaut I., Lallemand J., Burstin J. (2004) Genetic diversity within *Pisum sativum* using protein- and PCR-based markers. Theor. Appl. Genet. 108: 1309-1321. https://doi.org/10.1007/s00122-003-1540-5

Bernardo R. (2002) Breeding for Quantitative Traits in Plants. Stemma Press, Woodbury, MN.

Beyene T., Botha A.M., Myburg A.A. (2005) Phenotypic diversity for morphological and agronomic traits in traditional ethiopian highland maize accessions. South African J. Plant Soil. 22: 100-105. https://doi.org/10.1080/02571862.2005.10634689

Bradbury P., Zhang Z., Kroon D., Casstevens T., Ramdoss Y., Buckler E. (2007) TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633-2635.

Bramardi S.J., Bernet G.P., Asíns M.J., Carbonell E.A. (2005) Simultaneous agronomic and molecular characterization of genotypes via the generalised procrustes analysis: An application to cucumber. Crop Sci. 45: 1603-1609. https://doi.org/10.2135/cropsci2004.0633

Bruno C., Balzarini M. (2010) Ordenaciones de material genético a partir de información multidimensional. Rev. FCA UNCuyo. 42: 183-200.

Chen J., Zavala C., Ortega N., Petroli C., Franco J., Burgueño J., Costich D., Hearne, S.J. (2016) The development of quality control genotyping approaches: A case study using elite maize lines. PLoS ONE. 11: 1-17 https://doi.org/10.1371/journal.pone.0157236

Demey J. (2008) Diversidad genética en bancos de germoplasma: un enfoque biplot. Tesis de doctorado. Departamento de estadística. Universidad de Salamanca, España.

Dinesh A., Patil A., Zaidi P.H., Kuchanur P.H., Vinayan M.T., Seetharam K. (2016) Genetic diversity, linkage disequilibrium and population structure among CIMMYT maize inbred lines, selected for heat tolerance study. Maydica 61.

Di Renzo M.A., Bonamico N.C., Díaz D.D., Salerno J.C., Ibañez M.A., Gesumaría J.J. (2002) Inheritance of resistance to Mal de Río Cuarto (MRC) disease in *Zea mays* (L.). J. Agr. Sci. 139: 47-53.

Di Rienzo J., Casanoves F., Balzarini M., Gonzalez L., Tablada M., Robledo C. (2018) InfoStat. Grupo InfoStat, FCA, Universidad Nacional de Córdoba, Argentina. URL http://www.infostat.com.ar

Duvick D. (2004) Long-term selection in a commercial hybrid maize breeding program. Plant Breed. In: J. Janick (eds.) John Wiley & Sons, Inc., Hoboken, NJ, pp. 109-151.

Falconer D.S., Mackay T.F.C. (1996) An Introduction to Quantitative Genetics, Ed. 4. Prentice Hall, London.

Flint-Garcia S.A., Thuillet A.C., Yu J., Pressoir G., Romero S.M., Mitchel S.E., Doebley J., Kresovich S., Goodman M.M., Buckler E.S. (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J. 44: 1054-1064.

Gower J. (1975) Generalized procrustes analysis. Psychometrika 40: 33-51.

Hernández A., San Vicente F., Figueroa-Ruiz R. (2010) Evaluación y caracterización de líneas parentales de híbridos de maíz (*Zea mays* L.) en tres ambientes de Venezuela. Interciencia 35: 290-298.

Holland J.B., Nyquist W.E., Cervantes-Martínez C.T. (2003) Estimating and Interpreting Heritability for Plant Breeding: An Update. Plant Breed. Reviews. https://doi.org/10.1002/9780470650202.ch2

Mienie C.M.S., Fourie A.P. (2013) Genetic diversity in South African maize (*Zea mays* L.) genotypes as determined with microsatellite markers. African J. Biotechnol. 12: 123-133. https://doi.org/10.5897/AJB12.2282

Nyquist W.E. (1991) Estimation of heritability and prediction of selection response in plant populations. Crit. Rev. Plant Sci. 10: 235-322.

Tiwari A., Vajire D., Thakare K., Solunke R., Panche A. (2017) The Possibilities of the Prediction of Heterosis in Elite Lines of Onion Based on the Assessment of Genetic Diversity. EC Microbiology 10 (5): 211-219.

Warburton M.L., Xianchun X., Crossa J., Franco J., Melchinger A.E., Frisch M., Bohn M., Hoisington D. (2002) Genetic characterization of CIMMYT inbred maize lines and open pollinated populations using large scale fingerprinting methods. Crop Sci. 42: 1832-1840. https://doi.org/10.2135/cropsci2002.1832

Williams E., Piepho H.P., Whitaker D. (2011) Augmented p-rep designs. Biometrical J. 53: 19-27. https://doi.org/10.1002/bimj.201000102

Wu Y., San Vicente F., Huang K., Dhliwayo T., Costich D. E., Semagn K., Sudha N., Olsen M., Prasanna B.M., Zhang X., Babu R. (2016) Molecular characterization of CIMMYT maize inbred lines with genotyping-by-sequencing SNPs. Theor. Appl. Genet. 129: 753-765. https://doi.org/10.1007/s00122-016-2664-8

Yan J., Shah T., Warburton M.L., Buckler E.S., McMullen M.D., Crouch J. (2009) Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. PLoS One 4. https://doi.org/10.1371/journal.pone.0008451

Yang X., Yan J., Shah T., Warburton M.L., Li Q., Li L., Gao Y., Chai Y., Fu Z., Zhou Y., Xu S., Bai G., Meng Y., Zheng Y., Li J. (2010) Genetic analysis and characterization of a new maize association mapping panel for quantitative trait loci dissection. Theor. Appl. Genet. 121: 417-431. https://doi.org/10.1007/s00122-010-1320-y

Xia X.C., Reif J.C., Hoisington D.A., Melchinger A.E., Frisch M., Warburton M.L. (2004) Genetic diversity among CIMMYT maize inbred lines investigated with SSR markers: I. Lowland tropical maize. Crop Sci. 44: 2230-2237.

Xia X.C., Reif J.C., Melchinger A.E., Frisch M., Hoisington D.A., Beck D., Pixley K., Warburton M.L. (2005) Genetic diversity among CIMMYT maize inbred lines investigated with SSR markers: II. Subtropical, tropical midaltitude, and highland maize inbred lines and their relationships with elite U.S. and European maize. Crop Sci. 45: 2573-2582.

BAG | Journal of Basic & Applied Genetics

Vol XXX (1): 35-51; July 2019

Journal of the Argentine Society of Genetics

TAKING ADVANTAGE OF ORGANELLE GENOMES IN PLANT BREEDING: AN INTEGRATED APPROACH

APROVECHANDO LOS GENOMAS DE LAS ORGANELAS EN EL MEJORAMIENTO GENÉTICO DE PLANTAS: UN ENFOQUE INTEGRADO.

Colombo N.

Instituto de Genética "Ewald A. Favret", CIVyA, INTA, Nicolás Repetto y de los Los Reseros S/N (1686), Hurlingham, Provincia de Buenos Aires, Argentina

Corresponding author:
Colombo N.
colombo.noemi@inta.gob.ar

Cite this article as:
Colombo N. 2019. TAKING ADVANTAGE OF ORGANELLE GENOMES IN PLANT BREEDING: AN INTEGRATED APPROACH. BAG. Journal of Basic and Applied Genetics XXX (1): 31-55.

Received: 03/13/2019
Accepted: 06/21/2019

General Editor: Elsa Camadro

DOI: 10.35407/bag.2019.XXX.01.05

ISSN online version: 1852-6233

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

ABSTRACT

Plant cells carry their genetic information in three compartments: the nucleus, the plastids and the mitochondria. In last years, next-generation sequencing has allowed the development of genomic databases, which are increasingly improving our knowledge about the role of nuclear and cytoplasmic genes as well as their interactions in plant development. However, most plant breeding efforts consider the utilization of the nuclear genome, while less attention is given to plastid and mitochondrial genomes. The objective of this review is to present current knowledge about cytoplasmic and cytonuclear effects on agronomic traits bearing in mind the prospective utilization of all the genomes in plant breeding.

Key words: Cytoplasmic genes, cytoplasmic-nuclear interactions, plant breeding methods.

RESUMEN

La información genética de las células vegetales está contenida en tres compartimentos: el núcleo, los plástidos y las mitocondrias. En los últimos años, la secuenciación de última generación ha permitido desarrollar bases de datos genómicas que están aumentando progresivamente nuestro conocimiento sobre el rol de los genes nucleares y citoplásmicos y de sus interacciones durante el desarrollo de la planta. Sin embargo, la mayoría de los esfuerzos de la mejora vegetal se basan en el aprovechamiento del genoma nuclear y relegan a los genomas de los plástidos y las mitocondrias. El objetivo de esta revisión es actualizar el conocimiento sobre de los efectos citoplásmicos y las interacciones núcleo-citoplásmicas sobre caracteres interés agronómico, asumiendo la utilización potencial de todos los genomas en el mejoramiento vegetal.

Palabras clave: genes citoplásmicos, interacciones núcleo-citoplásmicas, métodos de mejoramiento vegetal.

PLANT ORGANELLE GENOMES

The plant cell is the result of endosymbiotic events which resulted in the evolution of the mitochondrion from an α-proteobacterium and, somewhat later, the evolution of the chloroplasts from a cyanobacterium. During co-evolution, the three genomes of plant cells have undergone significant structural changes that resulted in an optimized expression of the compartmentalized genetic material and cross-talk between the nucleus and the organelles. As a result of co-evolution, most genes from the symbionts were transferred to the nucleus of the host cell. Although mitochondria and plastids still retain their own, ancestral DNA, most proteins required for organelle function are encoded in the nucleus and must be imported (Allen, 2015; Archibald, 2015; Grainer and Bock, 2013; Smith and

Keeling, 2015). There are multiple copies of both plastid and mitochondrial DNA inside each organelle. The number of copies varies depending on the tissue type and it changes notably during development (Kumar *et al.*, 2015; Oldenburg and Bendich, 2015). Regarding their mode of inheritance, while nuclear genetic information is inherited biparentally, plastid and mitochondrial genomes of most land plants show predominantly maternal inheritance, with some cases of paternal and biparental inheritance (Birky, 1995; Xu, 2005; Greiner *et al.*, 2014).

High-throughput sequencing technologies have allowed rapid advance in organelle genetics and genomics. To date, 2257 complete chloroplast genomes and 246 plant mitochondrial genomes are available (https://www.ncbi.nlm.nih.gov/genome/browse#!/organelles).

The chloroplast genomes (plastomes) of land plants range between 107 kb (*Cathaya argyrophylla*) to 403 kb (*Codonopsis lanceolata*). They have highly conserved structures and organization of content, consisting of a single circular molecule with two copies of an Inverted Region (IR) that separate large and small single-copy (LSC and SSC) regions. Recent studies have identified considerable diversity within non-coding intergenic spacer regions, which often include important regulatory sequences. The chloroplast genome includes 120-130 genes, mainly participating in photosynthesis, transcription, and translation (Daniell *et al.,* 2016).

Plant mitochondria genomes range between 200 and 750 kb in angiosperms, but extreme sizes of 6.7 Mb and 11.3 Mb are found in *Silene noctiflora* and *Silene conica* respectively, resulting from massive proliferation of non-coding content. The number of genes usually ranges between 50 and 60, with multiple cis- or trans-spliced introns and large intergenic regions. Protein genes encode subunits of the oxidative phosphorylation chain complexes proteins involved in the biogenesis of these complexes and several ribosomal proteins. The physical organization of the plant mitochondrial DNA includes a set of sub-genomic forms resulting from homologous recombination between repeats, with a mixture of linear, circular and branched structures. Recombination appears to be an essential characteristic of plant mitochondrial genetic processes, both in shaping and maintaining the genome. In addition, autonomous plasmids of essentially unknown function are found, increasing the complexity of the genome (Gualberto *et al.*, 2014; Morley and Nielsen, 2017).

THE ROLES OF PLANT ORGANELLES

Chloroplasts and mitochondria are mainly known by their involvement in photosynthesis and ATP production, respectively. However, both organelles play a part in multiple metabolic pathways and are essential for normal growth and development of plants (van Dingenen *et al.*, 2016). Chloroplasts are part of the family of plastids, in which some components are interconvertible during development: proplastids, etioplasts, chloroplasts, chromoplasts, leucoplasts, elaioplasts, amyloplasts, proteinoplasts and gerontoplasts. All plastids perform house-keeping functions and basal metabolic functions essential to the cell metabolism and specific roles according to their differentiated type. Plastids are the site of carbon oxidation via photorespiration, chlorophyll synthesis, carotenoid, α-tocopherol (vitamin E), plastoquinone and phylloquinone (vitamin K) synthesis, fatty acid and lipid synthesis, nitrogen assimilation and aminoacid synthesis, sulfur metabolism, oxygen metabolism and chlororespiration (Wise, 2007; Rolland *et al.*, 2018).

Mitochondria are dynamic organelles, changing shape, number, size, composition and distribution inside the cell depending on developmental stage, type of tissue, cell cycle phase, energetic cell demand, and external stimuli. Mitochondria are involved in the synthesis of nucleotides, vitamins and cofactors, the metabolism of amino acids and lipids, the photorespiratory pathway and the export of organic acid intermediates for wider cellular biosynthesis (Welchen *et al.*, 2013; Rao *et al.*, 2017).

