



Introgression of Maize Diversity for Drought Tolerance: Subtropical Maize Landraces as Source of New Positive Variants

Pedro Augusto Medeiros Barbosa¹, Roberto Fritsche-Neto¹, Marcela Carvalho Andrade², César Daniel Petrolí², Juan Burgueño², Giovanni Galli¹, Martha C. Willcox², Kai Sonder², Víctor A. Vidal-Martínez³, Ernesto Sifuentes-Ibarra³ and Terence Luke Molnar^{2*}

¹ Allogamous Plant Breeding Laboratory, Department of Genetics, Luiz de Queiroz College of Agriculture, University of São Paulo, Piracicaba, Brazil, ² International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico, ³ Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP), Mexico City, Mexico

OPEN ACCESS

Edited by:

Roberto Papa,
Marche Polytechnic University, Italy

Reviewed by:

Salvatore Ceccarelli,
Bioversity International, Italy
Valerio Hoyos-Villegas,
McGill University, Canada

*Correspondence:

Terence Luke Molnar
t.molnar@cgiar.org

Specialty section:

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

Received: 05 April 2021

Accepted: 24 August 2021

Published: 23 September 2021

Citation:

Barbosa PAM, Fritsche-Neto R, Andrade MC, Petrolí CD, Burgueño J, Galli G, Willcox MC, Sonder K, Vidal-Martínez VA, Sifuentes-Ibarra E and Molnar TL (2021) Introgression of Maize Diversity for Drought Tolerance: Subtropical Maize Landraces as Source of New Positive Variants. *Front. Plant Sci.* 12:691211. doi: 10.3389/fpls.2021.691211

Current climate change models predict an increased frequency and intensity of drought for much of the developing world within the next 30 years. These events will negatively affect maize yields, potentially leading to economic and social instability in many smallholder farming communities. 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 introgression panel to identify chromosomal regions 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) of the index 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, four associated with GY in drought condition, and other four associated with GY in irrigated conditions each. Although each of these markers explained less than 0.1% of the phenotypic variation for the index and GY, 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 positive diversity delivered by landraces are still present on the backcrosses and this is a potential breeding strategy for improving maize for drought tolerance and for trait introgression bringing new superior allelic diversity from landraces to breeding populations.

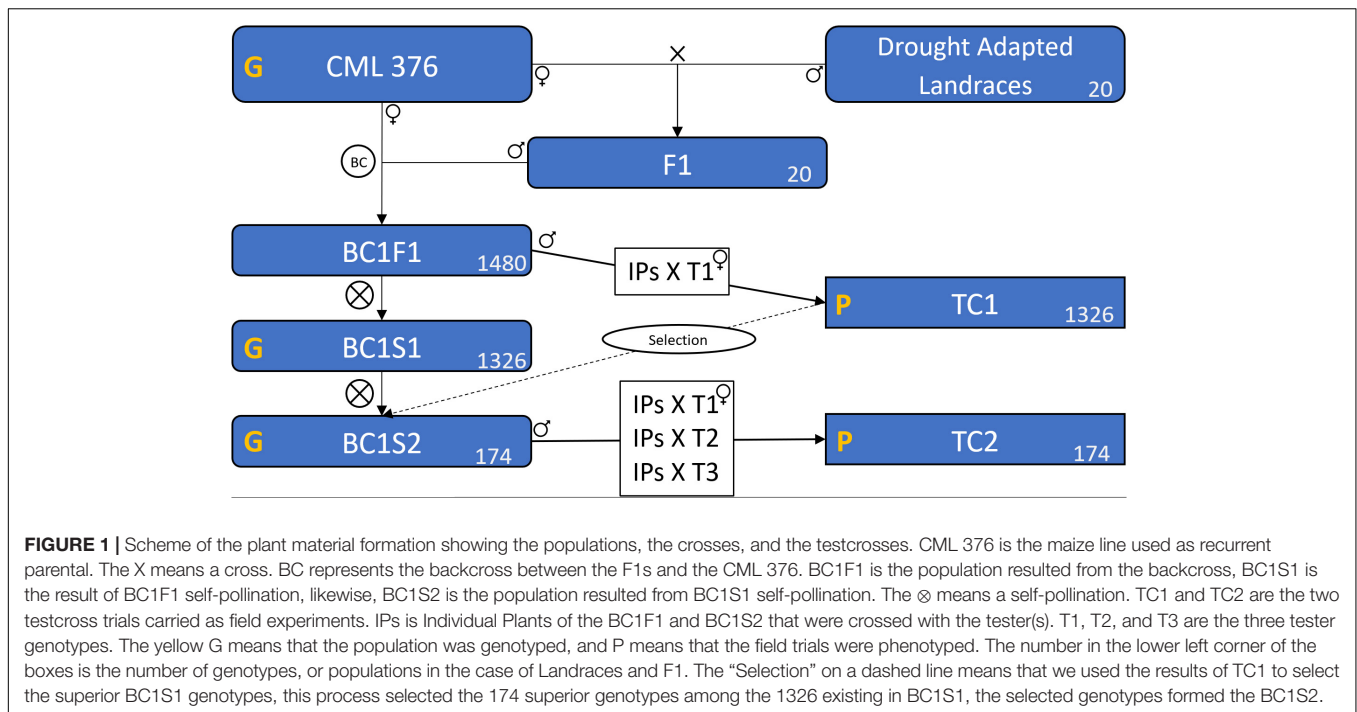
Keywords: plant breeding, GWAS, spatial analysis, abiotic stress, early testcross

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 193 million hectares in 2018 (Faostat, 2018). That corresponds to about 15% of the total area in agricultural use. Globally, the majority of maize hectares planted are in subtropical and tropical regions including many parts of Latin America, Africa, and Asia (Faostat, 2018). Millions of smallholders and subsistence farming communities in these regions rely on maize as a major source of calories as well as income to pay for basic needs and schooling for their children. For optimal yields, maize demands a large amount of water; therefore, drought events at key junctures in the growing cycle can cause significant losses in grain yield and year-to-year yield fluctuations (Harrison et al., 2014). This yield instability creates great financial uncertainty in the farming communities of developing countries where maize is important. Unfortunately, most climate change forecast models concur that an increase in the frequency and intensity of drought events is occurring and will continue to occur throughout the 21st century in many global regions (IPCC, 2007, 2014; Dai, 2013; Hoegh-Guldberg et al., 2018; Lu et al., 2019; Spinoni et al., 2020). These data suggest that if maize varieties with novel sources of drought tolerance are not soon developed, smallholder and subsistence farmers in many parts of the world will become even more vulnerable to yield fluctuations and the financial insecurity that entails. Fortunately, maize is a species with immense genetic variability and breeding for drought tolerance is an essential tool for overcoming the effects of climate change in maize production (Fisher et al., 2015). This genetic diversity has largely been collected in the world's germplasm banks and is additionally evident in the fields of many smallholder farmers that still plant landrace populations (Sanchez et al., 2000; Prasanna, 2012; Arteaga et al., 2016). In this context, the more than 25,000 maize landrace accessions maintained in the germplasm bank of the International Maize and Wheat Improvement Center (Spanish acronym, CIMMYT) are a tremendous resource for identifying novel resistance and developing new cultivars better able to maintain grain production in water-limited conditions. The reality, however, is that maize breeders in the private and public sector have for many years not used as new sources of variation the landrace accessions held in germplasm banks. This is mostly due to such factors as linkage drag, adaptation issues, poor agronomic characteristics and low yield potential (Goodman, 1985; Goodman et al., 2014). For that reason, in 2011 CIMMYT initiated the Seeds of Discovery project (SeeD), one of the components of MasAgro initiative, which has been largely funded by the Mexican government. The objective of SeeD called later MasAgro Biodiversidad, is to mine

the unexplored allelic variation in the maize germplasm bank for important abiotic and biotic pressure as well as nutritional and end-use characteristics.

Under abiotic stress conditions, tolerance is the ability of the plant and its yield component traits to maintain grain production during an extended period under that stress (Schafer, 1971; Miti et al., 2010). Between stand establishment and physiological maturity, maize has two periods in the growth cycle when the crop is most vulnerable to yield loss due to drought conditions: (i) from the late vegetative growth stage through to the end of flowering and (ii) during the grain filling period (Sayadi Maazou et al., 2016). Of these two, the flowering stress is considered the more devastating because of the greatly increased anthesis-silk interval (ASI) that can be provoked by drought conditions (Bolaños and Edmeades, 1996). ASI measures the number of days from when anthesis occurs to when receptive silks (stigmas) emerge. In non-stressed conditions, the maize male inflorescence typically begins releasing pollen one to two days before silk emergence and releases pollen for 4–7 days, existing a considerable variability in this period across genotypes (Hall et al., 1982; Struik and Makonnen, 1992; Uribealarea et al., 2002). However, under drought stress the interval between anthesis and silk emergence can be greatly increased resulting in significantly reduced yield because a reduced amount viable pollen is available when the silks finally do emerge. In more extreme drought conditions, silks may emerge too late to receive pollen or may not emerge at all resulting in barren plants and no grain production. Because of this vulnerability, ASI is one of the most studied traits in maize, and it has been used as a secondary trait for indirect selection for grain yield and drought tolerance (Bolaños and Edmeades, 1996; Xue et al., 2013). CIMMYT has conducted long-term recurrent selection breeding in two populations for drought tolerance focused on decreasing ASI and maintaining grain yield under drought stress (Bänziger et al., 2000). These two populations, La Posta Sequía (LPS) formed in the mid-1980s, and Drought Tolerant Population (DTP) formed in the early 1990s have now gone through recurrent selection for 7 and 9 cycles, respectively. In total, fifteen official CIMMYT maize inbred lines (CML) and several open-pollinated varieties (OPV) have been derived and released from the various cycles of LPS and DTP. These CML and OPV have been used extensively in east Africa, South Asia and Latin America and are adapted to either highland, mid-altitude subtropical or lowland tropical environments. While this breeding project was successful in developing new sources of elite drought tolerant material, the LPS and DTP source material represents only a fraction of the overall diversity in the CIMMYT germplasm bank. Much more of this diversity needs to be explored, using the latest techniques to unravel the genetic control of drought tolerance in maize. And



while ASI remains a trait of importance in developing drought tolerant maize, a plant's response to drought stress is based on physiological and morphological modifications that involve a complex number of genes and molecular pathways (Rengel et al., 2012; Min et al., 2016). By better comprehension, the genetics underlying drought tolerance, and by accessing more of the genetic diversity available for this trait, breeding programs working on this trait will improve selection precision and the speed of breeding cycles.

