University of São Paulo
"Luiz de Queiroz" College of Agriculture

Association mapping to exploit maize diversity for drought tolerance: landraces and early testcrosses as genetic resources

Pedro Augusto Medeiros Barbosa

Thesis presented to obtain the degree of Doctor in Science.
Area: Genetics and Plant Breeding

Piracicaba
2020

Pedro Augusto Medeiros Barbosa
Agronomist

Association mapping to exploit maize diversity for drought tolerance: landraces and early testcross as genetic resources

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Advisor:
Prof. Dr. **ROBERTO FRITSCHE NETO**

Thesis presented to obtain the degree of Doctor in Science. Area: Genetics and Plant Breeding

Piracicaba
2020

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Barbosa, Pedro Augusto Medeiros

Association mapping to exploit maize diversity for drought tolerance: landraces and early testcross as genetic resources / Pedro Augusto Medeiros Barbosa. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2020.

81 p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura “Luiz de Queiroz”.

1. Melhoramento de plantas 2. GWAS 3. Análise espacial 4. Estresse abiótico 5. Diversidade genética I. Título

AGRADECIMENTOS

À minha família. Meu pai Valmir e minha mãe Nice, meu irmão Victor e irmã Marília. Assim como todos meus demais familiares. Que me proporcionaram um ótimo ambiente familiar para meu crescimento e desenvolvimento pessoal.

À Daniela Vilela, que tem me acompanhado por todas alegrias, dificuldades e acasos dos últimos anos.

À Escola Superior de Agricultura "Luiz de Queiroz", que me trouxe inúmeros aprendizados e amizades para toda a vida. E ao Programa de Pós-Graduação em Genética e Melhoramento de Plantas e seus professores por todo o apoio e ensinamentos.

Ao Prof. Dr. Roberto Fritsche Neto, pela confiança, orientação e, principalmente, pela oportunidade especial de realizar esta tese.

Aos meus colegas do "Laboratório de Melhoramento de Plantas Alógamas", pelo grande apoio e compreensão nesta rápida e intensa passagem pelo laboratório.

Ao CIMMYT, especialmente ao melhorista Dr. Terry Molnar, à Dr. Marcela Andrade e equipe do Programa de Recursos Genéticos. E aos Drs. Juan Burgueño, José Crossa e Fernando Toledo da Unidade de Estatística e Biometria pelas ideias e sugestões nas análises dos dados.

Os experimentos de campo conduzidos por esta pesquisa e a coleta dos dados fenotípicos contaram com o apoio da empresa Corteva Agriscience, da *Secretaría de Agricultura Y Desarrollo Rural* e também do *Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias* do Governo Mexicano.

Aos grandes amigos e amigas que fiz nestes anos de doutorado, sem vocês este caminho teria sido impossível de percorrer.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES, pelo apoio financeiro ao projeto e pela bolsa de estudos.

CONTENTS

RESUMO

Mapeamento associativo em milho para tolerância a seca: variedades crioulas e *testcross* precoces como fonte de novo alelos

A produção de milho ao redor do mundo vem apresentando uma grande dificuldade nos últimos anos, eventos de seca estão se tornando mais frequentes e severos, principalmente em regiões tropicais e subtropicais. Além disso, as mudanças climáticas têm sugerido que isso é uma tendência real para os próximos anos. O estresse causado pela deficiência hídrica nas lavouras de milho pode resultar em grandes quebras de produtividade de grãos. O México é o centro de origem do milho, e diversos são os centros de diversidade genética espalhados pela América Latina. Toda essa diversidade deve ser explorada pelos programas de melhoramento, já que ela é uma extraordinária fonte de novos alelos que podem ter um papel fundamental nos ganhos genéticos alcançados pelo melhoramento nos próximos anos colaborando para o enfrentamento dos desafios que são previstos. Esta tese de doutorado aborda o tema em duas partes. Primeiro, apresentamos uma revisão sobre a abundante diversidade genética do milho, e o uso de variedades crioulas nos programas de melhoramento para tolerância a seca, evidenciando a importância de se desenvolver germoplasma tropical adaptado as condições de seca no intuito de contornar este problema. Ainda na revisão, nós discutimos o uso de estudos de Associação Genômica Ampla (GWAS) e seleção assistida por marcadores (MAS) neste contexto. Na segunda parte, apresentamos um trabalho onde utilizamos GWAS em um programa de pré-melhoramento de milho visando tolerância à seca, onde variedades crioulas foram usadas como fonte de diversidade genética. O objetivo foi estudar os recursos genéticos oferecidos por um painel inicial de crioulas para identificar as regiões cromossômicas associadas à tolerância à seca. Para tanto, foram utilizadas 1306 progênies oriundas de 20 populações crioulas selecionadas por seu desempenho em regiões secas da América Latina. Dados fenotípicos foram obtidos a partir de *testcross* precoces de duas gerações (RC1S1 em 2016 e RCS2 em 2017) conduzidos em dois regimes hídricos distintos, um irrigado e outro sob seca, em três locais no México. O índice de média harmônica do desempenho relativo (HMRP), calculado com base nos rendimentos de grãos em ambos os regimes de água, foi utilizado como medida de tolerância à seca dos genótipos. Os valores genotípicos foram estimados usando ajuste espacial em uma análise de dois estágios. Um conjunto final de 5695 marcadores de polimorfismo de nucleotídeo único (SNPs) foi considerado para o mapeamento associativo. Como resultados, pudemos detectar um total de 10 marcadores significativos associados ao rendimento de grãos e ao índice de tolerância à seca, e sugerimos dois genes putativos mapeados próximos a dois desses marcadores que podem ter importante participação na resposta da planta ao estresse hídrico. Além disso, para dois SNPs associados, alelos originários das variedades crioulas resultaram em um rendimento ligeiramente maior sob condições de seca. Nossos resultados indicam que a diversidade oferecida pelas variedades crioulas combinada com linhagens já melhoradas pode ter um importante papel no melhoramento de milho tropical para a tolerância à seca.

Palavras-chave: Melhoramento de plantas, GWAS, Análise espacial, Estresse abiótico, Diversidade genética

ABSTRACT

Association mapping to exploit maize diversity for drought tolerance: landraces and early testcross as genetic resources

Maize (Zea mayz L.) production worldwide have been facing a tremendous obstacle in the past years, drought events are increasing in frequency and severity, mainly in tropical and subtropical regions. Moreover, climate change forecasts this as a trend for the next few years as well. The maize production can be highly affected by water deficiency stress, resulting in losses in grain yield. Mexico is the center of origin of maize, and there are many diversity centers across Latin America. This diversity should be exploited by breeding programs once it is a source of new alleles that can be responsible for the needed genetic improvements to face the forecasting challenges. This doctoral thesis addresses the theme in two sections. First, we present a review of maize genetic diversity, the use of landraces introduction in breeding programs to improve for drought tolerance, highlighting the importance to develop improved tropical germplasm to face the drought issue. In this review we also discuss the opportunity to apply Genome-Wide Associations Studies (GWAS) and Marker-Assisted Selection (MAS) in this context. In the second section we present an original GWAS application in a pre-breeding program using selected landraces as genetic sources for drought tolerance maize improvement. The aim was to study the genetic resources of a landrace panel to identify maize chromosomal regions associated with drought tolerance. For that, we performed the GWAS in 1306 landraces progenies originated from 20 landraces populations selected due to its agricultural performance in dry regions of Latin America. Phenotypic data were obtained from early testcross trials of two generations (BC1S1 in 2016 and BCS2 in 2017) conducted in two water regimes, irrigated and drought condition, in three locations in Mexico. Harmonic Mean of Relative Performance (HMRP) of grain yield in both water regimes was used as a measure of drought tolerance of the genotypes. The genotypic values were estimated using a spatial adjustment in a two-stage analysis. A final set of 5,695 single-nucleotide polymorphism (SNPs) markers was considered for GWAS. We were able to detect a total of 10 significant markers associated with grain yield and drought tolerance index, and we suggest two putative genes mapped close to two of these markers that can be part of the plant's response to drought stress. Besides, for two associated SNPs, the alleles from landraces provided a slightly higher yield under drought conditions. Our results indicate that the diversity delivered by these landraces combined with commercial lines is an exciting strategy to improve maize for drought tolerance.

Keywords: Plant breeding, GWAS, Spatial analysis, Abiotic stress, Genetic diversity.

1. INTRODUCTION

Maize (*Zea mays L.*) is a major crop cultivated worldwide and it plays a central role in human food security and economic development. It corresponded to about 13,6% of the whole area destined for agriculture in 2018 (FAOSTAT, 2018). The production of this cereal is to a large extent concentrated in tropical and subtropical regions, such as Central and South America, Africa and Southeast Asia. However, studies coordinated by the Intergovernmental Panel on Climate Change forecasts an increase in the frequency and severity of drought events in these regions (IPCC, 2007). As maize is a high water-demanding plant, these drought events have caused significant losses in grain yield.

In this context, plant breeding has an important participation in the solution to provide new drought-tolerant cultivars. Maize is a highly variable species and genetic diversity is the base to develop improved crops, what enables breeding as an essential tool to overcome the effects of climate change in agriculture (Bedoya et al., 2017; Fisher et al., 2015; Warburton et al., 2008). A constant challenge is to develop and apply the best strategies to incorporate this diversity into breeding programs. Exploit the diversity of landraces is one exciting approach for that (Gorjanc et al., 2016; Mayer et al., 2017).

Villa et al., (2005) defined the term landrace as "a dynamic population of a cultivated plant that has a historical origin, a distinct identity and lacks formal crop improvement, as well as often being genetically diverse, locally adapted and associated with traditional agricultural systems". Due to these characteristics, landraces play an important role in studies of evolution, conservation, and diversity (Matsuoka et al., 2002), as well as being sources of alleles of interest for breeding programs. The terms local, traditional or creole varieties are sometimes used to characterize landraces populations, mainly in anthropological studies (Costa, 2013). These terms could be treated differently, depending on the area of study. In the present research, landraces will be a population of cultivated maize that had not undergone a formal breeding program.

Plant breeding, as genetics science in general, had rapid progress in recent decades. One of the advances was the use of Genome-wide Association Studies (GWAS) in plant breeding, which has the main purpose to find genomic regions associated with a phenotypic trait. With the advent of new high-throughput DNA sequencing technologies, GWAS become an extensively used tool for several studies in humans, animals, and plants (Brachi et al., 2011). For plant breeding studies it is useful for identifying natural genetic variants that are related to complex traits (Huang and Han, 2014), and has been used to a better understanding of drought tolerance in maize (Farfan et al., 2015; Hao et al., 2011; Setter et al., 2011; Xiao et al., 2017).

In the next sections, we will provide a review and discussion about the maize diversity and how the breeding programs have dealt with it, especially to improve for drought tolerance in tropical regions, as well as how GWAS have been used in this context. Afterward, we report an original GWAS performed with landrace progenies, using phenotypic data from early testcrosses of a current breeding program to drought tolerance. We aim to study the genetic resources of the landrace panel to identify new beneficial variants in chromosomal regions associated with drought tolerance.

REFERENCES

Bedoya, C.A., Dreisigacker, S., Hearne, S., Franco, J., Mir, C., Prasanna, B.M., Taba, S., Charcosset, A., Warburton, M.L., 2017. Genetic diversity and population structure of native maize populations in Latin America and the Caribbean. PLoS One 12, 1–21. https://doi.org/10.1371/journal.pone.0173488

Brachi, B., Morris, G.P., Borevitz, J.O., 2011. Genome-wide association studies in plants: The missing heritability is in the field. Genome Biol. 12. https://doi.org/10.1186/gb-2011-12-10-232

Costa, F.M., 2013. Diversidade genética e distribuição geográfica: uma abordagem para a conservação on farm e ex situ e o uso sustentável dos recursos genéticos de milho do Oeste de Santa Catarina. Universidade Federal de Santa Catarina.

FAOSTAT, 2018. Database of Food and Agriculture Organization of the United Nations Statistics Division [WWW Document]. URL http://www.fao.org/faostat (accessed 20.02.20).

Farfan, I.D.B., De La Fuente, G.N., Murray, S.C., Isakeit, T., Huang, P.-C., Warburton, M., Williams, P., Windham, G.L., Kolomiets, M., 2015. Genome Wide Association Study for Drought, Aflatoxin Resistance, and Important Agronomic Traits of Maize Hybrids in the Sub-Tropics. PLoS One 10, e0117737. https://doi.org/10.1371/journal.pone.0117737

Fisher, M., Abate, T., Lunduka, R.W., Asnake, W., Alemayehu, Y., Madulu, R.B., 2015. Drought tolerant maize for farmer adaptation to drought in sub-Saharan Africa: Determinants of adoption in eastern and southern Africa. Clim. Change 133, 283–299. https://doi.org/10.1007/s10584-015-1459-2

Gorjanc, G., Jenko, J., Hearne, S.J., Hickey, J.M., 2016. Initiating maize pre-breeding programs using genomic selection to harness polygenic variation from landrace populations. BMC Genomics 17, 1–15. https://doi.org/10.1186/s12864-015-2345-z

Hao, Z., Li, X., Xie, C., Weng, J., Li, M., Zhang, D., Liang, X., Liu, L., Liu, S., Zhang, S., 2011. Identification of Functional Genetic Variations Underlying Drought Tolerance in Maize Using SNP Markers. J. Integr. Plant Biol. 53, 641–652. https://doi.org/10.1111/j.1744-7909.2011.01051.x

Huang, X., Han, B., 2014. Natural Variations and Genome-Wide Association Studies in Crop Plants. Annu. Rev. Plant Biol. 65, 531–551. https://doi.org/10.1146/annurev-arplant-050213-035715

IPCC, 2007. Summary for Policymakers. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, United Kingdom. https://doi.org/10.1038/446727a

Matsuoka, Y., Vigouroux, Y., Goodman, M.M., Sanchez G., J., Buckler, E., Doebley, J., 2002. A single domestication for maize shown by multilocus microsatellite genotyping. Proc. Natl. Acad. Sci. 99, 6080–6084. https://doi.org/10.1073/pnas.052125199

Mayer, M., Unterseer, S., Bauer, E., de Leon, N., Ordas, B., Schön, C.C., 2017. Is there an optimum level of diversity in utilization of genetic resources? Theor. Appl. Genet. 130, 2283–2295. https://doi.org/10.1007/s00122-017-2959-4

Setter, T.L., Yan, J., Warburton, M., Ribaut, J.M., Xu, Y., Sawkins, M., Buckler, E.S., Zhang, Z., Gore, M.A., 2011. Genetic association mapping identifies single nucleotide polymorphisms in genes that affect abscisic acid levels in maize floral tissues during drought. J. Exp. Bot. 62, 701–716. https://doi.org/10.1093/jxb/erq308

Villa, T.C.C., Maxted, N., Scholten, M., Ford-Lloyd, B., 2005. Defining and identifying crop landraces. Plant Genet. Resour. Charact. Util. 3, 373–384. https://doi.org/10.1079/PGR200591

Warburton, M.L., Reif, J.C., Frisch, M., Bohn, M., Bedoya, C., Xia, X.C., Crossa, J., Franco, J., Hoisington, D., Pixley, K., Taba, S., Melchinger, A.E., 2008. Genetic diversity in CIMMYT nontemperate maize germplasm: Landraces, open pollinated varieties, and inbred lines. Crop Sci. 48, 617–624. https://doi.org/10.2135/cropsci2007.02.0103

Xiao, Y., Liu, H., Wu, L., Warburton, M., Yan, J., 2017. Genome-wide Association Studies in Maize: Praise and Stargaze. Mol. Plant 10, 359–374. https://doi.org/10.1016/j.molp.2016.12.008

2. MAIZE DIVERSITY AND BREEDING FOR DROUGHT TOLERANCE: EXPLOITING THE GENETIC RESOURCES AND ASSOCIATION MAPPING

ABSTRACT

Maize (*Zea mays L.*) is known as the most diverse species. Besides, it is also a staple food crop and responsible for a great portion of human nutrition worldwide. Differences in maize production are found around the world, ranging from high-technology farms to small fields destinated to family subsistence. Thus, specific germplasms are required in order to deal with this wide range of conditions. Perhaps one of the principal adverse of these conditions, drought causes substantial stress in maize, which is a high water-demand plant. The technology offered by breeding programs has a relevant impact on all kinds of maize production, and many strategies have been used to release drought tolerant cultivars. The high-quality genome information provided by molecular markers and the advanced statistical approaches had provided a novel set of tools that can assist breeders to find genomic regions associated with important traits, improving the efficiency of breeding strategies. In this review, we discuss the maize diversity constitution, the historical manner of how the breeding programs have taken advantage of it and the strategies that have been used to deal with the drought tolerance challenge, as well as, the role of genome-wide association studies in current research.

Keywords: Plant breeding, Genes, Heterosis, Hybrids, Tropical maize

2.1. INTRODUCTION

The advances in plant genetics and maize breeding have been tightly related, once maize is a model organism to genetic studies since the beginnings of this science. Barbara McClintock's studies of transposons and the evidence of chromosomal cross-over (Jain 2004), the commercial hybrid technology (Shull 1952) and the double-haploids, applied vastly in breeding programs (Liu et al. 2016b) are some of the discovery in plant genetics in which maize had a significant role. Moreover, maize had another important characteristic that differentiates it from other model organisms, such as *Arabidopsis thaliana* or *Drosophila melanogaster*. This crop is one of the most important for human food security.

Near to 45% of the world's maize areas are located in tropical zones, and despite its yield is lower than those obtained in temperate zones (Table 1 and Figure 1), 37% maize global production comes from tropical regions (FAOSTAT 2017). Furthermore, studies coordinated by the Intergovernmental Panel on Climate Change (IPCC) point to an increase in the frequency and severity of drought events in these regions, like Africa, southwest Asia and South America

(IPCC 2007). Therefore, there is a challenge to the maize breeding programs and the international science community to equalize this context.

Several studies have been showing the high degree of diversity in maize (Xia et al. 2004, 2005; Makumbi et al. 2011; Prasanna 2012; Masuka et al. 2017b) and the importance to introduce it in the commercial breeding programs. Among the tasks that can beneficiate from this diversity is the development of drought-tolerant cultivars.

The importance and necessity of drought-tolerant germplasm can be evidenced by the climate forecasts and the number of scientific studies considering this issue (Bänziger and Araus 2007; Araus et al. 2012; Masuka et al. 2012; Xu et al. 2014; Dias et al. 2018). During the last 100-150 years, many traditional approaches were considered, and recently, new technologies have emerged. The facilitation and reduced costs of molecular markers open a new era of genome analysis and allowed important tools as the genome-wide association studies (GWAS), which are used aiming to elucidate the genetic control of many traits of maize and, as well, to be an initial step to discover new genes related to major traits (Xiao et al. 2017).

In this review we discuss: *i*) the origin of such genetic diversity in maize; *ii*) its relationship with breeding programs and modern germplasm; *iii*) the maize breeding for drought tolerance and *iv*) the use of GWAS and Marker-assisted Selection (MAS) in maize breeding with special attention to drought tolerance.

2.2. MAIZE DIVERSITY AND BREEDING PROGRAMS

2.2.1. The maize origin

The origin and domestication process of Maize (Zea mays ssp. mays) are the focus of many research groups. Therefore, often, new finds and details are published about this theme (Vallebueno-Estrada et al. 2016; Kistler et al. 2018). The Zea genus is subdivided into five species, *Z. diploperennis*, *Z. perennis*, *Z. luxurians*, *Z. nicaraguensis*, and *Z. mays*. Within *Z. mays* species there are 4 sub-species, *Z. m. huehuetenangensis*, *Z. m. mexicana*, *Z. m. parviglumis*, and *Z.m. mays*, this last subspecies is the modern maize, all the other, 3 subspecies and the first 4 species, are considered kinds of teosintes (Edmeades et al. 2017). The modern maize evolved from the wild teosintes *Z. mays ssp. parviglumis* in Mexico around 9000 years ago (Piperno et al. 2009) and, by a gradual process has become the maize we know nowadays (Ramos-Madrigal et al. 2016). Maize already was an important food source for the American pre-Columbian society. After the European colonization, maize has spread throughout the world and rapidly had become a staple

food crop in all continents (McCann 2007; Mir et al. 2013). With more than 9000 years of domestication and 500 of global spread, currently, together with rice and wheat, it is one of the major crops for humankind (FAO 2016).

2.2.2. The establishment of maize diversity

Maize is known as a high diverse crop. During its domestication in Mexico, and through its spread across America, maize met a large assortment of environments (Kistler et al. 2018). This along with the outcrossing mating system of maize, an allogamous plant, and its genetic architecture of the flowering time, allowing a fast and local specific adaptation (Buckler et al. 2009) to such diverse environments, resulting in the known high genetic diversity of maize. The Europeans colonizers have already encountered a great maize diversity in the New World in the 15th and 16th centuries even with suggestions that *i*) a bottleneck event of short duration and very small size had an important role in the Teosinte-to-Maize evolution (Eyre-Walker et al. 1998); *ii*) a few loci of large effects are responsible for the differentiation between modern maize and teosinte (Doebley 2004); and finally, *iii*) a few thousand years is a short time in evolution scale. This remarkable maize diversity in the Americas made its adaptation to other environments around the world been relatively rapid (Edmeades et al. 2017).

Nowadays, maize is present on all continents. Thirty-three countries have sown more than a million hectares of maize in 2017, and between the ten more significant producers we can find averages yield of 3.12 t.ha^{-1} in India to 11.08 t.ha^{-1} in the USA (FAOSTAT 2017). That is, vast differences can be found in the technology applied to the maize production around the world from a high-technology hybrid production in the American corn-belt that reach average yields of more than 10 tonnes per hectare to small African communities that produce maize for local family consumption.

Furthermore, maize is the basis of the food security of many poor regions in Asia, Africa, and Latin America (McCann 2007; Shiferaw et al. 2013). Thus, the maize that grows in temperate regions of the USA is not the same that grows in Africa, neither that grows in India, for example. That is where maize breeding programs act to supply adapted genotypes for each specific situation.

2.2.3. Germplasm classification

As said, the range of environments faced by maize during its domestication in Mexico allowed the formation of high diversity maize populations. Currently, the maize genotypes are arranged by production environments or, as known as, by Megaenvironmets (ME). Regardless of a no universally recognized system for this classification (Gerpacio and Pingali 2007) the International Maize and Wheat Improvement Center (CIMMYT) had developed a system that has been used widely (Hartkamp et al. 2000). The MEs are defined by temperature, altitude, latitude (correlated with daylight length), and eventually by rainfall conditions and there is a basic division of 4 main groups: Tropical lowlands, Tropical highlands, Subtropical/Mid-altitude and Temperate. Where the altitude is grouped as follows: lowland, 0 to 1200 masl; mid-altitude, 1200 to 2000 masl; and highland >2000 masl.

There are some morphological and physiological differences between tropical and temperate maize germplasm. Tropical landraces usually are tall, the ear position is high, far from the ground, the plant may tiller easily and often is prolific, leaves are excessively and with heavy husks, which may help maize to survive in an intense insect and disease attack typical of tropical lands (Gowda et al. 2015). The tropical germplasm also responds when photoperiods are larger than 12h, delaying flowering. The temperate germplasm is adapted to a natural flower under long photoperiods, and the modern temperate maize, after the intense human selection from the last 140 years, presents strong and short stalks, small tassel and erect leaves, result from an intense selection under high plant density (Edmeades et al. 2017). The modern tropical maize hybrids are moving toward temperate phenotype, that is, the tropical maize breeders are making the selection in the direction of smaller plants, non-prolific, bigger and regular ears and with a high harvest index, i.e., more grains and less leaves and stalk, in a more beneficial source/sink ratio (Zaidi et al. 2003).

The temperate zones, basically the USA, Canada, Europe, represent the areas with high maize yield (Figure 2). There is a high correlation between this and investments on maize breeding that were done in these regions in the last century. Since the early 20th century, with the exploration of heterosis, the incorporation of hybrid technology and the intellectual property protection laws, maize seed market began a “fertile ground” for the private sector (Parentoni et al. 2015; Edmeades et al. 2017). Besides the scenario of more stable economy offered by these regions supported the early development of high yield genotypes together with better agronomic management and technologies as machinery, fertilizers, pest control, and water supply by irrigation (Duvick 2005). Moreover, with the biotechnology applied to plant breeding in the

1980s-1990s, as transgenics, the private sector had a new economic incentive to expand even more (Schimmelpfennig et al. 2004).

2.2.4. Modern maize breeding programs and diversity

Hence, with regard of the modern breeding done during the 20th century, the temperate and tropical maize breeding followed different paths. The temperate embark earlier than tropical in the hybrid trends, and the tropical remained longer using open-pollinated varieties (OPV) and landraces, and it has changed the genetic diversity in these two maize groups (Coors et al. 1999; Xia et al. 2005). Temperate germplasm shows a more restricted genetic base than the tropical (Warburton et al. 2008). Initiatives as Latin American Maize Project (LAMP) and Germplasm Enhancement of Maize (GEM) are examples of coordinated and cooperative efforts between the public and private sectors, from temperate and tropical regions. The aim of these projects were to obtain information from the maize germplasm and its diversity (landraces) to facilitate the access by breeders to this genetic resource and fight against a narrow genetic base faced by some modern maize hybrid breeding programs (Pollak 2003).

The use of landraces as a source of genetic diversity is reported as a strategy for combating the narrow genetic base problems normally found in commercial breeding programs, despite the complications of genetic load and non-adapted material, several ways to deal with it have been proposed (Miti et al. 2010; Lopes et al. 2015; Arteaga et al. 2016; Gorjanc et al. 2016; Böhm et al. 2017). In the 1960s and 1970s, CIMMYT had special attention to intrapopulation improvement via recurrent selection (Xia et al. 2004) and incorporate landraces to breeding programs creating populations, or "pools", arranged to support the diversity of global environments where maize was grown.

During the 1980s and 1990s, with the idea to be involved with a hybrid maize program, CIMMYT started to use these populations and “pools” to generate inbred lines that were released as the CMLs (CIMMYT maize lines) (Edmeades et al. 2017). Warburton et al. (2008) in a study with 261 CMLs showed that these inbred lines cover a considerable amount of the first landrace genetic variation. Nevertheless, there is considerable variation left to exploit from the landraces yet. So far, CIMMYT is still working on projects to exploit the landraces as a source of positive variants. In a recent study, Wu et al. (2016) show that the 539 CMLs collection released at that time present a wide genetic distance, and a structured population with a genetic divergence between the Temperate subgroup and Tropical subgroups. A similar study with CML used in Africa also shows a high genetic distance between the CMLs, but SNP markers could not show a

clear separation of the heterotic groups that were established based on combining ability tests (Semagn et al. 2012). The most recent CML catalog, from October 2018, lists 603 CML released (CIMMYT 2018), and they are recommended mainly to tropical and subtropical regions, like Latin America, Asia, and Africa.

