FEDERAL UNIVERSITY OF CEARÁ
AGRICULTURAL SCIENCES CENTER
DEPARTMENT OF PLANT SCIENCE
POSTGRADUATE COURSE IN AGRONOMY/PLANT SCIENCE

FERNANDA CARLA FERREIRA DE PONTES

COMBINING GENOTYPING APPROACHES IMPROVES RESOLUTION FOR ASSOCIATION MAPPING: A CASE STUDY IN TROPICAL MAIZE UNDER WATER STRESS CONDITIONS

FORTALEZA

2024

FERNANDA CARLA FERREIRA DE PONTES

COMBINING GENOTYPING APPROACHES IMPROVES RESOLUTION FOR ASSOCIATION MAPPING: A CASE STUDY IN TROPICAL MAIZE UNDER WATER STRESS CONDITIONS

Thesis presented to Postgraduate program in Agronomy/Plant Science of the Federal University of Ceará, as part of the requirements for obtaining the title of *Doctor Scientiae* in Agronomy/Plant Science. Concentration area: Plant Science. Research Line: Genetics and Plant Breeding.

Advisor: Prof. D.Sc. Júlio César do Vale Silva
Co-advisor: Prof. D.Sc Roberto Fritsche Neto

FORTALEZA

2024

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

P858c Pontes, Fernanda Carla Ferreira de.
Combining genotyping approaches improves resolution for association mapping: a case study in tropical maize under water stress conditions / Fernanda Carla Ferreira de Pontes. – 2024.
63 f. : il. color.

Tese (doutorado) – Universidade Federal do Ceará, Centro de Ciências Agrárias, Programa de Pós-Graduação em Agronomia (Fitotecnia), Fortaleza, 2024.
Orientação: Prof. Dr. Júlio César do Vale Silva.
Coorientação: Prof. Dr. Roberto Fritsche Neto.

1. SNP-Array. 2. Genotyping by Sequencing. 3. Simulated genome. 4. GWAS. I. Título.
CDD 630

FERNANDA CARLA FERREIRA DE PONTES

COMBINING GENOTYPING APPROACHES IMPROVES RESOLUTION FOR ASSOCIATION MAPPING: A CASE STUDY IN TROPICAL MAIZE UNDER WATER STRESS CONDITIONS

Thesis presented to Postgraduate program in Agronomy/Plant Science of the Federal University of Ceará, as part of the requirements for obtaining the title of *Doctor Scientiae* in Agronomy/Plant Science. Concentration area: Plant Science. Research Line: Genetics and Plant Breeding.

Approved in: 07/03/2024.

EXAMINATION BOARD

Prof. D. Sc. Júlio César do Vale Silva (Advisor)
Federal University of Ceará (UFC)

Prof. D. Sc Roberto Fritsche Neto (Co-advisor)
Louisiana State University (LSU)

D.Sc José Felipe Gonzaga Sabadin
Virginia Polytechnic Institute and State University (Virginia Tech)

D. Sc Rafael Massahiro Yassue
GDM Seeds

D. Sc Karina Lima Reis Borges
Louisiana State University (LSU)

Dedico este trabalho primeiramente a Deus, meu Senhor e meu tudo.

À Nossa Senhora, medianeira de todas as graças em minha vida.

À Maria do Rosário, minha mãe e melhor amiga, a quem devo a vida e todas as conquistas.

AGRADECIMENTOS

A Deus, criador de todas as coisas e dono de todo saber, pela coragem, paciência e força para enfrentar e superar as dificuldades me fazendo encontrar o caminho certo para alcançar meus objetivos. Sem a graça e misericórdia d'Ele não seria possível.

Aos meus pais, Carlos Fernando e Maria do Rosário, que apesar da distância sempre torcem e incentivam para que eu possa ir mais longe. Especialmente a minha mãe, pelo seu amor incondicional, por ser meu suporte, refúgio e impulso.

A Universidade Federal do Ceará (UFC) e ao programa de Pós-Graduação em Agronomia/Fitotecnia (PPGAF) pela oportunidade de realizar o curso de doutorado.

Ao professor Júlio César do Vale Silva pela orientação, disponibilidade, paciência e ajuda para tornar esta tese possível, muitas vezes em meio às adversidades. Obrigada pela convivência e por tudo que tive a oportunidade de aprender ao longo desses anos.

Ao professor Roberto Fritsche-Neto pela coorientação, disponibilidade e transmissão de conhecimentos. Obrigada por ter aceitado me receber em seu laboratório na mobilidade acadêmica, no qual fui muito bem acolhida.