In order to understand both the genetic diversity of a specie and the complex genetic factors controlling response to abiotic stress, molecular markers are an important tool for geneticists and breeders (Xu et al., 2017). Methodologies such as genome-wide association studies (GWAS) seek to relate molecular markers to observed phenotypes and thus discover the chromosomal regions associated with the trait of interest. In maize, this procedure has been used for many different traits, both abiotic and biotic (Xiao et al., 2017). In others important cultivated plants, GWAS has been used in studies on 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, seldomly, testcross hybrids. 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 the crop grown by 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 F₁ with a S₂ or S₃ experimental

line. The aim is to take advantage of early selection for yield potential, avoiding future testing costs on unpromising 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 genomic regions related to drought tolerance in two sequential inbreeding generations of early testcrosses obtained between landrace-derived semi-inbred lines and elite testers. The main reason to use landraces is to bring novel positive alleles to a breeding population, we intend to do that combining some breeding, genetics, and analytics techniques, as back-cross, early-testcross, GWAS, selection index and mixed models.

MATERIALS AND METHODS

Plant Material

Before the evaluations discussed in this study, 326 landraces of subtropical adaptation were evaluated in *per se* drought trials over 2 years (data not shown, see <https://seedsofdiscovery.org/catalogue/maize-drought-tolerant-landraces/> for more details). Subtropical adapted maize landraces are from areas between 800 and 1800 m above sea level (Edmeades et al., 2017). The 326 were selected from the CIMMYT International Germplasm Bank and the Mexican National Maize Collection collected and managed by Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP) using an aridity index described in Ruiz Corral et al. (2013).

TABLE 1 | 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).

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		

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.

They originated from dryland production areas, with 290 (89%) coming from Mexico and the rest from Argentina and Chile. At least 24 different maize races were represented in the selected set of accessions, however, 17% of them were not racially classified. Although latitude also plays an important role in adaptation, we did not consider latitude in the selection of the landraces as drought evaluations in Mexico are conducted during the winter dry season and much of the photoperiod sensitivity that could potentially be provoked by a change in latitude is negated. Based on the results of the *per se* evaluations, the 20 highest performing accessions were used to develop semi-inbred lines by the SeeD maize breeding project in the Genetic Resources Program of CIMMYT.

In deriving inbred lines from maize landraces, crossing to an elite line is a common practice to avoid severe inbreeding depression and to add favorable alleles for important agronomic traits (Tarter et al., 2004; Meseka et al., 2015; Navarro et al., 2017). In this case, the 20 selected landraces were crossed with the elite CIMMYT line CML376 to generate 20 segregating F1 populations (Figure 1). The F1 populations were then crossed again to CML376 (backcross) and ears were harvested and shelled individually creating 1480 BC1F1 lines with an average of 74 lines per landrace population and each line averaging 25% landrace genome. In the following cycle, individual plants were selfed to produce BC1S1 lines and at the same time the plant was crossed

TABLE 2 | Factors used in each model of first-stage spatial analysis (individual experiments).

Fixed spatial factors	Experiment (site/water management)
Row	SI/WW, SI/DR, TL/WW, LM/WW
Column	LM/DR
Row + Column	TL/DR

Decisions about the use or not of spatial factors on the models were taken based on variogram plots.

LM, SI, and TL are, respectively, the locations Los Mochis, Santiago del Ixcutlía, and Tlaltizápan.

WW and DR are the irrigation treatments, well-watered, and drought conditions.

to the F1 tester CML264/CML311 (hybrid produced by crossing CML264 and CML311) to produce testcross seed for yield trial evaluations (TC1) in the winter of 2015–2016. This tester was selected because it is a commonly used subtropical tester in Mexico and has the advantage of producing large quantities of testcross seed. Based on their TC1 yield performance under drought and well-watered conditions, 68 entries were selected for further testing (Table 1). The BC1S1 lines had been selfed in a nursery while the yield trials were conducted but only 64 of them produced harvestable BC1S2 ears. Two to three ears per BC1S1 line were harvested for a total of 174 BC1S2 ears. In the summer, these 174 BC1S2 lines were crossed to 3 testers for TC2 trials in

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).

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	

The accuracy (Acc%), coefficients of genetic (CVg), and error (CVe) variation and heritability are also presented.

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

the winter of 2016–2017. The testers used were the TC1 tester, CML264/CML311, plus CML373 and CL501801.

Testcross Phenotyping and Experimental Designs

From the 1480 BC1F1 individual plants selfed and crossed to the tester CML264/CML311, 1326 produced sufficient seed quantities of both selfed seed and testcross seed. Six check hybrids were added, so the TC1 experiment had a total of 1332 entries. The TC1 was carried out at three locations: 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 there were two treatments, drought (DR) and well-watered (WW). A sparse phenotyping arrangement of genotypes in location was used due to restrictions on field space and seed amount, but the landraces ancestry was balanced in each location. That means, we did not test the entire set of 1332 genotypes in all locations, using so an unbalanced distribution of genotypes across location. The distribution of the testcross entries was: 266 tested only in TL (~13 progenies 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 combinations of location and treatment (3 locations and 2 irrigation treatments). The experiment was designed in a way that each location contains three plots of the tested germplasm in that specific location, one plot in irrigated condition and two plots in drought condition. The irrigated (WW) treatments were planted in a modified augmented design with systematically distributed spatial check plots arranged in a chess knight move, the field implementation of this design were based mainly on the studies of Cullis et al. (1998); Muller et al. (2010), and Clarke and Stefanova (2011). The percentage of checks ranged from 11.7% in LM, 13.7% in TL, and 20.4% in SI. The DR treatments were conducted in a randomized complete block design with two replicates, except in Santiago del Ixcuintla, where even the

DR experiment was conducted as the WW experiments. The checks included in the trial were classified into three different categories: (i) the genetic check consisting of the recurrent parent crossed to the tester, CML264/CML311//CML376; (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) and finally, the spatial checks, hybrids CEBU, 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, and 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 m long with 0.75 m between the rows, and in LM, 4 m rows spaced by 1 m. The agronomic treatments (fertilizers, insecticides, and herbicides) were applied following common local practices. The irrigation in the drought treatments was interrupted 3 weeks before the expected anthesis date at each location, and after flowering terminated 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 ($\text{Mg}\cdot\text{ha}^{-1}$), considering the differences in plots size at each site.

The TC2 trials followed a similar protocol. The set of 174 BC1S2 lines selected using the TC1 data were tested in 18 location \times treatment \times tester combinations (3 locations, 2 irrigation treatments, and 3 testers). The 3 locations were TL, LM, and San Juan de Abajo (PV) – Nayarit State, replacing the SI TC1 location, the 2 irrigation treatments were WW and DR, and the 3 testers were the hybrid CML264/CML311 and the lines CML373 and CL501801. 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.