Results obtained in last years demonstrated that mitochondria and chloroplasts also play a crucial role in perceiving and responding to biotic and abiotic stress conditions and that both organelles participate in programmed cell death (Chi *et al.*, 2015;Liberatore *et al.*, 2016; Wang *et al.*, 2018; Beltrán *et al.,* 2018; Rolland *et al.,* 2018; Zhao *et al.*, 2018). Thousands of plastid and mitochondrial proteins needed to perform such a variety of functions are nuclear encoded and targeted to the organelles, making it crucial to ensure the coordinated expression of different genomes. A complex signaling network between the nucleus and the organelles, including both anterograde signaling (from the nucleus to the organelles) and retrograde signaling (from the organelles to the nucleus) as well as inter-organelle signaling mediates the communication between genomes and ensures proper gene expression (Blanco *et al.*, 2014; Kleine and Leister, 2016; van Aken and Pogson, 2017; de Souza *et al.*, 2017; Brunkard and Burch-Smith, 2018; Crawford *et al.*, 2018).

ORGANELLES GENOMES IN PLANT BREEDING

Breeders have been regularly aware of the contribution of cytoplasmic genomes to plant phenotype and have therefore chosen certain particular combinations of cytoplasmic and nuclear donor genetic materials. The

most direct way to uncover cytoplasmic genetic effects on a trait is by the use of reciprocal crosses, in which each individual is used both as a male and as a female parent. For any trait, differences observed between hybrids obtained from the same parents suggest that the cytoplasm plays a role in the considered trait. Reciprocal crosses included in some of the traditional mating designs have been used to estimate genetic effects in quantitative genetics (Hayman, 1954; Griffing, 1956). Fan *et al.* (2014) compared the results obtained using a diallel experiment with or without reciprocal crosses and found that including reciprocal crosses allowed for the recovery of more high yielding hybrids and influenced both the estimates of general combining ability (GCA) and specific combining ability (SCA) effects and the heterotic group classification in maize.

However, differences between direct and reciprocal hybrids may be also due to genomic imprinting and maternal effects, like endosperm dosage effects and maternal phenotypic effects resulting from the environment or genotype of the maternal parent. In order to avoid these confounding effects, new populations for QTL analysis have been proposed, like F2 reciprocal populations and their F2:F3 families (Tang *et al.*, 2013) or reciprocal RILs (McKay *et al.*, 2008).

Another reliable approach extensively used to identify independent cytoplasmic effects, consists in developing nuclear substitution lines or cytolines by backcrossing several times a cytoplasm donor by the recurrent male parent, in order to obtain isonuclear lines differing only in the cytoplasm type. Allen (2005) used a set of cytolines carrying the nuclear genome of a maize inbred line and the cytoplasm from different teosintes belonging to the sections *Zea* and *Luxuriantes*, and found that cytolines with cytoplasm from the more distantly related *Z. luxurians*, *Z. diploperennis*, or *Z. perennis* presented significant differences for 56 of the 58 characters studied, affecting growth, development, morphology, and function. Besides their usefulness to uncover cytoplasmic effects on agronomic traits, cytolines are also valuable tools to diversify the genetic basis of crops (Calugar *et al.*, 2016).

Sometimes it is necessary to bypass post- and pre-zygotic sexual incompatibilities that prevent the use of wide crosses. In this case, somatic hybrids can be developed via protoplast fusion, thus allowing the combination of nuclei and organelles from different origin. Practical results using somatic hybrids to improve agronomic traits have been recently reported in model families *Rutaceae, Brassicaceae* and *Solanaceae* (Xia, 2009; Eeckhaut *et al.*, 2013).

The magnitude of cytoplasmic genetic effects on phenotypic expression is still a matter of debate. Cytoplasmic genetic effects can be additive, due to mutations in organelle genes, or epistatic, resulting from interactions between organelle and nuclear genes. In a meta-analytic review Dobler *et al.* (2014) evaluated 521 effect-size estimates reported in 66 publications including animals, fungi and plants. These authors found that cytoplasmic effect sizes are generally moderate in size and associated with variation across a range of factors, like the analyzed trait type, the experimental design used, the gene action associated with the reported cytoplasmic effect (additive or epistatic) and the experimental scale (intrapopulation, interpopulation or interspecies).

A growing set of data obtained following different approaches show that organelle genomic variation can modulate the effects of nuclear genomic variation in plants. Cytonuclear genetic interactions are predictable considering the complex network of retrograde signaling existing between organelles and nuclear genomes which ensures normal plant development. Joseph *et al.* (2013) analyzed the effect of cytoplasmic genomes on quantitative variation within the metabolome using a reciprocal recombinant RILs population in *Arabidopsis*. These authors demonstrated that genetic variation in the organelles influenced the accumulation of over 80% of the detectable metabolites and that cytoplasmic background affected epistatic interactions between nuclear loci. Other studies using phenotypic, microarray, and metabolomics analyses as well as whole transcriptome sequencing of cytolines revealed cytonuclear effects in rice, maize and wheat (Tao *et al.*, 2004; Crosatti *et al.*, 2013; Soltani *et al.*, 2016; Miclaus *et al.*, 2016).

So far, the traits with major impact in plant breeding showing cytoplasmic effects have been cytoplasmic male sterility (CMS), caused by mitochondrial genes (Bohra *et al.*, 2016; Chen *et al.*, 2017) and herbicide resistance, codified by mutations in the chloroplast gene *psbA* (Greiner, 2012). However, many other characters like yield and quality parameters, disease resistance, chilling tolerance, tissue culture response and regeneration, combining ability and plant adaptation have been found to be associated with the effect of cytoplasmic genes and cytonuclear interactions (Chandra-Shekara *et al.*, 2007; Gordon and Staub, 2011; Reddy *et al.*, 2011; Bock *et al.*, 2014; Shen *et al.*, 2015; Roux *et al.*, 2016; Satyavathi *et al.*, 2016; Dey *et al.*, 2017b; Boussardon *et al.*, 2019). The most recent reviews on this subject have been published several years ago (Frei *et al.*, 2003; Dhillon *et al.*, 2008; Mackenzie, 2010) creating the need to bring together the new data obtained to date. In the next sections, an update on the effects of cytoplasmic genomes and their interactions on agronomic traits in different crops is presented, emphasizing the methodologies and plant materials employed.

Maize

In maize (*Zea mays* L.) male sterile cytoplasms have been classified in three major groups by their response to specific restorer genes: T (Texas), S (USDA), and

C (Charrua) (Gabay-Laughnan and Laughnan, 1994; Allen *et al.*, 2007; Su *et al.*, 2016; Li *et al.*, 2017). As in other species in which CMS is used to produce hybrid seeds, importance has been given to evaluate the effects associated to male sterile cytoplasms on agronomic traits. Cytoplasm T constitutes a typical case of the risks of genetic uniformity. This source of male sterility had been extensively adopted by breeders due to its reliability since the 1960's. However, cytoplasm T resulted susceptible to Southern Corn Leaf Blight caused by *Bipolaris maydis* (Nisikado and Miyake) Shoemaker; race T. As a result of the severe losses caused by the epidemic of 1970–1971 in USA and southern Canada, with >85% of the hybrids grown carrying cytoplasm T, this cytoplasm has been banned from hybrid seed production (Bruns, 2017). After cytoplasm T withdrawal, cytoplasms S and C have been adopted for hybrid seed production. Although cytoplasm C shows higher stability than cytoplasm S (Weider *et al.*, 2009), it should be kept in mind that C is also specifically susceptible to race C of *B. maydis*, which is only known to occur in China (Gao *et al.,* 2005). As in other crops, introduction of male sterile cytoplasms in maize breeding programs must be preceded by a careful evaluation of their associated defects on agronomic traits (Jovanovick *et al.*, 2017).

Apart from the case of male sterile cytoplasms, several studies have found cytoplasmic and cytonuclear effects in maize. A diallel analysis using nine quality protein maize (QPM) inbred lines evaluated over seven environments detected significant reciprocal effects for quality index, tryptophan, and anthesis date, which on the average accounted for <13% of the variation among hybrids (Machida *et al.,* 2010).

Tang *et al.* (2013) evaluated the cytoplasmic effects and cytonuclear interactions on plant height (PH) and ear height (EH), by using the joint analysis approach to both reciprocal F2 and F2: 3 families and incorporating the cytonuclear interaction mapping method. These authors identified six cytonuclear epistatic QTL affecting PH and five affecting EH. The average phenotypic variance explained by the genetic components of the QTL x cytoplasm interaction for each QTL was 18% for PH and 9% for EH. Regarding cytoplasmic effects, they reached 9% and 40% of the phenotypic contributions to PH and EH, respectively.

Flowering time in maize was analyzed applying the same approach (Tang *et al.*, 2014). In this case, the authors evaluated the days to tassel (DTT) and the days to pollen shed (DPS) and found that although the cytoplasmic effects were not significant between the direct and reciprocal populations, four and eight cytonuclear epistatic QTL significantly contributed to the variation in DTT and DPS, respectively. Most of the cytonuclear epistatic QTL cannot be detected when using the interval mapping method, evidencing the importance of proper statistical modeling. In a study carried out by Calugar *et al.* (2016) a set of cytolines, obtained after transferring the nucleus of five inbred lines on four cytoplasm sources by backcrossing for ten generations, was used to determine the cytoplasmic effect on the plant height, ear height, number of leaves/plants, leaf area and the tassel length on some maize inbred lines. Two cytoplasms (T 248 and TC 221) showed significant effect on plant and ear height, leaf area and the tassel length. Besides, the authors detected some interaction between the cytoplasm and the nucleus that caused significant differences in the analyzed traits when the cytoline was compared to the original inbred.

Several reports noted differential expression between reciprocal F1 hybrids in maize for various kernel and germination traits (Cervantes Ortiz *et al.*, 2007; Cabral *et al.*, 2013; de la Torre and Biasutti, 2015; Santos *et al.*, 2017; de Abreu *et al.*, 2019). In Angiosperms, double fertilization results in the development of the diploid embryo and triploid endosperm that are surrounded by the maternal seed coat derived from the ovule integuments. Therefore, communication between these three genetically distinct structures ensures viable seed development (Figueiredo and Kohler, 2016; Chettoor *et al.*, 2016). Differences observed in reciprocal F1 crosses may thus be due to epigenetic phenomena (imprinting and xenia), dosage effects (in case of triploid tissue such as endosperm) and cytoplasmic effects (associated to mitochondrial and chloroplast genomes). Interestingly, phenotypic and differential expression profiling carried out using reciprocal F1 hybrids to determine the genes associated to seed size (Zhang *et al.,* 2016), cold germination and desiccation tolerance (Kollipara *et al.*, 2002), suggest the role of gene imprinting and not cytoplasmic genetic effects as a molecular mechanism underlying the observed reciprocal effects.

Wheat

Pioneer research on alloplasmic lines in wheat (*Triticum aestivum* L.) led to the discovery of cytoplasmic male sterility associated to *Aegilops caudata* (Kihara, 1959). At present, wheat has a large set of alloplasmic lines providing an excellent tool for evaluating the genetic effects of different cytoplasms (Tsunewaki, 2009). Three alloplasmic wheat series involving *T. aestivum* nuclear genome and cytoplasms from *T. aestivum* subsp. *macha*, *Ae. ventricosa*, *Ae. squarrosa*, *Ae. uniaristata* and *Hordeum chilense,* were used by Atienza *et al.* (2007) who observed that plant height, flowering date and yield per plant were least affected by the donor cytoplasm of the *Triticum–Aegilops* complex than by *Hordeum chilense* cytoplasm, with the latter being associated with detrimental effects on agronomic traits. On the other hand, all the alloplasmic lines studied showed significant differences for seed lutein content relative to euplasmic controls, thus revealing the role of cytoplasm

genes on seed carotenoid content in wheat.

Soltani *et al.* (2016) used wheat alloplasmic lines carrying the cytoplasm of *Aegilops mutica* along with an integrated approach utilizing comparative quantitative trait locus (QTL) and epigenome analysis in order to evaluate the role of nuclear-cytoplasmic interactions upon interspecific hybridization. Results showed that cytoplasmic genomes modified the magnitude of QTL controlling plant height, dry matter weight and number of spikes per plant. Strikingly, when the methylation profiles were compared between alloplasmic and euplasmic lines, eight polymorphic regions associated with transposable elements, stress responsive, and metabolite pathways resulted affected by the cytoplasm type. Taken together, results suggest that novel nuclear-cytoplasmic interactions can trigger a potential epigenetic modification in the nuclear genomes and eventually change the genetic network controlling physiological traits.

In order to evaluate the effect of cytoplasmic diversity on traits related to heat tolerance during the reproductive stage Talukder *et al.* (2014) developed cytoplasmic near isogenic lines (NIL) using ten different cytoplasms and four different recurrent parents. Results showed that cytoplasmic variations can contribute to an increase in chlorophyll content and quantum efficiency of photosystem II during heat stress and detected interactions between cytoplasmic and nuclear genes, thus emphasizing the potential of cytoplasmic sources as components of any strategy to improve heat tolerance in wheat.