Although the genetic diversity into breeding programs in Africa is a concern (Masuka et al. 2017b), the genetic gains on maize yield in East and Southern Africa in the past ten years were comparable to the gain in other regions of the world (Masuka et al. 2017a). Thus, a significative increase in maize production in Africa is a real possibility that perhaps needs more international efforts to be realized.

It is notable that in the past 120 years breeding on temperate maize has received a more significant amount of investments than tropical maize, and the private sector in developed countries has driven the temperate maize investments. In contrast, to the tropical germplasm, that grows in the developing world, international efforts and public sector had and have the leading role in fomenting maize breeding (Coors et al. 1999). The hybrid adoption also has its part in the maize breeding history, and future advances in increase maize production in developing regions is related to make its use wide and feasible there (Edmeades et al. 2017). Maize diversity, although it is high, should always be a concern and efforts are necessary to introduce these genetic resources in commercial breeding programs.

2.3. FROM HETEROSIS TO HYBRID

2.3.1. Heterosis concept

Heterosis is the superiority of an offspring over its parents, and this superiority is used by breeders to release better genotypes to the farmers. Biologically, the phenomenon of heterosis is not well understood. Nevertheless, it has been extensively used (Bernardo 2010; Hallauer et al. 2010a). Although this phenomenon already was observed by scientists in the late 19th Century, as Darwin and Koelreuter, the concept and use of it for plant breeding proposes were not developed until the beginnings of the 20th century (Virmani et al. 2004). George Harrison Shull was a prominent scientist in the beginnings of 1900s that dedicated a great effort to, by his research and gathering studies of many others scientists, create the heterosis concept, describe it and publicize it (Shull 1908, 1914, 1952). In his study from 1914, Shull wrote:

"I suggest that instead of the phrases, "stimulus of heterozygosis", "heterozygotic stimulation", "the stimulating effects of hybridity", "stimulation due to differences in uniting gametes", etc. which have been used by myself and others, the word "heterosis" be adopted".

At the same time, Edward Murray East found similar conclusions but did not realize the great opportunity to use this idea in plant breeding, much because of the poor quantity of seeds that the pure inbred lines produced (Virmani et al. 2004). Since the beginning of Shull studies, or even when Darwin noticed the superiority of cross-fertilized progenies, the causes of heterosis never were wholly understood. Moreover, many suggestions were proposed, with two main genetic effects hypothesis: dominance and overdominance (Bernardo 2010). The dominance hypothesis is based on the idea of a masked effect of the deleterious allele in a heterozygous loci state presented by the hybrid. The overdominance hypothesis suggests that the superiority comes from a complementary effect of each allele, that is, both alleles of the heterozygous add a positive effect, and the heterozygous perform better than both homozygous states. With the advancements in epigenetic studies and "omics" sciences as, functional genomics, proteomics, transcriptomics, with the feasible use of molecular markers other suggestions of how heterosis works have been proposed (Feng et al. 2015; Li et al. 2016; Zhen et al. 2017; Fujimoto et al. 2018; Ryder et al. 2019). What looks clear is that heterosis has a trait dependence explanation, that is, as many genes can be responsible by the variations on a quantitative trait, a varied of genes and type of gene action could be responsible for a superior behavior of the hybrid depending on the trait we are considering (Khotyleva et al. 2016). The complete understanding of heterosis is not necessary to exploit it, as the breeders are doing for the past 110 years, but continuous discoveries and elucidative studies may be the key to new advances in food production, as the hybrid technology was a hundred years ago.

2.3.2. Using heterosis - hybrids

Thus, the heterosis is the vigor of hybrid. The concept of hybrid could be interpreted differently in some of the biology areas. For plant breeding, and more specifically in maize breeding, that superior offspring is named hybrid, and it is originated by a cross between populations or often between two inbred lines. There are three types of commercial hybrids, the single cross (the cross between two inbred lines), the three-way (a cross between a single cross with a third inbred line) and a double-cross hybrid (a cross between two single crosses).

The single-cross normally presents a higher yield than the other two. However, the shift from using open-pollinated varieties (OPV) to hybrid first pass through the double-cross and three-way cross during the 1930s, it happens because the reduced quantity of seed resulted from

a single cross, and the costs to produce a single cross hybrid seed was high enough to turn it unfeasible. Just in the 1960s, the seed use of single hybrid spread rapidly across USA maize fields, that was when the single cross hybrid seed production achieved a reasonable standard and costs (Crow and Dove 1998). In the USA, where this shift from maize OPV to maize hybrid first happened, the yields change from 2 $t.ha^{-1}$ before the double-cross hybrids introduction, in 1930 decade, to yields of 8 $t.ha^{-1}$ in 1980, when more than 90% of American maize fields were cultivated with single hybrids (Tomes 1998), and as already showed, nowadays it is around 11 $t.ha^{-1}$ (Table 1).

However, Troyer and Wellin (2009) found heterosis decreasing in recent hybrids, and the inbred yields increasing faster than heterosis yield, they also emphasize the necessity of a more efficient model to evaluate the inbred lines. Fu et al. (2014) had listed some problems in our capacity to utilize heterosis in plant breeding as the need to develop a more efficient pollination control technologies and to exploit a "wide hybridization", that is, the use of crosses between plants from related species.

Breed for hybrids is an activity that involves two main stages: *i*) generate inbred lines, and *ii*) find the best combination of two inbred lines. Thus, to create an elite hybrid is necessary first to breed at least two good inbred lines. For abiotic stress, as drought, for instance, to achieve a tolerant hybrid is strongly recommended that both parental lines have tolerance behavior (Betrán et al. 2003).

2.3.3. Choosing the allele sources combination – Heterotic groups

The choice of parental lines is a fundamental step to exploit heterosis properly. There is a differential level of heterosis depending on the parents chosen to breed, and the germplasm can be grouped in heterotic groups. A heterotic group is composed of genotypes that present a low specific combining ability between them; that is, there is a low level of heterosis. However, the crosses between plants from distinct heterotic groups will result in a high level of heterosis, and promising hybrids will be obtained. This concept has origin from Comstock (Comstock et al. 1949) studies of reciprocal recurrent selection and specific/general combining ability back in the 1940 decade (Reif et al. 2005).

The definition of which heterotic group each genotype belongs to is crucial germplasm information to a hybrid program. Morphological characteristics, pedigree, and diallel crosses used to be tools to do this discrimination (Hallauer et al. 1988; Melani and Carena 2005). Molecular markers can be used to this task (Lu and Bernardo 2001; Melani and Carena 2005; Tomkowiak et

al. 2019), although occasionally it has not been accurate to assign lines to appropriate heterotic groups (Hallauer et al. 2010a; Maruthi et al. 2019).

2.3.4. Testcross, an option to classify and evaluated lines

A common practice to evaluated new inbred lines (that will be used as parental lines to obtain hybrids) is using testcrosses. That is, cross the inbred lines with a tester, or a few of testers, and evaluated the progenies (the testcrosses) in field trials. The testers are already assigned to a specific heterotic group, and so the performance of the testcrosses will give an idea to which group the lines belong to (Melchinger and Gumber 1998; Bernardo 2010). Additionally, testcross also is relevant to evaluate the performance of the lines, and it is used for selection proposes as well (Hallauer et al. 2010b).

Except when using double haploids, to obtain an inbred line at least six generations of selfing (S0 to S6) should be done, and it takes time and resources. Wait this long to start the inbred line evaluations involves the expenditure on numerous low-performing individuals who would be easily excluded from the population in the final testcross. One strategy is to perform an early testcross, or early-generation test (Sprague 1946), that is, cross the early generation (S1-S4) with a tester (or testers) and do a selection during the inbreeding process, eliminating the very low-performance semi-endogamous lines, the efficiency of this approach is strongly related to the heritability of the early testcross trials, and the selection should be made prudently (Bernardo 2010).

In a study with testcrosses of a breeding population first generations set (P1, P2, F2, BC1, and BC2) Melchinger et al. (1988) conclude that the backcrosses and the F2 testcrosses are useful for selecting the best lines. As the effectiveness of early testing decreases drastically in low heritability situations, in this case maybe selfing could be done to try the early selection in next generation, or to apply a very weak selection intensity (Bernardo 1992). Incorporating molecular markers information with phenotypic data can improve the efficiency of selection in early testcross, allowing a more intensity selection (Eathington et al. 1997a). The allele (markers) effect depends on the allele frequencies, and it changes with selection, since this is the purpose of the selection, to increase the frequencies of favorable alleles. Hence, there is a dilemma in the use of molecular markers effects in selecting in an early generation. These effects should be correlated with its effects in later generations, if not, probably the real effectiveness of early selection will be low. However, a study regarding this aspect showed reasonable results by using marker

information in early testcross to predict lines performance in later generations (Eathington et al. 1997b).

Therefore, the several tasks and alternatives of strategies related to high exploitation of heterosis show how a maize breeding program is complex. To achieve elite hybrids, a long way to go should be done. Knowledge and technology are available to enhance breeding programs worldwide and to increase maize yield in regions where rapid improvement is necessary.

2.4. BREEDING FOR DROUGHT TOLERANCE

2.4.1. Drought stress in plants and the tolerance

Water deficit is a limiting factor for crop productivity, as it causes a reduction in photosynthetic activity (Zhang et al. 1999; Aslam et al. 2003; Rahman et al. 2004). The water deficit can reduce CO_2 arrival to Rubisco carboxylation sites, through stomatal closure, as well as losses in the photo and biochemistry processes (Ribaut et al. 2009). In the same way, it interferes with plant growth and may be irreversible depending on its duration and severity, and stage of development of the stressed plants (Chaves 1991; Bray 1993).

Drought tolerance is defined as an ability from genotype to be more productive under water stress conditions. The drought-tolerant genotype has a higher grain yield than the susceptible genotype in several environments under water-stress. The ideal genotype should present high grain yield in both conditions, i.e., under water-stress or not. The responses of plants to drought stresses are associated with their internal water content (Durães et al. 2004).

Some studies have classified drought tolerance by their possible mechanisms (Turner 1981; Kramer and Boyer 1995; Chaves et al. 2002; Blum 2005): *i*) drought escape being the ability of cultivars to complete their life cycle to coincide with water supply (mechanisms: precocity and recovery); *ii*) tolerance to drought in high water status in the plant, that is, the ability of cultivars to tolerate prolonged dry periods while maintaining a high internal water content (mechanisms: maintenance of water absorption and reduction of water loss); *iii*) tolerance to drought in low water status in the plant, the ability of cultivars to tolerate prolonged dry periods and low water contents in the tissue (mechanisms: maintenance of turgor and tolerance to dehydration or dissection).

Water stress can lead to physiological imbalance, which negatively influences the plant development resulting in reduced yields, and as discussed before, the level of loss depends largely on the physiological stage of the plant when stressed. The development of maize is divided into

two main stages: *i*) vegetative stage, from the emergence of seedlings to tassel emergence, *ii*) reproductive stage, from silks emergence to physiological maturity. Significant losses happen when drought stress occurs in reproductive stages when the plant is in flowering and grain filling (Blum 2011).

Water deficiency at the flowering stage can significantly reduce grain yield (Dwivedi et al. 2008). According to Bergamaschi et al. (2006), the susceptibility of maize occurs due to the physiological processes related to zygote formation and the beginning of grain filling, and also the high transpiration once it is when the maize plant presents its maximum leaf area. Moreover, when comparing this stage in maize with other cereals, except rice, this susceptibility is associated with the female floral structure itself. In maize we have female and male flowers separated on a single stem, developing almost at the same time, occurring competition between the male and female reproductive organs by translocated photosynthates (Monneveux et al. 2006; Bänziger and Araus 2007).

The water deficit during the flowering stage may also affect the synchronism between female and male flowers, as well as reduce the chance of appearance of a second ear in materials considered prolific. The period corresponding to 2 days before and about 25 days after silking was identified as the most sensitive to water deficit, drastically reducing the number of grains per ear (Barker et al. 2005).

2.4.2. Studying and evaluating the drought tolerance

Adaptation of plants to adverse environments or situations under suboptimal environmental factors involve adaptation to multidimensional stresses, with direct and indirect interactions. Thus, it is essential to identify and characterize genotypes via studies that consider different mechanisms as physiological, biochemical, and molecular.

Drought tolerance is a quantitative and complex character and is regulated by countless numbers of genes (Campos et al. 2004b; Kakumanu et al. 2012; Rengel et al. 2012a). In addition, for breeding evaluations, drought tolerance presents low heritability and has a high influence of genotype by environment interaction (Dias et al. 2013).

The breeding for drought tolerance is a difficult task, and its phenotyping is highly laborious. A strategy used in several breeding programs to select drought-tolerant genotypes has been to evaluate yield under non-stressed conditions and then select in locations that undergone through "random stress" (Magorokosho and Tongoona 2004). In this case, it is considered that genes for drought tolerance are related to productivity and that genotypes evaluated under

random stress can further enhance optimal performance conditions (Russel 1984). Another hypothesis also used is that the use of heterosis acts as an essential source of stress tolerance, with hybrids generally preferable to varieties under the drought (Bruce et al. 2002).

Another strategy widely used is to perform the selection of plants in water deficit conditions. An example of this is in Monneveux et al. (2006), they studied two maize populations from CIMMYT, the DTP1 and DTP2, which are adapted to the low altitude and medium altitude, respectively. The cycles C0, C3, C6 to DTP1 and C0, C3, C5, and C9 to DTP2 were evaluated for direct and correlated responses to drought tolerance under drought, low N and optimal conditions. The trials were submitted to one or two levels of water stress using controlled irrigation. During the flowering, and grain filling was induced severe water stress, being the yield of grains in these experiments more affected than the non-stressed ones. The same populations were also evaluated in well-watered conditions. The selection of genotypes in each cycle was based on an index that involved grain yield under drought and well-irrigated conditions, combined with anthesis-silking interval (ASI), sterility, leaf senescence and leaf rolling under drought conditions. Significant yield gains in both populations were observed under drought. When the C0 to C6 cycles were compared, the gain was 0.16 and 0.08 $t.ha^{-1}$ $Cycle^{-1}$ (14.3% and 4.0% $Cycle^{-1}$) in the DTP1 and DTP2 population, respectively. The authors further reported that selection for drought tolerance does not affect yield and number of grains per ear under optimal conditions. However, to the weight of 1000 grains in the DTP1 population, the selection results in gain under optimum conditions.

Since phenotyping for drought tolerance is a hamper, it becomes vital to understand the physiological and its genetic basis to select drought-tolerant cultivars. Some authors relate that the lack of understanding about molecular mechanisms of stable plant development under specific water stress hinders the selection of drought tolerant genotypes either via traditional or molecular breeding methods (Passioura 2010; Sinclair 2011).

Molecular biology and genomic approaches have assisted the breeders, collaborating with the identification of candidate genes and quantitative traits loci (QTLs) associated with drought tolerance and many other traits. In a brief review of biotechnology and plant breeding for abiotic stress, Varshney et al. (2011) relate the importance to combine molecular breeding, as QTL, GWAS and marker-assisted selection (MAS) with genetic engineering, that is cloning genes or change gene expression by engineering gene promoters. Those are advanced and possible strategies for improving abiotic stress tolerance, as was done in rice (Xiao et al. 2009) and maize (Nelson et al. 2007).

Some researchers have studied traits related to drought tolerance. Ribaut et al. (1997), for example, studying maize flowering time identified chromosomal regions related to water stress conditions, furthermore, probably, fewer genes are associated to flowering time than drought tolerance and flowering time may reflects the adaptation of a plant to its environment (Buckler et al. 2009). Ribaut et al. (1997) experiments were conducted in three ways: irrigated, intermediate stress, and severe stress. According to the authors, QTLs were mapped on all chromosomes except chromosome 7. Each QTL individually explained from 5 to 13%, 4.6 to 15.1% and from 4.4 to 15.2% for ASI trait, under the three imposed conditions, respectively. Some stable QTLs were also found for different conditions detected in different chromosomes for the male flowering (MF) and female flowering (FF) traits. Also, for ASI and FF traits, common QTLs were detected on chromosomes 1, 2, 5, and 8.

Sarli-gorla et al. (1999) studied male and female flowering, ASI, and plant height, through the QTL mapping. The experiment was conducted under two environmental conditions: under water deficit and well-watered. They used a population of 142 recombinant inbred lines (RILs) obtained by a cross between the genotypes B73 and H99. For male flowering and plant height, most of the QTLs detected were the same under control and stress conditions. Differently for female flowering and ASI, where many QTLs appeared to be expressed under control conditions or stress. In addition, QTLs conferring drought tolerance was in different chromosomes regions. Their research concluded that drought tolerance, in its different components, may not be a combination of favorable alleles, but rather based on physiological characteristics not directly associated with the trait control. Zhu et al. (2011) also studying drought tolerance, found QTLs for ASI, plant height, grain yield, ear height, and ear adjustment under total irrigation and severe conditions of late stress.

In another study, Almeida et al. (2013) considering three populations, found 83 and 64 QTLs for grain yield and ASI, respectively, when analyzed in two regimes: under water stress and well irrigated. Of these QTLs found in the individual environments, 56 were chosen to perform a consensus map and perform a meta-QTL analysis. In the mQTL analysis, seven QTLs were identified to grain yield on chromosomes 1 (2), 4, 5, 7 and 10 (2) and one for ASI on chromosome 3 over the three populations. The authors reported that these eight mQTL regions identified may be used in the future by MAS, in addition to marker-assisted recurrent selection within breeding programs.

2.5. GENOME-WIDE ASSOCIATION STUDIES AND ITS APPLICATION FOR DROUGHT TOLERANCE IN MAIZE

2.5.1. The GWAS context

Since the beginnings of the human interest in crop genetics, one of the major focuses of scientists has been to identify the genetic sources underlying the phenotypic variations in maize. This knowledge has been crucial in order to comprehend the evolution and the domestication process and contribute to the effective use of the genetic diversity for maize improvement. In the last decades, the advances in molecular approaches have led to the development of several types of genetic markers. Over the years, the continuous scientific progress has contributed to significant changes in the popularity of these markers, which have been applied to elucidate important biological challenges as mapping genomic regions related to traits of interest (Schlötterer 2004; Grover and Sharma 2016).

The progress of the genetic markers has initiated with morphological markers, passed to biochemical markers, and finally become the many classes of molecular markers. While the morphological and biochemical markers were considered limited in quantity and had high environmental influence, in the last years the techniques used to identify molecular markers evolved immensely and with the emergence of next-generation sequencing (NGS) technologies had a considerable reduction in cost, allowing the identification of a vast number of single nucleotide polymorphisms (SNP) (Yang et al. 2015).

Currently, the use of NGS for genotyping and SNPs discovery keep progressing as a consequence of the increasing quantity of sequencing services providers, the powerful bioinformatics approaches, and the dramatically drops of protocols costs. Briefly, the genotyping procedures based on NGS rely on enzymatic hydrolysis, used to reduce the genomic complexity, followed by sequencing of the fragments and usage of bioinformatics procedures to detect the SNPs. The most successful example of this strategy is the genotyping by sequencing (GBS) (Elshire et al. 2011) and its variations, which had been applied in numerous genetics and plant breeding studies as genomic selection, diversity analysis, genetic mapping and genomic association (Poland and Rife 2012; Deschamps et al. 2012).

The phenotypic variants of many complex crop traits are commonly controlled by several quantitative trait loci (QTLs), which interact with each other and are highly influenced by the environment. As it occurs for the majority of the crop species, in maize the most important traits are considered of high complexity (Xiao et al. 2017), as grain yield and the resistance to biotic and abiotic stresses.

2.5.2. Linkage mapping and association mapping

The most traditional approach for genetic mapping in plants has been the linkage or QTL mapping. This methodology is based on the generation of segregating populations, normally derived from crosses involving two genitors that display great phenotypic differences for the trait under interest. Within species that permit self-pollinations, like maize, the most common of these populations are F2, backcrosses, recombinant inbred lines (RILs) and double haploids (Collard et al. 2005). These populations are then genotyped and, using the information provided by the markers and the recombination events occurred within the population. It is possible to identify the genetic loci underlying the trait of interest, based on the principles of genetic linkage. Maize was one of the first species in which molecular markers were used to discover QTLs (Edwards et al. 1987; Stuber et al. 1987) and since then, advanced statistical approaches have been developed, allowing significant contribution to plant breeding (Lander and Botstein 1989; Stuber et al. 1992; Jansen and Stam 1994; Zhao-Bang Zeng and Zeng 1994; Kao et al. 1999; Zeng 2005; Silva et al. 2012).

Although QTL mapping is still a powerful approach to discover loci related to a given trait, it presents some important limitations: *i*) all the allelic diversity assessed by the study is that contemplated by the two genitors that originated the study population; *ii*) this studies demands substantial time and efforts to generate these populations, and *iii*) there are limited recombination events that occur during this process, what results in large blocks in linkage disequilibrium and limited mapping resolution (Xiao et al. 2017). As consequence of the size of these blocks, the QTL interval is also usually large and may contain a great number of genes (in maize, for example, even with few centimorgans, the sequence interval may represent millions of bases and consequently hundreds of genes) making it difficult point specifically the one responsible for the phenotypic variation (Flint-Garcia et al. 2005).

Despite based on the same fundamental principles of QTL mapping, the genome-wide association studies (GWAS), also known as linkage disequilibrium (LD) mapping or association mapping, has emerged as a powerful technique to identify genomic regions associated to phenotypic variation. This approach offers the benefit of exploiting evolutionary recombination events at the population level and thus can overcome some of the QTL mapping challenges. The advent of high-density genotyping of SNPs and the development of powerful statistical methods are the main reasons that boosted the GWAS use. The basic approach is focused on testing the associations between each genomic marker and the trait that has been evaluated across a great number of unrelated genotypes (Korte and Farlow 2013). As an alternative to traditional QTL mapping, at least three main conveniences were introduced by GWAS: *i*) its greater mapping

resolution, sometimes to the gene level, consequence of the many evolutionary recombination events that are exploited and the high-density genotyping; *ii*) the unnecessary construction of segregating population, and *iii*) the capacity to accomplish more significant allelic variation (Yu and Buckler 2006; Zhu et al. 2008).

2.5.3. The linkage disequilibrium and GWAS

One crucial factor that influences the resolution provided by GWAS is the LD within the genome of the species under study. The LD is defined as the non-random association of alleles between genetic loci, and its measurement depends on numerous factors, including genetic drift, natural and artificial selection, the population structure, and the genetic relationship. Furthermore, LD decay over sequence distance defines the marker density required to perform the association study. In maize, LD ranges from less than 1 kb in landraces and 2 kb for inbred lines and can reach over 500 kb in elite improved lines (Yu and Buckler 2006).

Initially proposed for human genetics, GWAS has been applied mainly in plants. Regarding its long historic as a genetic model and the outstanding agricultural relevance, maize was the plant species in which association studies were comprised for the first time (Thornsberry et al. 2001) and has been so far one of the species where GWAS has been mostly used (Xiao et al. 2017). Despite this original report, the first genome-wide scale research was conducted in maize only in 2008 and explored the variation of 8.590 loci in 553 elite maize inbred lines to discover the genes controlling the production of fatty acid content in grains (Beló et al. 2008).

2.5.4. GWAS in maize and mapping drought tolerance traits

In a recent review, Xiao et al. (2017) listed around a hundred publications that dissected dozens of maize traits via GWAS, including: molecular and cellular metabolites production, developmental and agronomic characteristics, yield traits, and biotic and abiotic stress resistance, which made possible the cloning of number of genes and pointing out many other candidate genes.

Specifically, regarding drought tolerance, several studies have been applying GWAS to study water-limited conditions using the traditional agronomic and yield traits (Hao et al. 2011; Thirunavukkarasu et al. 2014; Farfan et al. 2015a; Yuan et al. 2019). Anyway, although numerous genomic regions and putative genes have been discovered, the high environmental influence and

the complex quantitative mechanisms controlling this trait narrow the power of GWAS to dissect the drought tolerance (Xiao et al. 2017).

Thus, the strategy that has been usually employed is the evaluation of specific and more stable traits, which may have influence over the general tolerance, as for instance the survival of seedlings submitted to limited watering (Liu et al. 2013; Mao et al. 2015; Wang et al. 2016c) and drought-tolerance metabolic traits (Setter et al. 2011; Zhang et al. 2016). The associations obtained in the seedling studies were able to identify interesting genomic regions and candidate genes related to drought tolerance as: transposable elements inserted in the NAC gene promoter (*ZmNAC111*) (Mao et al. 2015), DREB transcription factors (Liu et al. 2013) and the gene ZmVPP1, responsible for encode a vacuolar-type H(+) pyrophosphatase (Wang et al. 2016c). In addition, the studies focusing on correlated metabolites evidenced many metabolite-associated loci (Zhang et al. 2016) as levels of sugars and abscisic acid (ABA) (Setter et al. 2011), this phytohormone plays an essential role in drought response, and increased production of ABA is one of the physiological changes that allow plant survival under drought condition (Seki et al. 2007).

Associations for biometrical traits related to drought tolerance, as relative ear position, hundred kernel weight, and timing to flowering also were obtained by GWAS. As a result of association studies, the maize gene GRMZM2G125777 was discovered as related to those traits (Xue et al. 2013a), this gene encodes a NAC membrane-bound transcription factor, this factor is responsible by plant development and abiotic stress response in Arabidopsis. Later, by functional genomics studies, it was elucidated that this gene also plays an important role in maize response to oxidative stress, as that produced by drought (Wang et al. 2016a). Despite the difficulties to finding associations between genomic regions and complex trait as yield, some exciting results were achieved (Farfan et al. 2015b): 10 quantitative traits variants (QTVs) for grain yield, plant, and ear height, days to anthesis and days to silk were found. Some of them were related to other QTLs previously reported, while others were new, demonstrating the utility of GWAS to solve and discover useful and new gene variants.

GWAS approach has become an outstanding tool for analyzing simple and complex traits. Anyway, for complex and quantitative characteristics, like drought tolerance, GWAS may fail to reveal the genes or the specific loci controlling the trait. This makes sense and is in accordance with the biological idea of a quantitative and complex trait, i.e., many different genes and gene actions are involved in trait variance. An exciting strategy to face this task is the use of correlated traits, which may decrease the number of loci under investigation, reduce the environmental influence and improve the detection power (Korte and Farlow 2013), or improve

the experimental/statistical capacity to detect variation caused by genetic factors in yield, in other words, increase the heritability of yield (Bernardo 2010) or index based on yield. Furthermore, it is important to emphasize that in general, the associated SNP is not the polymorphism responsible for the phenotype variation, but a genomic reference of its loci. As well, it is vital to note that GWAS is primary evidence that demands validation in other populations and genetic and statistical techniques.