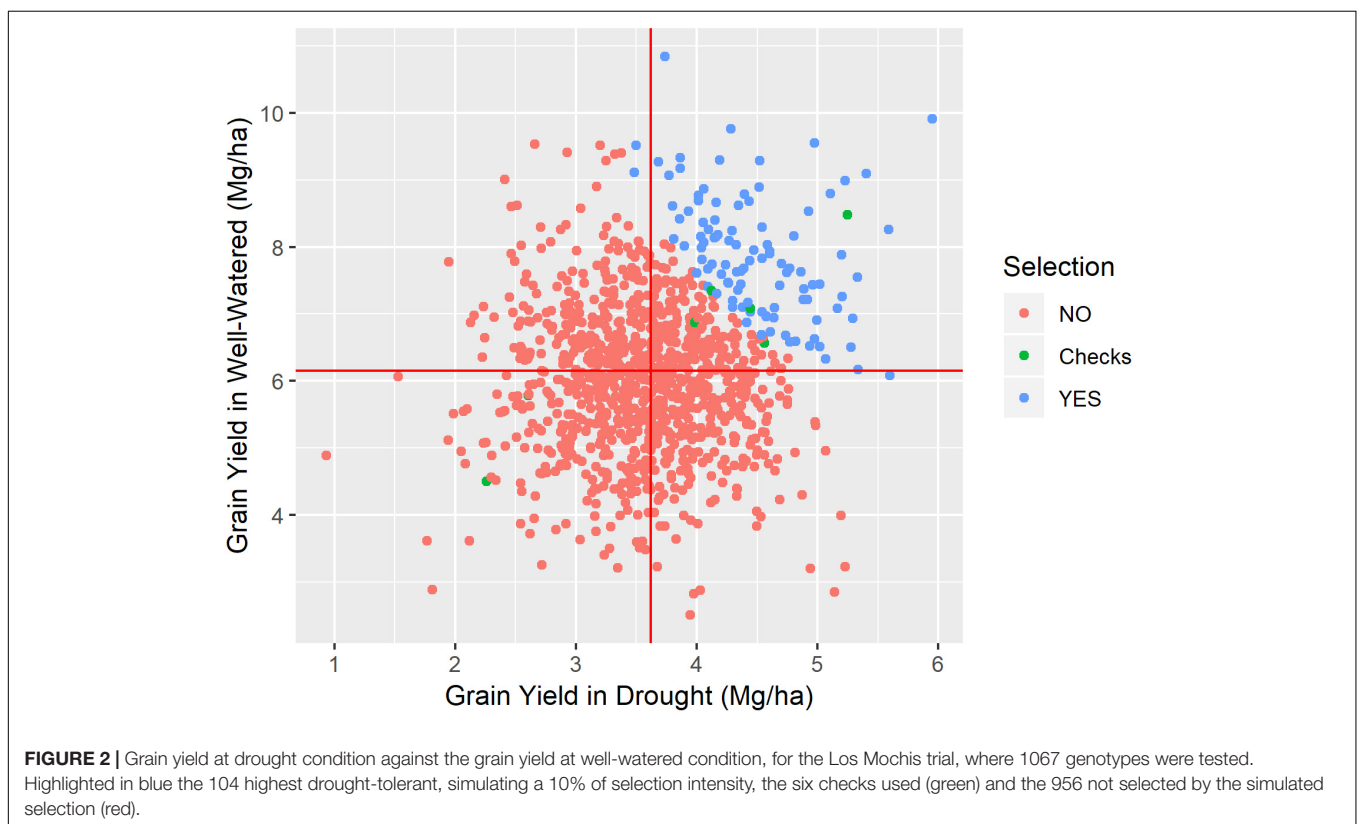
Genotypic Data

We used DArTseq technology (Sansaloni et al., 2011), developed by Diversity Array Technology company,¹ for genotyping all the 1326 BC1S1 and 174 BC1S2 lines and the recurrent parent CML376. The genotyping was carried out in the Genetic Analysis Service for Agriculture (Spanish acronym, 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, *Pst*I (CTGCAG) and *Nsp*I (CATG), and ligating individual barcodes adapters to identify the origin of each fragment after a samples pooling.

¹<http://www.diversityarrays.com>

TABLE 4 | 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

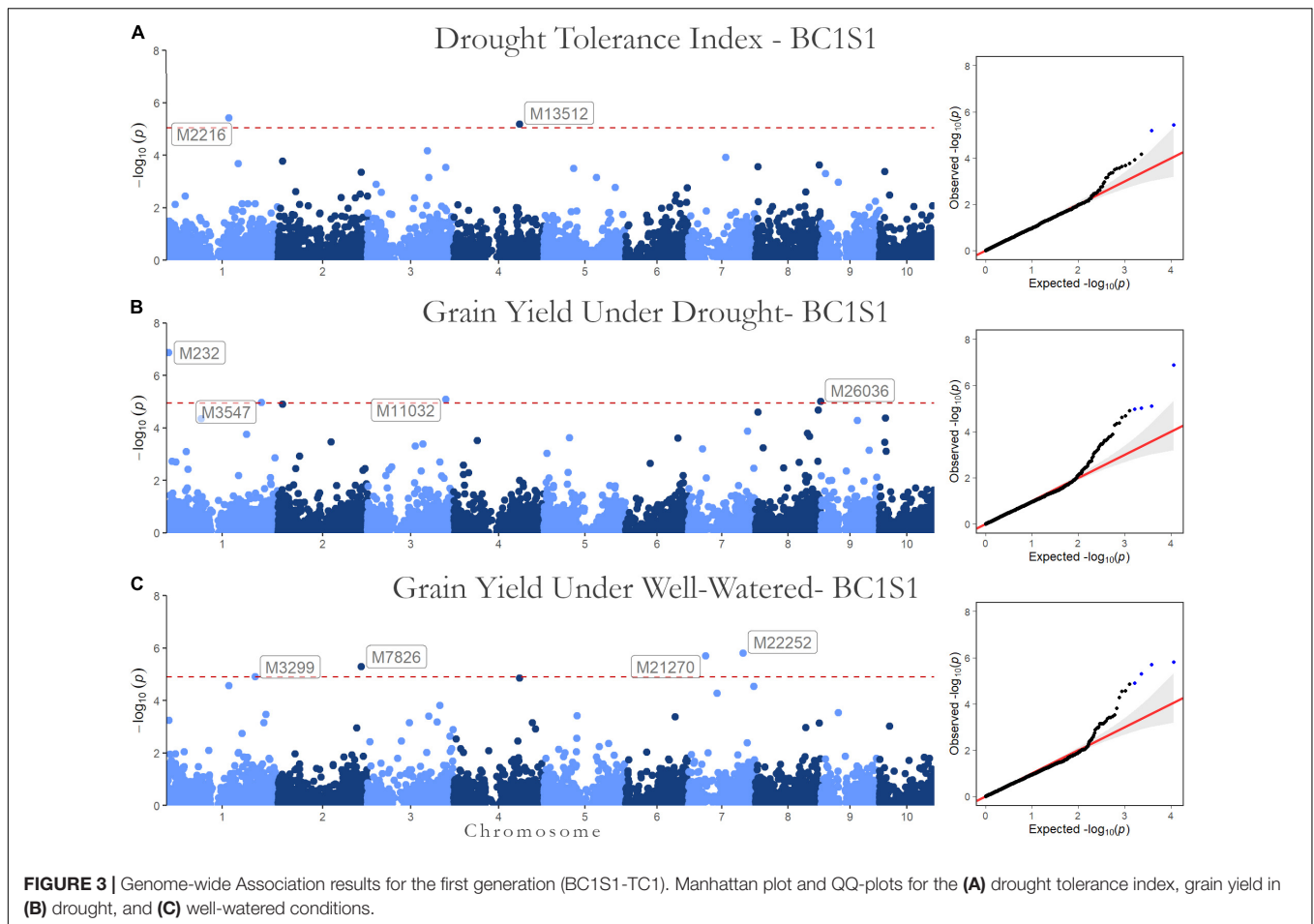


Successfully amplified fragments were sequenced using the sequencer Illumina HiSeq2500 (Illumina Inc., San Diego, CA). Then, the SNP calling for those fragments was produced by the DArTsoft analytical pipeline (DArTsoft; DArT P/L, Australia²) where the markers were identified *de novo* by comparing the sequences of fragments present in genomic libraries of samples previously processed, this process identifies and calls SNP markers against an existing library and it is totally independently of any reference genome. After this process, a set of filtering parameters were then applied to select high-quality markers for this study. To obtain markers positions on the chromosomes, the DNA fragment sequences were BLASTed against the *Zea mays L.* reference genome (B73 RefGen_v4).

²<https://www.diversityarrays.com/>

This procedure resulted in a genomic profile of 47,047 SNPs. Of this total, the markers of uncertain mapping or non-nuclear were still discarded, resulting in a profile of 32,592 SNPs with excellent quality that were positioned on the maize reference genome.

The 32,592 SNPs were named, for this study, by their order in the genome from M1 (the first marker in chromosome number 1) to M32592 (the last marker in chromosome number 10). A final quality control procedure was applied, and markers with minor allele frequency (MAF) <5% were excluded and a call rate $\geq 95\%$ used. Thus, markers with more than 5% of missing data were excluded. The imputation of missing data was performed by the Wright method carried out using the R package *snpReady* (Granato and Fritsche-Neto, 2018) using the function *raw.data*, and resulted in a final total of 5,695 SNPs



(Supplementary Table 1). The genome coverage was obtained by calculating the SNP density within a 100 Mb window, that is, the number of SNPs in each 100 Mb DNA segment for all the chromosomes, for this was used the R package rMVP (Yin et al., 2020).

Phenotypic Analysis

For analysis of the phenotypic data, a two-stage process was performed. In the first stage, adjusted yield means of genotypes were obtained by location, and then in the second stage a mixed model was fit jointly considering residual weights estimated in the first stage. The spatial corrections were applied in the first stage, using a first-order autoregressive process ($AR1 \otimes AR1$) to model the covariance structure of error. The spatial modeling approach was chosen due to some factors as the high number of genotypes tested, the augmented design with repeated check plots applied and the large size of the field trial.

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

$$y = X\tau + e$$

where y is the vector of phenotypic data, arranged by Row x Column (plot field position), X is the design matrix of fixed effects; τ is the vector genotypes effects (and tester in the case of TC2 trials), and eventual spatial factors row and columns (as explained below); and e vector of errors [$e \sim N(0, R\sigma^2)$]; R is the covariance matrix of e . The spatial analysis does a decomposition of the e into a vector ζ of spatial correlated effects, and a vector η of spatially independent residuals, in this spatial approach of modeling the residual we assume the error variance as:

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

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

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

where $AR1(\rho_{col})$ and $AR1(\rho_{row})$ are the first-order autoregressive correlation matrices for columns and rows. The components σ_ζ^2 and σ_η^2 are the spatial residual variance and the non-spatial residual variance, respectively, \otimes is the Kronecker product. The (ρ_{col}) and (ρ_{row}) are the autocorrelation parameter for field