In a recent work Takenaka *et al.* (2018) studied cytoplasmic genetic diversity affecting seedling emergence and growth under submergence stress. Using a set of 37 nucleo-cytoplasmic hybrids carrying the nuclear genome of the wheat cultivar Chinese Spring and different cytoplasms of the *Triticum-Aegilops* complex they found a significant diversity with divergent cytoplasmic effects on submergence response. While T^2 cytoplasm of *Aegilops mutica* showed a positive contribution to submergence tolerance, cytoplasms of *Aegilops umbellulata* and related species caused a greater inhibition. Evaluation of more nuclear genetic backgrounds is needed to detect nuclear-cytoplasmic interactions affecting this trait.

Using a different experimental approach, Bnejdi *et al.* (2010) studied cytoplasmic effects affecting grain resistance to yellow berry, a serious physiological disorder in wheat and triticale, characterized by softer, light colored and starchy endosperm, which lacks the vitreous texture characteristic of normal grains (Ammiraju *et al.,* 2002). In this case, the authors employed parental, F1, reciprocal F1 (RF1), F2, reciprocal F2 (RF2), BC1P1 and BC1P2 generations of four crosses involving four cultivars of durum wheat. Significant cytoplasmic genetic effects were found in all crosses, indicating that the choice of the female parent resistant to yellow berry could significantly contribute to an increase in resistance level.

Reciprocal crosses and the F1, F2, F3, BC1, and BC1F1 offspring were also used by Guo *et al.* (2017) to assess the effect of an *Aegilops* cytoplasm on the expression of the multi-ovary gene. Results showed that the heterogeneous cytoplasm could suppress the expression of the heterozygous, but not homozygous, dominant multi-ovary gene. In a subsequent research, Guo *et al.* (2018) used methylation-sensitive amplification polymorphisms (MSAP) to assess the DNA methylation status of the reciprocal crosses between *Aegilops* and common wheat. The authors found that heterogeneous cytoplasm significantly changed DNA methylation patterns between the reciprocal crosses and suggested that this epigenetic control plays a role in the suppression of the multi-ovary gene.

Rice

Although several CMS types have been described in rice (*Oryza sativa* L.), the majority of hybrids were developed using mainly WA (wild abortive type) and to a lesser extent, BT (Boro type) and HL (Hong-Lian type) male sterile cytoplasms (Tang *et al.*, 2017). In a study designed to analyze DNA methylation as affected by male sterile cytoplasms in rice, Xu *et al.* (2013) compared the extent and polymorphism of DNA methylation between male sterile lines (A) carrying four different cytoplasms and their maintainer lines (B) using the MSAP technique. Results showed identical differences in methylation between A and B lines at three sites in all the analyzed cytoplasms, suggesting a relationship of DNA methylation at these sites specifically with male sterile cytoplasms, since cytoplasm is the only difference between the A and B lines. Interestingly, it was also found that different cytoplasms affected DNA methylation to different levels, depending on the genetic distance between the nucleus and the cytoplasm of each cytoplasm type donor. Evidence of the effect of male sterile cytoplasms on nuclear gene expression has also been obtained by Hu *et al.* (2016) who analyzed the anther transcript profiles of three *indica* rice alloplasmic CMS lines and their maintainer line and found a set of differentially expressed genes (DEGs) involved in anther development.

A common drawback associated with different CMS in rice is panicle enclosure, in which part of the panicle fails to exert from the sheath of the flag leaves, leading to lower seed-setting rates and yield loss and demanding gibberellin application for hybrid seed production (Chen *et al.*, 2013). The effects of male sterile cytoplasms on quality traits of rice has been examined by Waza and Jaisbal (2015) who compared the difference in performance between 20A (WA-CMS line) x R (Restorer

line) hybrids and the corresponding B (Maintainer line) x R (Restorer line) hybrids. In this study, WA cytoplasmic influence for different traits was found to be highly cross-specific, depending on the nuclear background of the CMS line and the fertility restorer. Results showed that WA cytoplasm had no significant influence on some traits (head rice recovery, elongation ratio and aroma) but it negatively affected others (hulling recovery, milling recovery, kernel length before cooking, kernel breadth before cooking, kernel length after cooking, kernel breadth after cooking, alkali spread value and amylose content) and exhibited both favorable and unfavorable cytoplasmic effects depending upon the parental combination (kernel length/breadth ratio before and after cooking). The most significant effect of WA cytoplasm was reduction in length of cooked kernel followed by decrease in amylose content, which are two undesirable quality traits.

Narrow cytoplasmic genetic diversity observed both in rice hybrids and varieties has led breeders to look for new cytoplasmic resources and to characterize their effects on traits of agronomic value (Huang *et al.*, 2013; Kumar *et al.,* 2013; Toriyama and Kazama, 2016; El-Namaky, 2018). In order to study the effects of cytoplasm, nucleus, and interaction between nucleus and cytoplasm on agronomic traits in rice, Tao *et al.* (2004) evaluated fifteen isolines obtained by crossing five widely used *japonica* cytoplasm resources as females by three distinct *japonica* rice cultivars followed by several backcrosses to the male recurrent parent. Analysis of the results showed that cytoplasms had significant effects on yield, width of flag leaf, and low temperature tolerance. Besides, significant effects of cytoplasm-nucleus interaction on yield, plant height, and low temperature tolerance were also found. In a similar research undertaken to analyze eighteen isolines of *indica* rice obtained by backcrossing six different cytoplasmic sources with three cultivars as recurrent male parents, Tao *et al.* (2011) detected significant effects of cytoplasms on 1000-grain weight, which is a major component of yield and grain quality in rice production. Additionally, a three-way interaction between cytoplasms, nuclei and locations was found for filled-grain ratio, emphasizing the need to evaluate cytoplasmic effects in the nuclear backgrounds of interest and at multiple locations.

Sorghum

Evidence about cytoplasmic effects on agronomic traits in sorghum (*Sorghum bicolor* L. Moench) has been mainly obtained from research on male sterile cytoplasms. Although several types of male sterile cytoplasms are known in sorghum -A1, A2, A3, A4, A4M, A4VZM, A4G1, A5, A6, 9E, M35 and KS- A1 is the most widely used for commercial hybrid seed production followed by A2, due to adverse effects on agronomic traits and poor environmental stability of male fertility restoration observed in the other types (Reddy *et al.*, 2007; Kumar *et al.*, 2011; Elkonin and Tsvetova, 2012; Kozhemyakin *et al.*, 2017; Kante *et al.*, 2018). The effect of cytoplasms on performance of grain sorghum hybrids varies according to different authors, although A3 hybrids consistently showed reduced grain yield compared to A1 and A2 hybrids. On the other hand, no adverse effects associated with male sterile cytoplasms were observed in biomass sorghum hybrids (Hoffman and Rooney, 2013). Recently, Vacek and Rooney (2018) evaluated 16 isocytoplasmic bio-energy sorghum hybrids, each of them carrying three different male sterile cytoplasms A1, A2 and A3, to assess the effect of cytoplasm type on the agronomic performance and quality. Results showed that cytoplasms "per se" did not influence any agronomic or composition trait; however, hybrid by cytoplasm interactions were significant for several traits, showing the importance to identify the best cytoplasm and hybrid combination for sorghum use as a biomass source.

The effect of male sterile cytoplasms on resistance to diseases and insect pests in sorghum has been studied following different approaches. Durga *et al.* (2008) studied the influence of male sterile cytoplasm on the occurrence of leaf blight caused by *Exserohilum turcicum* (Pass) using paired cytoplasmic male-sterile (CMS) A lines and maintainer (B) lines, which were crossed with R-lines (restorers) to produce two types of hybrids: (A x R) and (B x R), differing only in the cytoplasm type. Although significant cytoplasmic effects were detected for some disease related parameters (reduced lesion length and lesion area), the overall disease damage was not significantly different between genotypes with male fertile and male sterile cytoplasm. Reddy *et al.* (2011) evaluated the effect of cytoplasms A1, A2, A3, A4M, 4G, 4VZM on grain mold resistance using a set of 72 hybrids obtained from the cross of 36 isonuclear alloplasmic lines (A lines) by two restorer lines (R). Results showed significant effects due to cytoplasms "per se" and to their interactions with A lines, R lines and years. A1 cytoplasm contributed to grain mold resistance, followed by A4VZM and A2, indicating that introduction of these two alternative cytoplasm to hybrid sorghum production should not increase the risk of grain mold. Insect resistance has also been evaluated as influenced by different cytoplasm types in sorghum, taking into account that A1 is highly susceptible to insect pests (Dhillon *et al.*, 2008). In the case of sorghum shoot fly (*Atherigona soccata* (Rondani)) Sharma *et al.* (2006) found that the expression of traits associated with resistance in the F1 hybrids depends on the interactions between cytoplasmic and nuclear genes and concluded that resistance to shoot fly is needed in both parents to develop shoot fly resistant hybrids. Akula *et al.* (2012) evaluated four isogenic lines in four male-sterile

backgrounds, A1, A2, A3 and A4, and their corresponding maintainer (B lines) lines. Results showed that the A4 cytoplasm was the least susceptible to sorghum shoot fly as it was comparatively less preferred for oviposition and had lower dead heart incidence than the other cytoplasms tested. Mohammed *et al.* (2016) studied the nature of gene action involved in shoot fly resistance using a complete diallel design and found significant reciprocal effects of combining abilities for oviposition, leaf glossy score and trichome density, thus supporting the influence of cytoplasmic factors in inheritance of shoot fly resistance.

Potato

A PCR-based classification method using chloroplast and mitochondrial DNA markers (Hosaka and Sanetomo, 2012; Sanetomo and Hosaka, 2013) distinguishes cytoplasms of cultivated potatoes and closely related wild species into six distinct types: M (an ancestral type of Andean cultivated potatoes), P (derived from *Solanum phureja*), A (the most prevalent *Solanum tuberosum* ssp. *andigena* type), W (wild species), T (the most prevalent *Solanum tuberosum* ssp. *tuberosum* type), and D (derived from *Solanum demissum*). Besides, potato mitochondrial genomes have been classified in five types: α, β, γ, δ, and κ (Lössl *et al.*, 2000). Cytoplasmic male sterility has been reported in T/β cytoplasm, as well as in D and W/γ-type derived from *S. stoloniferum*. T/β is the most extended cytoplasm type in potato cultivars all over the world except in German cultivars, due to the fact that *S. demissum* and *S. stoloniferum* were broadly used in German breeding programs for their resistance to late blight and potato virus Y, respectively. Knowledge of the cytoplasm types is necessary to prevent the cytoplasmic invasion of male sterile types, which severely limits the selection of male parents in breeding programs (Hosaka and Sanetomo, 2012; Mihovilovich *et al.*, 2015; Anisimova and Gabrilenko, 2017).

Sanetomo and Gebhardt (2015) estimated the correlation of cytoplasmic genomes with complex agronomic traits using 1,217 cultivars and breeding clones of 6 different populations. Results showed significant effects of cytoplasm type on traits such as resistance to late blight and tuber bruising, plant maturity, tuber shape, starch content and yield. On the contrary, no cytoplasmic difference was found for processing quality traits such as chip color and reducing sugar content. In particular, it was shown that the W/γ-type cytoplasm was correlated with increased tuber starch content and later plant maturity, while the D-type cytoplasm was correlated with increased foliage resistance to late blight. W/γ cytoplasm type was also found to be associated with potato tuber yield, starch content and/or starch yield when reproducibility of diagnostic nuclear DNA markers was evaluated using an association mapping approach (Schönhals *et al.*, 2016).

Reciprocal crosses between cultivated potatoes and diploid wild related species have been frequently used to evaluate the cytoplasmic effects on agronomic traits. Jansky (2011) found improved male fertility when *S. brevicaule* and *S. microdontum* were used as females instead of *S. tuberosum*, but lower percentages of selected clones and clones that tuberized when *S. chacoense* and *S. microdontum* were used as the female parents.

In the case of the crosses between *S. tuberosum* (T) and the hexaploid *S. demissum* (D), non-complete unilateral incompatibility determines that seed is obtained preferentially when the cultivated potato is used as pollen donor. Moreover, apparent size differences between DT and TD seeds are observed, the former being significantly larger than the latter (Sanetomo *et al.*, 2011). In order to shed light on the mechanisms involved in such behavior, Sanetomo and Hosaka (2011) compared reciprocal F1 hybrids TD and DT using methylation-sensitive amplified polymorphism (MSAP) analysis. Their results showed differences both in DNA sequences and in the DNA methylation level between TD and DT.

Somatic hybridization is a powerful tool to overcome the sexual barriers between the cultivated and wild species. In a recent review, Tiwari *et al.* (2018) examined research in somatic hybridization in potato during the past 40 years. Data show that majority of somatic hybrids follow recombination of mitochondrial genome from both parents, and chloroplast pattern from only one, except the recombination of the chloroplast genome observed once in *S. tuberosum*-*S. vernei* somatic hybrid. Given that the interaction between nuclear and cytoplasmic genes from different species can affect fertility and agronomic traits of somatic hybrids and progenies, information on such interactions could be useful when using somatic hybrids in breeding.