2.5.5. Marker-Assisted Selection for drought tolerance

Once found genome regions related to an interest trait, one can genotype a population for the genetic markers mapped in those regions and use them to assist the selection, that is the central idea of Marker-Assisted Selection (MAS), use genetic markers that are associated with a given trait variation to perform an indirect selection (Ben-Ari and Lavi 2012). The use of genetic markers to drive the selection was a great promise for plant breeding since the discovery of genetic markers (Mohler 2005). In 1990-2000s decades MAS start to be widely used in plant breeding (Bernardo 2008) and also for traits related to drought tolerance (Ribaut and Ragot 2007; Devi et al. 2017). The first step of MAS is finding the marker-trait associations and GWAS has joined the linkage mapping in this task, been another alternative to increase MAS using (Tsonev et al. 2009).

The most advantage of using MAS is the possibility to accelerate the selection process. In the early stages of selection, when there is still a large number of testing individuals and field evaluations are expenditure, the populations can be screened for a target marker, or for a set of them, and that genotypes without the desired markers profile are discarded with no pressure of selection applied outside the target regions (Ribaut and Hoisington 1998). The efficiency of MAS is slightly related to the quality of the marker-trait association, in other words, on the true relationship between the marker and the gene responsible by the desired phenotype. In the same way, markers with large effects (or major QTLs) are more conducive to positive results for this indirect selection. Therefore, for complex traits, with many loci acting with small effects, alternative MAS strategies should be considered (Ribaut and Hoisington 1998).

Many strategies, or approaches, were proposed using MAS in plant breeding, as marker-assisted backcross (MABC) and marker-assisted recurrent selection (MARS). As said, the MAS aims to accelerate the breeding process, but some of the advantages and disadvantages of each breeding scheme keep existing even when markers information is introduced. MABC is the simplest way to use MAS to transfer a gene, or a major effect QTL, from a donor to an elite

cultivar, and is based on backcross strategy but without field evaluations. Thus, the population is genotyped and the individuals that have the desired marker are backcrossed with the elite parental. Ribaut and Ragot (2007) detailed the use of MABC for drought-tolerant maize breeding. They select 5 QTLs for ASI (as already discussed, it is a drought tolerance correlated trait) as target loci to transfer favorable alleles from a drought-tolerant parent (Ac7643) to a CIMMYT elite line (CML247). Afterward, they proceeded early selection in each backcross population [(Ac7643 x CML243) x CML243] with no phenotypic selection or field evaluation applied during the MAS process (backcross and inbreeding). Their results reported an increased yield of the hybrids originated from this process.

MABC is a convenient tool for transfer one or a few markers (or genes) from one individual to another. For complex traits, or when the aim is to act on more than one trait at the same time, the MARS is more appropriate. MARS aims to accumulate a relatively large number of QTL in a population, considering that these are medium-low effect QTLs (Bernardo 2008). In a more recent study, Beyene et al. (2016) demonstrated the potential of MARS for improving maize grain yield under water-stress environments, showing gains of 51 kg ha^{-1} $year^{-1}$ under this condition.

Genomic Selection (GS) is also a MAS strategy, it uses all available markers of the study population to predict the individual phenotypes based on a training population, skipping the linkage and association mapping steps (the steps that aim to find trait-marker association) (Wang et al. 2018). The GS estimates an effect of each marker assigning to each of them a genetic value, and the prediction of the individuals is the sum of the genetic values of its markers (Crossa et al. 2017), this approach is more reliable with the concept of a complex trait, where is many loci have a small effect on the phenotypic variation.

The genetic gains should always be weight by the time to achieve it (Bernardo 2010). In the past 30 years, many different approaches have been proposed in this sense and the incorporation of molecular markers in this process is indispensable.

2.6. CONCLUSION

Despite the natural high diversity of maize, breeding programs should always be aware of the narrow genetic base resulted from consecutive years of intense selection. Tropical regions, thus developing countries, will face more often the increase in the frequency of drought events. In the same way, growing demand for food in this region is forecast, thus, it must be a constant

concern for the international scientific community, requiring that they keep going with coordinate actions and efforts to face this issue. Furthermore, the maize yield in developing countries is still much lower than in the developed ones, therefore, there is an opportunity for rapid increases. Many breeding strategies and genome approaches were developed. A combination of excellence phenotyping strategies and genome-wide associations studies for correct identification of genome regions associated with grain yield and drought tolerance is a useful tool available to assist the breeders with this task.

REFERENCES

Almeida, G.D., Makumbi, D., Magorokosho, C., Nair, S., Borém, A., Ribaut, J.M., Bänziger, M., Prasanna, B.M., Crossa, J., Babu, R., 2013. QTL mapping in three tropical maize populations reveals a set of constitutive and adaptive genomic regions for drought tolerance. Theor. Appl. Genet. 126, 583–600. https://doi.org/10.1007/s00122-012-2003-7

Araus, J.L., Serret, M.D., Edmeades, G.O., 2012. Phenotyping maize for adaptation to drought. Front. Physiol. 3 AUG, 1–20. https://doi.org/10.3389/fphys.2012.00305

Arteaga, M.C., Moreno-Letelier, A., Mastretta-Yanes, A., Vázquez-Lobo, A., Breña-Ochoa, A., Moreno-Estrada, A., Eguiarte, L.E., Piñero, D., 2016. Genomic variation in recently collected maize landraces from Mexico. Genomics Data 7, 38–45. https://doi.org/10.1016/j.gdata.2015.11.002

Aslam, M., Hammad, M., Tahir, N., 2003. Correlation and Path Coefficient Analysis of Different Morpho-Physiological Traits of Maize Inbreds under Water Stress Conditions. Int. J. Agric. Biol. 5, 446–448.

Bänziger, M., Araus, J.-L., 2007. Recent Advances in Breeding Maize for Drought and Salinity Stress Tolerance, in: Advances in Molecular Breeding Toward Drought and Salt Tolerant Crops. Springer Netherlands, Dordrecht, pp. 587–601. https://doi.org/10.1007/978-1-4020-5578-2_23

Barker, T., Campos, H., Cooper, M., Dolan, D., Edmeades, G., Habben, J., Schussler, J., Wright, D., Zinselmeier, C., 2005. Improving Drought Tolerance in Maize, in: Jnick, J. (Ed.), Plant Breeding Reviews. John Wiley & Sons, Inc., Oxford, UK, UK, pp. 173–253. https://doi.org/10.1002/9780470650301.ch7

Beló, A., Zheng, P., Luck, S., Shen, B., Meyer, D.J., Li, B., Tingey, S., Rafalski, A., 2008. Whole genome scan detects an allelic variant of fad2 associated with increased oleic acid levels in maize. Mol. Genet. Genomics 279, 1–10. https://doi.org/10.1007/s00438-007-0289-y

Ben-Ari, G., Lavi, U., 2012. Marker-assisted selection in plant breeding, in: Altman, A., Hasegawa, M.P. (Eds.), Plant Biotechnology and Agriculture. Elsevier, pp. 163–184. https://doi.org/10.1016/B978-0-12-381466-1.00011-0

Bergamaschi, H., Dalmago, G.A., Comiran, F., Bergonci, J.I., Müller, A.G., França, S., Santos, A.O., Radin, B., Menegassi Bianchi, C.A., Pereira, P.G., 2006. Deficit hídrico e produtividade na cultura do milho. Pesqui. Agropecu. Bras. 41, 243–249. https://doi.org/10.1590/S0100-204X2006000200008

Bernardo, R., 2010. Breeding for quantitative traits in plants, 2nd ed. Stemma press, Woodbury.

Bernardo, R., 2008. Molecular markers and selection for complex traits in plants: Learning from the last 20 years. Crop Sci. 48, 1649–1664. https://doi.org/10.2135/cropsci2008.03.0131

Bernardo, R., 1992. Retention of Genetically Superior Lines during Early-Generation Testcrossing of Maize. Crop Sci. 32, 933. https://doi.org/10.2135/cropsci1992.0011183X003200040019x

Betrán, F.J., Beck, D., Bänziger, M., Edmeades, G.O., 2003. Genetic Analysis of Inbred and Hybrid Grain Yield under Stress and Nonstress Environments in Tropical Maize. Crop Sci. 43, 807. https://doi.org/10.2135/cropsci2003.8070

Beyene, Y., Semagn, K., Crossa, J., Mugo, S., Atlin, G.N., Tarekegne, A., Meisel, B., Sehabiague, P., Vivek, B.S., Oikeh, S., Alvarado, G., Machida, L., Olsen, M., Prasanna, B.M., Bänziger, M., 2016. Improving maize grain yield under drought stress and non-stress environments in sub-saharan africa using marker-assisted recurrent selection. Crop Sci. 56, 344–353. https://doi.org/10.2135/cropsci2015.02.0135

Blum, A., 2011. Plant Breeding for Water-Limited Environments. Springer New York, New York, NY, NY. https://doi.org/10.1007/978-1-4419-7491-4

Blum, A., 2005. Drought resistance, water-use efficiency, and yield potential—are they compatible, dissonant, or mutually exclusive? Aust. J. Agric. Res. 56, 1159. https://doi.org/10.1071/AR05069

Böhm, J., Schipprack, W., Utz, H.F., Melchinger, A.E., 2017. Tapping the genetic diversity of landraces in allogamous crops with doubled haploid lines: a case study from European flint maize. Theor. Appl. Genet. 130, 861–873. https://doi.org/10.1007/s00122-017-2856-x

Bray, E.A., 1993. Molecular Responses to Water Deficit. Plant Physiol. 103, 1035–1040. https://doi.org/10.1104/pp.103.4.1035

Bruce, W.B., Edmeades, G.O., Barker, T.C., 2002. Molecular and physiological approaches to maize improvement for drought tolerance. J. Exp. Bot. 53, 13–25. https://doi.org/10.1093/jexbot/53.366.13

Buckler, E.S., Holland, J.B., Bradbury, P.J., Acharya, C.B., Brown, P.J., Browne, C., Ersoz, E., Flint-Garcia, S., Garcia, A., Glaubitz, J.C., Goodman, M.M., Harjes, C., Guill, K., Kroon, D.E., Larsson, S., Lepak, N.K., Li, H., Mitchell, S.E., Pressoir, G., Peiffer, J.A., Rosas, M.O., Rocheford, T.R., Romay, M.C., Romero, S., Salvo, S., Villeda, H.S., Sofia da Silva, H., Sun, Q., Tian, F., Upadyayula, N., Ware, D., Yates, H., Yu, J., Zhang, Z., Kresovich, S., McMullen, M.D., 2009. The Genetic Architecture of Maize Flowering Time. Science (80-.). 325, 714–718. https://doi.org/10.1126/science.1174276

Campos, H., Cooper, M., Habben, J.E., Edmeades, G.O., Schussler, J.R., 2004. Improving drought tolerance in maize: a view from industry. F. Crop. Res. 90, 19–34. https://doi.org/10.1016/J.FCR.2004.07.003

Chaves, M.M., 2002. How Plants Cope with Water Stress in the Field? Photosynthesis and Growth. Ann. Bot. 89, 907–916. https://doi.org/10.1093/aob/mcf105

Chaves, M.M., 1991. Effects of Water Deficits on Carbon Assimilation. J. Exp. Bot. 42, 1–16. https://doi.org/10.1093/jxb/42.1.1

CIMMYT, 2018. CIMMYT Maize Lines (CMLs) - Pedigree and characterization data [WWW Document]. website. URL https://data.cimmyt.org/dataset.xhtml?persistentId=hdl:11529/10246 (accessed 2.1.19).

Collard, B.C.Y., Jahufer, M.Z.Z., Brouwer, J.B., Pang, E.C.K., 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142, 169–196. https://doi.org/10.1007/s10681-005-1681-5

Comstock, R.E., Robinson, H.F., Harvey, P.H., 1949. A Breeding Procedure Designed To Make Maximum Use of Both General and Specific Combining Ability1. Agron. J. 41, 360. https://doi.org/10.2134/agronj1949.00021962004100080006x

Coors, James G., Pandey, Shivaji, Vasal, S.K., Cordova, H., Pandey, S., Srinivasan, G., 1999. Tropical Maize and Heterosis, in: Gerdes, J.T., Coors, J.G., Pandey, S. (Eds.), Genetics and Exploitation of Heterosis in Crops. American Society of Agronomy, p. 524. https://doi.org/10.2134/1999.geneticsandexploitation.c34

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de Los Campos, G., Burgueño, J., González-Camacho, J.M., Pérez-Elizalde, S., Beyene, Y., Dreisigacker, S., Singh, R., Zhang, X., Gowda, M., Roorkiwal, M., Rutkoski, J., Varshney, R.K., 2017. Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. Trends Plant Sci. 22, 961–975. https://doi.org/10.1016/j.tplants.2017.08.011

Crow, J.F., Dove, W.F., 1998. Anecdotal , Historical and Critical Commentaries on Genetics. 90 Years Ago: The Beginning of Hybrid Maize. Genetics 148, 923–928. https://doi.org/10.3109/10428199309067939

Deschamps, S., Llaca, V., May, G.D., 2012. Genotyping-by-Sequencing in Plants. Biology (Basel). 1, 460–483. https://doi.org/10.3390/biology1030460

Devi, E.L., Devi, C.P., Kumar, S., Sharma, S.K., Beemrote, A., Chongtham, S.K., Singh, C.H., Tania, C., Singh, T.B., Ningombam, A., Akoijam, R., Singh, I.M., Singh, Y.R., Monteshori, S., Omita, Y., Prakash, N., Ngachan, S. V., 2017. Marker assisted selection (MAS) towards generating stress tolerant crop plants. Plant Gene 11, 205–218. https://doi.org/10.1016/j.plgene.2017.05.014

Dias, G., Dan, A., Cosmos, M., 2013. QTL mapping in three tropical maize populations reveals a set of constitutive and adaptive genomic regions for drought tolerance 583–600. https://doi.org/10.1007/s00122-012-2003-7

Dias, K.O.D.G., Gezan, S.A., Guimarães, C.T., Nazarian, A., Da Costa E Silva, L., Parentoni, S.N., De Oliveira Guimarães, P.E., De Oliveira Anoni, C., Pádua, J.M.V., De Oliveira Pinto, M., Noda, R.W., Ribeiro, C.A.G., De Magalhães, J.V., Garcia, A.A.F., De Souza, J.C., Guimarães, L.J.M., Pastina, M.M., 2018. Improving accuracies of genomic predictions for drought tolerance in maize by joint modeling of additive and dominance effects in multi-environment trials. Heredity (Edinb). 121, 24–37. https://doi.org/10.1038/s41437-018-0053-6

Doebley, J., 2004. The Genetics of Maize Evolution. Annu. Rev. Genet. 38, 37–59. https://doi.org/10.1146/annurev.genet.38.072902.092425

Durães, F.O.M., Santos, M.X. dos, Gama, E.E.G. e, Magalhães, P.C., Albuquerque, P.E.P., Guimarães, C.T., 2004. Fenotipagem Associada a Tolerância a Seca em Milho. Circ. Técnica - EMBRAPA 39, 1–15.

Duvick, D.N., 2005. The Contribution of Breeding to Yield Advances in maize (Zea mays L.). pp. 83–145. https://doi.org/10.1016/S0065-2113(05)86002-X

Dwivedi, S., Perotti, E., Ortiz, R., 2008. Towards molecular breeding of reproductive traits in cereal crops. Plant Biotechnol. J. 6, 529–559. https://doi.org/10.1111/j.1467-7652.2008.00343.x

Eathington, S.R., Dudley, J.W., Rufener, G.K., 1997a. Usefulness of Marker-QTAL ssociations in Early Generation Selection. Crop Sci. 37, 1686. https://doi.org/10.2135/cropsci1997.0011183X003700060002x

Eathington, S.R., Dudley, J.W., Rufener, G.K., 1997b. Marker Effects Estimated from Testcrosses of Early and Late Generations of Inbreeding in Maize. Crop Sci. 37, 1679. https://doi.org/10.2135/cropsci1997.0011183X003700060001x

Edmeades, G.O., Trevisan, W., Prasanna, B.M., Campos, H., 2017. Tropical Maize (Zea mays L.), in: Campos, H., Caligari, P.D.S. (Eds.), Genetic Improvement of Tropical Crops. Springer International Publishing, Cham, p. 331. https://doi.org/10.1007/978-3-319-59819-2

Edwards, M.D., Stuber, C.W., Wendel, J.F., 1987. Molecular-marker-facilitated investigations of quantitative-trait loci in maize. I. Numbers, genomic distribution and types of gene action. Genetics 116, 113–125.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., Mitchell, S.E., 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6, 1–10. https://doi.org/10.1371/journal.pone.0019379

Eyre-Walker, A., Gaut, R.L., Hilton, H., Feldman, D.L., Gaut, B.S., 1998. Investigation of the bottleneck leading to the domestication of maize. Proc. Natl. Acad. Sci. U. S. A. 95, 4441–6.

FAO, 2016. Save and grow in practice. Maize. Rice. Wheat. Food and Agriculture Organiztion of the United Nations, Rome.

FAOSTAT, 2017. Database of Food and Agriculture Organization of the United Nations Statistics Division [WWW Document]. URL http://www.fao.org/faostat (accessed 5.5.19).

Farfan, I.D.B., De La Fuente, G.N., Murray, S.C., Isakeit, T., Huang, P.-C., Warburton, M., Williams, P., Windham, G.L., Kolomiets, M., 2015a. Genome Wide Association Study for Drought, Aflatoxin Resistance, and Important Agronomic Traits of Maize Hybrids in the Sub-Tropics. PLoS One 10, e0117737. https://doi.org/10.1371/journal.pone.0117737

Feng, S., Chen, Xiaoliu, Wu, S., Chen, Xuesen, 2015. Recent Advances in Understanding Plant Heterosis. Agric. Sci. 06, 1033–1038. https://doi.org/10.4236/as.2015.69098

Flint-Garcia, S.A., Thuillet, A.-C., Yu, J., Pressoir, G., Romero, S.M., Mitchell, 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. https://doi.org/10.1111/j.1365-313X.2005.02591.x

Fu, D., Xiao, M., Hayward, A., Fu, Y., Liu, G., Jiang, G., Zhang, H., 2014. Utilization of crop heterosis: A review. Euphytica 197, 161–173. https://doi.org/10.1007/s10681-014-1103-7

Fujimoto, R., Uezono, K., Ishikura, S., Osabe, K., Peacock, W.J., Dennis, E.S., 2018. Recent research on the mechanism of heterosis is important for crop and vegetable breeding systems. Breed. Sci. 68, 145–158. https://doi.org/10.1270/jsbbs.17155

Gerpacio, R. V., Pingali, P.L., 2007. Tropical and Subtropical Maize in Asia: Production Systems, Constraints, and Research Priorities, Cimmyt.

Gorjanc, G., Jenko, J., Hearne, S.J., Hickey, J.M., 2016. Initiating maize pre-breeding programs using genomic selection to harness polygenic variation from landrace populations. BMC Genomics 17, 1–15. https://doi.org/10.1186/s12864-015-2345-z

Gowda, M., Das, B., Makumbi, D., Babu, R., Semagn, K., Mahuku, G., Olsen, M.S., Bright, J.M., Beyene, Y., Prasanna, B.M., 2015. Genome-wide association and genomic prediction of resistance to maize lethal necrosis disease in tropical maize germplasm. Theor. Appl. Genet. 128, 1957–1968. https://doi.org/10.1007/s00122-015-2559-0

Grover, A., Sharma, P.C., 2016. Development and use of molecular markers: past and present. Crit. Rev. Biotechnol. 36, 290–302. https://doi.org/10.3109/07388551.2014.959891

Hallauer, A.R., Carena, M.J., Filho, J.B.M., 2010a. Heterosis, in: Hallauer, A.R., Carena, M.J., Filho, J.B.M. (Eds.), Quantitative Genetics in Maize Breeding. Springer New York, New York, NY, pp. 477–529. https://doi.org/10.1007/978-1-4419-0766-0_10

Hallauer, A.R., Carena, M.J., Filho, J.B.M., 2010b. Testers and Combining Ability, in: Hallauer, A.R., Carena, M.J., Filho, J.B.M. (Eds.), Quantitative Genetics in Maize Breeding. Springer New York, New York, NY, pp. 383–423. https://doi.org/10.1007/978-1-4419-0766-0_8

Hallauer, A.R., Russell, W.A., Lamkey, K.R., 1988. Corn Breeding, in: Sprague, G.F., Dudley, J.W. (Eds.), Corn and Corn Improvement. pp. 463–564. https://doi.org/10.2134/agronmonogr18.3ed.c8

Hao, Z., Li, X., Xie, C., Weng, J., Li, M., Zhang, D., Liang, X., Liu, L., Liu, S., Zhang, S., 2011. Identification of Functional Genetic Variations Underlying Drought Tolerance in Maize Using SNP Markers. J. Integr. Plant Biol. 53, 641–652. https://doi.org/10.1111/j.1744-7909.2011.01051.x

Hartkamp, A.D., White, J., Rodríguez, A., 2000. Maize production environments revisited: a GIS-based approach 37.

IPCC, 2007. Summary for Policymakers. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, United Kingdom. https://doi.org/10.1038/446727a

Jain, H.K., 2004. Hundred Years of Genetics: Revisiting some of the Landmarks, in: Jain, H.K., Kharkwal, M.C. (Eds.), Plant Breeding· Mendelian to Molecular Approaches. Springer Science+Business Media Dordrecht, New Delhi, pp. 1–16.

Jansen, R.C., Stam, P., 1994. High resolution of quantitative traits into multiple loci via interval mapping. Genetics 136, 1447–1455.

Kakumanu, A., Ambavaram, M.M., Klumas, C., Krishnan, A., Batlang, U., Myers, E., Grene, R., Pereira Virginia Bioinformatics Institute AKa, A., 2012. Effects of Drought on Gene Expression in Maize Reproductive and Leaf Meristem Tissue Revealed by RNA-Seq 1[W][OA]. https://doi.org/10.1104/pp.112.200444

Kao, C.H., Zeng, Z.B., Teasdale, R.D., 1999. Multiple interval mapping for quantitative trait loci. Genetics 152, 1203–16.

Khotyleva, L. V., Kilchevsky, A. V., Shapturenko, M.N., 2016. Theoretical aspects of heterosis. Vavilov J. Genet. Breed. 20, 482–492. https://doi.org/10.18699/vj16.174

Kistler, L., Maezumi, S.Y., Gregorio de Souza, J., Przelomska, N.A.S., Malaquias Costa, F., Smith, O., Loiselle, H., Ramos-Madrigal, J., Wales, N., Ribeiro, E.R., Morrison, R.R., Grimaldo, C., Prous, A.P., Arriaza, B., Gilbert, M.T.P., de Oliveira Freitas, F., Allaby, R.G., 2018. Multiproxy evidence highlights a complex evolutionary legacy of maize in South America. Science (80-.). 362, 1309–1313. https://doi.org/10.1126/science.aav0207

Korte, A., Farlow, A., 2013. The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9, 29. https://doi.org/10.1186/1746-4811-9-29

Kramer, P.J., Boyer, J.S., 1995. Water Relations of Plants and Soils, 2nd ed. Academic Press, Inc, San Diego, Californa.

Lander, E.S., Botstein, D., 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121, 185–199.

Li, D., Huang, Z., Song, S., Xin, Y., Mao, D., Lv, Q., Zhou, M., Tian, D., Tang, M., Wu, Q., Liu, X., Chen, T., Song, X., Fu, X., Zhao, B., Liang, C., Li, A., Liu, G., Li, S., Hu, S., Cao, X., Yu, J., Yuan, L., Chen, C., Zhu, L., 2016. Integrated analysis of phenome, genome, and transcriptome of hybrid rice uncovered multiple heterosis-related loci for yield increase. Proc. Natl. Acad. Sci. 113, E6026–E6035. https://doi.org/10.1073/pnas.1610115113

Liu, S., Wang, X., Wang, H., Xin, H., Yang, X., Yan, J., Li, J., Tran, L.-S.P., Shinozaki, K., Yamaguchi-Shinozaki, K., Qin, F., 2013. Genome-Wide Analysis of ZmDREB Genes and Their Association with Natural Variation in Drought Tolerance at Seedling Stage of Zea mays L. PLoS Genet. 9, e1003790. https://doi.org/10.1371/journal.pgen.1003790

Liu, Z., Wang, Y., Ren, J., Mei, M., Frei, U.K., Trampe, B., Lübberstedt, T., 2016. Maize Doubled Haploids, in: Plant Breeding Reviews. John Wiley & Sons, Inc., Hoboken, NJ, USA, pp. 123–166. https://doi.org/10.1002/9781119279723.ch3

Lopes, M.S., El-Basyoni, I., Baenziger, P.S., Singh, S., Royo, C., Ozbek, K., Aktas, H., Ozer, E., Ozdemir, F., Manickavelu, A., Ban, T., Vikram, P., 2015. Exploiting genetic diversity from landraces in wheat breeding for adaptation to climate change. J. Exp. Bot. 66, 3477–3486. https://doi.org/10.1093/jxb/erv122

Lu, H., Bernardo, R., 2001. Molecular marker diversity among current and historical maize inbreds. Theor. Appl. Genet. 103, 613–617. https://doi.org/10.1007/PL00002917

Magorokosho, C., Tongoona, P., 2004. Selection for drought tolerance in two tropical maize populations. African Crop Sci. J. 11, 151–161. https://doi.org/10.4314/acsj.v11i3.27566

Makumbi, D., Betrán, J.F., Bänziger, M., Ribaut, J.M., 2011. Combining ability, heterosis and genetic diversity in tropical maize (Zea mays L.) under stress and non-stress conditions. Euphytica 180, 143–162. https://doi.org/10.1007/s10681-010-0334-5

Mao, H., Wang, H., Liu, S., Li, Z., Yang, X., Yan, J., Li, J., Tran, L.S.P., Qin, F., 2015. A transposable element in a NAC gene is associated with drought tolerance in maize seedlings. Nat. Commun. 6, 1–13. https://doi.org/10.1038/ncomms9326

Maruthi, R.T., Gadag, R.N., Bhat, J.S., Karjagi, C.G, 2019. Heterotic grouping of late-maturing maize inbred lines based on combining ability and molecular marker studies. J. Environ. Biol., 40, 705-710.