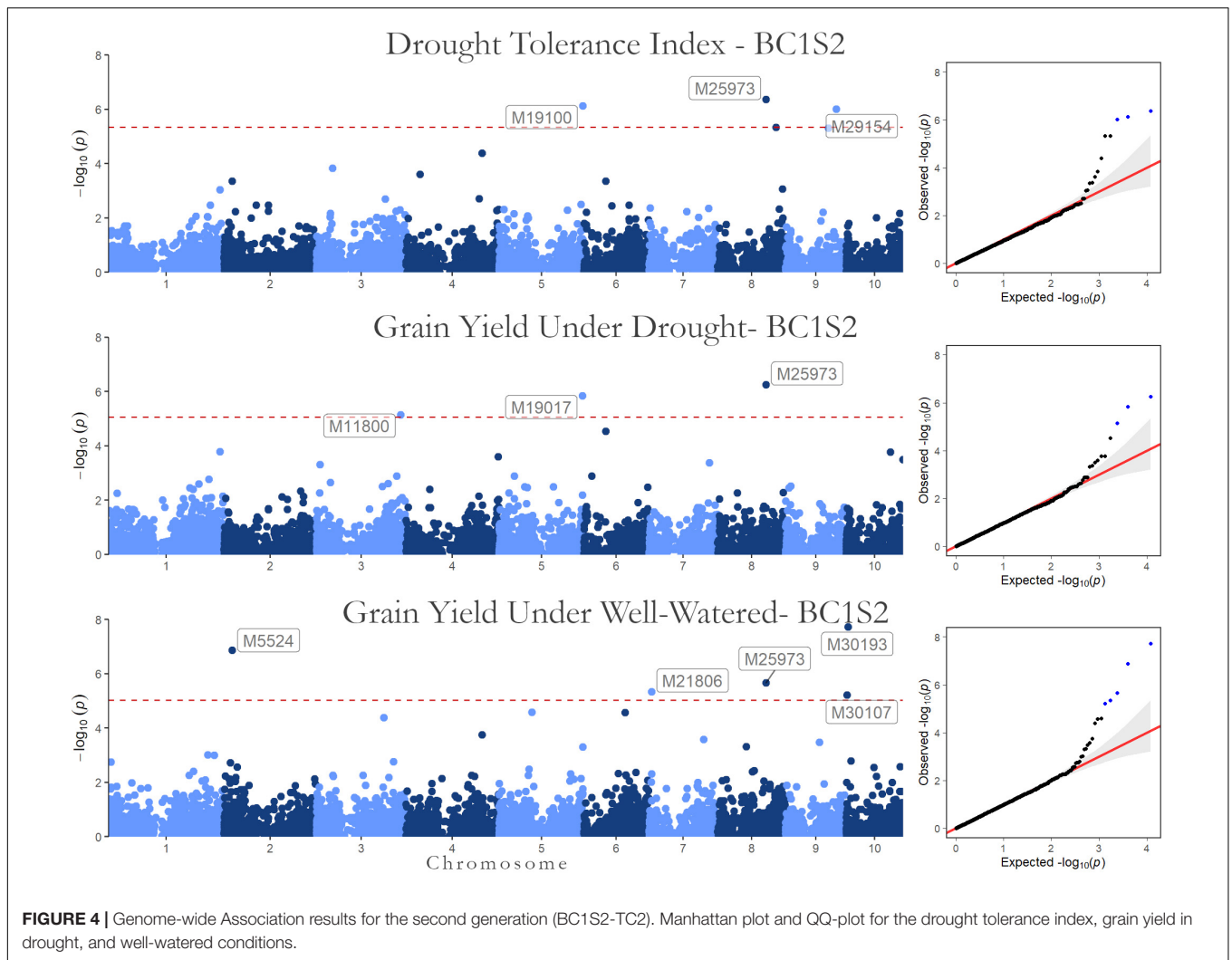


FIGURE 4 | Genome-wide Association results for the second generation (BC1S2-TC2). Manhattan plot and QQ-plot for the drought tolerance index, grain yield in drought, and well-watered conditions.

columns and field rows:

$$AR1_{row} = \begin{bmatrix} 1 & \rho_{row} & \rho_{row}^2 & \cdots & \rho_{row}^{r-1} \\ \rho_{row} & 1 & \rho_{row} & \cdots & \vdots \\ \rho_{row}^2 & \rho_{row} & 1 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{row}^{r-1} & \cdots & \cdots & \cdots & 1 \end{bmatrix}$$

$$AR1_{col} = \begin{bmatrix} 1 & \rho_{col} & \rho_{col}^2 & \cdots & \rho_{col}^{c-1} \\ \rho_{col} & 1 & \rho_{col} & \cdots & \vdots \\ \rho_{col}^2 & \rho_{col} & 1 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{col}^{c-1} & \cdots & \cdots & \cdots & 1 \end{bmatrix}$$

The Kronecker product of them results in the symmetric correlation matrix (size n by n, where n is the number of plots)

that captures the correlation between each pair of plots.

$$AR1_{row} \otimes AR1_{col} = \begin{bmatrix} 1 & cor_{1,2} & \cdots & cor_{1,n} \\ cor_{2,1} & 1 & \cdots & cor_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ cor_{n,1} & cor_{n,2} & \cdots & 1 \end{bmatrix}$$

Where $cor_{1,2}$ is the correlation between the plot 1 and plot 2, and η is the number of plots. Therefore, as said before, the final error variance structure for the first stage models is the sum of the correlated spatial error and the spatial independent error:

$$R = \sigma_{\zeta}^2 \begin{bmatrix} 1 & cor_{1,2} & \cdots & cor_{1,n} \\ cor_{2,1} & 1 & \cdots & cor_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ cor_{n,1} & cor_{n,2} & \cdots & 1 \end{bmatrix} + \sigma_{\eta}^2 I_{(n \times n)}$$

The DR treatments follow the same logic for the error variance structure. The difference between the analysis of WW treatments and DR treatments is in the fixed section of the model, in the

drought management there were replicates, so the factor replicate was added 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 replicates.

This analysis provides a variogram plot of the residuals that shows in a geographic form any spatial pattern that may be 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 fixed effect factor (Burgueño et al., 2000). The Wald test was used to verify the significance of adding this fixed effect in the model. And when significant they were retained 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.

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 is used to calculate the HMRP index:

$$HMRP_{ij} = \frac{2}{\left(\frac{GY_{wwij}}{\bar{y}_{wwj}}\right)^{-1} + \left(\frac{GY_{dr_{ij}}}{\bar{y}_{dr_j}}\right)^{-1}}$$

where $HMRP_{ij}$ is the harmonic mean relative performance of genotype i in site j ; GY_{wwij} is the BLUE of grain yield under well-irrigation treatment of genotype i in the site j ; $GY_{dr_{ij}}$ is the BLUE mean of grain yield in the drought condition of genotype i in site j ; \bar{y}_{wwj} and \bar{y}_{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:

$$y = X\beta + Zg + Wi + e$$

where y is the vector of BLUEs obtained in the first stage; β is the vector of fixed effects of site, g is the vector of the genotypic values of the lines, i is the vector of the random effects of the interaction sites by genotype, e is the random errors vector. X , Z , and W are the design matrix for β , g , and 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 location was estimated using Wald statistic, and for the random effects, genotypes, and genotypes x location interaction, by Likelihood Ratio Test (LRT). The

broad-sense heritability, for the joint analysis, was estimated as:

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

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 and r is the number of replicates. All the phenotypic analyses were performed using the ASReml-R v.3 package (Butler, 2009) in R, version 3.5.3 (R Core Team, 2019). In the **Supplementary Material** you can find R codes and some comments on how these analyses were carried out.

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., 2016) for 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:

$$g = X\beta + Zu + e$$

where g is the vector of adjusted phenotypic observation (BLUPs of HMRP obtained in the joint analysis presented above, or GYDR and GYWW); β is the vector of the fixed effects of intercept, single markers, and the three first principal components used for population structure control; u is the vector of random additive effects; 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 $u \sim N(0, G)$, 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 non-linear 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).

Gene Annotation

Through the MaizeGDB (Portwood et al., 2019) and National Center for Biotechnology Information (NCBI³) 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)

³www.ncbi.nlm.nih.gov

TABLE 5 | 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
GY/WW	M26036	8	178,731,659	A > T	AA	5.20%	0.052	0.011	0.00011
	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

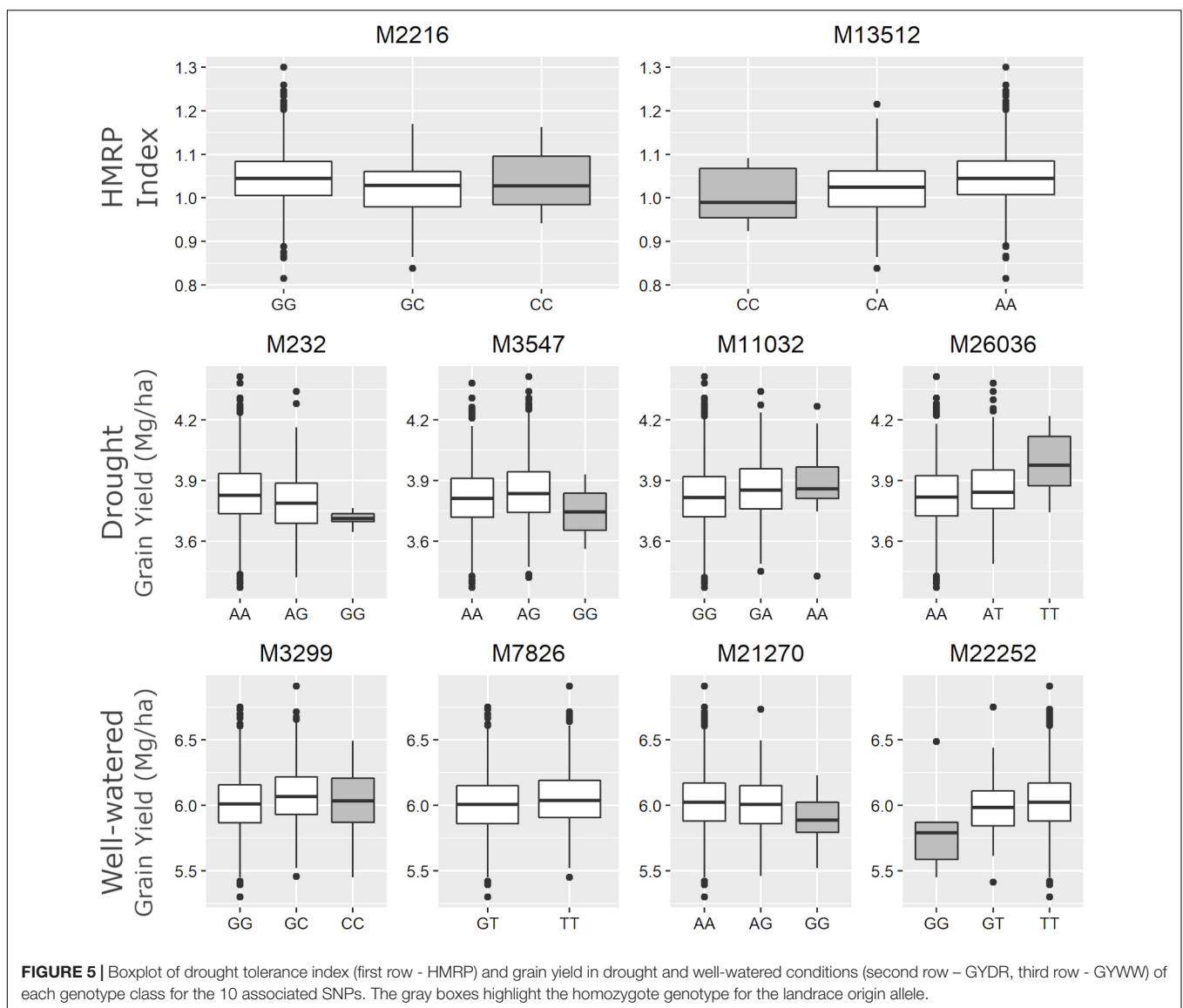


TABLE 6 | Significant SNPs for drought-tolerance index and grain yield in drought and well-watered conditions, with the SNP position (chromosome and position), minor frequency allele, effect, heritability, and r^2 for the GWAS of BC1S2 generation.