A todos os professores que passaram pela minha vida pelos ensinamentos e experiências transmitidas.

Aos membros da banca examinadora, José Felipe Gonzaga Sabadin, Karina Lima Reis Borges e Rafael Massahiro Yassue pela aceitação do convite e excelentes contribuições.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, for financial support through the granting of the scholarship.

To Limagrain, for genotyping the study material.

A toda minha família, especialmente primos (as), tios (as) e minha avó, Josefa Félix de Pontes, exemplo de força e perseverança. Agradeço as orações que com certeza me sustentaram.

Aos colegas do grupo de trabalho do NEMeV (Núcleo de Estudos em Melhoramento Vegetal) coordenado pelo professor Júlio César. Anderson Reinbold, Ingrid Pinheiro, Lucas Lobo, Maria Valnice e Marcos Adriano pela colaboração durante a condução e colheita dos experimentos.

Aos colegas do Laboratório de Melhoramento de Plantas Alógamas (ESALQ/USP), coordenado professor Roberto Fritsche-Neto, pela inspiração como profissionais e pelo acolhimento. Especialmente a Gabriela, Germano, Humberto, Karina, Lucas, Pedro, Rafael, Ronaldo, Sabadin, Raísa, Giovanni e Miguel.

Aos amigos que compartilharam comigo a jornada acadêmica, Andreza Melo, Arnaldo Sales, Beatriz Machado, Caris Viana, Charles Lobo, Francisco Linco, Ingrid Pinheiro, Jéssica Soares, Johny de Souza, Letícia Bessa, Liliane Nunes, Lucas Lobo, Rafael Santiago, Talytha Ravenna, Raylson Melo (*In memoriam*) e Wendson Moraes (*In memoriam*) pela convivência, cafés, conhecimentos compartilhados, palavras de ânimo e incentivo, dividindo angústias e alegrias deixando a pós-graduação mais leve. Em especial a minha amiga, Ingrid Pinheiro, pela parceria e companheirismo, sempre disposta a ajudar, esteve comigo dividindo momentos bons e tristes. A minha amiga, Jéssica Soares, sempre com palavras de apoio nas horas difíceis, obrigada pelo carinho e amizade. Aos amigos, Johny de Souza e Beatriz Machado, pelas conversas, companheirismo e amizade.

Aos funcionários da Fitotecnia, particularmente a Dona Elisa, sempre disposta a fazer o café, combustível diário das boas conversas. Ao secretário do PPGAF, Vitor Hugo, pela ajuda e conversas. Aos engenheiros agrônomos, Antônio Moreira, Francisco Vieira, Pedro D'jacy e Tomil Ricardo pela ajuda nos experimentos. Especialmente ao Vieira, o qual nutro uma amizade desde o mestrado e que ao longo dos anos se tornou um amigo fiel. Obrigada por sua ajuda, carinho e companheirismo.

Aos funcionários terceirizados pela ajuda durante a execução do experimento. Principalmente ao Seu Bezerra, pela ajuda dentro e fora do campo. A nossa amizade ultrapassou os portões da UFC. Obrigada por seu carinho e apoio.

A todos os meus amigos que apesar da distância estiveram ao meu lado em todas as horas, apoiando, rezando e torcendo pelo meu sucesso.

Enfim, obrigada a todos que contribuíram e participaram de alguma forma desta etapa.

Everything has its time; there is a time for every purpose under heaven. (Ecclesiastes 3,1).

Moreover, we know that all things work together for good to those who love God, to those who are called according to His purpose. (Romans 8,28).