Masuka, B., Araus, J.L., Das, B., Sonder, K., Cairns, J.E., 2012. Phenotyping for Abiotic Stress Tolerance in Maize. J. Integr. Plant Biol. 54, 238–249. https://doi.org/10.1111/j.1744-7909.2012.01118.x

Masuka, B., Atlin, G.N., Olsen, M., Magorokosho, C., Labuschagne, M., Crossa, J., Bänziger, M., Pixley, K. V., Vivek, B.S., Von Biljon, A., Macrobert, J., Alvarado, G., Prasanna, B.M., Makumbi, D., Tarekegne, A., Das, B., Zaman-Allah, M., Cairns, J.E., 2017a. Gains in maize genetic improvement in eastern and southern Africa: I. CIMMYT hybrid breeding pipeline. Crop Sci. 57, 168–179. https://doi.org/10.2135/cropsci2016.05.0343

Masuka, B.P., Van Biljon, A., Cairns, J.E., Das, B., Labuschagne, M., MacRobert, J., Makumbi, D., Magorokosho, C., Zaman-Allah, M., Ogugo, V., Olsen, M., Prasanna, B.M., Tarekegne, A., Semagn, K., 2017b. Genetic diversity among selected elite CIMMYT maize hybrids in east and Southern Africa. Crop Sci. 57, 2395–2404. https://doi.org/10.2135/cropsci2016.09.0754

McCann, J.C., 2007. Maize and Grace: Africa's Encounter with a New World Crop, 1500-2000. Harvard University Press.

Melani, M.D., Carena, M.J., 2005. Alternative Maize Heterotic Patterns for the Northern Corn Belt. Crop Sci. 45, 2186. https://doi.org/10.2135/cropsci2004.0289

Melchinger, A.E., Gumber, R.K., 1998. Overview of Heterosis and Heterotic Groups in Agronomic Crops, in: Larnkey, K.L., Staub, J.E. (Eds.), Concepts and Breeding of Heterosis in Crop Plants. Crop Science Society of America, Medison, pp. 29–44. https://doi.org/10.2135/cssaspecpub25.c3

Melchinger, A.E., Schmidt, W., Geiger, H.H., 1988. Comparison of Testcrosses Produced from F2 and First Backcross Populations in Maize. Crop Sci. 28, 743. https://doi.org/10.2135/cropsci1988.0011183X002800050004x

Mir, C., Zerjal, T., Combes, V., Dumas, F., Madur, D., Bedoya, C., Dreisigacker, S., Franco, J., Grudloyma, P., Hao, P.X., Hearne, S., Jampatong, C., Laloë, D., Muthamia, Z., Nguyen, T., Prasanna, B.M., Taba, S., Xie, C.X., Yunus, M., Zhang, S., Warburton, M.L., Charcosset, A., 2013. Out of America: Tracing the genetic footprints of the global diffusion of maize. Theor. Appl. Genet. https://doi.org/10.1007/s00122-013-2164-z

Miti, F., Pangirayi, T., Derera, J., 2010. S1 selection of local maize landraces for low soil nitrogen tolerance in Zambia. African J. Plant Sci.

Mohler, V., 2005. General Considerations : Marker-Assisted Selection. https://doi.org/10.1007/3-540-26538-4

Monneveux, P., Sánchez, C., Beck, D., Edmeades, G.O., 2006. Drought tolerance improvement in tropical maize source populations: Evidence of progress. Crop Sci. 46, 180–191. https://doi.org/10.2135/cropsci2005.04-0034

Nelson, D.E., Repetti, P.P., Adams, T.R., Creelman, R.A., Wu, J., Warner, D.C., Anstrom, D.C., Bensen, R.J., Castiglioni, P.P., Donnarummo, M.G., Hinchey, B.S., Kumimoto, R.W., Maszle, D.R., Canales, R.D., Krolikowski, K.A., Dotson, S.B., Gutterson, N., Ratcliffe, O.J., Heard, J.E., 2007. Plant nuclear factor Y (NF-Y) B subunits confer drought tolerance and lead to improved corn yields on water-limited acres. Proc. Natl. Acad. Sci. 104, 16450–16455. https://doi.org/10.1073/pnas.0707193104

Parentoni, S.N., Miranda, R.A. de, Garcia, J.C., 2015. Implications on the introduction of transgenics in Brazilian maize breeding programs. Crop Breed. Appl. Biotechnol. 13, 9–22. https://doi.org/10.1590/s1984-70332013000100002

Passioura, J.B., 2010. Scaling up: the essence of effective agricultural research. Funct. Plant Biol. 37, 585. https://doi.org/10.1071/FP10106

Piperno, D.R., Ranere, A.J., Holst, I., Iriarte, J., Dickau, R., 2009. Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. Proc. Natl. Acad. Sci. https://doi.org/10.1073/pnas.0812525106

Poland, J.A., Rife, T.W., 2012. Genotyping-by-Sequencing for Plant Breeding and Genetics. Plant Genome J. 5, 92. https://doi.org/10.3835/plantgenome2012.05.0005

Pollak, L.M., 2003. The History and Success of the public–private project on germplasm enhancement of maize (GEM), in: Advances in Agronomy. pp. 45–87. https://doi.org/10.1016/S0065-2113(02)78002-4

Prasanna, B.M., 2012. Diversity in global maize germplasm: Characterization and utilization. J. Biosci. 37, 843–855. https://doi.org/10.1007/s12038-012-9227-1

Rahman, M.U., Gul, S., Ahmad, I., 2004. Effects of Water Stress on Growth and Photosynthetic Pigments of Corn (Zea mays L.) Cultivars. Int. J. Agric. Biol. 6, 653–655.

Ramos-Madrigal, J., Smith, B.D., Moreno-Mayar, J.V., Gopalakrishnan, S., Ross-Ibarra, J., Gilbert, M.T.P., Wales, N., 2016. Genome Sequence of a 5,310-Year-Old Maize Cob Provides Insights into the Early Stages of Maize Domestication. Curr. Biol. 26, 3195–3201. https://doi.org/10.1016/j.cub.2016.09.036

Reif, J.C., Hailauer, A.R., Melchinger, A.E., 2005. Heterosis and heterotic patterns in maize. Maydica.

Rengel, D., Arribat, S., Maury, P., Martin-Magniette, M.-L., Hourlier, T., Laporte, M., Varès, D., Carrère, S., Grieu, P., Balzergue, S., Gouzy, J., Vincourt, P., Langlade, N.B., 2012. A Gene-Phenotype Network Based on Genetic Variability for Drought Responses Reveals Key Physiological Processes in Controlled and Natural Environments. PLoS One 7, e45249. https://doi.org/10.1371/journal.pone.0045249

Ribaut, J.-M., Betran, J., Monneveux, P., Setter, T., 2009. Drought Tolerance in Maize, in: Handbook of Maize: Its Biology. Springer New York, New York, NY, NY, pp. 311–344. https://doi.org/10.1007/978-0-387-79418-1_16

Ribaut, J.-M., Hoisington, D., 1998. Marker-assisted selection: new tools and strategies. Trends Plant Sci. 3, 236–239. https://doi.org/10.1016/S1360-1385(98)01240-0

Ribaut, J.-M., Jiang, C., Gonzalez-de-Leon, D., Edmeades, G.O., Hoisington, D.A., 1997. Identification of quantitative trait loci under drought conditions in tropical maize. 2. Yield components and marker-assisted selection strategies. Theor. Appl. Genet. 94, 887–896. https://doi.org/10.1007/s001220050492

Ribaut, J.M., Ragot, M., 2007. Marker-assisted selection to improve drought adaptation in maize: The backcross approach, perspectives, limitations, and alternatives. J. Exp. Bot. 58, 351–360. https://doi.org/10.1093/jxb/erl214

Russel, W.A., 1984. Agronomic performance of maize cultivars representing different eras of breeding. Maydica (Italy) 29, 375–390.

Ryder, P., McKeown, P.C., Fort, A., Spillane, C., 2019. Epigenetics and Heterosis in Crop Plants, in: Epigenetics in Plants of Agronomic Importance: Fundamentals and Applications. Springer International Publishing, Cham, pp. 129–147. https://doi.org/10.1007/978-3-030-14760-0_4

Sari-Gorla, M., Krajewski, P., Di Fonzo, N., Villa, M., Frova, C., 1999. Genetic analysis of drought tolerance in maize by molecular markers. II. Plant height and flowering. Theor. Appl. Genet. 99, 289–295. https://doi.org/10.1007/s001220051234

Schimmelpfennig, D.E., Pray, C.E., Brennan, M.F., 2004. The impact of seed industry concentration on innovation: a study of US biotech market leaders. Agric. Econ. 30, 157–167. https://doi.org/10.1016/j.agecon.2002.11.001

Schlötterer, C., 2004. The evolution of molecular markers — just a matter of fashion? Nat. Rev. Genet. 5, 63–69. https://doi.org/10.1038/nrg1249

Seki, M., Umezawa, T., Urano, K., Shinozaki, K., 2007. Regulatory metabolic networks in drought stress responses. Curr. Opin. Plant Biol. 10, 296–302. https://doi.org/10.1016/j.pbi.2007.04.014

Semagn, K., Magorokosho, C., Vivek, B.S., Makumbi, D., Beyene, Y., Mugo, S., Prasanna, B.M., Warburton, M.L., 2012. Molecular characterization of diverse CIMMYT maize inbred lines from eastern and southern Africa using single nucleotide polymorphic markers. BMC Genomics 13, 113. https://doi.org/10.1186/1471-2164-13-113

Setter, T.L., Yan, J., Warburton, M., Ribaut, J.M., Xu, Y., Sawkins, M., Buckler, E.S., Zhang, Z., Gore, M.A., 2011. Genetic association mapping identifies single nucleotide polymorphisms in genes that affect abscisic acid levels in maize floral tissues during drought. J. Exp. Bot. 62, 701–716. https://doi.org/10.1093/jxb/erq308

Shiferaw, B., Smale, M., Braun, H.J., Duveiller, E., Reynolds, M., Muricho, G., 2013. Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. Food Secur. 5, 291–317. https://doi.org/10.1007/s12571-013-0263-y

Shull, G.H., 1952. Beginnings of the heterosis concept. Biodivesity Herit. Libr. 1, 14–48.

Shull, G.H., 1914. Duplicate genes for capsule-form inBursa bursa-pastoris. Z. Indukt. Abstamm. Vererbungsl. 12, 97–149. https://doi.org/10.1007/bf01837282

Shull, G.H., 1908. The Composition of a Field of Maize. J. Hered. os-4, 296–301. https://doi.org/10.1093/jhered/os-4.1.296

Silva, L.D.C.E., Wang, S., Zeng, Z.-B., 2012. Composite Interval Mapping and Multiple Interval Mapping: Procedures and Guidelines for Using Windows QTL Cartographer, in: Rifkin, S.A. (Ed.), Quantitative Trait Loci (QTL). Springer, pp. 75–119. https://doi.org/10.1007/978-1-61779-785-9_6

Sinclair, T.R., 2011. Challenges in breeding for yield increase for drought. Trends Plant Sci. 16, 289–293. https://doi.org/10.1016/j.tplants.2011.02.008

Sprague, G.F., 1946. Early Testing of Inbred Lines of Corn1. Agron. J. 38, 108. https://doi.org/10.2134/agronj1946.00021962003800020002x

Stuber, C.W., Edwards, M.D., Wendel, J.F., 1987. Molecular Marker-Facilitated Investigations of Quantitative Trait Loci in Maize. II. Factors Influencing Yield and its Component Traits. Crop Sci. 27, 639. https://doi.org/10.2135/cropsci1987.0011183X002700040006x

Stuber, C.W., Lincoln, S.E., Wolff, D.W., Helentjaris, T., Lander, E.S., 1992. Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics 132, 823–839.

Thirunavukkarasu, N., Hossain, F., Arora, K., Sharma, R., Shiriga, K., Mittal, S., Mohan, S., Namratha, P., Dogga, S., Rani, T., Katragadda, S., Rathore, A., Shah, T., Mohapatra, T., Gupta, H., 2014. Functional mechanisms of drought tolerance in subtropical maize (Zea mays L.) identified using genome-wide association mapping. BMC Genomics 15, 1182. https://doi.org/10.1186/1471-2164-15-1182

Thornsberry, J.M., Goodman, M.M., Doebley, J., Kresovich, S., Nielsen, D., Buckler, E.S., 2001. Dwarf8 polymorphisms associate with variation in flowering time. Nat. Genet. 28, 286–289. https://doi.org/10.1038/90135

Tomes, D.T., 1998. Heterosis: Performance Stability, Adaptability to Changing Technology, and the Foundation of Agriculture as a Business, in: Larnkey, K.L., Staub, J.E. (Eds.), Concepts and Breeding of Heterosis in Crop Plants. Crop Science Society of America, Madison, pp. 13–27. https://doi.org/10.2135/cssaspecpub25.c2

Tomkowiak, A., Bocianowski, J., Radzikowska, D., & Kowalczewski, P. Ł., 2019. Selection of Parental Material to Maximize Heterosis Using SNP and SilicoDarT Markers in Maize. Plants. 8, 349-. https://doi.org/10.3390/plants8090349

Troyer, A.F., Wellin, E.J., 2009. Heterosis decreasing in hybrids: Yield test inbreds. Crop Sci. 49, 1969–1976. https://doi.org/10.2135/cropsci2009.04.0170

Tsonev, S., Todorovska, E., Avramova, V., Kolev, S., Abu-Mhadi, N., Christov, N.K., 2009. Genomics assisted improvement of drought tolerance in maize: QTL approaches. Biotechnol. Biotechnol. Equip. 23, 1410–1413. https://doi.org/10.2478/V10133-009-0004-8

Turner, N.C., 1981. Techniques and experimental approaches for the measurement of plant water status. Plant Soil 58, 339–366. https://doi.org/10.1007/BF02180062

Vallebueno-Estrada, M., Rodríguez-Arévalo, I., Rougon-Cardoso, A., Martínez González, J., García Cook, A., Montiel, R., Vielle-Calzada, J.-P., 2016. The earliest maize from San Marcos Tehuacán is a partial domesticate with genomic evidence of inbreeding. Proc. Natl. Acad. Sci. 113, 14151–14156. https://doi.org/10.1073/pnas.1609701113

Varshney, R.K., Bansal, K.C., Aggarwal, P.K., Datta, S.K., Craufurd, P.Q., 2011. Agricultural biotechnology for crop improvement in a variable climate: hope or hype? Trends Plant Sci. 16, 363–371. https://doi.org/10.1016/j.tplants.2011.03.004

Virmani, S.S., Pandey, M., Singh, I.S., Xu, W.J., 2004. Classical and Molecular Concepts of Heterosis, in: H.K., J., Kharkwal, M.C. (Eds.), Plant Breeding: Mendelian to Molecular Approaches. Springer Science+Business Media Dordrecht, New Delhi, pp. 407–418.

Wang, D., Yu, Y., Liu, Z., Li, S., Wang, Z., Xiang, F., 2016. Membrane-bound NAC transcription factors in maize and their contribution to the oxidative stress response. Plant Sci. 250, 30–39. https://doi.org/10.1016/j.plantsci.2016.05.019

Wang, X., Wang, H., Liu, S., Ferjani, A., Li, J., Yan, J., Yang, X., Qin, F., 2016. Genetic variation in ZmVPP1 contributes to drought tolerance in maize seedlings. Nat. Genet. 48, 1233–1241. https://doi.org/10.1038/ng.3636

Wang, X., Xu, Y., Hu, Z., Xu, C., 2018. Genomic selection methods for crop improvement: Current status and prospects. Crop J. 6, 330–340. https://doi.org/10.1016/j.cj.2018.03.001

Warburton, M.L., Reif, J.C., Frisch, M., Bohn, M., Bedoya, C., Xia, X.C., Crossa, J., Franco, J., Hoisington, D., Pixley, K., Taba, S., Melchinger, A.E., 2008. Genetic diversity in CIMMYT nontemperate maize germplasm: Landraces, open pollinated varieties, and inbred lines. Crop Sci. 48, 617–624. https://doi.org/10.2135/cropsci2007.02.0103

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

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. https://doi.org/10.2135/cropsci2004.0326

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. https://doi.org/10.2135/cropsci2005.0246

Xiao, B.Z., Chen, X., Xiang, C. Bin, Tang, N., Zhang, Q.F., Xiong, L.Z., 2009. Evaluation of seven function-known candidate genes for their effects on improving drought resistance of transgenic rice under field conditions. Mol. Plant 2, 73–83. https://doi.org/10.1093/mp/ssn068

Xiao, Y., Liu, H., Wu, L., Warburton, M., Yan, J., 2017. Genome-wide Association Studies in Maize: Praise and Stargaze. Mol. Plant 10, 359–374. https://doi.org/10.1016/j.molp.2016.12.008

Xu, J., Yuan, Y., Xu, Y., Zhang, G., Guo, X., Wu, F., Wang, Q., Rong, T., Pan, G., Cao, M., Tang, Q., Gao, S., Liu, Y., Wang, J., Lan, H., Lu, Y., 2014. Identification of candidate genes for drought tolerance by whole-genome resequencing in maize. BMC Plant Biol. 14, 1–15. https://doi.org/10.1186/1471-2229-14-83

Xue, Y., Warburton, M.L., Sawkins, M., Zhang, Xuehai, Setter, T., Xu, Y., Grudloyma, P., Gethi, J., Ribaut, J.-M., Li, W., Zhang, Xiaobo, Zheng, Y., Yan, J., 2013. Genome-wide association analysis for nine agronomic traits in maize under well-watered and water-stressed conditions. Theor. Appl. Genet. 126, 2587–2596. https://doi.org/10.1007/s00122-013-2158-x

Yang, H., Li, C., Lam, H.-M., Clements, J., Yan, G., Zhao, S., 2015. Sequencing consolidates molecular markers with plant breeding practice. Theor. Appl. Genet. 128, 779–795. https://doi.org/10.1007/s00122-015-2499-8

Yu, J., Buckler, E.S., 2006. Genetic association mapping and genome organization of maize. Curr. Opin. Biotechnol. https://doi.org/10.1016/j.copbio.2006.02.003

Yuan, Y., Cairns, J.E., Babu, R., Gowda, M., Makumbi, D., Magorokosho, C., Zhang, A., Liu, Y., Wang, N., Hao, Z., San Vicente, F., Olsen, M.S., Prasanna, B.M., Lu, Y., Zhang, X., 2019. Genome-Wide Association Mapping and Genomic Prediction Analyses Reveal the Genetic Architecture of Grain Yield and Flowering Time Under Drought and Heat Stress Conditions in Maize. Front. Plant Sci. 9, 1919. https://doi.org/10.3389/fpls.2018.01919

Zaidi, P.H., Srinivasan, G., Sanchez, C., 2003. Morpho-physiological traits associated with variable field performance of different types maize germplasm across multiple environments. Maydica 207–220.

Zeng, Z.-B., 2005. QTL mapping and the genetic basis of adaptation: recent developments, in: Mauricio, R. (Ed.), Genetics of Adaptation. Springer-Verlag, Berlin/Heidelberg, pp. 25–37. https://doi.org/10.1007/1-4020-3836-4_4

Zhang, J., Nguyen, H.T., Blum, A., 1999. Genetic analysis of osmotic adjustment in crop plants. J. Exp. Bot. 50, 291–302. https://doi.org/10.1093/jxb/50.332.291

Zhang, X., Warburton, M.L., Setter, T., Liu, H., Xue, Y., Yang, N., Yan, J., Xiao, Y., 2016. Genome-wide association studies of drought-related metabolic changes in maize using an enlarged SNP panel. Theor. Appl. Genet. 129, 1449–1463. https://doi.org/10.1007/s00122-016-2716-0

Zhao-Bang Zeng, Zeng, Z.B., 1994. Precision mapping of Quantitative trait loci. Genetics 136, 1457–1468. https://doi.org/10.1093/aob/mcr323

Zhen, G., Qin, P., Liu, K.Y., Nie, D.Y., Yang, Y.Z., Deng, X.W., He, H., 2017. Genome-wide dissection of heterosis for yield traits in two-line hybrid rice populations. Sci. Rep. 7, 1–14. https://doi.org/10.1038/s41598-017-06742-7

Zhu, C., Gore, M., Buckler, E.S., Yu, J., 2008. Status and Prospects of Association Mapping in Plants. Plant Genome J. 1, 5. https://doi.org/10.3835/plantgenome2008.02.0089

Zhu, J.J., Wang, X.P., SUN, C.X., ZHU, X.M., LI, M., Zhang, G.D., TIAN, Y.C., WANG, Z.L., 2011. Mapping of QTL associated with drought tolerance in a semi-automobile rain shelter in maize (Zea mays L.). Agric. Sci. China 10, 987–996. https://doi.org/10.1016/S1671-2927(11)60085-0

TABLE AND FIGURES

Table 1: Average of maize yields, a total of maize area and total maize production in Africa, America, Asia and Europe. Only countries that harvested more than 100,000 ha were considered. Oceania is not in the list once no country there achieve it. "America – USA & Brazil" and "Asia – China" are the values of the continents discounting USA, Brazil, and China, as they were discrepant data of their continents. The second section presents the 10 countries with higher yields individually.

Continents	Yield ($t.ha^{-1}$)	Area (ha)	Production (tonnes)
Africa	1,89	40.069.054	83.619.170
America	4,21	71.294.740	575.985.861
America - USA & Brazil	3,66	20.432.245	107.303.611
Asia	5,00	66.832.585	359.542.826
Asia - China	4,93	24.433.522	100.471.826
Europe	6,98	17.147.016	107.378.921
Country			
United States of America	11,08	33.469.003	370.960.390
Canada	10,52	1.339.323	14.095.300
France	8,75	1.614.109	14.121.680
Argentina	7,58	6.530.695	49.475.895
South Africa	6,40	2.628.618	16.820.000
China, mainland	6,11	42.399.064	259.071.000
Romania	5,96	2.405.242	14.326.100
Brazil	5,62	17.393.493	97.721.860
Ukraine	5,51	4.480.665	24.668.750
Indonesia	5,20	5.374.971	27.952.000

Data from FAOSTAT (2017)

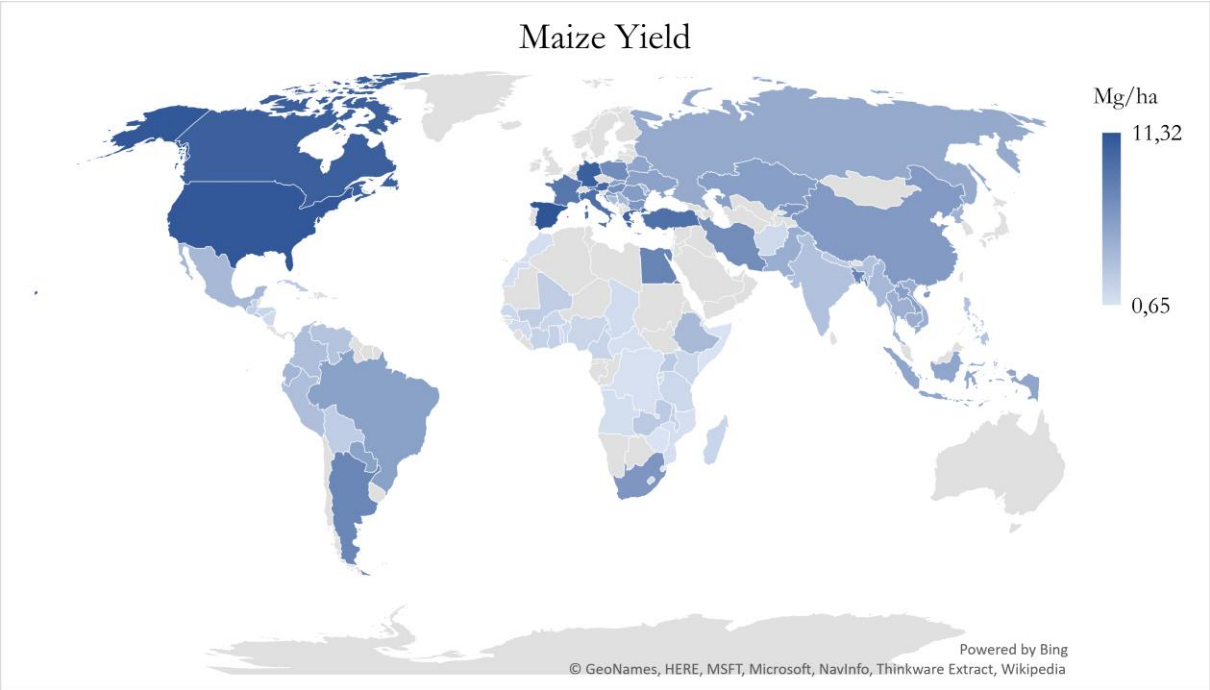

Figure 1: Maize yield in $Mg.ha^{-1}$ in 83 countries that have cultivated more than 100.000 ha in 2017. Countries in gray are those not considered. Data from FAOSTAT.

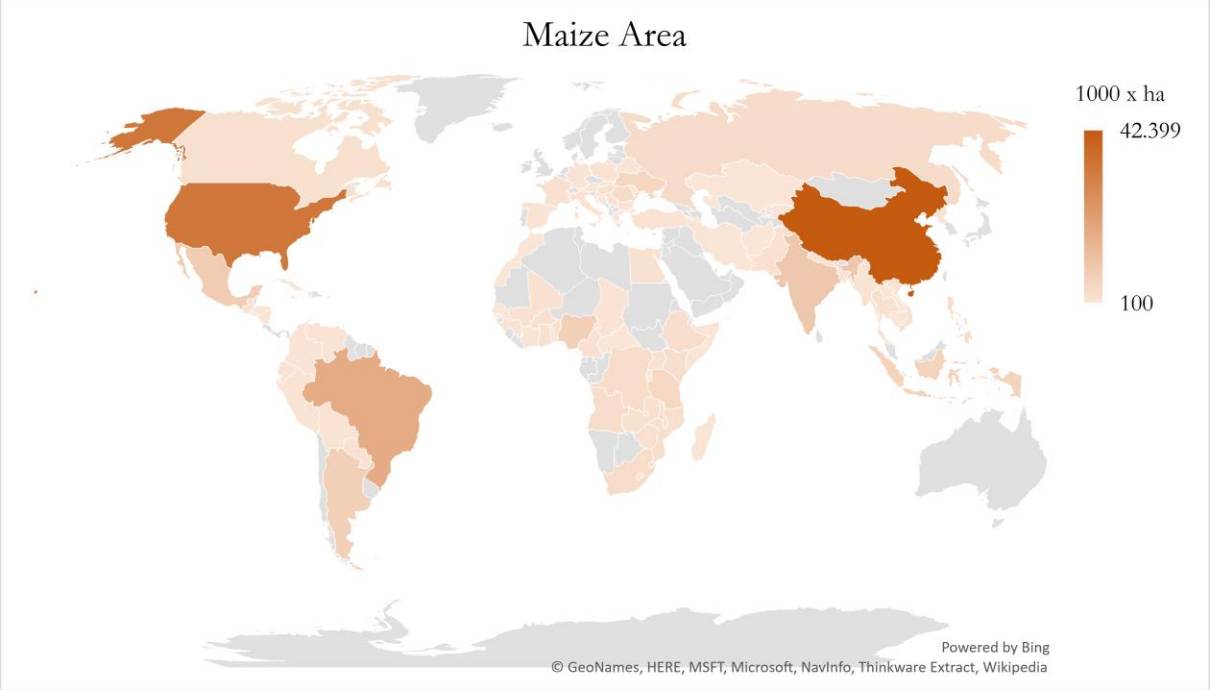

Figure 2: Maize area (1000 ha) in 83 countries that have cultivated more than 100.000 ha in 2017. Countries in gray are those not considered. Data from FAOSTAT.