Brassicaceas

This family includes several vegetable crops: cabbage, cauliflower and broccoli (*Brassica oleracea* var. *capitata* L., *var. botrytis* L. and *var. italica* Plenck, respectively), turnip (*Brassica rapa* L. spp. *rapa*) and radish (*Raphanus sativus* L.) and an important oil crop, oilseed rape (*Brassica napus* L.). Hybrid seed production in these crops has been developed worldwide using Ogura male sterile cytoplasm discovered in a Japanese radish (*Raphanus sativus* L.) and introgressed in *B. oleracea* by repeated backcrosses (Yamagishi and Bhat, 2014; Kaminski *et al.*, 2016; Sekhon *et al.*, 2018). However, these initial alloplasmic male sterile lines carrying Ogura cytoplasm presented chlorophyll deficiency at low temperatures, underdeveloped nectaries and malformed ovaries and pods which reduced the seed set. Additionally, the same defects were observed after Ogura cytoplasm transfer into *B. napus*. It was then assumed that undesirable effects

were due to negative interactions between the *Brassica* nucleus and the *Raphanus* chloroplasts. Protoplasts from a normal *B. napus* line were fused with protoplasts from a CMS (Ogura radish cytoplasm) *B. napus*, and protoplasts from a normal *B. oleracea* line were fused with protoplasts from a CMS (Ogura radish cytoplasm) *B. oleracea*, in order to select cybrids carrying only *Brassica* chloroplasts that grew normally. These improved CMS lines are known as Ogu-INRA and are widely used to produce hybrids in *Brassicaceae* (Pelletier and Budar, 2015). Similar advances have been achieved by Indian breeders who developed and characterized several Ogu-CMS lines in cauliflower and cabbage (Dey *et al.*, 2017a; Bathia *et al.*, 2015; Parkash *et al.,* 2015). Recently, Dey *et al.,* (2017b) reported that the introgression of Ogura cytoplasm into the nuclear background of cauliflower genotypes significantly affected nutritional traits. However, the effects were genotype specific suggesting the role of nuclear–cytoplasmic interaction in expression of different quality traits. Thus, Ogura cytoplasm interacted favorably in particular nuclear backgrounds in expression of antioxidant capacities. On the other hand, it had adverse effects for anthocyanin, total chlorophyll content in most of the genotypes, which is desirable in cauliflower but not in cabbage. Besides, while ascorbic acid concentration was adversely affected, total carotenoids and β-carotene concentration were higher in most of the genotypes after introgression of Ogura cytoplasm. Following a novel approach Singh *et al.* (2018) combined CMS and doubled haploid inbred lines to determine heterotic combinations for important antioxidant compounds such as CUPRAC, FRAP, phenols, carotenoids, anthocyanins and ascorbic acid in cauliflower.

Interspecific and intergeneric crosses are commonly used to exploit alloplasmic effects in plant breeding. Chang *et al.* (2015) investigated the alloplasmic effect of the cytoplasm of *B. juncea* and *B. napus* on heat and cold toleranceof *B. carinata*, by comparing the performance of alloplasmic and euplasmic lines of *B. carinata* for a variety of physiological parameters. While plants with cytoplasm of *B. napus* showed little difference inheat tolerance, those with the cytoplasm of *B. juncea* displayed higher heat injury than the euplasmic lines. Moreover, both alloplasmic lines showed decreased cold tolerance than the euplasmic lines. Results suggested that tolerance of extreme temperature stress was controlled by the nucleus, the cytoplasm and the interaction between maternal and nuclear genomes.

In the case of oilseed rape, recent data have revealed cytoplasmic effects on yield related traits and quality traits, like number of seeds per pod, oil content, protein content, glucosinolates, oleic acid, linolenic acid and erucic acid (Ishaq *et al.*, 2016; Guo *et al.,* 2017; Szała *et al.,* 2018).

Cucumber

In cucumber (*Cucumis sativus* L.) plastids and mitochondria are inherited maternally and paternally, respectively (Corriveau and Coleman, 1988; Havey 1997). Chilling tolerance was reported to display maternal inheritance (Chung *et al.*, 2003) so it was postulated that chilling tolerance was associated with the plastid genome. In order to identify candidate plastid genomic regions, Chung *et al.* (2007) carried out a comparative complete sequencing of chloroplast DNA of a susceptible and a tolerant cucumber line and found three polymorphic sites associated with the trait. Afterwards, sdCAPS (simply derived cleaved amplified polymorphic sequence) were developed converting sequence data in PCR-based markers that were successfully used to distinguish plastid types (Ali *et al.*, 2013; 2014). Therefore, breeding chilling tolerance into elite cultivars by backcrossing may be effective for the rapid introduction of plastomes conferring a tolerant phenotype (Gordon and Staub, 2011; 2014).

Diallel mating designs have been used to estimate GCA, SCA, and reciprocal-cross effects for agronomic traits in cucumber. Significant reciprocal effects were reported for fresh and dry weight per plant (Shen *et al.*, 2015) and for internode length, leaf length, leaf width, fruit length, fruit diameter, number of fruits per plant, yield per fruit and yield per plant (Golabadi *et al.*, 2015).

Organelle omics generate a novel and promising field for cucumber breeding. In this regard, a tiling microarray comprising the whole cucumber chloroplast genome has been developed and it has been used to study chloroplast responses to abiotic stresses (Żmieńko *et al.*, 2011). Besides, cucumber plants regenerated from cell cultures occasionally originate paternally transmitted mosaic (MSC) phenotypes, characterized by slower growth, chlorotic patterns on the leaves and fruit, lower fertility, and rearrangements in their mitochondrial DNAs (Malepszy *et al.*, 1996; Lilly *et al.*, 2001; Bartoszewski *et al.*, 2004; 2007; Del Valle-Echeverria *et al.*, 2015). Analysis of nuclear gene expression in MSC mutants will help to understand mitochondrial retrograde signaling and will allow to identify genes associated with stress responses and use them as potential selection targets for breeding (Pawełkowicz *et al.*, 2016; Mróz *et al.*, 2018).

Onion

In onion (*Allium cepa* L.) two main sources of CMS -S and T- have been mainly used in hybrid seed production. S type results from the interaction of a cytoplasmic factor S and a single nuclear restorer gene Ms (Jones and Emsweller, 1936; Jones and Clarke, 1943). T type is determined by the interaction of the cytoplasmic factor T and two to three complementary restorer genes

(Berninger, 1965; Schweisguth, 1973). While S type is the most widely used due to the relatively common occurrence of the recessive allele at Ms, the stability of male sterility over environments and no reduction of female fertility (Goldman *et al.*, 2000; Leite *et al.*, 1999), T cytoplasm is commercially used in Europe and Japan (Havey, 2000) and is present in Brazilian onion populations (Fernandes Santos *et al.*, 2010). Moreover, new sources of CMS from *Allium galanthum* (Havey, 1999) and *Allium rolyei* (Vu *et al.*, 2012) have been identified making it possible to diversify the cytoplasms used in hybrid seed production, to reduce the risks of genetic uniformity associated to the major use of S type (Havey, 2018). Recently, complete sequencing of mitochondrial genomes of S, T and normal cytoplasms has been achieved and a chimeric gene encoded by *orf 725* has been postulated as the common causal gene for CMS induction in onions (Kim *et al.*, 2016; Kim *et al.,* 2019).

Carrot

The main types of CMS in carrot (*Daucus carota* ssp. *sativus* L.) are "brown anther" (Sa), characterized by shriveled, yellow-to-brown anthers with no pollen (Welch and Grimball, 1947) and "petaloid" (Sp), in which anthers are replaced by a whorl of petals (Thompson 1961; Peterson and Simon, 1986). While "brown anther" type was found in a lot of cultivars as well as in wild relatives, "petaloid" type was only identified in wild relatives and has been introduced into the nuclear genetic background of the cultivated carrot (Linke *et al.*, 2019). In addition to the two main types, CMS-GUM, CMS-MAR and CMS-GAD (from *D. carota subsp. gummifer, D. carota subsp. maritimus and D. carota subsp. gadecaei, respectively)* have been described (Linke *et al.*, 1999; Nothnagel *et al.*, 2000).

Different hypotheses have been postulated to explain restoration of fertility in carrot involving single or multiple nuclear genes with complex interactions (Thompson, 1961; Hansche and Gabelman, 1963; Börner *et al.*, 1995; Wolyn and Chahal, 1998). Alessandro *et al.* (2013) found that restoration of "petaloid" cytoplasmic male sterility was due to a single dominant gene, *Rf1*, and developed a linkage map using molecular markers, some of which can be applied in marker assisted selection (MAS) in hybrid breeding programs.

Both "brown anther" and "petaloid" systems show instability due to high temperatures, dry conditions, growing time or long day conditions. Although hybrid seed production is mainly based on the use of petaloid CMS type because of less frequent reversion to male fertility, seed yields on the brown-anther CMS are generally higher (Havey, 2004; Dhall, 2010).

In a recent study carried out to determine the genetic basis of carrot shoot growth, Turner *et al.* (2018) analysed a diallel mating design and found highly significant reciprocal effects in all the evaluated traits: canopy height and width, shoot biomass, root biomass, and the ratio of shoot: root biomass. Besides, significant Reciprocal x E interactions were observed for canopy height at harvest and fresh shoot biomass.

Sunflower

Although 72 sources of cytoplasmic male sterile cytoplasm have been described by different authors in sunflower (*Helianthus annuus*), nearly all hybrid seed production relies on the use of a single male sterile cytoplasm, PET1, derived from *Helianthus petiolaris* ssp. *petiolaris* (Leclerq, 1969; Sabar *et al.*, 2003; Serieys and Christov, 2004). In this context, diversification of the cytoplasmic background is desirable to avoid the risks of genetic uniformity (Jan and Vick, 2007; Zhang *et al.*, 2010; Christov, 2013; Reddermann and Horn, 2018; Makarenko *et al.*, 2019).

Several reports have discussed the effects of different cytoplasm sources on agronomic traits as a prerequisite to their introgression in sunflower breeding programs. Jan *et al.* (2014) evaluated twenty diverse cytoplasmic substitution lines from six annual and six perennial wild diploid *Helianthus* species for agronomic and oil traits. Results showed that cytoplasms of perennial species *H. mollis*, *H. grosseserratus*, *H. divaricatus* and *H. angustifolius* had more adverse cytoplasmic effects affecting agronomic traits. In contrast, cytoplasms from annual species had no adverse effects. Additionally, ten alien CMS sources from annual species, wild *H. annuus* accessions and perennial species were tested and yield-reducing cytoplasmic effects were only observed in perennial *H. maximiliani* and annual *H. annuus* PI 413178 and PI 413024. No significant cytoplasmic effects were detected in oil percentage and fatty acid composition. These data support the exploitation of wild annual *Helianthus* species to broaden cytoplasmic diversity in sunflower breeding.

Tyagi *et al.* (2015a) used nine CMS sources to develop CMS alloplasmic lines, here designated as CMS analogues, by crossing them by a common maintainer line followed by repeated backcrossing. The CMS analogues, carrying the same nuclear genotype and different cytoplasmic genomes, were evaluated in the field for 21 morphological, agronomic, physiological and quality traits. Significant differences between CMS analogues were detected for all the traits. The genetic parameters analysis indicated that selection for grain yield accompanied with high harvest index, large head size and biological yield can be effectively used from these sources for genetic improvement in sunflower. The same CMS sources were evaluated under water stress conditions (Tyagi *et al.*, 2015b) and it was found that CMS-XA (unknown origin), E002-91 (*H. annuus*), ARG-3A (*H. argophyllus*) PHIR-27A (*H. praecox* ssp *hirtus*)

and PRUN-29A (*H. praecox* ssp. *runyonii*) presented significantly higher yield than the common maintainer line, making them potentially useful to develop efficient water use CMS lines. Furthermore, the effects of cytoplasmic sources on the estimation of combining ability for agronomic traits and stability under different environments were also evaluated (Tyagi and Dhillon, 2017; Tyagi *et al.*, 2018).

A research by Velasco *et al.* (2007) analyzed the relationships between fatty acid profile and seed oil content in F1s and F2s of reciprocal crosses between CAS-3, a high stearic acid mutant and ADV-37, a high seed oil content inbred line. Results demonstrated the existence of cytoplasmic effects in the genetic control of oil content both at the F1 and F2 plant level. On the contrary, cytoplasmic effects on stearic acid content were only observed at the F1 but not at the F2 plant level, a difference which may be due to small environmental influence, sampling deviations or to the effect of maternal rather than cytoplasmic genetic effects. Ferfuia and Vannozzi (2015) studied seed fatty acid composition in seeds from reciprocal F1s, F2s and BC populations between two high oleic inbred lines under different environmental conditions. Results showed that oleic acid percentage was affected by cytoplasmic or cytoplasmic x nuclear interaction. In particular, the expression of nuclear genes affecting oleic acid percentage, *OLs* and/or *Olm*, was modified by temperature and cytoplasm genotype.

Another trait of interest for sunflower breeding is the regeneration ability for "*in vitro*" culture. In order to evaluate the effect of different cytoplasmic background on the regeneration ability in sunflower, Cravero *et al.* (2012) tested seven alloplasmic CMS lines introgressed into the inbred line HA89 and one fertile cytoplasm (*H. annuus*) under different *in vitro* culture conditions, detecting cytoplasm by culture media interaction for the regeneration percentage and productivity rates. The authors concluded that the non-nuclear genome could be considered as another source of variability modifying the regeneration ability of recalcitrant sunflower genotypes.