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 7 | 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.

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 <i>cis-trans</i> isomerase CYP21

The most described genes founded in each search and the protein type/family that these most described genes encode.

* Finds described in the discussion and detailed in **Table 8**.

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.

Origin of Alleles for Significant Markers and Superior Allele

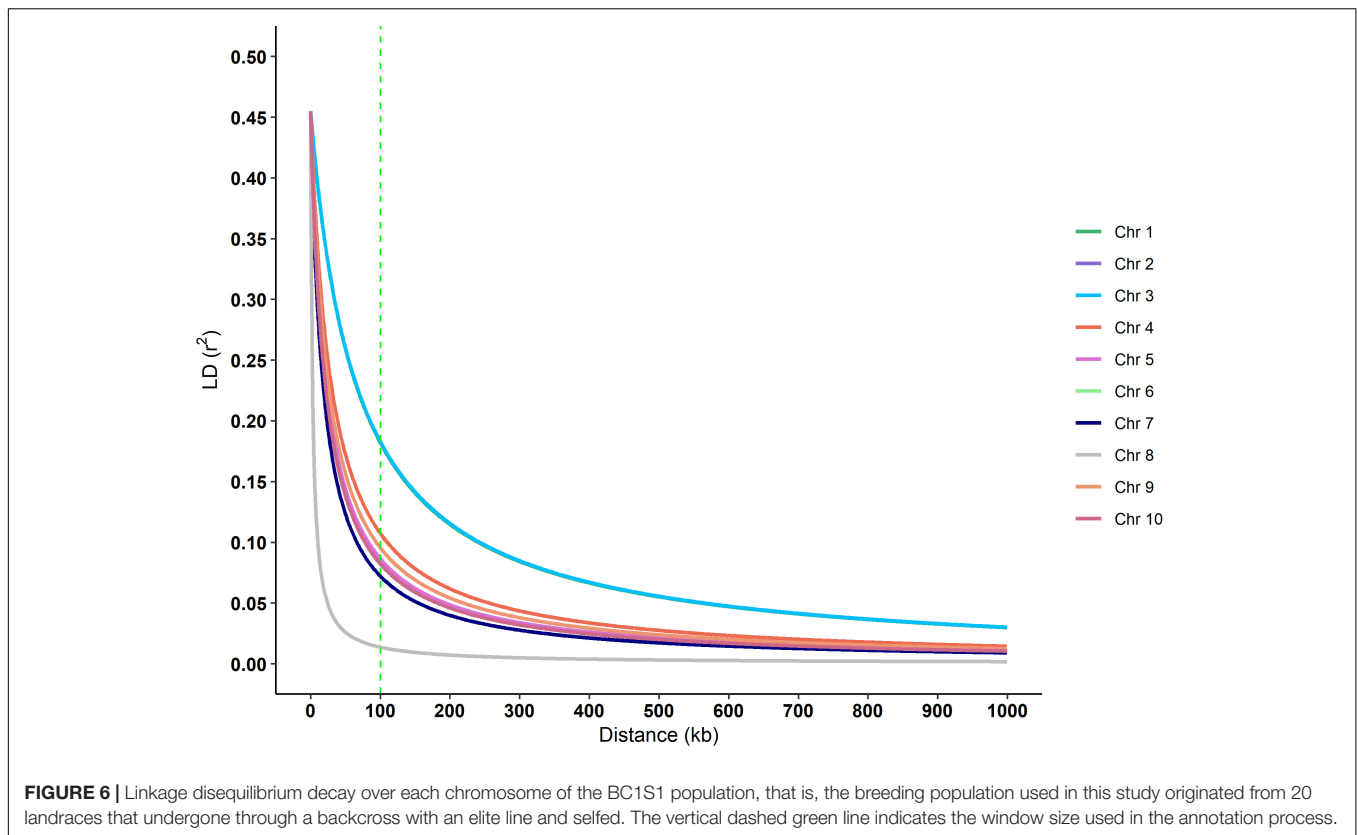
The purpose of using landraces in a breeding program is to bring novel positive alleles to a breeding population. As the recurrent parent CML376 was also genotyped, it was possible to check if favorable alleles in our semi inbred experimental lines are derived from the CML376 or the landraces. The allele frequency of SNPs associated with the traits, and derived from the landraces, was calculated for the entire population, and stratified for each landrace family group. This allowed verifying if specific landrace is a potential donor for the allele of interest.

On the other hand, regarding selection, it is also important to know the distribution of the superior allele across the landrace progenies in the breeding population, regardless of its origin, landraces or the elite recurrent parent line. The superior allele is the one that causes a positive variation in the trait of interest. Therefore, we also calculated, for the significant SNPs, the frequency of the alleles with positive effects for each landrace progeny.

To evaluate the population in terms of favorable alleles content, the 20 families were ranked by the sum of the ranks of the superior allele frequency for each SNPs. Two rankings 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.

GWAS Validation

Validation was performed by two approaches. In the first, we compared the results of GWAS over BC1S1 and BC1S2. In the



second, we compare the BLUPs obtained from the TC2 yield trials with the genetic value estimative of BC1S2 individuals based on the effects of significant SNPs, obtained from BC1S1 GWAS. The genetic values estimative were calculated by multiplying the allele substitution effect vector by the marker matrix of BC1S2 individuals, as shown in the model below:

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

where \hat{y}_2 is the vector (size n) of estimated genetic values in generation BC1S2, X is the genomic matrix of BC1S2 individuals, with dimension of n by m , where, n is equal to 174, that is, all the genotypes of BC1S2 generation, and m is the significant SNPs found in the BC1S1 GWAS, and finally, α_1 is the vector of allele substitution value for each significant SNP. 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.

RESULTS

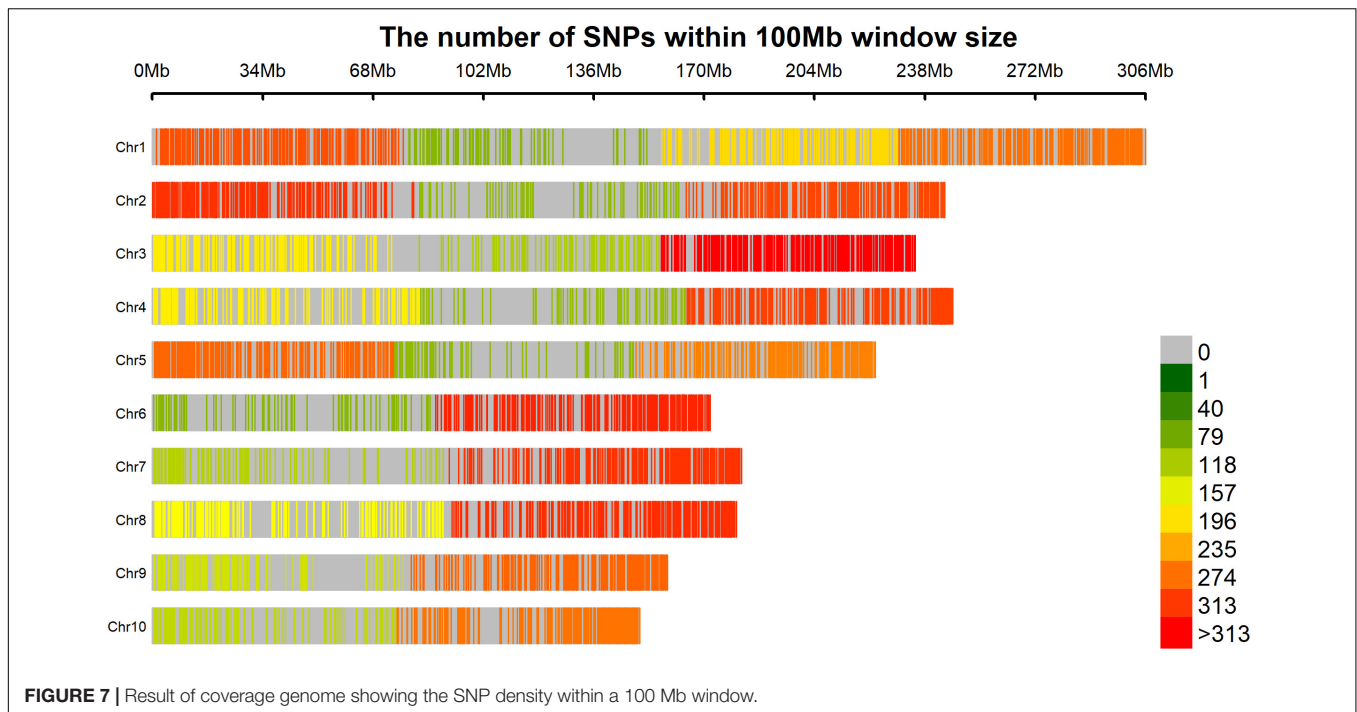
Phenotypic Analysis

Based on the variograms surface trends, specific spatial factors were chosen as fixed effects to better fit the model to each site/irrigation treatment (Table 2). In all the individual trials, the Wald test confirmed the significance of using the spatial factors.