3. ON THE INTROGRESSION OF MAIZE DIVERSITY FOR DROUGHT TOLERANCE: LANDRACES AND EARLY TESTCROSSES AS GENETIC SOURCE

ABSTRACT

In a scenario of climate change and increased frequency of drought events, maize production can be negatively affected resulting in yield reductions. In this context, knowledge about the genetic resources available for traits related to drought tolerance has great importance in developing breeding program strategies. The aim of this research was to study a maize landrace panel to identify chromosomal regions (SNP) of maize associated with a drought tolerance index. For that, we performed Genome-Wide Association Study (GWAS) on 1326 landrace progenies developed by the CIMMYT Genetic Resources Program, originating from 20 landraces populations collected in arid regions. Phenotypic data were obtained from early testcross trials conducted in three sites and two contrasting irrigation environments, full irrigation (well-watered) and reduced irrigation (drought). The populations were genotyped using the DArTSeq™ platform, and a final set of 5,695 SNPs markers was used. The genotypic values were estimated using spatial adjustment in a two-stage analysis. First, we performed the individual analysis for each site/irrigation treatment combination. The best linear unbiased estimates (BLUEs) were used to calculate the Harmonic Mean of Relative Performance (HMRP) as a drought tolerance index for each testcross. The second stage was a joint analysis, which was performed using the HMRP to obtain the best linear unbiased predictions (BLUPs) for each genotype. Then, GWAS was performed to determine the marker-index associations and the marker-Grain Yield (GY) associations for the two irrigation treatments. We detected two significant markers associated with the drought-tolerance index, and four marker-GY associations for drought and irrigated conditions each. Although each of these markers explained less than 0.1% of the phenotypic variation for the index and GY, through the annotation process, we found two genes likely related to the plant response to drought stress. For these markers, alleles from landraces provide a slightly higher yield under drought conditions, our results indicate that the diversity delivered by landraces combined with commercial lines is an exciting strategy to improve maize for drought tolerance.

3.1. INTRODUCTION

Maize is a major crop cultivated in many regions of the world and is one of the most important cereals for food production, occupying around 197 million hectares in 2017 (FAOSTAT 2017). That corresponds to about 15% of the whole area destined for agriculture. Maize production is to a large extent concentrated in tropical and subtropical regions of the globe, including many parts of Central and South America, Africa and Southeast Asia. Millions of smallholder and subsistence farming communities in these regions rely on maize as a major source of calories and income to pay for basic necessities and schooling for their children. As maize is a high water-demanding plant drought events cause significant losses in grain yield and introduces year-to-year yield instability. This yield instability creates a financial instability in the farming communities of developing countries where maize is important. Most climate change

forecast models concur that an increase in the frequency and intensity of drought events will occur in the coming years (IPCC 2007, 2014). If maize varieties with novel sources of drought tolerance are not developed in the near future, smallholder and subsistence farmers will become even more vulnerable to yield fluctuations and the financial insecurity that comes with it. Fortunately, maize is a species with vast genetic variability, enabling breeding as an essential tool to overcome the effects of climate change in agriculture (Fisher et al. 2015). In this context, the more than 25,000 maize landrace accessions maintained in the maize germplasm bank of the International Maize and Wheat Improvement Center (Spanish acronym, CIMMYT) is a tremendous resource for identifying and using available maize genetic resources for developing cultivars able to maintain grain production in water limited conditions. However, the reality is that breeders in elite maize breeding programs in the private and public sector have, for many years, stopped using non-inbred materials in the germplasm banks as new sources of variation. This is mostly due to such factors as linkage drag, adaptation issues, poor agronomic characteristics and low yield potential (Goodman et al. 2014). For that reason, in 2011 CIMMYT initiated the Seeds of Discovery project (SeeD), largely funded by the Mexican government through MasAgro Biodiversidad program. The objective of the SeeD project and MasAgro Biodiversidad is to explore unexplored allelic variation in the bank for important abiotic and biotic stresses, such as drought tolerance, as well as nutritional and end-use characteristics.

Tolerance to abiotic or biotic stress is related to the effects of this stress on crop productivity, or another trait of interest (Schafer 1971; Miti et al. 2010). CIMMYT has been conducting a long-term recurrent selection program to improve maize for drought tolerance, the traditional selection process usually applied in it considers traits as maintain time from sown to anthesis, yield in well-watered conditions, maintain or increase grain yield under drought and decrease the interval between anthesis and silk (Bänziger et al. 2000). Anthesis-Silk Interval (ASI) is one of the most studied traits, and it has been used as a secondary trait to the indirect selection for grain yield and drought tolerance (Bolaños and Edmeades 1996; Xue et al. 2013b). Nevertheless, it is known that selecting progenies that escape drought stress by flowering early can result in yield decrease (Bolaños and Edmeades 1996).

Regarding grain yield, drought tolerance is the ability of the plant to maintain grain production when affected by a dry period. Grain yield, by itself, is a complex trait, profoundly affected by the environment, and it requires evaluations in many trials and different locations to achieve proper experimental parameters for selection. Hence, considering that drought tolerance is a trait that relates the grain yield, is expected to be even more complex (Campos et al. 2004a), once the plant responses to drought stress are based in plant physiology and morphology

modifications that involve a large number of genes and molecular pathways (Rengel et al. 2012b; Min et al. 2016).

The existence of genetic diversity is one of the first requirements for the success of a plant breeding program. Mexico is the center of origin and domestication for maize, and high genetic diversity can still be found in germplasm banks or even in small local producers that use landrace population (Sanchez et al. 2000; Prasanna 2012; Arteaga et al. 2016). Studies that quantify and characterize existing diversity are of great importance for decisions regarding the methodologies to be used by maize breeders, whether for drought tolerance or any other trait. Understanding the genetic control of drought tolerance traits is essential for the success of a breeding program that is involved in this matter. In this context, molecular markers are an important tool in the analysis of genetic diversity and, consequently, in obtaining gains in breeding programs(Xu et al. 2017). Further, genome-wide association studies (GWAS) seek to relate molecular markers to observed phenotypes and thus discovery of chromosomal regions associated with the trait of interest. This procedure has been used for many different traits in maize (Xiao et al. 2017), both abiotic and biotic stresses and as well in others plants as: millet (Jaiswal et al. 2019), common beans (Fritsche-Neto et al. 2019), rice (Huang et al. 2010), soybeans (Hwang et al. 2014).

In a review on the use of GWAS in maize, Xiao et al. (2017) showed that most studies use a panel of inbred lines with *per se* phenotypic data or, seldom, testcrosses by means inbred lines. For crops such as maize where hybrids are used to exploit heterosis, the testcross is an important strategy in research for the development of improved hybrids, as it represents the real genotypic constitution of crop as grown by most farmers. Additionally, early generation testing (EGT) of semi-inbred lines as testcrosses is a common strategy used in maize breeding programs, being a cross between a tester inbred or F_1 with a S_2 or S_3 experimental line. The aim is to take advantage of early selection for yield potential, avoiding future testing costs on unpromising plants or families (Bernardo 2010). EGT takes advantage of the fact that yield potential can be identified early in the inbreeding process and shortens the time to line release because the inbreeding and testing process can be done in tandem instead of sequentially. Early testcross populations to date have not been commonly used in GWAS studies. Therefore, this study aims to identify and validate genomic regions related to drought tolerance in two sequential inbreeding generations of early testcrosses obtained between landrace-derived semi-inbred lines and elite testers.

3.2. MATERIAL AND METHODS

3.2.1. Plant Material

We used a breeding population developed by the CIMMYT Genetic Resources Program. Basically, this breeding population was generated from a first process of test and selection of the 20 more adapted landraces over an initial collection of 400 landraces. These 20 landraces were the initial plant material that generate the testcross trials used by this research (Figure 3**Error! Reference source not found.**). Following we briefly describe the selection process of the 20 landraces and the process to generate the testcrosses.

Landraces are known as populations locally adapted, and Subtropical adapted maize landraces are from areas between 800 and 1800 meters above sea level (Edmeades et al. 2017). Although latitude also plays an important role in adaptation, we did not consider latitude in the selection of the landraces for this study. Twenty landraces accessions were selected due to their agricultural performance in dry regions. This selection was done over 400 subtropical adaptation landrace accessions from the CIMMYT Maize Germplasm Bank and the Mexican National Maize Collection collected and managed by Instituto Nacional de Investigación Forestal Agrícola y Pesquera de Mexico (INIFAP). The selected landraces originate from dryland production areas, with 290 (89%) coming from Mexico and the rest from Argentina and Chile. At least 24 maize races are represented in the selected although 17% of them have not been racially classified. A seed increase was conducted in the summer of 2013 to produce seed for drought evaluations and after the increase, 326 landraces had produced enough seed for multi-location, multi-replicate yield trials. The 326 accessions were evaluated in the winters of 2014 and 2015 under both managed drought and normal irrigation conditions. More details on the evaluated landrace accessions (that was not by itself the plant material of this study) and the 2014 and 2015 yield trials can be found at https://seedsofdiscovery.org/catalogue/. From this process we selected the 20 most drought adapted landrace populations and those were used in this research.

These 20 landraces were crossed with the CIMMYT Maize Line number 376 (CML376), to generate 20 F1 segregating populations. The reason for mating the landraces with a CIMMYT line is to provide agronomic adaptation, once those are non-suitable for most of maize production regions and have a strong local adaptation to an environment normally managed with less agriculture technology. This is a common way to deal with maize landraces (Meseka et al. 2015; Navarro et al. 2017). These F1 populations have undergone a backcross, using CML376 as recurrent parent, forming the BC1F1 population, composed of 1480 plants (74 per landrace family on average). These plants were selfed [BC1S1 (backcross 1, selfing 1)] and, that same BC1F1 plants, crossed to a female tester (CML264/CML311). This tester is a standard female in commercial hybrids for subtropical areas of Mexico and has the advantage of producing large

quantities of hybrid seed. The seeds generated by this cross were used in testcross trials (TC1). The data from the TC1 trials was used to proceed a selection **Error! Reference source not found.**based on the yield at the drought conditions trials, discarding those that present a very low production on well-watered trials (Table 1**Error! Reference source not found.**). In addition, the selection was stratified by landrace family, trying to do not exclude an entire landrace progeny (even it happens, for an extra reason, as seed quality) or do not select more than 30% of the same landrace family. BC1S1 were self-crossed again, at the same time of growing the testcrosses, the seeds from 64 selected BC1S1 were used to compose the next generation, progenies of these 64 plants compound the 174 BC1S2 (backcross 1, selfing 2). The selected BC1S2 were planted and undergone through another testcross process (TC2), this with three different testers (CML264/CML311, CML373, CL501801).

3.2.2. Testcross and phenotyping

From the 1480 BC1F1 crossed plants (BC1F1 x Tester), 1326 produced both enough selfed seed and testcross seed for the TC1 trials. Six genotypes were added to be used as checks, so the TC1 had a total of 1332 genotypes. This set of 1332 genotypes were unbalanced planted in five locations in Mexico, only three of the locations were used for this study: the INIFAP experiment station at Santiago del Ixcuintla (SI) – Nayarit State, the CIMMYT experiment station at Tlaltizapán (TL) – Morelos State and the INIFAP station at Los Mochis (LM) – Sinaloa State. In each location, two trials were conducted in contrasting water supply conditions, well-watered (WW), and drought (DR). The unbalanced design was due to restrictions on field space and seed amounts, but the landraces ancestry was balanced in each location. The distribution of the testcross entries was: 266 tested only in TL (~13 of each 20 landraces), 266 tested only in LM (~13 of each 20 landraces), 265 tested in TL and LM (~13 of each 20 landraces), 529 tested in the 3 locations (~26 of each 20 landraces). Thus, the TC1 experiment consisted of 6 trials (3 locations and 2 irrigation treatments each). The WW trials were planted in a spatial design with systematically distributed spatial check plots, with no replicates of the 1326 progenies, i.e., in an unreplicated design scheme. The DR trials were conducted in a randomized complete block design with two replicates, except in Santiago del Ixcuintla, where even the DR experiment was conducted in an unreplicated design, and check plots with ~25 replication. The checks included in the trial were of three categories: i) CML264/CML311//CML376, the genetic check consisting of the recurrent parents crossed to the tester; ii) the comparative checks consisting of the commercially sold hybrid CML264/CML311//CL106951 and the CIMMYT standard check

for drought evaluations LPSC7-F64/CML550; iii) the spatial checks used for the spatial adjustment, Asgrow hybrid CEBU and DuPont-Pioneer hybrids P3055W and P4082W..

The TC1 trials were planted on January 14, 20 and 30 of 2016 at TL, SI and LM, respectively. Harvest dates for the trial, in the same order, were June 4, 10, 27. Winter planting dates were chosen because this is the dry season for much of Mexico, especially the western half of the country. Using drip irrigation in all the sites, we were able manage the amount of water delivered to each treatment. In all locations we used a two-row plot, in TL and SI were 4.5 meters long with 0.75 m between the rows, and in LM, 4 meters rows spaced by 1 m. The agronomic treatments (fertilizers, insecticides, and herbicides) were applied following common local practices. The irrigation in drought trials was interrupted before the anthesis expected date for each trail, and after the flowering one more irrigation was done to ensure the grain filling. The plots were harvested by hand, the ears threshed by machine, the weight and moisture grain of each plot were taken, and the plot grain production was adjusted to 13% of grain moisture and converted to 1000 kg per hectare ($t.ha^{-1}$), considering the differences in plots size of each site.

Implementation of the TC2 trials followed a similar process. The set of 174 genotypes selected by TC1 data were tested in 18 trials, 3 locations, 2 irrigation treatment (drought and well-watered), but now with 3 testers (CML264/CML311, CML373, CL501801). The 3 locations were TL, LM and Puerto Vallarta (PV) – Jalisco State, replacing SI. That is, one location for the TC2 trials was different from TC1 location. All TC2 trials used two-rows plot of 4 m length, with 0.75 m between rows, harvest was accomplished using a Wintersteiger Classic plot combine at TL and PV location, and a New Holland TR 88 2-plot combine at LM, both combines provide measures of the grain weight and grain moisture of each plot. The agronomic treatments and water supply were managed the same as for the TC1 trials.

3.2.3. Genotypic data

DArTseq technology (Sansaloni et al., 2011), from Diversity Array Technology company (http://www.diversityarrays.com), was used for genotyping all the 1326 BC1S1 and 174 BC1S2 samples and the recurrent genitor CML376, it was carried out in the Genetic Analysis Service for Agriculture (SAGA) facility at CIMMYT, Mexico. A bulk of 10 seeds of each BC1S1 and BC1S2 line was sampled. A genomic representation was generated by digesting nuclear DNA with a combination of two restriction enzymes, *PstI* (CTGCAG) and *NspI* (CATG), and ligating individual barcodes adapters to identify the origin of each fragment after a samples pooling. Successfully amplified fragments were sequenced using the sequencer Illumina HiSeq2500 (Illumina Inc., San Diego, CA). Then, the SNP calling for those fragments was

produced by the analytical pipeline (DArTsoft; DArT P/L, Australia http://www.diversityarrays.com/software.html#dartsoft), and a set of filtering parameters were then applied to select high-quality markers for this study. In order to obtain the positions of the markers on the chromosomes, the sequences of the DNA fragments were BLASTed against the *Zea mays L.* reference genome (**B73 RefGen_v4)**. This procedure resulted in a genomic profile of 47,074 SNPs with excellent quality and positioned on the maize reference genome.

The 47,074 SNPs were named, for this study, by the order in the genome from M1 (the first mark in chromosome number 1) to M47074 (the last one at chromosome number 10). A new quality control procedure was applied, and markers with minor allele frequency (MAF) <5% excluded and a call rate of 95% used, thus, markers with more than 5% of missing data were excluded as well. The imputation of missing data was performed by Wright method carried out using the R package snpReady (Granato and Fritsche-Neto 2018) by the function raw.data, and it results in a profile of 5,695 SNPs.

3.2.4. Phenotypic Analysis

For analysis of the phenotypic data, a two-stage process was performed. In the first, adjusted means of genotypes were obtained by location, and then a mixed model was fit jointly (second stage) considering residual weights estimated in the first stage. The spatial corrections were applied in the first stage, using a first-order autoregressive process (AR1⊗AR1) to model the covariance structure of error. The spatial correction approach was chosen due to some factors as the high number of genotypes tested, the non-replicated experimental design applied and the large size of the field trial, in this case, field trends may imply serious bias in the treatment estimates.

The WW trials were designed in a complete spatial model, with no replicates of the test genotypes, and the checks were systematically distributed in the field, so these trials were fitted with the following model in the first stage:

$$\boldsymbol{y} = X\boldsymbol{\tau} + \boldsymbol{e}$$

where $\boldsymbol{y}$ is the vector of phenotypic data, arranged by Row x Column (plot field position), X is the design matrix of fixed effects; $\boldsymbol{\tau}$ is the vector genotypes effects (and tester in the case of TC2 trials), and eventual spatial factors row and columns (as explained below); and $\boldsymbol{e}$ vector of errors ($\boldsymbol{e} \sim \boldsymbol{N}(0, \boldsymbol{R}\sigma^2)$). In the spatial model $\boldsymbol{e}$ is partitioned into a vector ζ of spatial correlated effects, and a vector η of independent errors, and we assume the error variance as:

$$\boldsymbol{R} = var[\zeta] + var[\eta]$$

The $var[\eta]$ is assumed to be independent $\sigma^2 \boldsymbol{I}_n$, and the $var[\zeta]$ is the correlated spatial error, and we assume a first-order autoregressive, (AR1⊗AR1) (Gilmour et al. 1997), so the error variance is the sum of them, and we can show it as (Dutkowski et al. 2002):

$$\boldsymbol{R} = \sigma_\zeta^2[AR1(\rho_{col}) \otimes AR1(\rho_{row})] + \sigma_\eta^2 I$$

where AR1(ρ_{col}) and AR1(ρ_{row}) are the first-order autoregressive correlation matrices for columns and rows. The components σ_ζ^2 and σ_η^2 are the spatial and non-spatial residual variances, respectively.

The DR trials follow the same logic, but with the factor block as fixed effects. Therefore, the model is similar, but X also contains the block design, and the vector τ also contains the fixed effects of blocks.

This analysis provides a variogram plot of the residuals that shows in a geographic form if there is a spatial pattern that can be interpreted as a result of systematic variation due to the row and columns effects. Depending on the variogram of each trial, the row and column may be included as a factor of fixed effect (Burgueño et al. 2000). The Wald test was used to verify the significance of adding this fixed effect in the model.

Once all the individual trials were fitted to its specific model, we obtained the genotypes adjusted means (BLUE) of each location and irrigation treatment irrigation treatment.

3.2.5. Harmonic Mean Relative Performance Index

With the adjusted mean of each genotype in each combination of location and irrigation treatment irrigation treatment, we applied the Harmonic Mean of Relative Performance Index (HMRP) as a measure of genotype drought tolerance in each location. This index, proposed by Resende (2004), allows selection of genotypes that had good performance in both contrasting environments, in our case, in the drought and well-watered fields. The following equation obtains the HMRP index:

$$\mathrm{HMR}P_{ij} = \frac{2}{\left(\frac{GYww_{ij}}{\bar{x}_{ww_j}}\right)^{-1} + \left(\frac{GYdr_{ij}}{\bar{x}_{dr_j}}\right)^{-1}}$$

where $HMRP_{ij}$ is the harmonic mean relative performance of genotype *i* in site *j*; $GYww_{ij}$ is the BLUE of grain yield under well-irrigation treatment of genotype *i* in the site *j*; $GYdr_{ij}$ is the BLUE mean of grain yield in the drought condition of genotype *i* in site *j*; $\bar{x}_{ww\ j}$ and $\bar{x}_{dr\ j}$ are the overall BLUE means of the well-watered and drought condition experiments, respectively, in site *j*.

For the joint analysis (second stage), we used a linear mixed model to compute the best linear unbiased predictions (BLUPs) of the HMRP, grain yield in drought (GYDR) and well-watered (GYWW) conditions for each genotype fitting the following model:

$$\mathbf{y} = X\boldsymbol{\beta} + Z\mathbf{g} + S\boldsymbol{i} + \mathbf{e}$$

where $\mathbf{y}$ is the vector of BLUEs obtained in the first stage; $\boldsymbol{\beta}$ is the vector of fixed effects of site, $\mathbf{g}$ is the vector of the genotypic values of the lines, $\boldsymbol{i}$ is the vector of the random effects of the interaction sites by genotype, $\mathbf{e}$ is the random errors vector. *X*, *Z*, and *S* are the design matrix for $\boldsymbol{\beta}$, $\mathbf{g}$, and $\boldsymbol{i}$. The weighting method was based on squared standard errors from stage one BLUES. As the weight of HMRP, we used the mean of squared standard errors of the well-watered and drought experiments, as the index use both adjusted means (WW and DR BLUES) as its components. The significance of the fixed effect of local was estimated using Wald statistic, and for the random effects, genotypes and genotypes x local interaction, by Likelihood Ratio Test (LRT). The broad-sense heritability was estimated as:

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{ge}^2}{l} + \frac{\sigma_e^2}{l}}$$

where σ_g^2 is the genotypic variance, σ_{ge}^2 is the variance due to genotype by site interaction, σ_e^2 is the residual variance, *l* is the number of sites, 3, in this study. All the phenotypic analyses were performed using the ASReml-R package (Butler 2009) in R, version 3.5.3 (R Core Team 2019).

3.2.6. Genome-wide association studies

Genome-wide association analysis based on Mixed Linear Model (MLM) was performed using the Fixed and Random Model Circulating Probability Unification in the R package FARMCPU (Liu et al. 2016a) to the three traits, grain yield in drought and well-watered conditions and the drought tolerance index (HMRP). Principal Component Analysis was done using genomic information to assess population structure. The genome-wide association model (MLM equation) used was:

$$\boldsymbol{g} = X\boldsymbol{\beta} + Z\boldsymbol{u} + \boldsymbol{e}$$

where $\boldsymbol{g}$ is the vector of adjusted phenotypic observation (BLUPs of HMRP obtained in the joint analysis presented above, or GYDR and GYWW); $\boldsymbol{\beta}$ is the vector of the fixed effects of intercept, single markers, and the three first principal components used for population structure control; $\boldsymbol{u}$ is the vector of random additive effects; $\boldsymbol{e}$ is the vector of random residual; the X is the design matrix of fixed effects; Z is the genotype incidence matrix driven by VanRaden's genomic relationship matrix (GRM) (VanRaden 2008), so $\boldsymbol{u}$~N(0,$\mathbf{G}$), where G is the

VanRaden's GRM. A Multiple-test threshold adjustment was performed by a permutation method to determine the SNP significance level. This procedure was carried out in FARMCPU using the FarmCPU.P.Threshold function. The same GWAS process was done with the BC1S1 and BC1S2 populations, using the phenotypic data from TC1 and TC2, respectively.

The linkage disequilibrium for the BC1S1 data was measured by the r^2 parameter (square of the correlation coefficient between two loci), using Synbreed R package (Wimmer et al. 2012). The LD decay pattern for each chromosome was evaluated by nonlinear regression fitting expectation of r^2 of each pair of markers with the distance between them to generate a curve of LD decay (Hill and Weir 1988; Remington et al. 2001).

3.2.7. Gene annotation

Through the MaizeGDB (Portwood et al. 2019) and National Center for Biotechnology Information (NCBI, www.ncbi.nlm.nih.gov) databases, a search into maize reference genome (B73 RefGen_v4) was carried out to check the surroundings of GWAS-significant markers. This search was done inside a window of 200k base pairs (bp) centered on the markers, i.e., in a chromosome segment of 100k bp going up and downstream of the markers, we used this window size based on the LD decay obtained to our population and on the maize LD decay reported in the literature. Gene ontology is not thoroughly consistent across platforms, and MaizeGDB notation was chosen as primary. The functional annotation of the genes found inside those windows was analyzed and related to a drought tolerance process.

3.2.8. Origin of alleles for significant markers and superior allele

The recurrent genitor CML376 was also genotyped. Thus, it is possible to check if an allele that drives the phenotype to the desired form is derived from the CML376 or the landraces. The frequency of alleles derived from the landraces for the SNPs associated with the traits was calculated for the entire population and stratified for each landrace family group. It allowed verifying if specific landrace is a potential donor for the allele of interest. This study is critical because the main reason to use landraces is to bring new positive variants (genes) to a breeding population. Thus, it is essential to verify if the detected favorable genes in our breeding population are coming from the landraces or from the CML376.

On the other hand, as regards selection, it is also important to know the distribution of the superior allele across the landrace progenies in the breeding population, regardless if it is from landraces or the backcrossed elite line. The superior allele is the one that causes a positive

variation in the trait of interest. Therefore, we also calculate the frequency of the alleles, to the significant SNPs, with positive effects on the traits for each landrace progeny.

To evaluate the landraces regards their content of favorable alleles they were ranked by the sum of the ranks of the superior allele frequency for each SNPs. And two ranks were made, one considering all significant SNPs, regardless of the origin of the superior allele, and the other considering only the markers where the superior allele is from landraces. Thus, with the former we assess which landraces could be more suitable for selection, and with the latter which landraces would be the best source of superior new alleles.