Soybean

Several sources of CMS have been described in soybean (*Glycine max* L. Merr): RNTED, ZD 83-19, N8855, N21566, N23168 and N23661. Several hybrid soybean cultivars developed in China using CMS yielded 20% more than control varieties. However, both a low out cross pod-set rate in the existing CMS lines and the influence of day length and temperature on male sterility of some CMS lines and male fertility of F1 hybrids have delayed hybrid seed production in soybean to date (Bai and Gai, 2006; Zhao and Gai, 2006; Dong *et al.*, 2012; Nie *et al.*, 2017).

Stay-green mutants show impaired chlorophyll degradation during leaf senescence and seed maturation and they can affect seed maturation, seed oil quality, and meal quality in oil crops (Delmas *et al.*, 2013). Among the stay-green mutants described in soybean, green cotyledon gene *cytG* is maternally inherited (Terao, 1918). Sequencing of the chloroplast genome revealed a 5-bp insertion causing a frame-shift in *psbM* gene, which encodes one of the small subunits of photosystem II, thus linking photosynthesis in pre-senescent leaves with chlorophyll degradation during leaf senescence and seed maturation (Kohzuma *et al.*, 2017).

Significant reciprocal effects on physiological characters such as CO_2 exchange rates, intercellular CO_2 concentration, stomatal conductance, transpiration, plant height, number of branches, fertile nodes, filled pods, seeds per plant, weight of seeds per plant and weight of 100 seeds were detected in soybean by diallel analysis (Karyawati *et al.*, 2015). Using the same analysis Cruz *et al.* (2011) found significant reciprocal effects on resistance to Asian soybean rust (*Phakopsora pachyrhizi*) and suggested the involvement of cytoplasmic or maternal effects on this trait.

Xu *et al.* (2011) looked for QTLs for the seed size traits in soybean using F2:3, F2:4 and F2:5 populations from the direct and reciprocal crosses and employing a multi-QTL joint analysis (MJA) along with composite interval mapping (CIM). These authors detected cytoplasmic effects on seed length, seed width, seed thickness and the ratios length to thickness and width to thickness, but not on the ratio length to width. Besides, 92 cytoplasm-by-QTL interactions were detected, 28 of which were consistent with main effect QTLs detected by CIM.

Cotton

Most recognized CMS systems in cotton are CMS-D2 and CMS-D8, that were developed by transferring the cytoplasm of wild species *Gossypium harknessii* Brandegee and *Gossypium trilobum* (DC) Skovst, respectively, into tetraploid upland cotton *Gossypium hirsutum* (Wang *et al.*, 2010; Wu *et al.*, 2017; Yang *et al.*, 2018). Regarding the effect of male sterile cytoplasms on agronomic traits Tuteja and Banga (2011) evaluated four D2 type male sterile lines (A) and their corresponding maintainer lines (B) crossed as paired crosses with eight restorer lines (R). Cytoplasmic effects were estimated by comparing (A × R) and (B × R) hybrids combinations. Results indicated that although male sterile cytoplasm had a significant unfavorable effect on number of bolls, boll weight, yield and fiber quality traits in some of the cross combinations, performance of male sterility-based hybrids in cotton is governed by the interaction of nuclear genes with the sterility-inducing cytoplasm, making it more appropriate to test the CMS lines in

newer combinations rather than converting the female parents of released hybrids into male sterile lines. Recently, a comparison between cytoplasmic effects of D2 and D8 on lint yield and fiber quality was done by Zhang *et al.* (2019) using eight pairs of reciprocal hybrids obtained from crosses between two restorer lines carrying D2 and D8 cytoplasm and four commercial cotton cultivars. Analysis of results demonstrated that D2 cytoplasm had mild negative effects on lint yield and its component, but it had beneficial effects on most of the fiber quality traits. On the other hand, the negative effects of D8 cytoplasm on lint yield and its components were more profound than D2 cytoplasm, with no effect on quality traits (except for a reduction in micronaire), thus presenting important challenges for hybrid cotton breeding.

Complete diallel designs have frequently been used to determine the mode of gene action for agronomic traits in cotton. Using this approach, significant reciprocal effects have been detected for fiber length, fiber strength, fiber elongation and fiber fineness and lint percentage (Shaukat *et al.*, 2013), monopodia branch length (Zangi *et al.*, 2010), days to first flowering, seeds, locule^{-1} and lint percentage (Khan *et al.*, 2011).

Using another approach, Wu *et al.* (2010) designed an additive and dominance (AD) genetic model with cytoplasmic effects to estimate genetic effects on several seed traits in F3 hybrids of 13 cotton chromosome substitution lines crossed with five elite cultivars. Significant cytoplasmic effects were detected for seed oil content, oil index, seed index, seed volume, and seed embryo percentage.

PERSPECTIVES

Current knowledge and methods available make it possible to envisage greater opportunities to select not just the best nuclear genotypes but also the best cytonuclear interactions, by considering the information of all the genomes present in different plant cell compartments. Characterization of genetic diversity of chloroplast and mitochondrial genomes can be easily achieved by modern "omics" technologies. Besides, assessment of the additive and epistatic effects of cytoplasmic genes on traits of interest is facilitated by the development of specific genetic designs and statistical models. In addition, molecular markers associated with cytoplasm types can be applied in Marker Assisted Selection (MAS) schemes to increase the efficiency of their incorporation in breeding programs. As it has been suggested by Kersten *et al.* (2016), the availability of the genomic information of all three DNA-containing cell organelles will allow a holistic approach in plant breeding in the future. This perspective will contribute to optimize the use of genetic resources and will allow increasing genetic cytoplasmic diversity to reduce the vulnerability of crops to potential biotic and abiotic risks.

REFERENCES

Akula U.V., Poluru P.G., Jangam A.K., Jagannath P.V. (2012) Influence of types of sterile cytoplasm on the resistance to sorghum shoot fly (*Atherigona soccata*). Plant Breed. 131: 94-99.

Ali A., Yang E.M., Lee S.Y., Chung S.M. (2013) Evaluation of chloroplast genotypes of Korean cucumber cultivars (*Cucumis sativus* L.) using sdCAPS markers related to chilling tolerance. Kor. J. Hort. Sci. Technol. 31: 219-223.

Ali A., Yang E.M., Bang S.W., Chung S. M., Staub J.E. (2014) Assessment of chilling injury and molecular marker analysis in cucumber cultivars (*Cucumis sativus* L.). Kor. J. Hort. Sci. Technol. 32: 227-234.

Allen J.F. (2015) Why chloroplasts and mitochondria retain their own genomes and genetic systems: Colocation for redox regulation of gene expression. Proc. Natl. Acad. Sci. U.S.A. 112: 10231-10238.

Allen J.O. (2005) Effect of Teosinte Cytoplasmic Genomes on Maize Phenotype. Genetics 169: 863-880.

Allen J.O., Fauron C.M., Minx P., Roark L., Oddiraju S., Lin G.N., Meyer L., Sun H., Kim K., Wang C., Du F., Xu D., Gibson M., Cifrese J., Clifton S.W., Newton K.J. (2007) Comparisons among two fertile and three male-sterile mitochondrial genomes of maize. Genetics 177: 1173-1192.

Alessandro M.S., Galmarini C.R., Iorizzo M., Simon P.W. (2013) Molecular mapping of vernalization requirement and fertility restoration genes in carrot. Theor. Appl. Genet. 126: 415-423.

Ammiraju J.S.S., Dholakia B.B., Jawdekar G., Santra D.K., Gupta V.S., Roder M.S., Singh H., Lagu M.D., Dhaliwa H.S., Rao V.S.,Ranjekar P.K. (2002) Inheritance and identification of DNA markers associated with yellow berry tolerance in wheat (*Triticum aestivum* L.). Euphytica 123: 229–223.

Anisimova I.N., Gavrilenko T.A. (2017) Cytoplasmic male sterility and prospects for its utilizationin potato breeding, genetic studies and hybrid seed production. Russ. J. Genet. Appl. Res.7: 721-735.

Archibald, J.M. (2015) Endosymbiosis and eukaryotic cell evolution. Curr. Biol. 25: R911-R921.

Atienza S.G., Martin A., Pecchioni N., PlataniC., Cattivelli L. (2007) The nuclear–cytoplasmic interaction controls carotenoid content in wheat. Euphytica 159: 325-331.

Bai Y.N., Gai J.Y. (2006) Development of a new cytoplasmic-nuclear male-sterility line of soybean and inheritance of its male-fertility restorability. Plant Breed. 125: 85-88.

Bartoszewski G., Malepszy S., Havey M.J. (2004) Mosaic (MSC) cucumbers regenerated from independent cell cultures possess different mitochondrial rearrangements. Curr. Genet. 45: 45–53.

Bartoszewski G., Havey M.J., Ziółkowska A., Długosz M., Malepszy S. (2007) The selection of mosaic (MSC) phenotype after passage of cucumber (*Cucumis sativus* L.) through cell culture—a method to obtain plant mitochondrial mutants. J. Appl. Genet. 48: 1–9.

Beltrán J., Wamboldt Y., Sanchez R., LaBrant E.W., Kundariya H., Virdi K.S., Elowsky C., Mackenzie S.A. (2018) Specialized plastids trigger tissue-specific signaling for systemic stress response in plants. Plant Physiol. 178: 672-683.

Berninger E. (1965) Contribution à l'étude de la stérilité mâle de l'oignon (*Allium cepa* L.). Ann. Amélior. Plant. 15: 183-199.

Bhatia R., Dey S.S., Sharma P.C., Kumar R. (2015) *In v*itro maintenance of CMS lines of Indian cauliflower: an alternative for conventional CMS-based hybrid seed production. J. Hortic. Sci. Biotechnol. 90: 171-179.

Birky C.W. (1995) Uniparental inheritance of mitochondrial and chloroplast genes. Mechanisms and evolution. Proc. Natl. Acad. Sci. U.S.A. 92: 11331-11338.

Blanco N.E., Guinea-Dıaz M., Whelan L., Strand A. (2014) Interaction between plastid and mitochondrial retrograde signaling pathways during changes to plastid redox status. Philos. Trans. R. Soc. Lond. B Biol. Sci. 369:20130231.

Bnejdi F., Saadoun M., El Gazzah M. (2010) Cytoplasmic effects on grain resistance to yellowberry in durum wheat. Czech. J. Genet. Plant Breed. 46:145-148.

Bock D.G., Andrew R.L., Rieseberg L.H. (2014) On the adaptive value of cytoplasmic genomes in plants. Mol. Ecol. 23:4899-4911.

Bohra A., Jha U.C., Adhimoolam P., Bisht D., Singh N.P. (2016) Cytoplasmic male sterility (CMS) in hybrid breeding in field crops. Plant Cell Rep. 35: 967-993.

Börner T., Linke B., Nothnagel T., Scheike, Schulz B., Steinborn R., Brennicke A., Stein M., Wricke G. (1995) Inheritance of nuclear and cytoplasmic factors affecting male sterility in *Daucus carota*. In; Kück U., Wricke G. (Eds.) Genetic mechanisms for hybrid breeding. Adv. Plant Breeding 18. Blackwell Science, Berlin, pp. 111-122.

Botterman J., Leemans J. (1988) Engineering of herbicide resistance in plants. Biotech. Gen. Eng. Rev. 6:321-340.

Boussardon C.,Martin-Magniette M-L., Godin B.,Benamar A., Vittrant B., Citerne S.,Mary-Huard T., Macherel D., Rajjou L., Budar F. (2019) Novel cytonuclear combinations modify*Arabidopsis thaliana*seed physiologyand vigor. Front. Plant Sci. 10:32.

Brunkard J.O., Burch-Smith T.M. (2018) Ties that bind: the integration of plastid signaling pathways in plant cell metabolism. Essays Biochem. 62: 95-107.

Bruns H.A. (2017) Southern corn leaf blight: astory worth retelling. Agron. J. 109:1-7.

Cabral P.D.S., do Amaral Júnior A.T.; Duarte Vieira H., Saltires Santos J., Freitas I. L. de J., Gonzaga Pereira M. (2013). Genetic effects on seed quality in diallel crosses of popcorn. Ciênc. Agrotec. 37: 502-511.

Calugar R., Rotar I., Has V. (2016) Cytoplasmic diversification effect on certain plant vegetative traits of some maize (*Zea mays* l.) inbred lines. Res. J. Agric. Sci. 48:28-35.

Cervantes Ortiz F., García De los Santos G., Carballo-Carballo A., Bergvinson D., Crossa J.L., Mendoza-Elos M., Moreno-Martínez E. (2007) Herencia del vigor de plántula y su relación con caracteres de planta adulta en líneas endogámicas de maíz tropical. Agrociencia 41: 425-433.

Chandra-Shekara A.C., Prasanna B.M., Singh B.B., Unnikrishnan K.V., Seetharam A. (2007) Effect of cytoplasm and cytoplasm-nuclear interaction on combining ability and heterosis for agronomic traits in pearl millet *Pennisetum glaucum* (L) Br. R. Euphytica 153:15-26.

Chang C., Sun D., Kakihara F., Hondo K. (2015) Cytoplasmic effects of *Brassica napus* and *B. juncea* on extreme temperature stresses of *B. carinata*. Euphytica 204:335-342.