Landraces progenies and the genetic check (Tester \times CML376) distributions of grain yield means (Supplementary Figure 1) and the drought tolerance index (Supplementary Figure 2) show that there are landrace progeny genotypes performing better than the genetic check, which is a desirable situation for selection. Significant differences among genotypes were detected for 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 for GYDR (25.5%).

The Harmonic Mean of the Relative Performance index measures drought tolerance. Its values are around 1.0, with higher values being more tolerant genotypes and lower values 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 4). The genotypes with higher drought tolerance index values are also those that presented higher yield under both irrigation treatments (Figure 2). 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 (**Supplementary Figure 3**) what was expected as the genetic diversity was higher in the first generation, before applying selection.

GWAS

The use of principal components (PC) as covariates into the GWAS model was necessary as the population structure might result in a spurious marker-trait association. This necessity was confirmed by comparing the QQ-plots of the GWAS model with no PC as covariates and the QQ-plots of models containing 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 PCs 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 (**Supplementary Figure 4**), despite 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 analysis, 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 association results (p -values of the marker-trait association in a Manhattan plot and the quantile-quantile plots) are shown in **Figure 3** for BC1S1 and in **Figure 4** for BC1S2. The quantile-quantile plots (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).

Regarding the BC1S1 results, 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 (**Figure 3A**), one on chromosome 1 (M2216) and another on chromosome 4 (M13512). For drought treatment, four associated markers were found (**Figure 3B**), two on chromosome 1 (M232 and M3547), one on chromosome 3 (M11032), and the last on chromosome 8 (M26036). For the well-watered treatment, four markers were significant (**Figure 3C**), on chromosomes 1 (M3299), 2 (M7826), and a two on chromosome 7 (M21270 and M22252). 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 5**).

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 5**). 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.

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 (Table 5) were significant in the BC1S2-TC2 GWAS (Table 6 and Figure 4). 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 the BC1S2 estimates 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.

Gene Annotation and Allele Origin

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 7). 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 6), 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 50–300k 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 other chromosomes presented 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 4 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. Clearly, 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.

Regarding the allele origin, the frequency of landrace alleles in the BC1S1 lines ranges from 5 to 31%. The SNPs M3547 and M7826 present the highest frequencies for the landrace allele among the ten significant SNPs (Supplementary Figure 5). The favorable allele for four SNPs (M3547, M11032, M26036, and M3299) came from the landraces. Populations originating from the NVOL46, COAH20, COAH21, CHIH87, and COAH78 landraces are those with a higher frequency of the superior allele for these 4 SNPs (Landrace origin rank on Table 1 and Supplementary Figure 5). Regardless of the origin of the superior allele, the five landrace populations that present a higher frequency of superior allele were those obtained from SNLP17, NVOL46, COAH78 CHIH85, and COAH20 (general rank on Table 1 and Supplementary Figure 6). This suggests that these populations have a higher frequency of favorable alleles, being a good source for drought tolerance.

DISCUSSION

Field Trials and Phenotypes

All the field trials presented strong spatial (row and columns) trends. Therefore, the spatial modeling used in the first stage analysis was crucial to avoid, or minimize, their effects on the means, mainly in the WW treatments, where augmented design was carried out. Thus, we expected lower heritabilities in these trials (WW) than in the DR treatments, even though drought-stressed trials usually are more affected by field spatial trends (Badu-Apraku et al., 2004), resulting in lower heritability for yield. That was one of the reasons we chose to use two repetitions in the DR treatments, to be able to estimate better the residual variance and to avoid the possible complete loss of genotypes, since under stress the number of missing plots is often higher, which would greatly affect the genetic estimates in case there is only one plot of each genotype. Another reason that would explain the greater heritability in drought than in the irrigated condition may be the fact that a diverse population when tested in a stressful environment may have its individuals better discriminated than if tested in a stress-free environment, that is, genetic diversity may be more evident in the phenotypic response when in a stressful environment, which may be reflected in heritability.

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 is not feasible to perform a testcross trial with more than 1,000 genotypes crossed to a group of testers. In the second testing year this was minimized, as with only 174 BC1S2 lines, testcrossing with three testers was possible.

Similar values are found when comparing the grain yield means and the drought tolerance index for the BC1 progeny testcrosses and the genetic checks (Tester crossed with CML376). However, the distribution suggests a superior performance for the BC1 progenies for grain yield (**Supplementary Figure 1**), and for the drought tolerance index (**Supplementary Figure 2**). It can be inferred 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, landraces or CML.

The usage of HMRP and other index based on harmonic mean is recognized for stress tolerance studies (Jafari et al., 2009; de Mendonça et al., 2017; Lyra et al., 2017). The Harmonic Mean Relative Performance was confirmed as a simple and efficient index that allows the selection of lines that perform well in both the DR and WW irrigation treatments, achieving the objective of these “stressed x not stressed” trials (**Figure 2**). Using the HMRP index, the breeder can select the highest performing lines in both treatments or those that perform very well in one and close to the average in the another. Additionally, the breeder can avoid selecting those that underperform in at least one of the treatments. Sites and line by site interaction effects were significant regarding the index (**Table 3**). Therefore, even if the means across the sites are similar, differential expression of the drought tolerance was identified, and selection based on results by location should be avoided.

Genotyping

We believe that the final number of SNPs (5,695) was relatively low considering the start point of 47,074 markers comparing with others maize GWAS published (Xiao et al., 2017), mainly because the strict quality control filters in terms of missing data, however, the markers were widely distributed across the genome, covering all chromosome arms without any large uncovered segment and the low density areas are usually around the centromere region (**Figure 7**). This good genome coverage is important for this kind of study and considering the good quality of the final set of 5,695 SNPs due the strict quality control process, the relative low number is not a limitation. 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 originated from 20 different landrace parents, and for the low minor allele frequency (MAF) of the SNPs associated to the traits (**Table 5**). Another important detail that

should be considered is that the germplasm of the reference genome used to obtain markers positions (B73 RefGen_v4) has a temperate origin and certainly presents incongruities with landrace derived materials, however, it is the best reference genome available at that moment. It is important to remember that maize landraces, due to heavy transposon activity and other factors, often contain large amounts of insertions, deletions, and reorganized chromosomes. Perhaps, if a closer genome had been used to position the markers, we may have an even better coverage, however, because our population derived from Landraces crosses would be difficult to get any closer reference genome. The drastic reduction on the final set of markers, from 32,592 to 5,695, occurred during the quality control process; the call rate used was high (95%) to ensure a high-quality set of markers over a large set with many imputations.

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 for the annotation of the maize gene Zm00001d034385, a validated gene associated with drought tolerance. It is mapped 24M bp from the M3547 SNP in the present study that was associated with GYDR. Although it is on the same chromosome arm (chromosome 1, long arm), it is a considerable long distance away, given a LD decay of ~100k bp.

The LD decay reported in maize varies considerably among 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 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 were at the midpoint as the initial subtropical 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 material that are neither landrace nor elite inbred line, 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 that are not completely described. These two cases are SNP M232 that is associated with GYDR and is

TABLE 8 | Two significant SNPs located close to putative genes that encode proteins associated with plant response to drought.

SNP	Distance	Putative gene and protein type	References
M232	84.3k bp	Zm00001d027539 Glutathione S-transferase	McGonigle et al., 2002; Chen et al., 2012; Xu et al., 2015; Min et al., 2016
M11032	80.7k bp	Zm00001d043929 Fasciclin-like arabinogalactan	Johnson et al., 2003; Faik et al., 2006; Finkelstein, 2013; Zang et al., 2015; Wang et al., 2016; Cagnola et al., 2018

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.

located 85k bp close to the gene; and Zm00001d027539 that belongs to the glutathione S-transferase (GST) gene family (GST, EC 2.5.1.18). This second gene family plays an important role in plant response to environmental conditions (McGonigle et al., 2002) plus 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) leading to a reduction in kernel number and thus lowering yield.

These two findings (Table 8) 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 their 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 basis for further investigations. This is especially so for the M11032 SNP, where the alternative allele originates from the landrace (Figure 5) is the superior one, and the homozygous state for this allele had a slightly higher grain yield in drought treatments. The landraces ARZM12193, CHIH338, CHIH87, COAH19 are not always good sources for the superior allele, 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 NVO146, COAH20, COAH21, CHIH87, and COAH78 would be the main sources for exploring this diversity since they have a higher frequency of these favorable alleles.

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. This is to be expected since selection was made in BC1S1 to generate the BC1S2 we induced a non-random allele frequency change between the generations that can affect the results as the new generation 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 (0.04–0.17), 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 to use it to predict the genotypic values of a second testcross involving 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 still needs to be developed.

Favorable Alleles and Their Origin

The frequency of allele from landrace over the entire population are the same values of the MAF (Table 5), once the allele from landraces is, as expected, the allele in minor frequency. Landrace origin rank (Table 1 and Supplementary Figure 5) gives the idea of which landraces are the most promising source for positive variants, and it suggests that the landraces NVOL46, COAH20, COAH21, CHIH87, and COAH78 are those that could be used for further investigations regarding new variants, at the same time, the populations originated from them could be more explored by breeders to exploit this new variant superiority.

The general rank (Table 1 and Supplementary Figure 6) gives the idea of which landrace populations are the promising to selection regards the combination with adapted elite genitor. This suggests that the SNLP17, NVOL46, COAH78 CHIH85, and COAH20 populations have a higher frequency of favorable alleles and/or their combination, being a good source for drought tolerance.