3.2.9. GWAS validation

Validation was performed by two approaches. In the first, we compare the results of GWAS over BC1S1 and BC1S2. In the second, the effects of allelic substitution of significant SNPs (from BC1S1 GWAS) was used for predicting BC1S2 individuals by:

$$\hat{y}_2 = \mathrm{X}\alpha_1$$

where $\hat{y}_2$ is the vector (size n) of estimated genetic values in generation BC1S2, X is the n x m genomic matrix of BC1S2 individuals by BC1S1 significant markers and α_1 is the vector of allele substitution value for each significant SNP. n is the number of genotypes in BC1S2 and m the number of significant SNPs. The prediction ability was measured by Pearson Product-Moment correlation $r_{(\hat{y}_2, y_2)}$ where y_2 is the vector of BLUPS of the genotypes in the TC2. This procedure was carried out to the tolerance index and grain yield in both irrigation treatments.

3.3. RESULTS

3.3.1. Phenotypic analysis

Based on the variograms surface trends, specific factors (row, column or a combination of them) were chosen, as fixed effects, to fit a better model to each site/irrigation treatment (Table 4). In all the individual trials, the Wald test confirmed the significance of using the spatial factors.

Landraces progenies and genetic check (Tester x CML376) distributions of grain yield means (Figure 4) and drought tolerance index (Figure 5) show that there are in the landrace progenies genotypes performing better than the check, what is a desirable situation for a selection. Significance differences among genotypes were detected to grain yield under well-

watered conditions (Table 3). Under drought and for the index, genetic differences were detected only to a 0.2 p-value threshold, both with a p-value of 0.15 (data not shown).

Overall, the heritabilities were low, considering the joint analysis for the drought-tolerance index, GY in drought (GYDR) and well-watered (GYWW) conditions as expected for such complex traits (Table 3). The heritability of HRMP (0.06) and GYWW (0.04) were lower than for GYDR (0.17). The accuracy was similar for the three traits, with a higher value to GYDR (25.5%).

The Harmonic Mean of Relative Performance index measures drought tolerance. Its values are around 1.0, being higher values more tolerant genotypes and lower less tolerant. Regardless of site, the HMRP mean is always close to 1.0 because it is relative to the population in each site (Table 5). The genotypes with higher drought tolerance index values are also those that presented higher yield under both irrigation treatments (Figure 7). In BC1S1, the tolerance index to the location TL presents a higher variance than the other locations, and LM the lower. In contrast, in BC1S2 generation the LM presents a variance higher than TL and PV, and in general, the amplitude of the drought tolerance index was lower in BC1S2 than in BC1S1 (Figure 6), what was expected as the genetic diversity was higher in the first generation and before a selection.

3.3.2. GWAS

The use of principal components (PC) as covariates into the GWAS model was necessary once the population structure might result in a spurious association between the markers and the trait. This necessity was confirmed comparing the QQ-plots of the GWAS model with no PC as covariates and the QQ-plots of models with it (QQ-plots not shown). The QQ-plots showed a better adjustment in the models that consider the correction for the first PCs. However, the sum of variance explained by the first three PC was just 5.58%, distributed by 2.76%, 1.44% and, 1.38%, evidence that the population was not very structured. Plotting the first two PCs no clear stratification in the population is noted (Figure 8), even knowing that the landraces families exist in the population. Only when considering the third PC is it possible to separate the group of landraces into SNLP166 and CHIH87 progenies. The lack of population structure gives one positive side to the analyzes, since there is a lower risk of false-positive due to a possible structure, which is reduced further with the use of PCs as a covariate in the model.

The final Quantile-quantile plots (QQ-plots) for the drought tolerance index and for grain yield in drought and irrigated fields are presented in Figure 9. The QQ-plot is a plot of the negative log observed p-values against the distribution of expected p-values, i.e., under the null

hypothesis, that there is no association between the SNP and the trait. This is a simple tool to check how well the chosen GWAS model absorbed the population structure (covariates in the model). The dots on the upper right section of the graph are the SNPs with some association with HMRP. We expected that most markers would not be associated with the trait. Looking at the QQ-plot, most markers have a p-value observed close to the expected null hypothesis, which can be verified by the line-shaped markers' p-values that are aligned to the red line (observed equals the expected).

The thresholds calculated from the permutation (8.78×10^{-6} for HMRP, 1.13×10^{-5} for GY in DR and 1.28×10^{-5} in WW) were very close to the Bonferroni correction ($0.05/N = 9.01 \times 10^{-6}$) for the three GWAS procedures, considering the logarithmic transformation $-log_{10}$. Two markers associated with HRMP were found, one on chromosome 1 (M2216) and another on chromosome 4 (M13512). For drought treatment, four associated markers were found, two on chromosome 1 (M232 and M3547), one on chromosome 3 (M11032), and the last on chromosome 8 (M26026). For the well-watered treatment, four markers were significant, on chromosomes 1 (M3299), 2 (M7826), and a two on chromosome 7 (M21270 and M22252) (Figure 9). The minor allele frequency of these 10 SNPs ranged from 5.20% (M26036) to 31.24% (M7826). The significant SNPs together explain just 0.04% of the drought tolerance index phenotypic variation (r^2), and 0.07% and 0.08% of the grain yield phenotypic variation under drought and well-watered treatments (Table 6).

The SNP M7826, associated with GYWW, was the only marker (among the ten) that did not have any individual in one of the homozygous classes (Figure 11). The alternative homozygous at M13512 showed a slight increase in the HMRP. In general, the heterozygous class tended to be between the homozygous classes, following a trend of additive substitution effect, except to the SNPs M3547 and M3299, where the heterozygous class had a higher grain yield (in DR and WW respectively) than the homozygous classes. The SNPs M26036 and M11032, associated with GYDR, suggest a higher GY when homozygous for the alternative allele. Conversely, the SNPs M232 and M3547 showed lower yield when homozygous for the alternative allele. When homozygous for the alternative allele, the SNP M21270 is less productive in well-watered treatments. In contrast, the SNP M22252 shows an opposite direction, higher yields when homozygous to the alternative.

3.3.3. Validation of significant markers over irrigation treatments and generations

The significant markers found using in this research were not stable across traits in the first generation (BC1S1). The two SNPs associated with the drought tolerance index were not

associated with grain yield. Likewise, none of the SNPs associated with grain yield were in both irrigation treatments. Furthermore, none of the 10 SNPs found by GWAS in BC1S1-TC1 were significant in the BC1S2-TC2 GWAS (Table 6 and Table 7). The BC1S2-TC2 GWAS had assigned 3 SNPs for drought tolerance index and 3 for grain yield under drought treatments and 5 for grain yield in well-watered treatments. In the second generation, one SNP was stable across the traits. The SNP M25973 was associated with the drought tolerance index and with the grain yield in both irrigation treatments, although it was not detected in the first generation.

The correlation between the predicted values of BC1S2, estimated based on the effects of allelic substitution of significant SNPs obtained from BC1S1 GWAS, and the observed BLUPs was -0.11 for the drought tolerance index, -0.20 for grain yield under drought treatments and -0.23 under well-watered.

3.3.4. Gene annotation

Putative genes for drought tolerance were identified by searching the B73 genome in the regions around the ten significant SNPs identified in the analysis (Table 8). The search was performed inside a 100k bp window up and down-stream of the markers. The linkage disequilibrium (LD) decay estimative from the BC1S1 data can be observed by plotting the r^2 against the distance of base pairs (Figure 12), a rapid decay occurs while increasing the chromosome distance between two loci for all the chromosomes. The overall LD decay extent was in a range of 50k-300k bp and just for the chromosome number 3 the r^2 does not decrease to less than 0.13 within 100k bp. A more abrupt decrease of the LD was observed in chromosome 8, and in contrast, a smoother decay was observed in chromosome 3, all the others present a similar decay pattern.

The SNP M2216 is 60k bp upstream from a predicted gene that is still not characterized. M13512 is in a region of chromosome four that has nine putative protein coding sequences. The closest gene to M232 SNP is located 85k bp upstream from it, and other three putative genes are less than 60k bp away. M3547 SNP is in an intron of a putative gene and there are five more probable genes inside its search window. M11032 is also in an intron of an uncharacterized putative gene, with eight more putative genes around it. The SNP M26036 is in a region with 14 possible genes. M3299 is in an exon of an uncharacterized predicted gene and with two more in its search window. M7826 is 1.5k bp only from a putative gene, and in less than 50k bp there are seven more predicted genes. The SNP M21270 is on a chromosome region with 20 putative genes. Moreover, in the M22252 window, we could find six predicted genes. It is clear that a

window of 100k bp can host a large number of genes, and possibly many more than those listed here, as annotation and gene discovery is a work in process.

3.3.5. Favorable alleles and their origin

The frequency of landrace alleles in the BC1S1 lines ranges from 5% to 31%. It should be noted that these are the same value as the presented MAF in Table 6, as expected once the landraces were the donor parent in the backcross and so the landrace allele is the one in minor frequency. The SNPs M3547 and M7826 present the highest frequencies for the landrace allele among the ten significant SNPs (Figure 13). Moreover, the superior allele (favorable) of four SNPs (M3547, M11032, M26036, and M3299) came from the landraces. Furthermore, five landraces populations that present a higher frequency of the superior allele for these 4 SNPs were those originated from NVOL46, COAH20, COAH21, CHIH87, and COAH78 populations (Figure 11**Error! Reference source not found.**). This suggests that these populations have a higher frequency of favorable alleles, being a good source for drought tolerance. However, regardless of the origin of the superior allele, the five landrace populations that present higher frequency of superior allele were obtained from SNLP17, NVOL46, COAH78 CHIH85, and COAH20.

3.4. DISCUSSION

3.4.1. Field trials and phenotypes

All the field trials presented strong spatial (row and columns) trends. Therefore, the spatial corrections used in the first stage was crucial to avoid, or minimize, their effects on the means, mainly in the WW trials, which were not replicated. Thus, we expected lower heritabilities in these trials (WW) than that from DR, as happened, even knowing that such drought-stressed trials usually are more affected by spatial trends of field variations (Badu-Apraku et al. 2004), what eventually results in low heritability for yield. That was one of the reasons we chose to use two repetitions on the reduced irrigation trial. However, the heritability also depends on the genetic variance, and one argument to explain the higher heritability under DR treatments could be that the drought might, in a diverse population, discriminate between the genotypes better than when it is tested in a watered and more homogeneous treatment.

As the TC1 used a single tester, the tester by genotype interaction could not be separated from the estimated genotypic value (BLUEs and BLUPs). Thus, the first generation (BC1S1) data are biased by this interaction, which is acceptable, since it would not be feasible to

perform a testcross trial with more than 1,000 genotypes crossed to a group of testers. In the second generation (BC1S2) this was minimized, as fewer (170) genotypes were tested, allowing a testcross with more testers (three).

Comparing the grain yield means and the drought tolerance index of the BC1 progeny testcrosses and the genetic checks (Tester/CML376), similar values were found. However, the distribution suggests a superior performance of BC1 progenies (Figure 4), and the same for the drought tolerance index (Figure 5). It can infer that there are novel alleles originating from the landrace donor parent that contribute to drought tolerance and/or a positive interaction between genes from the different sources, landrace and CML.

The Harmonic Mean Relative Performance was confirmed as a simple and efficient index that allows the selection of genotypes that perform well in both irrigation treatments, achieving the objective of these "stressed x not stressed" trials. The HMRP selects the highest performing genotypes in both treatments (Figure 7, upper right) and, depending on the selection intensity, also those that perform very well in one and close to the average in the another, but does not select those that under perform in at least one of the treatments. Sites and genotype by site interaction effects were significant in regard to the index (Table 3). Therefore, even if the means across the sites are similar, differential expression of the drought tolerance was identified, and the selection based on results by location should be avoided.

3.4.2. Genotyping

We believe that the final number of SNPs (5,695) was relatively low, however it shows a spread distribution across the genome, what is positive for this kind of study. This number of markers may be due to the genetic mating design of the population. Crossing and backcrossing to a homozygous line (CML376) could dilute allele frequency. It could also be a reason for the restricted population structure (estimated by the PCA), even though the lines were originated from 20 different landrace parents, and also for the low MAF of the SNPs associated to the traits (Table 6). Another important details that should be considered is that the reference genome used to this study (B73 RefGen_v4) does not fit non-temperate germplasm and certainly is not ideal for working with landrace derived materials, however, it is the best reference genome that can be used nowadays.

3.4.3. GWAS

Many researchers use correlated traits to study drought tolerance such as seedling survival rate under drought (Mao et al. 2015), ear length and kernel number per row (Lu et al.

2006), and one of the most used, Anthesis-Silk Interval (ASI) (Yu 1994; Bolaños and Edmeades 1996; Vargas et al. 2006; Wang and Qin 2017). This is due to the low heritability of grain yield compared with the correlated traits. Xue et al. (2013) using a diversity panel of CIMMYT lines and studying nine traits related to drought tolerance found 42 SNPs associated with those traits, any one of the SNPs listed by the authors was detected by our study as significant for the HRMP index. Setter et al. (2011) in an association mapping study of abscisic acid levels during drought stress in maize found 8 polymorphisms related with drought tolerance. One of those polymorphisms was used to the annotation of the maize gene Zm00001d034385, a validated gene associated to drought tolerance. It is mapped 24M bp from the M3547 SNP in the present study which is associated with GYDR., Although it is on the same chromosome arm (chromosome 1, long arm), that is a considerable long distance away, considering a LD decay of ~100k bp.

The LD decay reported in maize varies considerably among the populations (Flint-Garcia et al. 2003). In landraces or a highly diverse panel, the LD decay is reported to be smaller than in elite inbred lines, by a magnitude of 1k bp for landraces and ~500k bp for elite inbred lines (Yan et al. 2009). Additionally, the average of LD decay distance in tropical germplasm (~10k bp) is shorter than in temperate germplasm (~100k bp) (Lu et al. 2011). The populations used in this study was at the midpoint as the initial sub-tropical landraces were the donor parent in a backcross where an elite inbred line was the recurrent parent.

We focused our search for genes in a window of 100k bp around the 10 trait-associated SNPs. The LD extent for our population (50- 300k bp) is in accordance with those observed in the literature. This is the expectation for a population that is neither a set of landraces nor elite inbred lines, since the former normally has a shorter LD decay than evidenced in our experimental lines and the latter, a larger LD decay (Remington et al. 2001; Flint-Garcia et al. 2003). Except for two cases, our genome search found putative genes not completely described. These two cases are described below: SNP M232 associated with GYDR is located 85k bp close to the gene Zm00001d027539 that belongs to the glutathione S-transferase (GST) gene family (GST, EC 2.5.1.18). It plays an important role in plant response to environmental conditions (McGonigle et al. 2002) and some abiotic stresses such as herbicide (Cataneo et al. 2003), heat (Cetinkaya et al. 2014) and drought (Xu et al. 2015; Min et al. 2016). In Arabidopsis, plants with mutated GST gene were more tolerant to drought stress than wild-type plants (Chen et al. 2012). Additionally, there is the predicted gene Zm00001d043929 associated with GYDR in the present study and mapped to ~80k bp upstream from the M11032 SNP. This gene encodes a protein with a fasciclin-like domain (FAS1) and this domain is found in Fasciclin-like arabinogalactan (FLAs) proteins that are involved in cell adhesion. Also, some FLAs have their expression

regulated by the phytohormone abscisic acid (ABA) (Johnson et al. 2003). Both FLAs and ABA play an important role in plant response to stress (Faik et al. 2006; Finkelstein 2013; Zang et al. 2015). Wang et al. (2016) points out that the participation of FLAs in enhancing cell wall synthesis in response to drought stress, which could minimize water loss and cell dehydration. Reduction of some FLA gene expressions is also associated with kernel abortion in maize (Cagnola et al. 2018) what represents a reduction of kernel numbers thus lowering yield.

These two findings are promising for further maize breeding studies. In our experimental lines, we detected markers within ~80k bp of these genes. Thus, fine-mapping and allele variant studies in these specific *loci* could help to elucidate its role and could be included in a genomic selection set. These two markers (M232 and M11032) could be used in our experimental lines to drive the selection in the next generations and be a base for these further investigations. This is especially so for the M11032 SNP, where the alternative allele originates from the landrace (Figure 11) and the homozygous state for this allele had a slightly higher grain yield in drought treatments. The landraces ARZM12193, CHIH338, CHIH87, COAH19 are not good sources for all the superior alleles, since the frequency of M11032 is null in these populations.

There were three more significant SNPs (M3547, M26036, and M3299) where the superior allele is from the landraces, which demonstrates new diversity for drought tolerance alleles in the experimental lines. In this context, the landraces NVOL46, COAH20, COAH21, CHIH87, and COAH78 are the main sources for exploring this diversity since they have a higher frequency of these favorable alleles.

3.4.4. Validation and significant markers over irrigation treatment and generations

For the BC1S1 testcrosses, the lack of significant SNP correlation (?) among the irrigation treatments may be a sign of differential expression of genes under contrasting water environments. For example, Cagnola et al. (2018), demonstrated that drought stress could induce a reduction in FLA gene expressions, which may cause kernel abortion in maize resulting in reduced grain yield. As discussed above, there is a putative gene inside the SNP M11032 window that belongs to this gene family.

Likewise, there was no correlation of significant SNPs between the BC1S1 and BC1S2 generations. However, selection was made in BC1S1 to generate the BC1S2. In this way, we induced a nonrandom allele frequency change between the generations, it can affect the results as the new population is diverging from the first (Henshall 2013). Thus, in our study, the use of the BC1S2 generation to validate the GWAS results of BC1S1 generation was not effective.

Additionally, the use of allelic substitution effects estimated using the BC1S1 data to predict the phenotype of BC1S2 generation as a validation method was not successful. The correlations between the predicted values and the observed BLUPs were low and negative to the three cases (-0.11 for the drought tolerance index, -0.20 for GYDR treatments and -0.23 for GYWW). The allele frequency also interferes with the average effect of the allele. Therefore, the selection and its incumbent frequency allele changes, also results in allelic effect deviations estimated in BC1S1, which could not be constant through generations of selection. Moreover, the heritabilities were low in BC1S1 (0.04 to 0.17), and the significant SNPs explain less than 0.1% of phenotypic variation, that is indicative of the complexity of quantitative traits as drought tolerance and grain yield, that reduces the efficiency of using a few SNPs for prediction, even if they are significant. Finally, there is the effect of the tester on the BC1S1 phenotypic data, since the interaction tester by landrace progenies could not be estimated in this case. Thus, the allelic substitution effect is biased for this specific testcross and use it to predict the genotypic values of a second testcross, that involves another tester, may be unfeasible.

Early testcross is a strategy to take advantage of early selection for yield and other traits that avoids excessive resource spending on unpromising families, and it can also be applied in a pre-breeding program. However, as Bernardo (2010) concludes, the low heritability limits its effectiveness. The GWAS in early generations achieved promising results and was demonstrated to be a useful tool for identifying new genetic variants in our landrace population. Nevertheless, a practical and effective method to validate a GWAS in this situation is still needs to be developed.

3.5. CONCLUSION

The genome-wide association study (GWAS) was able to identify genomic regions in early testcrosses associated with drought tolerance, although they were not stable over generations and irrigation treatments. Furthermore, two promising putative genes that encode proteins associated with physiological plant response to abiotic stress are located close to the significant SNPs found in this study. Additionally, some alleles from the landraces provide a slightly higher yield under drought treatments. Thus, these results indicate that the diversity delivered from landraces x elite inbred line crosses is an exciting breeding strategy for improving maize for drought tolerance and for trait introgression bringing new superior allelic diversity from landraces to breeding populations.

REFERENCES

Arteaga, M.C., Moreno-Letelier, A., Mastretta-Yanes, A., Vázquez-Lobo, A., Breña-Ochoa, A., Moreno-Estrada, A., Eguiarte, L.E., Piñero, D., 2016. Genomic variation in recently collected maize landraces from Mexico. Genomics Data 7, 38–45. https://doi.org/10.1016/j.gdata.2015.11.002

Badu-Apraku, B., Fakorede, M.A.B., Menkir, A., Kamara, A.Y., Adam, A., 2004. Effects of drought screening methodology on genetic variances and covariances in Pool 16 DT maize population. J. Agric. Sci. 142, 445–452. https://doi.org/10.1017/S0021859604004538

Bänziger, M., Edmeades, G.O., Beck, D., Bellon, M., 2000. Breeding for Drought and Nitrogen Stress Tolerance in Maize: From Theory to Practice. CIMMYT Rep. 68. https://doi.org/633.1553

Bernardo, R., 2010. Breeding for quantitative traits in plants, 2nd ed. Stemma press, Woodbury.

Bolaños, J., Edmeades, G.O., 1996. The importance of the anthesis-silking interval in breeding for drought tolerance in tropical maize. F. Crop. Res. 48, 65–80. https://doi.org/10.1016/0378-4290(96)00036-6

Burgueño, J., Cadena, A., Crossa, J., Bänziger, M., Gilmour, A.R., Cullis, B., 2000. Users's guide for spatial analysis of field variety trials using ASREML. CIMMYT.

Butler, D., 2009. ASREML R package: asreml() fits the linear model.

Cagnola, J.I., Dumont de Chassart, G.J., Ibarra, S.E., Chimenti, C., Ricardi, M.M., Delzer, B., Ghiglione, H., Zhu, T., Otegui, M.E., Estevez, J.M., Casal, J.J., 2018. Reduced expression of selected FASCICLIN-LIKE ARABINOGALACTAN PROTEIN genes associates with the abortion of kernels in field crops of Zea mays (maize) and of Arabidopsis seeds. Plant Cell Environ. 41, 661–674. https://doi.org/10.1111/pce.13136

Campos, H., Cooper, M., Habben, J.E., Edmeades, G.O., Schussler, J.R., 2004. Improving drought tolerance in maize: A view from industry. F. Crop. Res. 90, 19–34. https://doi.org/10.1016/j.fcr.2004.07.003

Cataneo, A.C., Déstro, G.F.G., Ferreira, L.C., Chamma, K.L., Sousa, D.C.F., 2003. Atividade de glutationa S-transferase na degradação do herbicida glyphosate em plantas de milho (Zea mays). Planta Daninha 21, 307–312. https://doi.org/10.1590/S0100-83582003000200017

Cetinkaya, H., Tasci, E., Dinler, B.S., 2014. Regulation of glutathione S -transferase enzyme activity with salt pre-treatment under heat stress in maize leaves 4, 45–56.

Chen, J.-H., Jiang, H.-W., Hsieh, E.-J., Chen, H.-Y., Chien, C.-T., Hsieh, H.-L., Lin, T.-P., 2012. Drought and Salt Stress Tolerance of an Arabidopsis Glutathione S-Transferase U17 Knockout Mutant Are Attributed to the Combined Effect of Glutathione and Abscisic Acid. Plant Physiol. 158, 340–351. https://doi.org/10.1104/pp.111.181875

Dutkowski, G.W., Silva, J.C. e, Gilmour, A.R., Lopez, G.A., 2002. Spatial analysis methods for forest genetic trials. Can. J. For. Res. 32, 2201–2214. https://doi.org/10.1139/x02-111

Edmeades, G.O., Trevisan, W., Prasanna, B.M., Campos, H., 2017. Tropical Maize (Zea mays L.), in: Campos, H., Caligari, P.D.S. (Eds.), Genetic Improvement of Tropical Crops. Springer International Publishing, Cham, p. 331. https://doi.org/10.1007/978-3-319-59819-2

Faik, A., Abouzouhair, J., Sarhan, F., 2006. Putative fasciclin-like arabinogalactan-proteins (FLA) in wheat (Triticum aestivum) and rice (Oryza sativa): Identification and bioinformatic analyses. Mol. Genet. Genomics 276, 478–494. https://doi.org/10.1007/s00438-006-0159-z

FAOSTAT, 2017. Database of Food and Agriculture Organization of the United Nations Statistics Division [WWW Document]. URL http://www.fao.org/faostat (accessed 5.5.19).

Finkelstein, R., 2013. Abscisic Acid Synthesis and Response. Arab. B. 11, e0166. https://doi.org/10.1199/tab.0166

Fisher, M., Abate, T., Lunduka, R.W., Asnake, W., Alemayehu, Y., Madulu, R.B., 2015. Drought tolerant maize for farmer adaptation to drought in sub-Saharan Africa: Determinants of adoption in eastern and southern Africa. Clim. Change 133, 283–299. https://doi.org/10.1007/s10584-015-1459-2

Flint-Garcia, S.A., Thornsberry, J.M., Buckler, E.S., 2003. Structure of linkage disequilibrium in plants. Annu. Rev. Plant Biol. 54, 357–74. https://doi.org/10.1146/annurev.arplant.54.031902.134907

Fritsche-Neto, R., Souza, T.L.P.O. de, Pereira, H.S., Faria, L.C. de, Melo, L.C., Novaes, E., Brum, I.J.B., Jannink, J.-L., 2019. Association mapping in common bean revealed regions associated with Anthracnose and Angular Leaf Spot resistance. Sci. Agric. 76, 321–327. https://doi.org/10.1590/1678-992x-2017-0306

Gilmour, A.R., Cullis, B.R., Verbyla, Arūnas P, Journal, S., Statistics, E., Gilmour, A.R., Cullis, B.R., Verbyla, Artunas P, 1997. Accounting for Natural and Extraneous Variation in the Analysis of Field Experiments Published by: International Biometric Society Stable URL: http://www.jstor.org/stable/1400446 Your use of the JSTOR archive indicates your acceptance of the Terms & Con 2, 269–293.

Goodman, M.M., Holland, J.B., Sanchez Gonzalez, J.J., 2014. Breeding and diversity., in: Wusirika, R., Bohn, M., Lai, J., Kole, C. (Eds.), Genetics, Genomics and Breeding of Maize. CRC Press, New York, NY, pp. 15–40. https://doi.org/10.1201/b17274

Granato, I., Fritsche-Neto, R., 2018. snpReady: Preparing Genotypic Datasets in Order to Run Genomic Analysis.

Henshall, J.M., 2013. Genome-Wide Association Studies and Genomic Prediction 1019, 411–421. https://doi.org/10.1007/978-1-62703-447-0

Hill, W.G., Weir, B.S., 1988. Variances and covariances of squared linkage disequilibria in finite populations. Theor. Popul. Biol. https://doi.org/10.1016/0040-5809(88)90004-4

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., Zhu, C., Lu, T., Zhang, Z., Li, M., Fan, D., Guo, Y., Wang, A., Wang, L., Deng, L., Li, Wenjun, Lu, Y., Weng, Q., Liu, K., Huang, T., Zhou, T., Jing, Y., Li, Wei, Lin, Z., Buckler, E.S., Qian, Q., Zhang, Q.F., Li, J., Han, B., 2010. Genome-wide asociation studies of 14 agronomic traits in rice landraces. Nat. Genet. 42, 961–967. https://doi.org/10.1038/ng.695

Hwang, E.-Y., Song, Q., Jia, G., Specht, J.E., Hyten, D.L., Costa, J., Cregan, P.B., 2014. A genome-wide association study of seed protein and oil content in soybean. BMC Genomics 15, 1. https://doi.org/10.1186/1471-2164-15-1

IPCC, 2014. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Geneva, Switzerland.