Chen H., Jiang S., Zheng J., Lin Y. (2013) Improving panicle exertion of rice cytoplasmic male sterile line by combination of artificial microRNA and artificial target mimic. Plant Biotechnol. J. 11:336-343.

Chen Z., Zhao N., Li S., Grover C.E., Nie H., Wendel J.F., Hua J. (2017) Plant mitochondrial genome evolution and cytoplasmic male sterility. Crit. Rev. Plant Sci. 36: 55-69.

Chettoor A. M., Phillips A. R., Coker C. T., Dilkes B., Evans M. M. (2016) Maternal gametophyte effects on seed development in maize. Genetics 204: 233-48.

Chi W., Feng P., Ma J., Zhang L. (2015) Metabolites and chloroplast retrograde signaling. Curr. Opin. Plant Biol.25:32–38.

Christov M. (2013) Contribution of interspecific and intergeneric hybridization to sunflower breeding. Helia 36: 1-18.

Chung S.M., Staub J.E., Fazio G. (2003) Inheritance of chilling injury: a maternally inherited trait in cucumber. J. Amer. Soc. Hort. Sci. 128:526-530.

Chung S.M., Gordon V.S., Staub J.E. (2007) Sequencing cucumber (*Cucumis sativus* L.) chloroplast genomes identifies difference between chilling-tolerant and -susceptible cucumber lines. Genome 50:215-225.

Corriveau J.P., Coleman A.W. (1988) Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 Angiosperm species. Am. J. Bot. 75: 1443-1458.

Cravero V., Nestares G., Zorzoli R., Picardi L. (2012) Verifying the influence of the cytoplasmon the regeneration ability in sunflower. HELIA 35: 9-18.

Crawford T., Lehotai N., Strand A. (2018) The role of retrograde signals during plant stress responses. J. Exp. Bot. 69: 2783-2795.

Crosatti C., Quansah L., Maré C., Giusti L., Roncaglia E., Atienza S. G., Cattivelli L., Fait A. (2013) Cytoplasmic genome substitution in wheat affects the nuclear-cytoplasmic cross-talk leading to transcript and metabolite alterations. BMC genomics 14: 868.

Cruz M.F. A. da., Souza G. A. de., Rodrigues F. Á., Sediyama C. S., Barros E. G. de. (2011) Reação de genótipos de soja à infecção natural por ferrugem asiática. Pesq. Agropec. Bras. 46: 215-218.

Daniell H., Lin C. S., Yu M., Chang W. J. (2016) Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 17:134.

de Abreu, V. M., Von Pinho E. V. de R., Mendes-Resende M.P., Balestre M., Lima A.C., Santos H. O., Von Pinho R.G. (2019) Combining ability and heterosis of maize genotypes under water stress during seed germination and seedling emergence. Crop Sci. 59: 33-43.

de la Torre M.V., Biasutti C.A. (2015) Efectos recíprocos y aptitud combinatoria en caracteres relacionados al vigor temprano y al rendimiento en grano en maíz. AgriScientia 32: 41-53

de Souza A., Wang J.Z., Dehesh K. (2017) Retrograde signals: integrators of interorganellar communication and orchestrators of plant development. Annu. Rev. Plant Biol. 68:85-108.

Del Valle-Echevarria A., Kiełkowska A., Bartoszewski G., Havey M.J. (2015) The mosaic mutants of cucumber: a method to produce knock-downs of mitochondrial transcripts. G3 5: 1211-1221.

Delmas F., Sankaranarayanan S., Deb S., Widdup E., Bournonville C., Bollier N., Northey J.G., McCourt P., Samuel M.A. (2013) ABI3 controls embryo degreening through Mendel's I locus. Proc. Natl. Acad. Sci. U.S.A. 110: E3888-E3894.

Dey S., Bhatia R., Bhardwaj I., Mishra V., Sharma K., Parkash C., Kumar S., Sharma V.K., Kumar R. (2017a) Molecular-agronomic characterization and genetic study reveals usefulness of refined Ogura cytoplasm based CMS lines in hybrid breeding of cauliflower (*Brassica oleracea* var. *botrytis* L.).Sci. Hort.224: 27-36.

Dey S.S., Bhatia R., Prakash C., Sharma S., Dabral M., Mishra V.. Bhardwaj I., Sharma K., Sharma V.K., Kumar R. (2017b) Alteration in important quality traits and antioxidant activities in *Brassica oleracea* with Ogura cybrid cytoplasm. Plant Breed. 136: 400-409.

Dhall R.K. (2010) Status of male sterility in vegetables for hybrid development. A review. Adv. Hort. Sci. 24: 263-279.

Dhillon M.K., Sharma H.C., Smith C.M. (2008) Implications of cytoplasmic male-sterility systems for development and deployment of pest resistant hybrids in cereals. CAB Rev. Perspect. Agric. Vet. Sci. Nutrit. Nat. Res. 3 (068): 1-16.

Dobler R., Rogell B., Budar F., Dowling D.K (2014) A meta-analysis of the strength and nature of cytoplasmic genetic effects. J. Evol. Biol. 27: 2021-2034.

Dong D., Li Z., Yuan F., Zhu S., Chen P., Yu W., Yang Q.H., Fu X.J., Yu X.M., Li B.Q., Zhu, D.H. (2012) Inheritance and fine mapping of a restorer-of-fertility (*Rf*) gene for the cytoplasmic male sterility in soybean. Plant Sci. 188: 36-40.

Durga K.K., Reddy B.V.S., Reddy M.S.S., Ganesh M. (2008) Influence of cytoplasm on the occurrence of leaf blight (*Exserohilum turcicum* (PASS.)) in sorghum (*Sorghumbicolor* (L.) MOENCH). Indian J. Agric. Res., 42: 97-101.

EeckhautT., Lakshmanan P.S., Deryckere D., Van Bockstaele E., Van Huylenbroeck J. (2013) Progress in plant protoplast research. Planta 238: 991-1003.

Elkonin L.A., Tsvetova M.I. (2012) Heritable effect of plant water availability conditions on restoration of male fertility in the "9E" CMS-inducing cytoplasm of sorghum. Front. Plant Sci. 3: 91.

El-Namaky R. (2018) The genetic variability of floral and agronomic characteristics of newly-bred cytoplasmic male sterile rice. Agriculture 8: 68.

FanX.M., ZhangY.D., YaoW.H., BiY.Q., LiuL., ChenH.M., KangM.S.(2014) Reciprocal diallel crosses impact combining ability, variance estimation, and heterotic group classification. Crop Sci. 54: 89-97.

Ferfuia C., Vannozzi G.P. (2015) Maternal effect on seed fatty acid composition in a reciprocal cross of high oleic sunflower (*Helianthus annuus* L.). Euphytica 205: 325-336.

Fernandes Santos C.A., Lopes Leite D., Rodrigues Oliveira V., Amorim Rodrigues M. (2010) Marker-assisted selection of maintainer lines within and onion tropical population. Sci. Agric. (Piracicaba, Braz.) 67: 223-227.

Figueiredo D. D., Köhler C. (2016) Bridging the generation gap: communication between maternal sporophyte, female gametophyte and fertilization products. Curr. Opin. Plant Biol. 29: 16-20.

Frei U., Peiretti E. G., Wenzel G. (2003) Significance of cytoplasmic DNA in plant breeding. Plant Breed. Rev. 21:175-210.

Gabay-Laughnan S., Laughnan J.R. (1994) The male sterility and restorer genes in maize. In: Freeling M., Walbot V. (Eds.) The maize handbook. Springer, New York, pp. 418-423.

Gao Z.S., Cai H.W., Liang G.H. (2005) Field assay of seedling and adult-plant resistance to southern leaf blight in maize. Plant Breed. 124: 356-360.

Golabadi M., Golkar P., Eghtedary A. (2015) Combining ability analysis of fruit yield and morphological traits in greenhouse cucumber (*Cucumis sativus* L.). Can. J. Plant Sci. 95: 377-385.

Goldman L., Schroeck G., Havey M.J. (2000) History of public onion breeding programs and pedigree of public onion germplasm releases in the United States. Plant Breed. Rev. 20: 67-103.

Gordon V.S., Staub J.E. (2011) Comparative analysis of chilling response in cucumber through plastidic and nuclear genetic effects component analysis. J. Amer.Soc.Hort.Sci. 136: 256-264.

Gordon V.S., Staub J.E. (2014) Backcross introgression of plastomic factors controlling chilling tolerance into elite cucumber (*Cucumis sa*tivus L.) germplasm: early generation recovery of recurrent parent phenotype. Euphytica 195:217-234.

Greiner S. (2012) Plastome mutants of higher plants. In: Bock R., Knoop V. (Eds.) Genomics of chloroplasts and mitochondria. Advances in Photosynthesis and Respiration, 35. Springer, Dordrecht, pp. 237-266.

Greiner S., Bock R. (2013) Tuning a menage a trois: co-evolution and co-adaptation of nuclear and organellar genomes in plants. Bioessays 35: 354-365.

Greiner S., Sobanski J., Bock R. (2014) Why are most organelle genomes transmitted maternally? Bioessays 37: 80-94.

Griffing B. (1956) Concept of general and specific combining ability in relation to a diallel crossing system. Aust. J. Biol. Sci. 9: 463-493.

Gualberto J.M., Mileshina D., Wallet C., Niazi A.K., Weber-Lofti F., Dietrich A. (2014) The plant mitochondrial genome: dynamics and maintenance. Biochimie 100: 107-120.

Guo J., Li Z., Zhang G., Tang H., Song O., Song Y., Ma S-., Niu N., Wang J. (2017) Special heterogeneous cytoplasm suppresses the expression of the gene producing multi-ovary in common wheat. Euphytica 213:247.

Guo J., Zhang G., Tang H., Song Y., Ma S., Niu N., Wang J. (2018) Changes in DNA methylation are associated with heterogeneous cytoplasm suppression of the multi-ovary gene in wheat (*Triticum aestivum*). Crop Pasture Sci. 69: 354-361.

Guo Y., Si P., Wang N., Wen J., Yi B., Ma B., Tu J., Zou J., Fu T., Shen J. (2017) Genetic effects and genotype× environment interactions govern seed oil content in *Brassica napus* L. BMC Genet. 18: 1.

Hansche P.E., Gabelman W.H. (1963) Digenic control of male sterility in carrots, *Daucus carota* L. Crop Sci. 3: 383-386.

Havey M.J. (1997) Predominant paternal transmission of the mitochondrial genome in cucumber. J. Hered. 88: 232-235.

Havey M.J. (1999) Seed yield, floral morphology and lack of male-fertility restoration of male-sterile onion (*Allium cepa*) populations possessing the cytoplasm of *Allium galanthum*. J. Amer. Soc. Hort. Sci. 124: 626-629.

Havey M.J. (2000) Diversity among male-sterility-inducing and male-fertile cytoplasms of onion. Theor. Appl. Genet. 101: 778-782.

Havey M.J. (2004) The use of cytoplasmic male sterility for hybrid seed production. In: Daniell H., Chase C. (Eds.) Molecular biology and biotechnology of plant organelles. Springer, Dordrecht, pp. 623-634.

Havey M.J. (2018) Onion breeding. In: Goldman I. (Ed.) Plant breeding reviews, 42. John Wiley & Sons Inc., New Jersey. pp. 39-85.

Hayman B. I. (1954) The theory and analysis of diallel crosses. Genetics 39: 789-809.

Hoffmann L. Jr, Rooney W.L. (2013) Cytoplasm has no effect on the yield and quality of biomass sorghum hybrids. J. Sustain. Bioenergy Syst. 3: 129-134.

Hosaka K., Sanetomo R. (2012) Development of a rapid identification method for potato cytoplasm and its use for evaluating Japanese collections. Theor. Appl. Genet. 125: 1237-1251.

Hu J., Chen G., Zhang H., Qian Q., Ding Y. (2016) Comparative transcript profiling of alloplasmic male-sterile lines revealed altered gene expression related to pollen development in rice (*Oryza sativa* L.). BMC Plant Biol. 16: 175.

Huang F., Fu X., Efisue A., Zhang S., Jin D. (2013) Genetically characterizing a new indica cytoplasmic male sterility with cytoplasm for its potential use in hybrid rice production. Crop Sci. 53: 132-140.

Ishaq M., Razia R., Khan S.A. (2016) Exploring genotypic variations for improved oil content and healthy fatty acids composition in rapeseed (*Brassica napus* L.). J. Sci. FoodAgric. 97: 1924-1930.

Jan C. C., Seiler G.J., Hammond J.J. (2014) Effect of wild *Helianthus* cytoplasms on agronomic and oil characteristics of cultivated sunflower (*Helianthus annuus* L.). Plant Breed. 133: 262-267.

Jan C.C., Vick B.A. (2007) Inheritance and allelic relationships of fertility restoration genes for seven newsources of male-sterile cytoplasm in sunflower. Plant Breed. 126: 213-217.

Jansky S. (2011) Parental effects on the performance of cultivated and wild species hybrids in potato. Euphytica 178: 273-281.

Jones H., Emsweller S. (1936) A male sterile onion. Proc. Amer. Soc. Hort. Sci. 34: 582-585.

Jones H., Clarke A. (1943) Inheritance of male sterility in the onion and the production of hybrid seed. Proc. Am. Soc. Hort. Sci. 43: 189-19.