CONCLUSION

The GWAS analysis was able to identify genomic regions associated with drought tolerance in early generation testcrosses, although they were not stable over generations or irrigation treatments. Two promising putative genes that encode proteins related with the physiological plant response to abiotic stress are located close to the significant SNPs found in this study. Some alleles from the landraces provide a slightly higher yield under drought treatments. These results indicate that the diversity delivered from landraces x elite inbred line backcrosses can play an important role for improving the maize tolerance to drought, and for trait introgression bringing new superior allelic diversity from landraces to breeding populations. As we expected for a complex trait such as drought tolerance, we did not find a unique

REFERENCES

Arteaga, M. C., Moreno-Letelier, A., Mastretta-Yanes, A., Vázquez-Lobo, A., Breña-Ochoa, A., Moreno-Estrada, A., et al. (2016). Genomic variation in recently collected maize landraces from Mexico. *Genomics Data* 7, 38–45. doi: 10.1016/j.gdata.2015.11.002

marker with a large genetic effect, but a handful of regions that can be transferred to advanced and adapted breeding populations using landraces as origin of positive variants.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: dataverse: <https://hdl.handle.net/11529/10548540>.

AUTHOR CONTRIBUTIONS

PB wrote the manuscript, analyzed the genetic and field data, prepared the genetic markers data, and participated in the conceptual design of the research. TM and MW created the conception of the study. TM and MA designed and carried out the experiments. VV-M and ES-I were responsible for the field operations. TM, RF-N, KS, and GG reviewed the manuscript. JB supported and reviewed the statistical and spatial analysis. CP conducted the genotyping process in the laboratory and the first marker data preparation. All authors contributed to the article and approved the submitted version.

FUNDING

This study was financed by the Sustainable Modernization of Traditional Agriculture (MasAgro) project supported by the Ministry of Agriculture and Rural Development (SADER) of the Government of Mexico. This study was also financed by the Brazilian institutions: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) – Finance Code 001, and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

ACKNOWLEDGMENTS

We acknowledge the support of all CIMMYT and Allogamous Plant Breeding Laboratory staff who were somehow involved in this project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.691211/full#supplementary-material>

Badu-Apraku, B., Fakorede, M. A. B., Menkir, A., Kamara, A. Y., and 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. doi: 10.1017/S0021859604004538

Bänziger, M., Edmeades, G. O., Beck, D., and Bellon, M. (2000). *Breeding for Drought and Nitrogen Stress Tolerance in Maize: From Theory to Practice*. Mexico: CIMMYT.

- Bernardo, R. (2010). *Breeding for Quantitative Traits in Plants*, 2nd Edn. Stemma press: Woodbury.
- Bolaños, J., and 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. doi: 10.1016/0378-4290(96)00036-36
- Burgueño, J., Cadena, A., Crossa, J., Bänziger, M., Gilmour, A. R., and Cullis, B. (2000). *Users's Guide for Spatial Analysis of Field Variety Trials using ASREML*. Mexico: CIMMYT.
- Butler, D. (2009). *ASREML R Package: Asreml() Fits the Linear Model*.
- Cagnola, J. L., Dumont, de Chassart, G. J., Ibarra, S. E., Chimenti, C., Ricardi, M. M., et al. (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. doi: 10.1111/pce.13136
- Cataneo, A. C., Déstro, G. F. G., Ferreira, L. C., Chamma, K. L., and Sousa, D. C. F. (2003). Atividade de glutatona S-transferase na degradação do herbicida glyphosate em plantas de milho (*Zea mays*). *Planta Daninha* 21, 307–312. doi: 10.1590/S0100-83582003000200017
- Cetinkaya, H., Tasci, E., and Dinler, B. S. (2014). *Regulation of Glutathione S-transferase Enzyme Activity with Salt Pre-treatment Under Heat Stress in Maize Leaves*. *Res. Plant Biol.* 4, 45–56.
- Chen, J.-H., Jiang, H.-W., Hsieh, E.-J., Chen, H.-Y., Chien, C.-T., Hsieh, H.-L., et al. (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. doi: 10.1104/pp.111.181875
- Clarke, G. P. Y., and Stefanova, K. T. (2011). Optimal design for early-generation plant-breeding trials with unreplicated or partially replicated test lines. *Aust. New Zeal. J. Stat.* 53, 461–480. doi: 10.1111/j.1467-842X.2011.00642.x
- Cullis, B., Gogel, B., Verbyla, A., and Thompson, R. (1998). Spatial analysis of multi-environment early generation variety trials. *Biometrics* 54, 1–18.
- Dai, A. (2013). Increasing drought under global warming in observations and models. *Nat. Clim. Chang.* 3, 52–58. doi: 10.1038/nclimate1633
- de Mendonça, L. F., Granato, Í. C., Alves, F. C., Morais, P. P. P., Vidotti, M. S., and Fritsche-Neto, R. (2017). Accuracy and simultaneous selection gains for N-stress tolerance and N-use efficiency in maize tropical lines. *Sci. Agric.* 74, 481–488. doi: 10.1590/1678-992x-2016-2313
- Dutkowski, G. W., Silva, J. C. E., Gilmour, A. R., and Lopez, G. A. (2002). Spatial analysis methods for forest genetic trials. *Can. J. For. Res.* 32, 2201–2214. doi: 10.1139/x02-111
- Edmeades, G. O., Trevisan, W., Prasanna, B. M., and Campos, H. (2017). “Tropical Maize (*Zea mays* L.),” in *Genetic Improvement of Tropical Crops*, eds H. Campos and P. D. S. Caligari (Cham: Springer International Publishing), 331. doi: 10.1007/978-3-319-59819-2
- Faik, A., Abouzouhair, J., and 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. doi: 10.1007/s00438-006-0159-z
- Faostat, (2018). *Database of Food and Agriculture Organization of the United Nations Statistics Division*. Rome: FAO.
- Finkelstein, R. (2013). Abscisic acid synthesis and response. *Arab. B* 11:e0166. doi: 10.1199/tab.0166
- Fisher, M., Abate, T., Lunduka, R. W., Asnake, W., Alemayehu, Y., and 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. doi: 10.1007/s10584-015-1459-1452
- Flint-Garcia, S. A., Thornsberry, J. M., and Buckler, E. S. (2003). Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.* 54, 357–374. doi: 10.1146/annurev.arplant.54.031902.134907
- Fritsche-Neto, R., Souza, T. L. P. O., de, Pereira, H. S., Faria, L. C., de, et al. (2019). Association mapping in common bean revealed regions associated with anthracnose and angular leaf spot resistance. *Sci. Agric.* 76, 321–327. doi: 10.1590/1678-992x-2017-2306
- Gilmour, A. R., Cullis, B. R., Verbyla, A. P., Journal, S., Statistics, E., Gilmour, A. R., et al. (1997). *Accounting for Natural and Extraneous Variation in the Analysis of Field Experiments*. Washington, DC: International Biometric Society Stable.
- Goodman, M. M. (1985). Exotic maize germplasm: status, prospects, and remedies. *Iowa State J. Res.* 59, 497–527.
- Goodman, M. M., Holland, J. B., and Sanchez Gonzalez, J. J. (2014). “Breeding and diversity,” in *Genetics, Genomics and Breeding of Maize*, eds R. Wusirikia, M. Bohn, J. Lai, and C. Kole (New York, NY: CRC Press), 15–40. doi: 10.1201/b17274
- Granato, I., and Fritsche-Neto, R. (2018). *snpReady: Preparing Genotypic Datasets in Order to Run Genomic Analysis. R package version 0.9.6*. Available online at: <https://cran.r-project.org/package=snpReady>.
- Hall, A. J., Vilella, F., Trapani, N., and Chimenti, C. (1982). The effects of water stress and genotype on the dynamics of pollen-shedding and silking in maize. *F. Crop. Res.* 5, 349–363. doi: 10.1016/0378-4290(82)90036-90033
- Harrison, M. T., Tardieu, F., Dong, Z., Messina, C. D., and Hammer, G. L. (2014). Characterizing drought stress and trait influence on maize yield under current and future conditions. *Glob. Chang. Biol.* 20, 867–878. doi: 10.1111/gcb.12381
- Henshall, J. M. (2013). *Genome-Wide Association Studies and Genomic Prediction*. Berlin: Springer. doi: 10.1007/978-1-62703-447-440
- Hill, W. G., and Weir, B. S. (1988). Variances and covariances of squared linkage disequilibria in finite populations. *Theor. Popul. Biol.* 33, 54–78. doi: 10.1016/0040-5809(88)90004-90004
- Hoegh-Guldberg, O., Jacob, D., Taylor, M., Bindi, M., Brown, S., Camilloni, I., et al. (2018). “Impacts of 1.5°C global warming on natural and human systems,” in *Global warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*, eds V. Masson-Delmotte, P. Zhai, H. O. Pörtner, D. Roberts, J. Skea, P. R. Shukla, et al. (Geneva: IPCC)
- Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., et al. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* 42, 961–967. doi: 10.1038/ng.695
- Hwang, E.-Y., Song, Q., Jia, G., Specht, J. E., Hyten, D. L., Costa, J., et al. (2014). A genome-wide association study of seed protein and oil content in soybean. *BMC Genomics* 15:1. doi: 10.1186/1471-2164-15-11
- IPCC (2007). *Summary for Policymakers*. Cambridge: IPCC.
- 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: IPCC.
- Jafari, A., Paknejad, F., and Jami Al-Ahmadi, M. (2009). Evaluation of selection indices for drought tolerance of corn (*Zea mays* L.) hybrids. *Int. J. Plant Prod.* 3, 33–38. doi: 10.22069/ijpp.2012.661
- Jaiswal, V., Gupta, S., Gahlaut, V., Muthamilarasan, M., Bandyopadhyay, T., Ramchiary, N., et al. (2019). Genome-Wide association study of major agronomic traits in foxtail millet (*Setaria italica* L.) using ddRAD sequencing. *Sci. Rep.* 9:5020. doi: 10.1038/s41598-019-41602-41606
- Johnson, K. L., Jones, B. J., Bacic, A., and Schultz, C. J. (2003). The fasciclin-like arabinogalactan proteins of *Arabidopsis*. a multigene family of putative cell adhesion molecules. *Plant Physiol.* 133, 1911–1925. doi: 10.1104/pp.103.031237
- Liu, X., Huang, M., Fan, B., Buckler, E. S., and Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet.* 12:e1005767. doi: 10.1371/journal.pgen.1005767
- Lu, G. H., Tang, J. H., Yan, J. B., Ma, X. Q., Li, J. S., Chen, S. J., et al. (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. doi: 10.1111/j.1744-7909.2006.00289.x
- Lu, Y., Cai, H., Jiang, T., Sun, S., Wang, Y., Zhao, J., et al. (2019). Assessment of global drought propensity and its impacts on agricultural water use in future climate scenarios. *Agric. For. Meteorol.* 278:107623. doi: 10.1016/j.agrformet.2019.107623
- Lu, Y., Shah, T., Hao, Z., Taba, S., Zhang, S., Gao, S., et al. (2011). Comparative SNP and haplotype analysis reveals a higher genetic diversity and rapid LD decay in tropical than temperate germplasm in maize. *PLoS One* 6:e24861. doi: 10.1371/journal.pone.0024861
- Lyra, D. H., de Freitas Mendonça, L., Galli, G., Alves, F. C., Granato, Í. C., and Fritsche-Neto, R. (2017). Multi-trait genomic prediction for nitrogen response indices in tropical maize hybrids. *Mol. Breed.* 37:80. doi: 10.1007/s11032-017-0681-1
- Mao, H., Wang, H., Liu, S., Li, Z., Yang, X., Yan, J., et al. (2015). A transposable element in a NAC gene is associated with drought tolerance in maize seedlings. *Nat. Commun.* 6:8326. doi: 10.1038/ncomms9326