IPCC, 2007. Summary for Policymakers. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, United Kingdom. https://doi.org/10.1038/446727a

Jaiswal, V., Gupta, S., Gahlaut, V., Muthamilarasan, M., Bandyopadhyay, T., Ramchiary, N., Prasad, M., 2019. Genome-Wide Association Study of Major Agronomic Traits in Foxtail Millet (Setaria italica L.) Using ddRAD Sequencing. Sci. Rep. 9, 1–11. https://doi.org/10.1038/s41598-019-41602-6

Johnson, K.L., Jones, B.J., Bacic, A., Schultz, C.J., 2003. The Fasciclin-Like Arabinogalactan Proteins of Arabidopsis. A Multigene Family of Putative Cell Adhesion Molecules. Plant Physiol. 133, 1911–1925. https://doi.org/10.1104/pp.103.031237.kinases

Liu, X., Huang, M., Fan, B., Buckler, E.S., Zhang, Z., 2016. Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. PLOS Genet. 12, e1005767. https://doi.org/10.1371/journal.pgen.1005767

Lu, G.H., Tang, J.H., Yan, J.B., Ma, X.Q., Li, J.S., Chen, S.J., Ma, J.C., Liu, Z.X., E, L.Z., Zhang, Y.R., Dai, J.R., 2006. Quantitative trait loci mapping of maize yield and its components under different water treatments at flowering time. J. Integr. Plant Biol. 48, 1233–1243. https://doi.org/10.1111/j.1744-7909.2006.00289.x

Lu, Y., Shah, T., Hao, Z., Taba, S., Zhang, S., Gao, S., Liu, J., Cao, M., Wang, J., Prakash, A.B., Rong, T., Xu, Y., 2011. Comparative SNP and haplotype analysis reveals a higher genetic diversity and rapider LD decay in tropical than temperate germplasm in maize. PLoS One 6. https://doi.org/10.1371/journal.pone.0024861

Mao, H., Wang, H., Liu, S., Li, Z., Yang, X., Yan, J., Li, J., Tran, L.S.P., Qin, F., 2015. A transposable element in a NAC gene is associated with drought tolerance in maize seedlings. Nat. Commun. 6, 1–13. https://doi.org/10.1038/ncomms9326

McGonigle, B., Keeler, S.J., Lau, S.-M.C., Koeppe, M.K., O'Keefe, D.P., 2002. A Genomics Approach to the Comprehensive Analysis of the Glutathione S -Transferase Gene Family in Soybean and Maize. Plant Physiol. 124, 1105–1120. https://doi.org/10.1104/pp.124.3.1105

Meseka, S., Menkir, A., Obeng-Antwi, K., 2015. Exploitation of beneficial alleles from maize (Zea mays L.) landraces to enhance performance of an elite variety in water stress environments. Euphytica 201, 149–160. https://doi.org/10.1007/s10681-014-1214-1

Min, H., Chen, C., Wei, S., Shang, X., Sun, M., Xia, R., Liu, X., Hao, D., Chen, H., Xie, Q., 2016. Identification of Drought Tolerant Mechanisms in Maize Seedlings Based on Transcriptome Analysis of Recombination Inbred Lines. Front. Plant Sci. 7. https://doi.org/10.3389/fpls.2016.01080

Miti, F., Pangirayi, T., Derera, J., 2010. S1 selection of local maize landraces for low soil nitrogen tolerance in Zambia. African J. Plant Sci.

Navarro, J.A.R., Willcox, M., Burgueño, J., Romay, C., Swarts, K., Trachsel, S., Preciado, E., Terron, A., Delgado, H.V., Vidal, V., Ortega, A., Banda, A.E., Orlando, N., Montiel, G., Ortiz-monasterio, I., Vicente, F.S., Espinoza, A.G., Atlin, G., Wenzl, P., Hearne, S., Buckler, E.S., 2017. A study of allelic diversity underlying flowering-time adaptation in maize landraces. Nat. Publ. Gr. 49, 476–480. https://doi.org/10.1038/ng.3784

Portwood, J.L., Woodhouse, M.R., Cannon, E.K., Gardiner, J.M., Harper, L.C., Schaeffer, M.L., Walsh, J.R., Sen, T.Z., Cho, K.T., Schott, D.A., Braun, B.L., Dietze, M., Dunfee, B., Elsik, C.G., Manchanda, N., Coe, E., Sachs, M., Stinard, P., Tolbert, J., Zimmerman, S., Andorf, C.M., 2019. MaizeGDB 2018: the maize multi-genome genetics and genomics database. Nucleic Acids Res. 47, D1146–D1154. https://doi.org/10.1093/nar/gky1046

Prasanna, B.M., 2012. Diversity in global maize germplasm: Characterization and utilization. J. Biosci. 37, 843–855. https://doi.org/10.1007/s12038-012-9227-1

R Core Team, 2019. R: A Language and Environment for Statistical Computing.

Remington, D.L., Thornsberry, J.M., Matsuoka, Y., Wilson, L.M., Whitt, S.R., Doebley, J., Kresovich, S., Goodman, M.M., Buckler, E.S., 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc. Natl. Acad. Sci. 98, 11479–11484. https://doi.org/10.1073/pnas.201394398

Rengel, D., Arribat, S., Maury, P., Martin-Magniette, M.-L., Hourlier, T., Laporte, M., Varès, D., Carrère, S., Grieu, P., Balzergue, S., Gouzy, J., Vincourt, P., Langlade, N.B., 2012. A gene-phenotype network based on genetic variability for drought responses reveals key physiological processes in controlled and natural environments. PLoS One 7, e45249. https://doi.org/10.1371/journal.pone.0045249

Resende, M.D.V., 2004. Métodos Estatísticos ótimos na análise de experimentos de campo. Documentos (Embrapa) 100, 65.

Sanchez, J.J., Goodman, M.M., Stuber, C.W., 2000. Isozymatic and morphological diversity in the races of maize of Mexico. Econ. Bot. 54, 43–59. https://doi.org/10.1007/BF02866599

Schafer, J.F., 1971. Tolerance to plant disease. Annu. Rev. Phytopathol. 9, 235–252. https://doi.org/10.1146/annurev.py.09.090171.001315

Setter, T.L., Yan, J., Warburton, M., Ribaut, J.M., Xu, Y., Sawkins, M., Buckler, E.S., Zhang, Z., Gore, M.A., 2011. Genetic association mapping identifies single nucleotide polymorphisms in genes that affect abscisic acid levels in maize floral tissues during drought. J. Exp. Bot. 62, 701–716. https://doi.org/10.1093/jxb/erq308

VanRaden, P.M., 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91, 4414–23. https://doi.org/10.3168/jds.2007-0980

Vargas, M., Van Eeuwijk, F.A., Crossa, J., Ribaut, J.M., 2006. Mapping QTLs and QTL x environment interaction for CIMMYT maize drought stress program using factorial regression and partial least squares methods. Theor. Appl. Genet. 112, 1009–1023. https://doi.org/10.1007/s00122-005-0204-z

Wang, H., Qin, F., 2017. Genome-Wide Association Study Reveals Natural Variations Contributing to Drought Resistance in Crops. Front. Plant Sci. 8, 1–12. https://doi.org/10.3389/fpls.2017.01110

Wang, X., Cai, X., Xu, C., Wang, Q., Dai, S., 2016. Drought-responsive mechanisms in plant leaves revealed by proteomics. Int. J. Mol. Sci. 17. https://doi.org/10.3390/ijms17101706

Wimmer, V., Albrecht, T., Auinger, H.-J., Schoen, C.-C., 2012. synbreed: a framework for the analysis of genomic prediction data using R. Bioinformatics 28, 2086–2087.

Xiao, Y., Liu, H., Wu, L., Warburton, M., Yan, J., 2017. Genome-wide Association Studies in Maize: Praise and Stargaze. Mol. Plant 10, 359–374. https://doi.org/10.1016/j.molp.2016.12.008

Xu, J., Xing, X.J., Tian, Y.S., Peng, R.H., Xue, Y., Zhao, W., Yao, Q.H., Zhang, H., 2015. Transgenic Arabidopsis plants expressing tomato glutathione S-transferase showed enhanced resistance to salt and drought stress. PLoS One 10, 1–16. https://doi.org/10.1371/journal.pone.0136960

Xu, Y., Li, P., Zou, C., Lu, Y., Xie, C., Zhang, X., Prasanna, B.M., Olsen, M.S., 2017. Enhancing genetic gain in the era of molecular breeding. J. Exp. Bot. 68, 2641–2666. https://doi.org/10.1093/jxb/erx135

Xue, Y., Warburton, M.L., Sawkins, M., Zhang, X.X., Setter, T., Xu, Y., Grudloyma, P., Gethi, J., Ribaut, J.-M.J., Li, W., Zhang, X.X., Zheng, Y., Yan, J., 2013. Genome-wide association analysis for nine agronomic traits in maize under well-watered and water-stressed conditions. Theor. Appl. Genet. 126, 2587–2596. https://doi.org/10.1007/s00122-013-2158-x

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

Yu, Y.G., 1994. RFLP and Microsatellite Mapping of a Gene for Soybean Mosaic Virus Resistance. Phytopathology 84, 60. https://doi.org/10.1094/Phyto-84-60

Zang, L., Zheng, T., Chu, Y., Ding, C., Zhang, W., Huang, Q., Su, X., 2015. Genome-Wide Analysis of the Fasciclin-Like Arabinogalactan Protein Gene Family Reveals Differential Expression Patterns, Localization, and Salt Stress Response in Populus. Front. Plant Sci. 6, 1–16. https://doi.org/10.3389/fpls.2015.01140

TABLES AND FIGURES

Table 2: 20 Landraces from CIMMYT germplasm bank used as a source for the breeding populations studied in this research, number of progenies of each landrace evaluated in generation 1 (backcross1, selfing 1 - BC1S1) and generation 2 (backcross 1, selfing 2- BC1S2). The rank of allele frequency of superior alleles considering the 10 significant SNPs (General) and the 4 SNPs with superior allele originated from landraces (Landrace origin), 1 is the first position, i.e., with the highest frequency and 20 the last, with the lowest. Highlighted the top five landraces for each rank of superior alleles frequency.

Landrace	Number of progenies		Rank for superior allele frequency	
	BC1S1	BC1S2	General	Landrace origin
ARZM12193	68	22	17	20
ARZM12236	70	0	6	14
ARZM12237	71	12	12	16
ARZM12240	70	11	15	12
CHIH333	60	6	14	13
CHIH338	73	8	19	19
CHIH59	33	2	13	11
CHIH85	68	0	4	9
CHIH87	74	11	6	4
COAH117	68	17	9	6
COAH18	75	4	16	7
COAH19	62	2	20	17
COAH20	71	5	5	3
COAH21	68	14	11	2
COAH78	66	6	3	5
NVOL19	58	7	18	15
NVOL46	70	15	2	1
SNLP166	66	5	8	18
SNLP169	71	17	10	10
SNLP17	64	10	1	8
Total	1326	174		

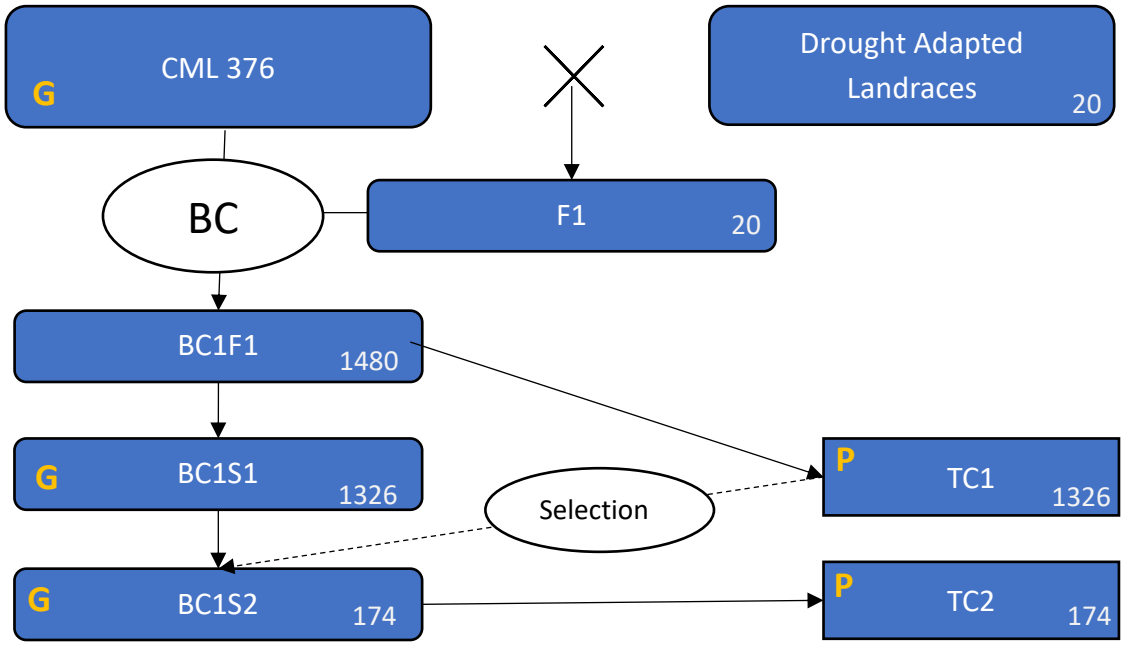

Figure 3: Scheme of the plant material formation showing the populations, the crosses, and the testcrosses. BC is the backcross, TC1 and TC2 are the two testcross trials. The X means the first cross between the 20 landrace populations and the CML376. The ⊗ means a selfing. The yellow G means that the population was genotyped, and P means the phenotyped. The number in the lower left corner is the number of genotypes, or populations in the case of Landraces and F1. The "Selection" means where the selection was applied.

Table 3: Wald statistic for fixed effects and variance components for random effects to the joint analysis (second-stage) for the three traits, grain yield under well-watered (GY WW) and drought (GY DR) and the drought tolerance index (HMRP). The accuracy (Acc%), coefficients of genetic (CVg) and error (CVe) variation and heritability are also presented.

Factor	GY WW		GY DR		HMRP	
Fixed effects (Wald)						
Location	364	***	201	***	8	*
Random effects (LRT)						
Genotype	0.03	**	0.03	+	0.0007	+
Genotype x Site	0.58	***	0.00	ns	0.03	***
Acc. (%)	19.9		25.5		20.1	
CVg	2.98		6.14		3.64	
CVe	20.76		32.85		25.13	
Heritability	0.04		0.17		0.06	

ns, not significant; + p<0.2; * p<0.05; ** p<0.01; *** p<0.001

Table 4: Factors used in each model of first-stage spatial analysis (individual experiments). Decisions about the use or not of spatial factors on the models were taken based on variograms plots. LM, SI and TL are, respectively, the locations Los Mochis, Santiago del Ixcutlia, and Tlaltizápan. WW and DR are the irrigation treatments, well-watered and drought conditions

Fixed spatial factors	Experiment (Site/Water Management)
Row	SI/WW, SI/DR, TL/WW, LM/WW
Column	LM/DR
Row + Column	TL/DR

Table 5: Minimum, mean, maximum and variance of drought tolerance index for each site (LM, Los Mochis; TL, Taltizapan and SI, Santiago del Ixcuintla) and overall sites for the generation BC1S1 and BC1S2

	BC1S1				BC1S2			
	Min.	Mean	Max.	Var	Min.	Mean	Max.	Var
Overall	0.320	0.983	1.710	0.146	0.745	0,998	1.308	0.077
LM	0.389	0.984	1.627	0.143	0.546	0,994	1.459	0.141
TL	0.196	0.967	1.635	0.207	0.794	0,998	1.337	0.096
SI	0.303	0.966	1.551	0.198	-	-	-	-
PV	-	-	-	-	0.637	0.992	1.278	0.096

Table 6: Significant SNPs for drought-tolerance index and grain yield in drought and well-watered conditions, with the SNP position (chromosome and position), polymorphism (PM) of the SNP (Reference > Alternative), genotype of CML376 (CML) for the given SNP, minor frequency allele, effect of allele substitution, heritability, and r^2 for the GWAS of BC1S1 generation

Trait	SNP	Chr	Position	PM	CML	MAF	Effect	h^2	r^2
HMRP	M2216	1	171,883,409	G>C	GG	7.31%	-0.020	0.014	0.00020
	M13512	4	180,305,244	C>A	AA	9.37%	0.018	0.014	0.00019
GY / DR	M232	1	7,760,066	A>G	AA	6.47%	-0.059	0.017	0.00027
	M3547	1	261,017,573	A>G	AA	22.12%	0.035	0.016	0.00026
	M11032	3	214,051,229	G>A	GG	8.00%	0.040	0.009	0.00087
	M26036	8	178,731,659	A>T	AA	5.20%	0.052	0.011	0.00011
GY / WW	M3299	1	243,650,161	G>C	GG	10.52%	0.052	0.009	0.00009
	M7826	2	227,370,542	G>T	TT	31.24%	0.051	0.021	0.00044
	M21270	7	46,333,211	A>G	AA	7.31%	-0.064	0.011	0.00011
	M22252	7	149,023,564	G>T	TT	5.55%	0.084	0.013	0.00019

Table 7: Significant SNPs for drought-tolerance index and grain yield in drought and well-watered conditions, with the SNP position (chromosome and position), minor frequency allele, effect, heritability, and r^2 for the GWAS of BC1S2 generation

Trait	SNP	Chr	Position	MAF	Effect	h^2	r^2
HMRP	M19100	5	221,581,023	7.35%	-0.028	0.039	0.0015
	M25973	8	128,738,707	7.06%	0.030	0.042	0.0018
	M29154	9	133,967,900	6.47%	-0.031	0.041	0.0017
GY / DR	M11800	3	222,428,531	7.06%	-0.137	0.027	0.0007
	M19017	5	220,204,698	7.94%	0.155	0.039	0.0015
	M25973	8	128,738,707	7.06%	0.188	0.051	0.0026
GY / WW	M5524	2	21,392,902	11.76%	0.350	0.056	0.0031
	M21806	7	7,471,998	5.88%	0.358	0.031	0.0010
	M25973	8	128,738,707	7.06%	0.366	0.038	0.0015
	M30107	10	3,024,706	14.71%	-0.216	0.026	0.0007
	M30193	10	5,754,305	5.59%	-0.485	0.054	0.0029

Table 8: Description of SNPs, their positions, number of genes (putative and described) found in a search window of 100k bp up and downstream from the mark positions. The most described genes founded in each search and the protein type/family that these most described genes encode

SNP	Chr	Position	Number of finds	Most described promising finds	Protein Type/Family
M2216	1	171.883.409	1	Zm00001d030982	Ankyrin repeat family protein
M13512	4	180.305.244	9	Zm00001d052123	Thioredoxin-like 5
M232	1	7.760.066	4	Zm00001d027539	Glutathione transferase 1*
M3547	1	261.017.573	6	Zm00001d033375	Golgi-body localization protein domain
M11032	3	214.051.229	9	Zm00001d043929	Fasciclin-like arabinogalactan*
M26036	8	178.731.659	14	Zm00001d012693	Transmembrane transport protein
				Zm00001d012699	Crystal structure of auxin-binding protein 1
M3299	1	243.650.161	3	Zm00001d032931	Oxoglutarate/iron-dependent dioxygenase
M7826	2	227.370.542	8	Zm00001d007301	Protein disulfide isomerase
				Zm00001d007307	Magnesium transporter NIPA4
M21270	7	46.333.211	20	Zm00001d019612	Protein transport protein SEC31
M22252	7	149.023.564	6	Zm00001d032932	Agmatine coumaroyltransferase
				Zm00001d021332	Peptidyl-prolyl cis-trans isomerase CYP21

* Finds described in the discussion and detailed in Table 9.

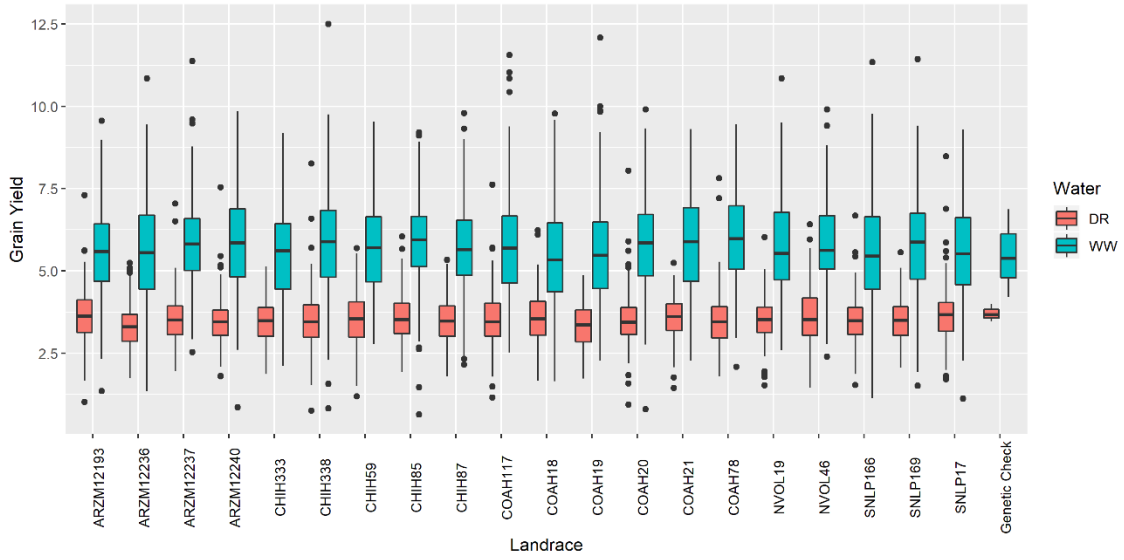

Figure 4: Boxplot of landraces progenies and genetic check (TesterxCML376) for grain yield (t.ha^{-1}) at both irrigation treatments.

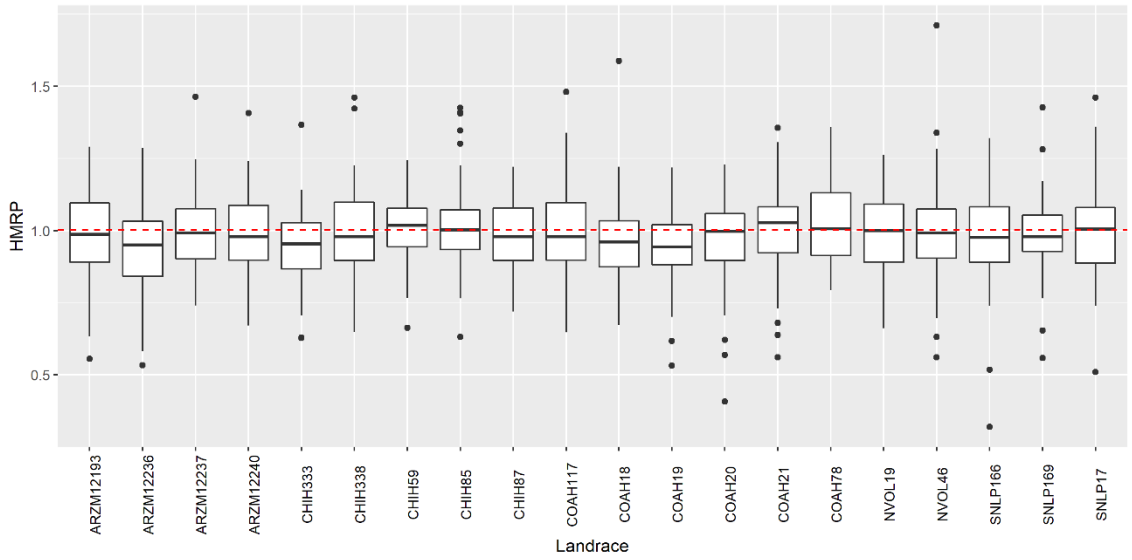

Figure 5: Boxplot of landraces progenies and genetic check (Tester x CML376) for drought-tolerance index (HMRP) at both irrigation treatments. The red dashed line is the global mean HMRP of the genetic check (1.002616).

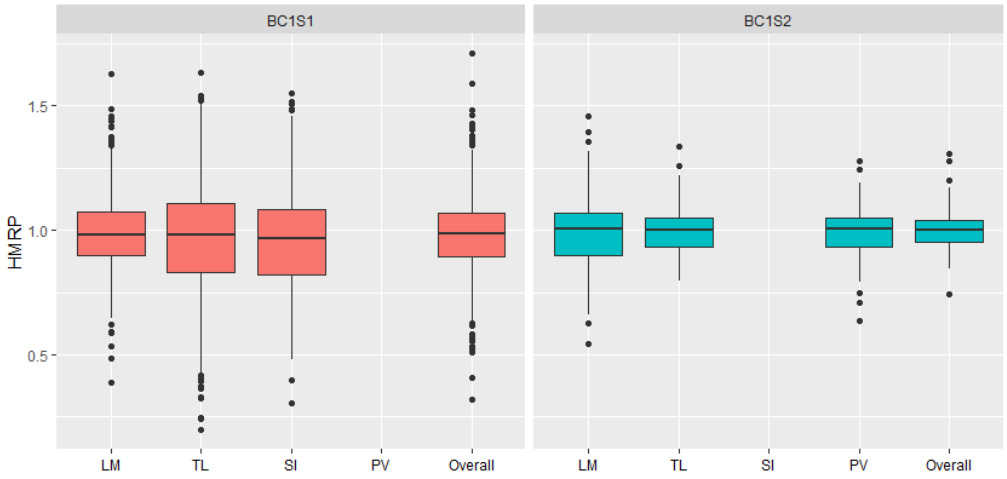

Figure 6: Boxplot of drought tolerance index for each site (LM, Los Mochis; TL, Taltizapan and SI, Santiago del Ixcuintla, PV, Puerto Vallarta) and overall sites to the generations BC1S1 and BC1S2.