Joseph, B., Corwin, J.A., Li, B., Atwell, S., Kliebenstein, D.J. (2013) Cytoplasmic genetic variation and extensive cytonuclear interactions influence natural variation in the metabolome. eLife 2: e00776.

Jovanović S.V., Todorović G., Grčić N., Štrbanović R., Stanisavljević R., Brkić J., Krizmanić G. (2017) Effects of different types of cytoplasm on the kernel row number of maize inbred lines. AGROFOR Int. J. 2: 155-161.

Kaminski P., Podwyszynska M., Starzycki M., Starzycka-Korbas E. (2016) Interspecific hybridization of cytoplasmic male-sterile rapeseed with Ogura cytoplasm and *Brassica rapa* var. *pekinensis* as a method to obtain male-sterile Chinese cabbage inbred lines. Euphytica 208: 519-534.

Kante M., Rattunde H.F.W., Nébié B., Weltzi E., Haussmann B.I.G., Leiser W.L. (2018) QTL mapping and validation of fertility restoration in West African sorghum A1 cytoplasm and identification of a potential causative mutation for *Rf2*. Theor. Appl. Genet.131: 2397-241.

Karyawati A.S., Sitompul S. M., Basuki N., Sitawati (2015) Combining ability analysis for physiological characters of soybean (*Glycine max* L. Merrill).Int. J. Plant Res. 5: 113-121.

Kersten B., Faivre Rampant P., Mader M., Le Paslier M-C., Bounon R., Berard A., Vettori C., Schroeder H., Leplé J-C., Fladung M (2016) Genome sequences of *Populus tremula* chloroplast and mitochondrion: implications for holistic poplar breeding. PLoSONE11: e0147209.

Khan S.A., Khan N.U., Mohammad F., Ahmad M., Khan I.A., Bibi Z., Khan I.U. (2011) Combining ability analysis in intraspecific F1 diallel cross of upland cotton. Pak. J. Bot. 43: 1719-1723.

Kihara H (1959) Fertility and morphological variation in the substitution backcrosses of the hybrid *Triticum vulgare* x *Aegilops caudata* . Proc. X Int. Congr. Genet. 1:.142–171.

Kim B., Kim K., Yang T., Kim S. (2016) Completion of the mitochondrial genome sequence of onion (*Allium cepa* L.) containing the CMS-S male-sterile cytoplasm and identification of an independent event of the *ccmFN* gene split. Curr. Genet. 62: 873-885.

Kim B., Yang T.J., Kim S. (2019) Identification of a gene responsible for cytoplasmic male-sterility in onions (*Allium cepa* L.) using comparative analysis of mitochondrial genome sequences of two recently diverged cytoplasms. Theor. Appl. Genet. 132: 313-322.

Kleine T., Leister D. (2016) Retrograde signaling: organelles go networking. Biochim. Biophys.Acta 1857: 1313-1325.

Kohzuma K., Sato Y., Ito H., Okuzaki A., Watanabe M., Kobayashi H., Nakano M., Yamatani H., Masuda Y., Nagashima Y., Fukuoka H., Yamada T., Kanazawa A., Kitamura K., Tabei Y., Ikeuchi M., Sakamoto W., Tanaka A., Kusaba M. (2017) The non-Mendelian green cotyledon gene in soybean encodes a small subunit of photosystem II. Plant Physiol. 173: 2138-2147.

Kollipara K.P., Saab I.N., Wych R.D., Lauer M.J., Singletary G.W. (2002) Expression profiling of reciprocal maize hybrids divergent for cold germination and desiccation tolerance. Plant Physiol. 129: 974-992.

Kozhemyakin V.V., Elkonin L.A., Dahlberg J.A. (2017) Effect of drought stress on male fertility restoration in A3 CMS-inducing cytoplasm of sorghum. Crop J. 5: 282-289.

Kumar A.A., Reddy B.V.S., Sharma H.C., Hash C.T., Rao P.S., Ramaiah B., Reddy P.S. (2011) Recent advances in sorghum genetic enhancement research at ICRISAT. Am. J. Plant Sci. 2: 589-600.

Kumar V., Apte U.B., Bhagwat S.G., Jadhav B.B., Sawant D.S., Das B.K. (2013) Phenotypic and molecular characterization of diversified cytoplasmic male sterile lines of rice (*Oryza sativa* L.). Electron. J. Plant Breed. 4: 1193-1200.

Kumar R.A., Oldenburg D.J., Bendich A.J. (2015) Molecular integrity of chloroplast DNA and mitochondrial DNA in mesophyll and bundle sheath cells of maize. Planta 241: 1221-1230.

Leclercq P. (1969) Cytoplasmic male sterility in sunflower. Ann. Amelior. Plant 19: 99-106.

Leite D.L., Galmarini C.R., Havey M.J. (1999) Cytoplasms of elite open-pollinated onions from Argentina and Brazil. Allium Improvement Newsletter 9: 1-4.

Li C., Zhao Z., Liu Y., Liang B., Guan S., Lan H., Wang J., Lu Y., Cao M. (2017) Comparative transcriptome analysis of isonuclear-alloplasmic lines unmask key transcription factor genes and metabolic pathways involved in sterility of maize CMS-C. PeerJ. 5: e3408.

Liberatore K.L., Dukowic-Schultze S., Miller M.E., Chen C., Kianian S.F. (2016) The role of mitochondria in plant development and stress tolerance. Free Radic. Biol. Med. 100: 238-256.

Lilly J.W., Bartoszewski G., Malepszy S., Havey M.J. (2001) A major deletion in the cucumber mitochondrial genome sorts with the MSC phenotype. Curr. Genet. 40: 144-151.

Linke B., Nothnagel T., Börner T. (1999) Morphological characterization of modified flower morphology of three novel alloplasmic male sterile carrot sources. Plant Breed. 118: 543-548.

Linke B., Alessandro M.S., Galmarini C.R., Nothnagel T. (2019) Carrot floral development and reproductive biology. In Simon P. *et al.* (Eds.) *The Carrot Genomes*. Compendium of Plant Genomes. Springer Nature, Switzerland, pp. 27-57.

Lössl A., Götz M., Braun A., Wenzel G. (2000) Molecular markers for cytoplasm in potato: male sterility andcontribution of different plastid-mitochondrial configurations to starch production. Euphytica 116: 221-230.

Machida L., Derera J., Tongoona P., MacRobert J. (2010) Combining ability and reciprocal cross effects of elite quality protein maize inbred lines in subtropical environments. Crop Sci. 50: 1708-1717.

Mackenzie S. A. (2010). The influence of mitochondrial genetics on crop breeding strategies. In: Janick J. (Ed.) Plant Breeding Reviews. John Wiley and Sons, New Jersey, pp. 115-138.

Makarenko M.S., Usatov A.V., Tatarinova T.V., Azarin K.V., Logacheva M.D., Gavrilova V.A., Horn R. (2019) Characterization of the mitochondrial genome of the MAX1 type of cytoplasmic male-sterile sunflower. BMC Plant Biol. 19 (Suppl 1): 51.

Malepszy S., Burza W., Smiech M. (1996) Characterization of a cucumber (*Cucumis sativus* L.) somaclonal variant with paternal inheritance. J. Appl. Genet. 37: 65-78.

McKay J.K., Richards J.H., Nemali K.S., Sen S., Mitchell-Olds T., Boles S., Stahl E.A.,Wayne T., Juenge T.E. (2008) Genetics of drought adaptation in *Arabidopsis thaliana* II. QTL analysis of a new mapping population Kas-1 x Tsu-1. Evolution 62:3014-26.

Miclaus M., Balacescu O., Has I., Balacescu L., Has V., Suteu D., Neuenschwander S., Keller I., Bruggmann R. (2016) Maize cytolines unmask key nuclear genes that are under the control of retrograde signaling pathways in plants. Genome Biol. Evol. 8: 3256-3270.

Mihovilovich E., Sanetomo R., Hosaka K., Ordoñez B., Aponte M., Bonierbale M. (2015) Cytoplasmic diversity in potato breeding: case study from the International Potato Center. Mol. Breed. 35: 137.

Mohammed R., Are A.K., Munghate R.S., Bhavanasi R., Polavarapu K.K.B., Sharma H.C. (2016) Inheritance of resistance to sorghum shoot fly *Atherigona soccata* in sorghum, *Sorghum bicolor* (L.) Moench. Front. Plant Sci. 7: 543.

Morley S.A., Nielsen B.L (2017) Plant mitochondrial DNA. Front Biosci 22: 1023-1032.

Mróz T.L., Eves-van den Akker S., Bernat A., Skarzyńska A., Pryszcz L., Olberg M., Havey M.J., Bartoszewski G. (2018) Transcriptome analyses of mosaic (MSC) mitochondrial mutants of cucumber in a highly inbred nuclear background. G3 8: 953-965.

Nie Z., Zhao T., Yang S., Gai J. (2017) Development of a cytoplasmic male-sterile line NJCMS4A for hybrid soybean production. Plant Breed. 136: 516-525.

Nothnagel T., Straka P., Linke B. (2000) Male sterility in populations of *Daucus* and the development of alloplasmic male-sterile carrot lines. Plant Breed. 119: 145-152.

Oldenburg D.J., Bendich A.J. (2015) DNA maintenance in plastids and mitochondria of plants. Front. Plant Sci. 6: 883.

Parkash C., Dey S.S., Bhatia R., Dhiman M.R. (2015) Indigenously developed SI and CMS lines in hybrid breeding of cabbage. Indian J. Hortic. 72: 212-217.

Pawełkowicz M., Zieliński K., Zielińska D., Pląder W., Yagi K., Wojcieszek M., Siedlecka E., Bartoszewski G., Skarzyńska A., Przybecki Z. (2016) Next generation sequencing and omics in cucumber (*Cucumis* sativus L.) breeding directed research. Plant Sci. 242: 77-88.

Pelletier G., Budar F. (2015) Brassica Ogu-INRA cytoplasmic male sterility: an example of successful plant somatic fusion for hybrid seed production. In: Li X-Q., Donnelly D.J., Jensen T.G. (Eds.) Somatic Genome Manipulation. Springer, New York, Heidelberg, Dordrecht, London, pp. 199-216.

Peterson C.E., Simon P.W. (1986) Carrot breeding. In: Bassett M.J. (Ed.) Breeding vegetable crops, AVI, Westport, pp. 321-356.

Rao R.S.P., Salvato F., Thalc B., Eubelc H., Thelend J.J., Møller I.M. (2017) The proteome of higher plant mitochondria. Mitochondrion 33: 22-37.

Reddermann A., Horn R. (2018) Recombination events involving the *atp9* gene are associated with male sterility of CMS PET2 in Sunflower. Int. J. Mol. Sci. 19: 806.

Reddy B.V.S., Ramesh S., Reddy S.P., Ramaiah B. (2007) Combining ability and heterosis as influenced by male-sterility inducing cytoplasms in sorghum [*Sorghum bicolor* (L.) Moench]. Euphytica 154: 153-164.

Reddy S.P., Rao M.D., Reddy B.V.S., Are A.K., Thakur R.P., Rao V.P. (2011) Evaluation of A1, A2, A3, A4(M), A4 (G), A4 (VZM) cytoplasm in isonuclear backgrounds for grain mold resistance. Crop Prot. 30: 658-662.

Rolland N., Bouchnak I., Moyet L., Salvi D., Kuntz M. (2018) The main functions of plastids. In: Maréchal E. (Ed.) Plastids. Methods in Molecular Biology, Humana Press, New York, pp. 73-85.

Roux F., Mary-Huard T., Barillot E., Wenes E., Botran L., Durand S., Villoutreix R., Martin-Magniette M.L., Camilleri C., Budar F. (2016) Cytonuclear interactions affect adaptive traits of the annual plant *Arabidopsis thaliana* in the field. Proc. Natl. Acad. Sci. U.S.A. 113:3687-3692.

Sabar M., Gagliardi D., Balk J., Leaver C.J. (2003) ORFB is a subunit of F1FO-ATP synthase: insight into the basis of cytoplasmic male sterility in sunflower. EMBO Rep. 4: 381-386.

Sanetomo R., Gebhardt C. (2015) Cytoplasmic genome types of European potatoes and their effects on complex agronomic traits. BMC Plant Biol. 15: 162.

Sanetomo R., Hosaka K. (2011) Reciprocal differences in DNA sequence and methylation status of the pollen DNA between F1 hybrids of *Solanum tuberosum* x *S. demissum*. Euphytica 182: 219-229.

Sanetomo R., Ono S., Hosaka K. (2011) Characterization of crossability in the crosses between *Solanum demissum* and *S. tuberosum*, and the F1 and BC1 progenies. Am. J. Pot. Res. 88: 500-510.

Sanetomo R., Hosaka K. (2013) A recombination-derived mitochondrial genome retained stoichiometrically only among *Solanum verrucosum* Schltdl. and Mexican polyploid wild potato species. Genet. Resour. Crop Evol. 60: 2391-2404.

Santos J.F., Dirk L.M.A., Downie A.B., Sanches M.F.G., Vieira R.D. (2017) Reciprocal effect of parental lines on the physiological potential and seed composition of corn hybrid seeds. Seed Sci. Res. 27: 206-216.