ABSTRACT

Genome-Wide Association Studies (GWAS) identify genome variations related to specific phenotypes, typically analyzed by Single Nucleotide Polymorphism (SNP) markers. Genotyping platforms such as those involving genomic hybridization microarray (SNP-Chip or SNP-Array) or sequencing-based genotyping techniques (GBS) are effective in genotyping various samples with hundreds of thousands of SNPs. However, these approaches can introduce bias in tropical maize germplasm analyses, as the temperate line B73 is commonly used as the reference genome. Therefore, an alternative to overcome this limitation is using a simulated genome called "Mock," which is adapted to the population and created with bioinformatics tools. A few recent studies have shown that SNP-Array, GBS, and Mock yield similar results concerning population structure, definition of heterotic groups, tester selection, and genomic hybrid prediction. However, no studies have been identified thus far regarding the results generated by these different genotyping approaches for GWAS. Therefore, this study aims to test the equivalence among the three genotyping scenarios in identifying significant effect genes in GWAS. To achieve this, maize was used as the model species, where 360 inbred lines from a public panel were genotyped by SNP-Array via the Affymetrix platform and GBS. The GBS data were used to perform SNP calling using the temperate inbred line B73 as the reference genome (GBS-B73) and a simulated genome "Mock" obtained *in-silico* (GBS-Mock). The study encompassed four above-ground traits with plants grown under two levels of water supply: well-watered (WW) and water-stressed (WS). In total, 46, 34, and 31 SNP were identified in the SNP-Array, GBS-B73, and GBS-Mock scenarios, respectively, across the two water levels. Overall, the candidate genes identified varied along the scenarios but had the same functionality. Regarding SNP-Array and GBS-B73, genes with functional similarity were identified even without coincidence in the physical position of the SNPs. These genes and regions are involved in various processes and responses with applications in plant breeding. In terms of accuracy, the combination of genotyping scenarios compared to those isolated is feasible and recommended, as it increased all traits under both water supply conditions. In this sense, it is worth highlighting the combination of GBS-B73 and GBS-Mock scenarios, not only due to the increase in the resolution of GWAS results but also due to the reduction of costs associated with genotyping as well as the possibility of conducting genomic breeding methods.

Keywords: SNP-Array; genotyping by sequencing; simulated genome; GWAS.

RESUMO

Estudos de genética de associação (GWAS) identificam variações no genoma relacionadas a fenótipos específicos, geralmente analisadas por marcadores SNP (*Single nucleotide polymorphisms*). Plataformas de genotipagem como aquelas que envolvem a hibridização genômica de *microarray* (SNP-Chip *ou* SNP-*Array*) ou técnicas de genotipagem por sequenciamento (GBS) são eficazes para genotipar várias amostras com centenas de milhares de SNP. No entanto, essas abordagens podem causar viés em análises de germoplasma de milho tropical, pois geralmente se utiliza a linhagem temperada B73 como genoma de referência. Assim, uma alternativa para contornar esse entrave é o uso de um genoma simulado denominado "Mock", adaptado à população e criado com ferramentas de bioinformática. Alguns poucos estudos demonstraram recentemente que SNP-*Array*, GBS e Mock geram resultados semelhantes no que diz respeito a estruturação de população, definição de grupos heteróticos, escolha de testadores até a predição genômica de híbrido. Contudo, não foram identificados estudos até o momento sobre os resultados gerados por essas diferentes abordagens de genotipagem quanto a GWAS. Portanto, o objetivo do estudo foi verificar a equivalência entre os três cenários de genotipagem na identificação de genes de efeito significativo em GWAS. Para isso, usou-se o milho como espécie modelo, na qual 360 linhagens endogâmicas de um painel público foram genotipadas por SNP-*Array* via plataforma Affymetrix e GBS. Os dados de GBS foram usados para realizar a chamada SNP utilizando a linhagem endogâmica temperada B73 (GBS-B73) como genoma de referência e, um genoma simulado "Mock" obtido *in sílico* (GBS-Mock). O estudo contemplou quatro caracteres da parte aérea com plantas crescidas em dois níveis de suprimento de água: bem irrigado (WW) e estresse hídrico (WS). No total, foram identificados 46, 34 e 31 SNP nos cenários SNP-*Array*, GBS-B73 e GBS-Mock, respectivamente, nos dois níveis de suplementação hídrica. De forma geral, observou-se entre os cenários a identificação de genes candidatos diferentes, mas que apresentam a mesma funcionalidade. Em relação a SNP-*Array* e GBS-B73, foram identificados genes com semelhança funcional mesmo sem coincidência na posição física dos SNP. Esses genes ou regiões estão envolvidos em diversos processos e respostas com aplicações no melhoramento vegetal. Em termos de acurácia, a combinação de cenários de genotipagem em comparação a aqueles isolados, é viável e recomendada, pois resultou em aumento para todos os caracteres nas duas condições de suprimento hídrico. Neste sentido, vale destacar a combinação dos cenários GBS-B73 e GBS-Mock, não apenas devido ao incremento na

resolução dos resultados de GWAS, mas também pela redução de custos associados à genotipagem bem como a possibilidade de conduzir métodos de melhoramento genômico.

Palavras-chave: SNP-*Arrray*; genotipagem por sequenciamento; genoma simulado; GWAS.