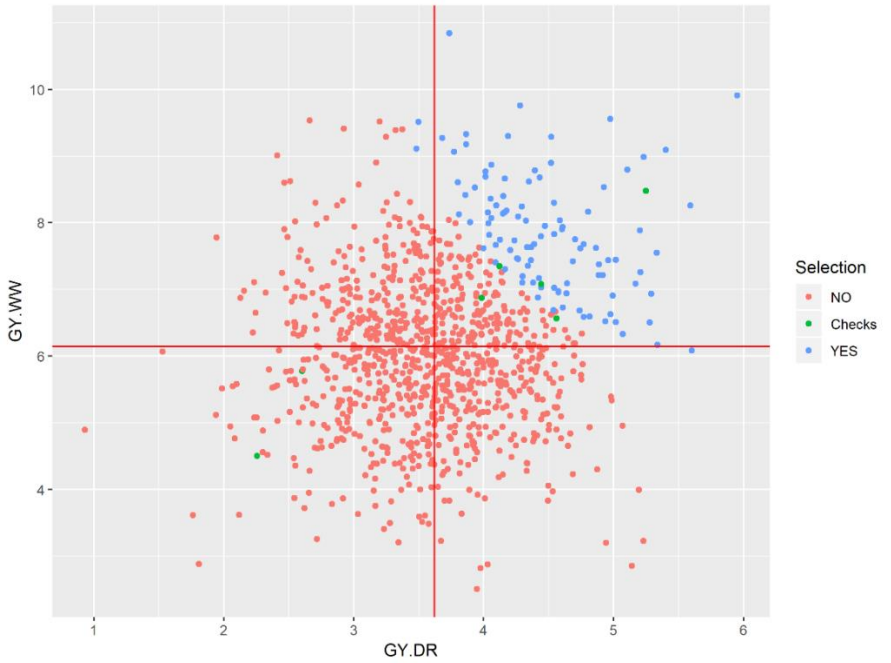

Figure 7: Grain yield at drought condition against the grain yield at well-watered condition, for the Los Mochis trial, where 1067 genotypes were tested. Highlighted in blue the 104 highest drought-tolerant, simulating a 10% of selection intensity, the 6 checks used (green) and the 956 not selected by the simulated selection (red).

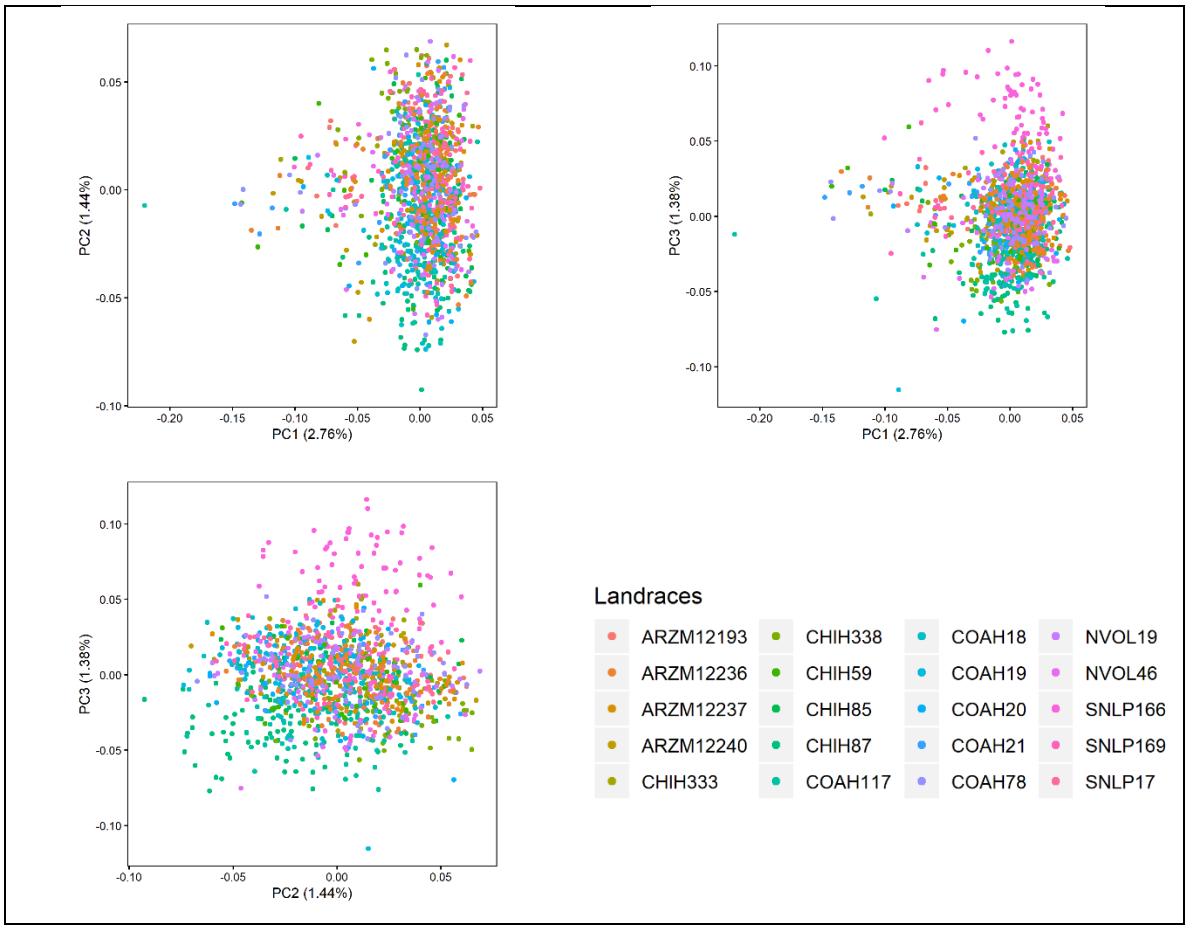

Figure 8: Plots of the three first principal components. The 20 landraces families are distinguished by the colors.

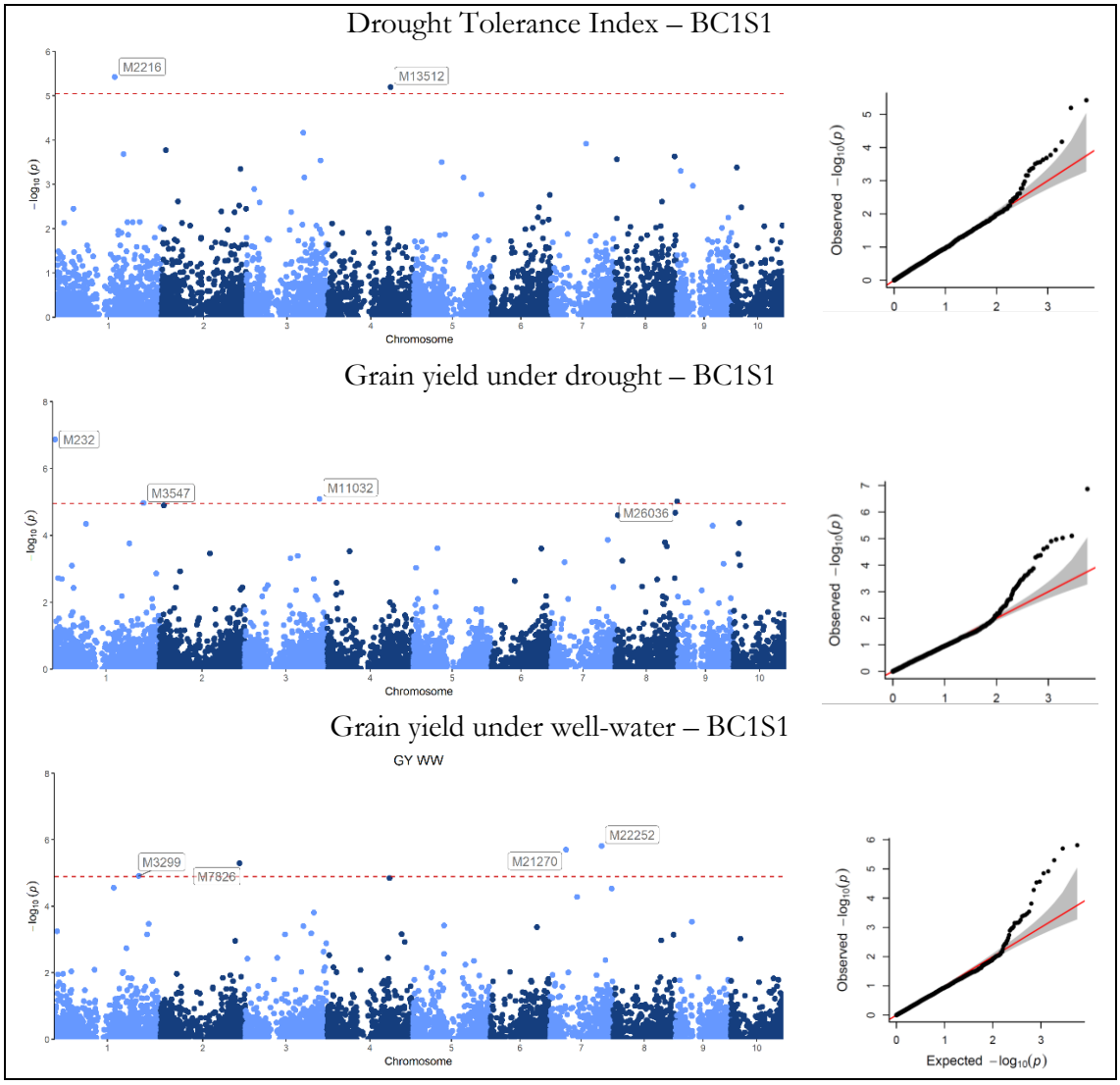

Figure 9: Genome-wide Association results for the first generation (BC1S1-TC1). Manhattan plot and QQ-plots for the drought tolerance index, grain yield in drought and well-watered conditions.

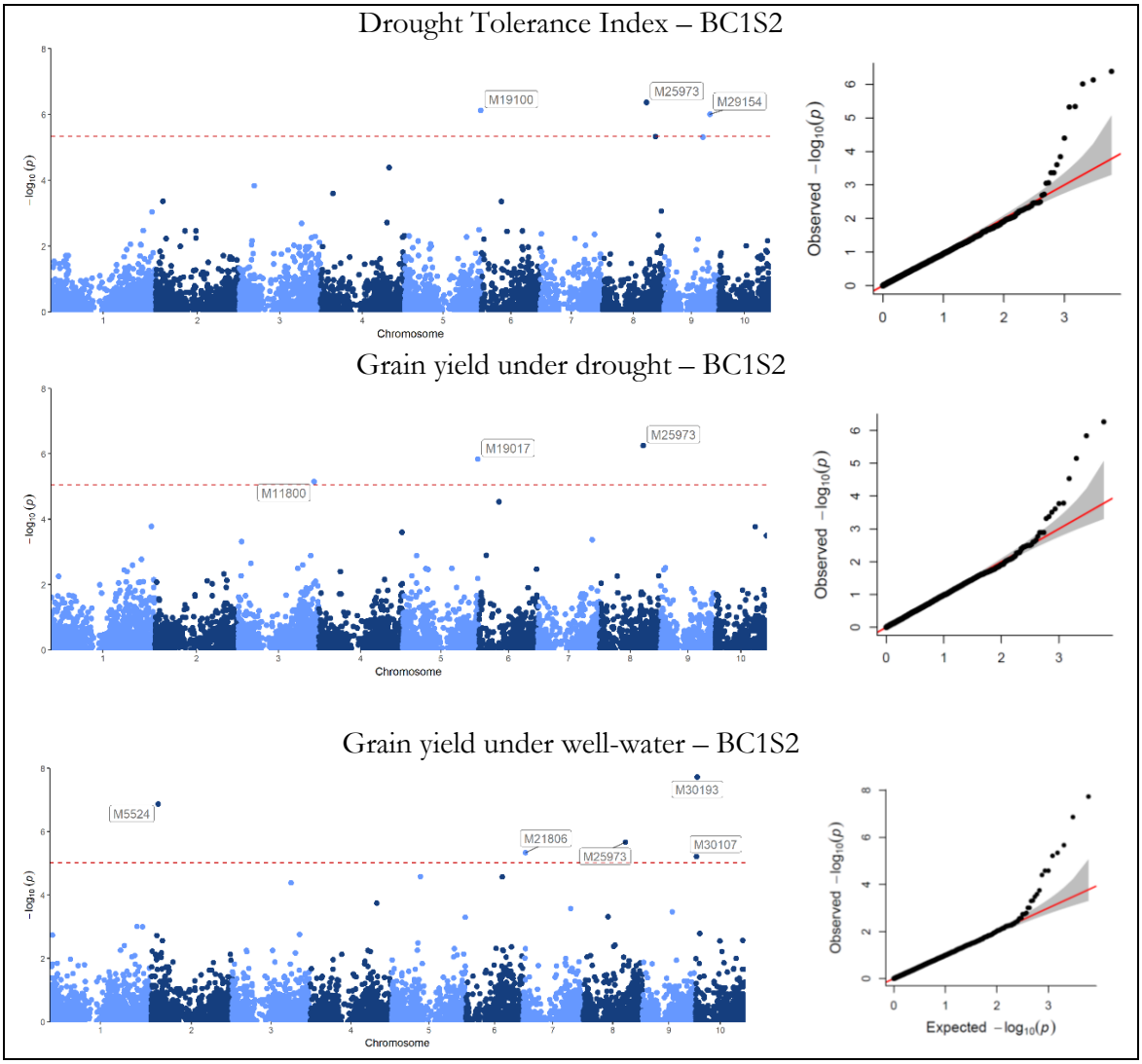

Figure 10: Genome-wide Association results for the second generation (BC1S2-TC2). Manhattan plot and QQ-plot for the drought tolerance index, grain yield in drought and well-watered conditions.

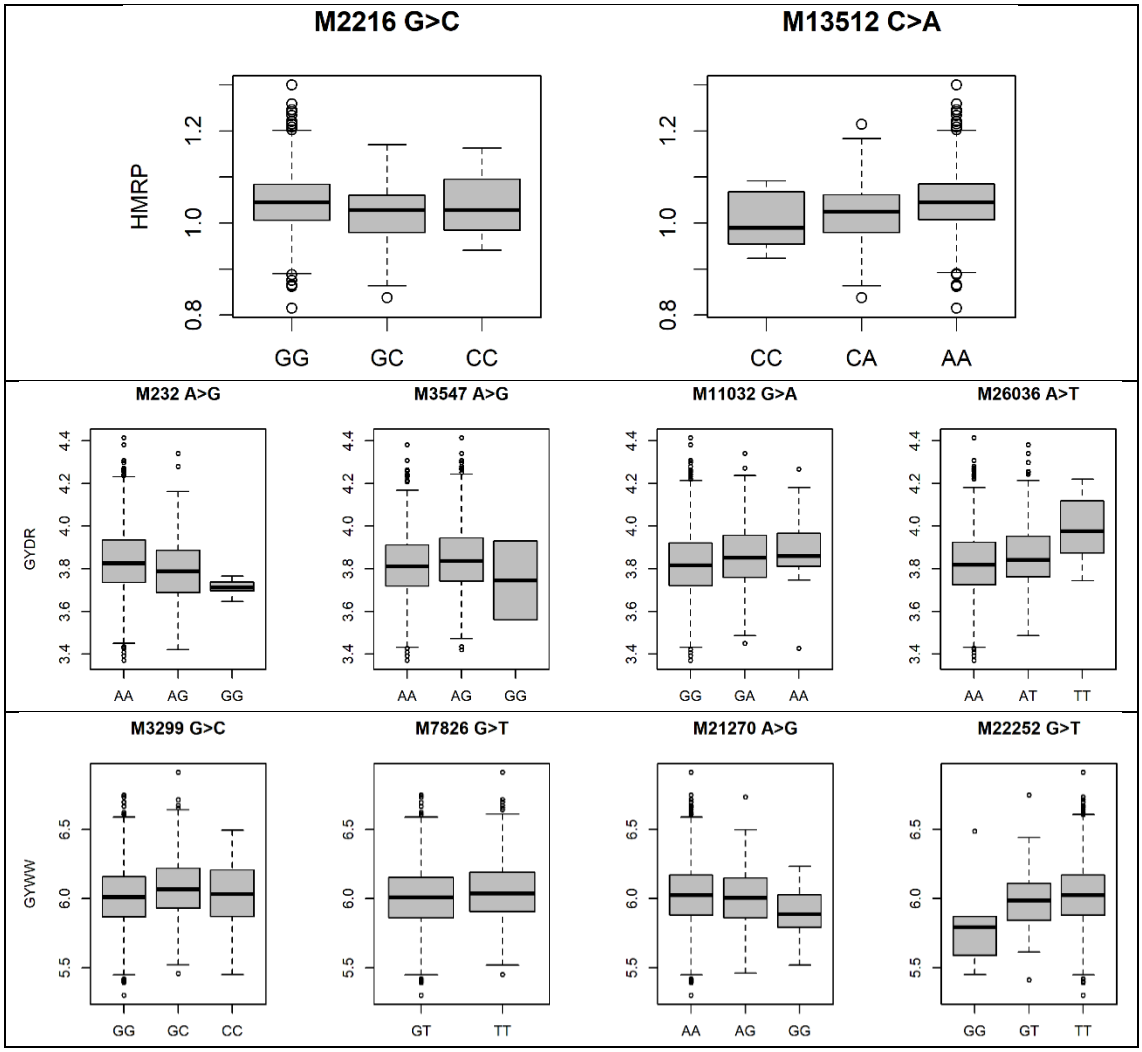

Figure 11: Boxplot of drought tolerance index (first row - HMRP) and grain yield in drought and well-watered conditions (second row – GYDR, third row - GYWW) of each genotype class for the 10 associated SNPs.

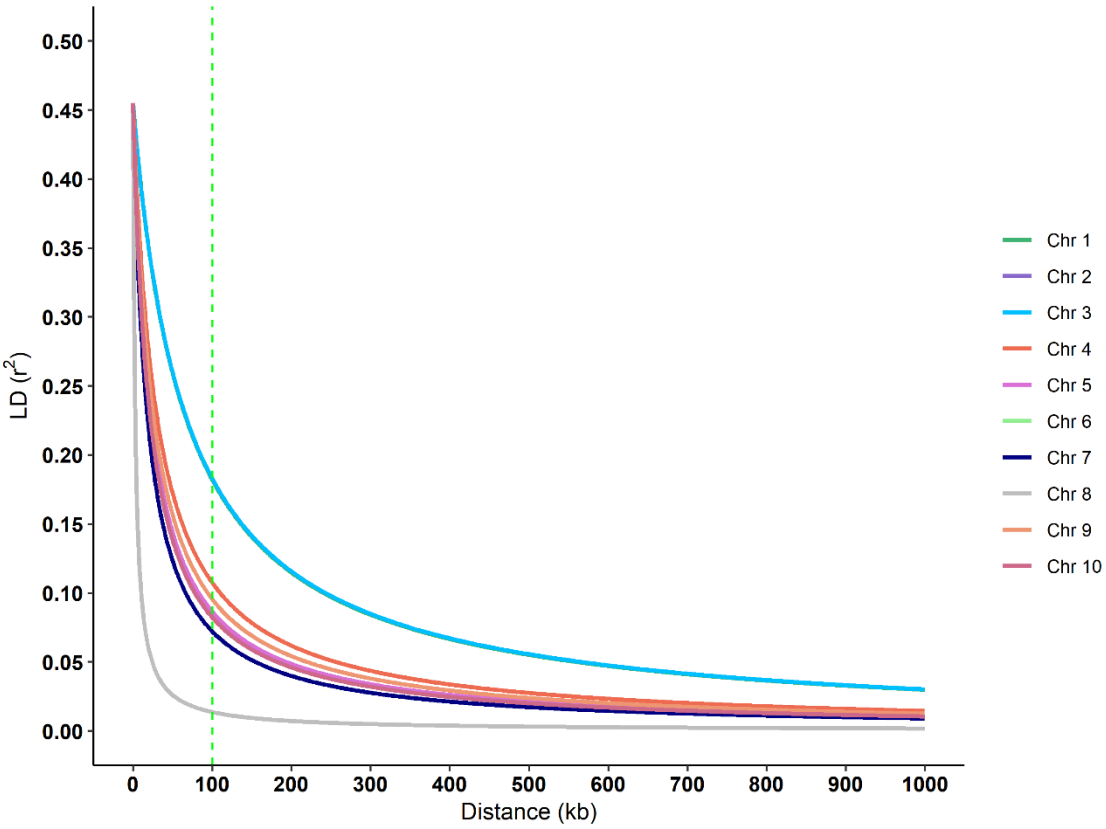

Figure 12: Linkage disequilibrium decay over each chromosome of the BC1S1 population, that is, the breeding population used in this study originated from 20 landraces that undergone through a backcross with an elite line and selfed. The vertical dashed green line indicates the window size used in the annotation process.

Table 9: Two significative SNPs located close to putative genes that encode proteins associated with plant response to drought. The distances between the mark and the putative gene in pair of bases, the putative gene name, the protein encoded type/family and references that studied these proteins with drought stress in plants.

SNP	Distance	Putative gene and protein type	References
M232	84.3M bp	Zm00001d027539 Glutathione S-transferase	MCGONIGLE et al.,2002; MIN et al., 2016; XU et al., 2015; CHEN et al., 2012
M11032	80.7M bp	Zm00001d043929 Fasciclin-like arabinogalactan	JOHNSON et al., 2003; FAIK; ABOUZOUHAIR, 2006; FINKELSTEIN, 2013; ZANG et al., 2015; WANG et al, 2016; CAGNOLA et al., 2018

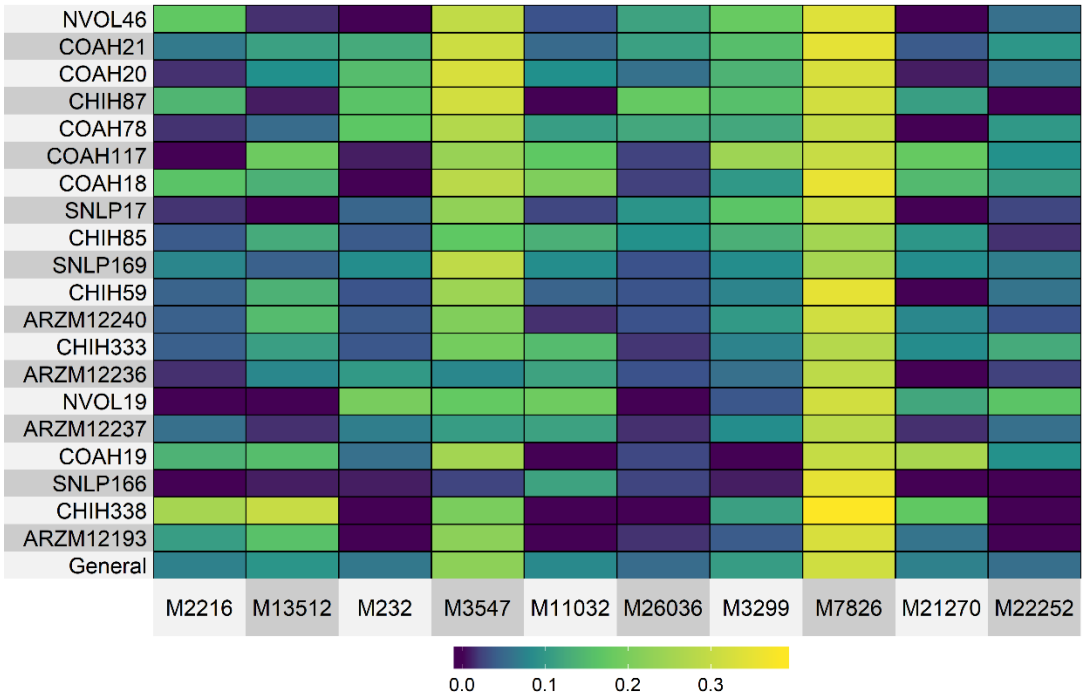

Figure 13: Frequency of allele from landrace for the significant markers (columns) by landrace family (rows) and in all the population (general). The frequency of allele from landrace over the entire population are the same values of the MAF presented in Table 6, once the allele from landraces is, as expected, the allele in minor frequency. Landraces are ranked by the higher frequency of superior allele considering the four SNPs where the landrace allele have shown positive effect on the traits (M3547, M11032, M26036, M3299, highlighted by the red rectangle). This rank gives the idea of which landraces are the most promising source for positive variants.

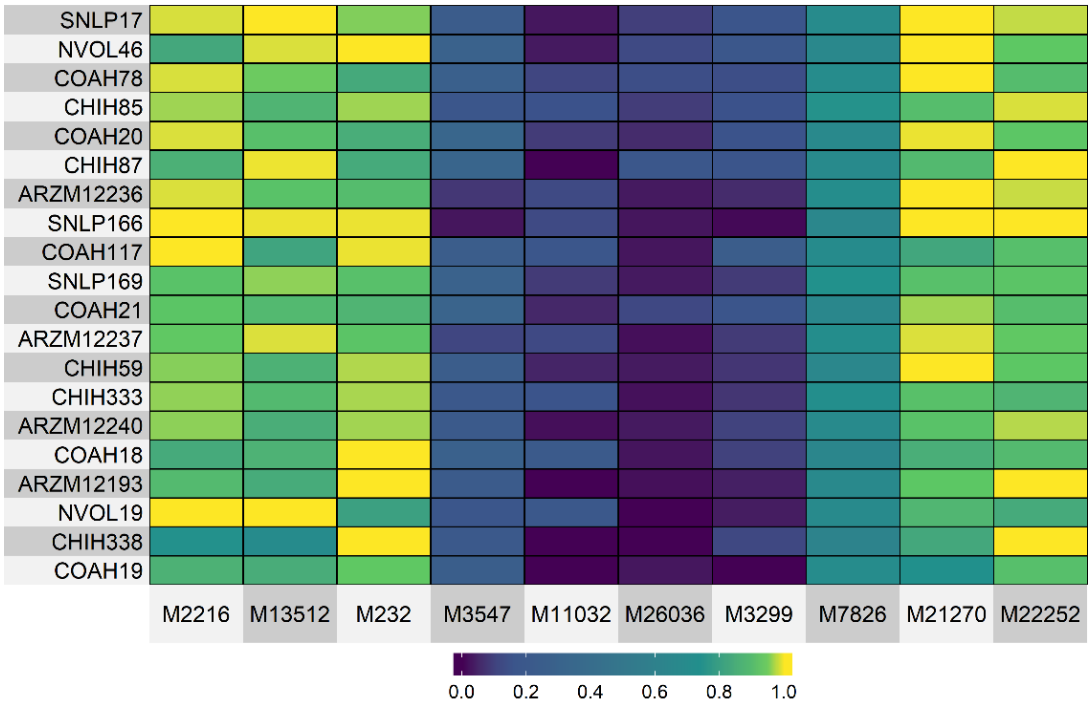

Figure 14: Frequency of superior allele (positive effects) for significant markers (columns) by landrace family (rows). Landraces are ranked by the higher frequency of superior allele considering all the 10 SNPs, regardless their origin. This rank gives the idea of which landrace populations are the promising to selection in this breeding population regards the combination with adapted elite genitor.