Satyavathi V.V., Manga V., Rao M.V., Chittibabu M. (2016) Genetic analysis of reciprocal differences in the inheritance of *in vitro* characters in pearl millet. Genet. Mol. Biol. 39: 54-61.

Schönhals E.M., Ortega F., Barandalla L., Aragones A., Ruiz de Galarreta J.I., Liao J.-C., Sanetomo R., Walkemeier B., Tacke E., Ritter E., Gebhardt C. (2016) Identification and reproducibility of diagnostic DNA markers for tuber starch and yield optimization in a novel association mapping population of potato (*Solanum tuberosum* L.). Theor. Appl. Genet.129: 767-785.

Schweisguth B. (1973) Étude d'un nouveau type de stérilité mâle chez l'oignon, *Allium cepa* L. Ann. Amélior. Plant. 23: 221-233.

Sekhon B.S., Singh Y., Sharma S., Sharma S., Verma A., Vishalakshi, Singh H. (2018) Cytoplasmic male sterility (CMS) in cauliflower breeding: are view. Adv. Res. 15: 1-8.

Serieys H., Christov M. Identification, study and utilization in breeding programs of new CMS sources (1999 - 2004). Proc. FAO Consultation Meeting, 17-20 July 2004, Novi Sad, Serbia, pp. 1-63.

Sharma H.C., Dhillon M.K., Reddy B.V.S. (2006) Expression of resistance to *Atherigona soccata* in F1 hybrids involving shoot fly resistant and susceptible cytoplasmic male-sterile and restorer lines of sorghum. Plant Breed. 125: 473-477.

Shaukat S., Khan T.M., Shakee L. A., Ijaz S. (2013) Estimation of best parents and superior cross combinations for yield and fiber quality related traits in upland cotton (*Gossypium hirsutum* L.). Sci. Tech. Dev. 32: 281-284.

Shen J., Dirks R., Havey M.J. (2015) Diallel crossing among doubled haploids of cucumber reveals significant reciprocal-cross differences. J. Amer. Soc. Hort. Sci. 140:178-182.

Singh S., Bhatia R., Kumar R., Sharma K., Dash S., Dey S.S. (2018) Cytoplasmic male sterile and doubled haploid lines with desirable combining ability enhances the concentration of important antioxidant attributes in *Brassica oleracea*. Euphytica 214: 207.

Smith D. R., Keeling P. J. (2015). Mitochondrial and plastid genome architecture: Reoccurring themes, but significant differences at the extremes. Proc. Natl. Acad. Sci. U.S.A., 112: 10177-10184.

Soltani A., Kumar A., Mergoum M., Pirseyedi S.M., Hegstad J.B., Mazaheri M., Kianian S.F. (2016) Novel nuclear-cytoplasmic interaction in wheat (*Triticum aestivum*) induces vigorous plants. Funct. Integr. Genomics16: 171-182.

Su A., Song W., Xing J., Zhao Y., Zhang R., Li C., Duan M., Luo M., Shi Z., Zhao J. (2016) Identification of genes potentially associated with the fertility instability of S-type cytoplasmic male sterility in maize via bulked segregant RNA-Seq. PLoS ONE 11: e0163489.

Szała L., Cegielska-Taras T., Adamska E., Kaczmarek Z. (2018) Assessment of genetic effects on important breeding traits in reciprocal DH populations of winter oilseed rape (*Brassica napus* L.). J. Integr. Agr. 17:76-85.

Takenaka S., Yamamoto R., Nakamura C. (2018) Genetic diversity of submergence stress response in cytoplasms of the *Triticum*-

Aegilops complex. Sci. Rep: 1-13.

Talukder S.K., Prasad P.V., Todd T., Babar M.A., Poland J., Bowden R., Fritz A. (2014) Effect of cytoplasmic diversity on post anthesis heat tolerance in wheat. Euphytica, 204: 383-394.

Tang H., Xie Y., Liu Y-G-., Chen L. (2017) Advances in understanding the molecular mechanisms of cytoplasmic male sterility and restoration in rice. Plant Reprod. 30: 179-184.

Tang Z., Yang Z., Hu Z., Zhang D., Lu X., Jia B., Deng D., Xu C. (2013) Cytonuclear epistatic quantitative trait locus mapping for plant height and ear height in maize. Mol. Breed. 31: 1-14.

Tang Z., Hu W., Huang J., Lu X., Yang Z., Lei S., Zhang Y., Xu C. (2014) Potential involvement of maternal cytoplasm in the regulation of flowering time via interaction with nuclear genes in maize. Crop Sci. 54: 544-553.

Tao D., Hu F., Yang J., Yang G., Yang Y., Xu P., Li J., Ye C., Dai L. (2004) Cytoplasm and cytoplasm-nucleus interactions affect agronomic traits in japonica rice. Euphytica 135: 129-134.

Tao D., Xu P., Zhou J., Deng X., Li J., Deng W., Yang J., Yang G., Li Q., Hu F. (2011) Cytoplasm affects grain weight and filled-grain ratio in indica rice. BMC Genet. 12: 53.

Terao H. (1918) Maternal inheritance in the soybean. Am. Nat. 52: 51-56.

Thompson D.J. (1961) Studies on the inheritance of male sterility in the carrot (*Daucus carota L.* var. *sativa*). Proc. Am. Soc. Hort. Sci. 78: 332-338.

Tiwari J.K., Devi S., Ali N., Luthra S.K., Kumar V., Bhardwaj V., Singh R.K., Rawat S. K., Chakrabarti S.K. (2018) Progress in somatic hybridization research in potato during the past 40 years. Plant Cell Tissue Organ Cult. 132: 225-238.

Toriyama K., Kazama T. (2016) Development of cytoplasmic male sterileIR24 and IR64 using CW-CMS/Rf17 system. Rice 9: 22.

Tsunewaki T. (2009) Plasmon analysis in the *Triticum-Aegilops* complex. Breed. Sci. 59: 455-470.

Turner S.D., Maurizio P.L., Valdar W., Yandell B.S., Simon P.W. (2018) Dissecting the genetic architecture of shoot growth in carrot (*Daucus carota* L.) using a dialle mating design. G3 8: 411-426.

Tuteja O.P., Banga M. (2011) Effects of cytoplasm on heterosis for agronomic traits in upland cotton (*Gossypium hirsutum*). Indian J. Agr. Sci. 81: 1001-1007.

Tyagi V., Dhillon S.K., Bajaj R.K., Gupta S. (2015a) Phenotyping and genetic evaluation of sterile cytoplasmic male sterile analogues in sunflower (*Helianthus annuus* L.). Bangladesh J. Bot. 44: 23-30.

Tyagi V., Dhillon S.K., Gill B.S. (2015b) Morphophysiological expression in cms analogues of sunflower (*Helianthus annuus* L.) under water stress environment. Electron. J. Plant Breed.6: 1150-1156.

Tyagi V., Dhillon S.K. (2017) Effect of alien cytoplasm on combining ability for earliness and seed yield in sunflower under irrigation and drought stress. HELIA 40: 71-84.

Tyagi V., Dhillon S.K., Kaushik P. (2018) Stability analysis of some novel cytoplasmic male sterile sourcesof sunflower and their hybrids. HELIA 41: 153-200.

Vacek L.A., Rooney W.L. (2018) Effects of cytoplasm, male and female parents on biomass productivity in sorghum (*Sorghum bicolor* L. Moench). J. Crop Improv. 32: 635-647.

Van Aken O., Pogson B. J. (2017) Convergence of mitochondrial and chloroplastic ANAC017/PAP-dependent retrograde signaling pathways and suppression of programmed cell death. Cell Death. Differ. 24: 955-960.

Van Dingenen J., Blomme J., Gonzalez N., Inzé D. (2016) Plants grow with a little help from their organelle friends. J. Exp. Bot. 67: 6267-6281.

Velasco L., Pérez-Vich B., Fernández-Martínez J.M. (2007) Relationships between seed oil content and fatty acid composition in high stearic acid sunflower. Plant Breed. 126: 503-508.

Vu H.Q., Yoshimatsu Y., Khrustaleva L.I., Yamauchi N., Shigyo M. (2012) Alien genes introgression and development of alien monosomic addition lines from a threatened species, *Allium roylei* Stearn, to *Allium cepa* L. Theor. Appl. Genet. 124: 1241-1257.

Wang Y., Berkowitz O., Selinski J., Xu Y., Hartmann A., Whelan J (2018) Stress responsive mitochondrial proteins in *Arabidopsis thaliana*. Free Radic. Biol.Med. 122: 28–39.

Wang F., Feng C., O'Connell M.A., Stewart J.McD., Zhang J. (2010) RFLP analysis of mitochondrial DNA in two cytoplasmic male sterility systems (CMS-D2 and CMS-D8) of cotton. Euphytica 172: 93.

Waza S.A., Jaiswal H.K. (2015) Effects of WA cytoplasm on various quality characteristics of rice hybrids. J. Anim. Plant. Sci. 25: 1693-1698.

Weider C., Stamp P., Christov N., Hüsken A., Foueillassar X., Camp K.H.,Munsch M., (2009) Stability of cytoplasmic male sterility in maize under different environmental conditions. Crop Sci. 49: 77-84.

Welch J.E., Grimball E.L. (1947) Male sterility in carrot. Science 106: 594.

Welchen E., García L., Mansilla N., Gonzalez D.H. (2013) Coordination of plant mitochondrial biogenesis: keeping pace with cellular requirements. Front. Plant Sci. 4: 551.

Wise R.R. (2007) The diversity of plastid form and function. In: Wise R.R., Hoober J.K. (Eds.) The Structure and Function of Plastids. Springer, Dordrecht, pp. 3-26.

Wolyn D., Chahal A. (1998) Nuclear and cytoplasmic interactions for petaloid male-sterile accessions of wild carrot (*Daucus carota* L.). J. Am. Soc. Hort. Sci. 123: 849-853.

Wu J., McCarty J.C., Jenkins J.N. (2010) Cotton chromosome substitution lines crossed with cultivars: genetic model evaluation and seed trait analyses. Theor. Appl. Genet. 120: 1473-1483.

Wu J., Zhang M., Zhang B., Zhang X., Guo L., Qi T., Wang H., Zhang J., Xing C. (2017) Genome-wide comparative transcriptome analysis of CMS-D2 and its maintainer and restorer lines in upland cotton. BMC Genomics 18: 454.

Xia G. (2009) Progress of chromosome engineering mediated by asymmetric somatic hybridization. J. Genet. Genomics 36: 547-556.

Xu J. (2005) The inheritance of organelle genes and genomes: patterns and mechanisms. Genome 48: 951-958.

Xu P., Yan W., HeJ., Li Y., Zhang H., Peng H., Wu X. (2013) DNA methylation affected by male sterile cytoplasm in rice (*Oryza sativa* L.). Mol. Breed. 31: 719-727.

Xu Y., Li H.-N., Li G.-J., Wang X., Cheng L.-G., Zhang Y.-M. (2011) Mapping quantitative trait loci for seed size traits in soybean (*Glycine max* L. Merr.). Theor. Appl. Genet. 122: 581-594.

Yamagishi H., Bhat R.S. (2014) Cytoplasmic male sterility in *Brassicaceae* crops. Review. Breed. Sci. 64: 38-47.

Yang L., Wu Y., Zhang M., Zhang J., Stewart J.McD., Xing C., Wu J., Jin S. (2018) Transcriptome, cytological and biochemical analysis of cytoplasmic male sterility and maintainer line in CMS-D8 cotton. Plant Mol. Biol. 97: 537.

Zhang J., Wang L., Zhao A., Liu H., Jan C.-C., Qi D., Liu G. (2010) Morphological and cytological study in a new type of cytoplasmic male-sterile line CMSGIG2 in sunflower (*Helianthus annuus*). Plant Breed. 129: 19-23.

Zhang J., Abdelraheem A.,Stewart J. McD. (2019) A comparative analysis of cytoplasmic effects on lint yield and fiber quality between CMS-D2 and CMS-D8 systems in upland cotton. Crop Sci. 59: 624-631.

Zhang X., Hirsch C.N., Sekhon R.S., de Leon N., Kaeppler S.M. (2016) Evidence for maternal control of seed size in maize from phenotypic and transcriptional analysis. J. Exp. Bot. 67: 1907-1917.

Zangi M. R., Jelodar N.B., Kazemitabar S.K., Vafaei-tabar M. (2010) Cytoplasmic and combining ability effects on agro-morphological characters in intra and inter crosses of Pima and upland cottons (G. *hirsutum* and G. *barbadense*). Int. J. Biol. 2: 94-102.

Zhao T.J., Gai J.Y. (2006) Discovery of new male-sterile cytoplasm sources and development of a new cytoplasmic-nuclear male-sterile line NJCMS3A in soybean. Euphytica 152: 387-396.

Zhao Y., Luo L., Xu J., Xin P., Guo H., Wu J., Bai L., Wang G., Chu J., Zuo J., Yu H., Huang X., Li J. (2018) Malate transported from chloroplast to mitochondrion triggers production of ROS and PCD in *Arabidopsis thaliana*. Cell Res. 28: 448-461.

Żmieńko A., Guzowska-Nowowiejska M., Urbaniak R., Pląder W., Formanowicz P., Figlerowicz M. (2011) A tiling microarray for global analysis of chloroplast genome expression in cucumber and other plants. Plant Methods 7: 29.