- McGonigle, B., Keeler, S. J., Lau, S.-M. C., Koeppe, M. K., and 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. doi: 10.1104/pp.124.3.1105
- Meseka, S., Menkir, A., and 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. doi: 10.1007/s10681-014-1214-1211
- Min, H., Chen, C., Wei, S., Shang, X., Sun, M., Xia, R., et al. (2016). Identification of drought tolerant mechanisms in maize seedlings based on transcriptome analysis of recombination inbred lines. *Front. Plant Sci.* 7:1080. doi: 10.3389/fpls.2016.01080
- Miti, F., Pangirayi, T., and Derera, J. (2010). S1 selection of local maize landraces for low soil nitrogen tolerance in Zambia. *African J. Plant Sci.* 8, 001–015.
- Muller, B. U., Schutzenmeister, A., and Piepho, H. P. (2010). Arrangement of check plots in augmented block designs when spatial analysis is used. *Plant Breed.* 129, 581–589. doi: 10.1111/j.1439-0523.2010.01803.x
- Navarro, J. A. R., Willcox, M., Burgueño, J., Romay, C., Swarts, K., Trachsel, S., et al. (2017). A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nat. Publ. Gr.* 49, 476–480. doi: 10.1038/ng.3784
- Portwood, J. L., Woodhouse, M. R., Cannon, E. K., Gardiner, J. M., Harper, L. C., Schaeffer, M. L., et al. (2019). MaizeGDB 2018: the maize multi-genome genetics and genomics database. *Nucleic Acids Res.* 47, D1146–D1154. doi: 10.1093/nar/gky1046
- Prasanna, B. M. (2012). Diversity in global maize germplasm: characterization and utilization. *J. Biosci.* 37, 843–855. doi: 10.1007/s12038-012-9227-9221
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna: R Core Team.
- Remington, D. L., Thornsberry, J. M., Matsuoka, Y., Wilson, L. M., Whitt, S. R., Doebley, J., et al. (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl. Acad. Sci. U.S.A.* 98, 11479–11484. doi: 10.1073/pnas.201394398
- Rengel, D., Arribat, S., Maury, P., Martin-Magniette, M.-L., Hourlier, T., Laporte, M., et al. (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. doi: 10.1371/journal.pone.0045249
- Resende, M. D. V. (2004). Métodos Estatísticos ótimos na análise de experimentos de campo. *Documentos (Embrapa)* 100:65.
- Ruiz Corral, J. A., Sánchez González, J. de J., Hernández Casillas, J. M., Willcox, M. C., Ramírez Ojeda, G., Ramírez Díaz, J. L., et al. (2013). Identificación de razas mexicanas de maíz adaptadas a condiciones deficientes de humedad mediante datos biogeográficos. *Rev. Mex. Ciencias Agrícolas* 4, 829–842. doi: 10.29312/remexca.v4i6.1152
- Sanchez, J. J., Goodman, M. M., and Stuber, C. W. (2000). Isozymatic and morphological diversity in the races of maize of Mexico. *Econ. Bot.* 54, 43–59. doi: 10.1007/BF02866599
- Sansaloni, C., Petrolí, C., Jaccoud, D., Carling, J., Detering, F., Grattapaglia, D., et al. (2011). Diversity arrays technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. *BMC Proc.* 5:P54. doi: 10.1186/1753-6561-5-S7-P54
- Sayadi Maazou, A.-R., Tu, J., Qiu, J., and Liu, Z. (2016). Breeding for drought tolerance in maize (*Zea mays* L.). *Am. J. Plant Sci.* 07, 1858–1870. doi: 10.4236/ajps.2016.714172
- Schafer, J. F. (1971). Tolerance to plant disease. *Annu. Rev. Phytopathol.* 9, 235–252. doi: 10.1146/annurev.py.09.090171.001315
- Setter, T. L., Yan, J., Warburton, M., Ribaut, J. M., Xu, Y., Sawkins, M., et al. (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. doi: 10.1093/jxb/erq308
- Spinoni, J., Barbosa, P., Bucchignani, E., Cassano, J., Cavazos, T., Christensen, J. H., et al. (2020). Future global meteorological drought hot spots: a study based on CORDEX data. *J. Clim.* 33, 3635–3661. doi: 10.1175/jcli-d-19-0084.1
- Struik, P., and Makonnen, T. (1992). Effects of timing, intensity and duration of pollination on kernel set and yield in maize (*Zea mays* L.) under temperate conditions. *Netherlands J. Agric. Sci.* 40, 409–429.
- Tarter, J. A., Goodman, M. M., and Holland, J. B. (2004). Recovery of exotic alleles in semixotic maize inbreds derived from crosses between Latin American accessions and a temperate line. *Theor. Appl. Genet.* 109, 609–617. doi: 10.1007/s00122-004-1660-1666
- Uribelarrea, M., Cárcova, J., Otegui, M. E., and Westgate, M. E. (2002). Pollen production, pollination dynamics, and kernel set in maize. *Crop Sci.* 42, 1910–1918. doi: 10.2135/cropsci2002.1910
- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-2980
- Vargas, M., Van Eeuwijk, F. A., Crossa, J., and 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. doi: 10.1007/s00122-005-0204-z
- Wang, H., and Qin, F. (2017). Genome-Wide association study reveals natural variations contributing to drought resistance in crops. *Front. Plant Sci.* 8:110. doi: 10.3389/fpls.2017.01110
- Wang, X., Cai, X., Xu, C., Wang, Q., and Dai, S. (2016). Drought-responsive mechanisms in plant leaves revealed by proteomics. *Int. J. Mol. Sci.* 17:1706. doi: 10.3390/ijms17101706
- Wimmer, V., Albrecht, T., Auinger, H.-J., and 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., and Yan, J. (2017). Genome-wide association studies in Maize: praise and stargaze. *Mol. Plant* 10, 359–374. doi: 10.1016/j.molp.2016.12.008
- Xu, J., Xing, X. J., Tian, Y. S., Peng, R. H., Xue, Y., Zhao, W., et al. (2015). Transgenic *Arabidopsis* plants expressing tomato glutathione S-transferase showed enhanced resistance to salt and drought stress. *PLoS One* 10:e0136960. doi: 10.1371/journal.pone.0136960
- Xu, Y., Li, P., Zou, C., Lu, Y., Xie, C., Zhang, X., et al. (2017). Enhancing genetic gain in the era of molecular breeding. *J. Exp. Bot.* 68, 2641–2666. doi: 10.1093/jxb/erx135
- Xue, Y., Warburton, M. L., Sawkins, M., Zhang, X. X., Setter, T., Xu, Y., et al. (2013). Genome-wide association analysis for nine agronomic traits in maize under well-watered and water-stressed conditions. *Theor. Appl. Genet.* 126, 2587–2596. doi: 10.1007/s00122-013-2158-x
- Yan, J., Shah, T., Warburton, M. L., Buckler, E. S., McMullen, M. D., and Crouch, J. (2009). Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PLoS One* 4:e8451. doi: 10.1371/journal.pone.0008451
- Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., et al. (2020). *rMVP: Memory-Efficient, Visualize-Enhanced, Parallel-Accelerated GWAS Tool*. R package version 1.0.1. Available online at: <https://cran.r-project.org/package=rMVP> (accessed July 4, 2020).
- Yu, Y. G. (1994). RFLP and microsatellite mapping of a gene for soybean mosaic virus resistance. *Phytopathology* 84:60. doi: 10.1094/Phyto-84-60
- Zang, L., Zheng, T., Chu, Y., Ding, C., Zhang, W., Huang, Q., et al. (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.* 6:1140. doi: 10.3389/fpls.2015.01140

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Barbosa, Fritsche-Neto, Andrade, Petrolí, Burgueño, Galli, Willcox, Sonder, Vidal-Martínez, Sifuentes-Ibarra and Molnar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.