LIST OF FIGURES

LIST OF TABLES

LIST OF SUPPLEMENTARY TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

ASE	Average Effect of Allele Substitution
BLAST	Basic Local Alignment Search Tool
BLUP	Best Linear Unbiased Prediction
CR	Call Rate
dBLUPs	de-regressed BLUPs
FarmCPU	Fixed and Random Model Circulating Probability Unification
GBS	Genotyping by Sequencing
GWAS	Genome-Wide Association Study
LD	Linkage Disequilibrium
LRT	Likelihood-Ratio Test
MAF	Minor Allele Frequency
NGS	Next Generation Sequencing
PCA	Principal Component Analysis
PH	Plant Height
QTL	Quantitative Trait Locus
REML	Restricted Maximum Likelihood
SD	Stalk Diameter
SDM	Shoot Dry Matter
SPAD	Portable Chlorophyll Meter
SNP	Single Nucleotide Polymorphism
UFC	Federal University of Ceará
V2	Two Fully Developed Leaves
V6	Six Fully Developed Leaves
WA	Water applied
WW	Well-watered
WS	Water Stressed

CONTENTS

1 INTRODUCTION

Water is the most abundant and often the most limiting of all the resources plants need to grow and function (TAIZ et al., 2015). Water availability is considered one of the most influential factors in agricultural productivity, controlling species distribution in different climatic zones on Earth (Turner; Jones, 1980). In the tropical zone, characterized by relatively high temperatures and low rainfall compared to other zones, plants thriving in these environments are often more exposed to prolonged periods of water scarcity, especially in arid and semi-arid regions. According to climate change projections, this scenario will likely continue or worsen over the years, with potentially more drastic effects on plants (Raza *et al.*, 2019).

Stress can be considered a significant deviation from optimal life conditions (LARCHER, 2003) inducing changes and responses as the plant fails to complete its physiological processes for growth and production. The lack of adequate water supply causes greater expansion of the root system into deeper and moister zones of the soil profile, reduction in the development of cells in the aerial tissues, resulting in decreased growth and stomatal closure to reduce transpiration rate and, consequently, photosynthetic activity (FRENSCH; HSIAO, 1994; HSIAO, 1973). Control measures are complex and difficult to manage by humans, and the search for genotypes that will perform better and economically viable yields in water-limited environments has been increasingly important for genetic improvement.

Conventional breeding for water deficit conditions is still time-consuming, laborious, and costly, as experimental conditions must be carefully managed. However, in recent years, with advances in molecular biology, the development of high-throughput genotyping technologies, and progress in platform development, new opportunities have emerged to enhance this process. This is partly due to cost reduction, which has consequently driven advances in genomic sequencing; another factor is the versatility of SNP (Single Nucleotide Polymorphism) markers, most commonly used in this process (Ingvarsson; Street, 2011). SNPs are abundant markers in crop genomes and are ideal for genetic discovery research and molecular improvement (Rasheed *et al.*, 2017). According to the same authors, genotyping platforms involving Next Generation Sequencing (NGS) and SNP-Array technologies are suitable for genotyping hundreds to thousands of samples with many SNP markers in a single assay much more quickly, revolutionizing the study of genomics and molecular biology.

Genotyping techniques by sequencing or GBS (Genotyping by Sequencing) are simple and highly multiplexed systems used for constructing libraries intended for next-generation sequencing. SNP-Array is a technique that uses microarrays designed to pre-select previously identified genetic markers characterized by wide polymorphism. These markers are then incorporated into a specific platform. GBS-scored SNP platforms provide a large number of markers, although with high rates of missing data. On the other hand, Array-scored SNP platforms are of high quality but have relatively high costs (Elbasyoni *et al.*, 2018) and possible ascertainment bias if the genetic material used for array development is not related with the tested germplasm (Heslot *et al.*, 2013).

Arrays are well-designed and established in the market to assist studies and breeding programs of major commodity crops (GANAL et al., 2011). For minor crops, arrays are still rarely available, and researchers often rely on information from other crops that is already accessible. However, due to the high cost associated with array development, these platforms are preferably employed when it is possible to use a "universal" approach that applies to a wide variety of germplasms. However, this can be challenging if researchers are attempting to identify rare SNPs across various germplasms; a universal design can become large and expensive, resulting in a large number of monomorphic loci for non-target germplasm groups (THOMSON, 2014).

The advancement of model genome knowledge and the advent of next-generation sequencing techniques open up the possibility of a great leap in understanding the genome of relatively lesser-known species. The GBS pipelines are based on a reference genome or assembly of a new genome, applied to model organisms and species lacking pre-existing genomic information (DAVEY et al., 2011; POLAND et al., 2012). In cases where a reference genome is not yet available, a simulated genome can be employed for SNP discovery, which can serve as a valid alternative (MELO et al., 2016). The same authors developed a bioinformatics pipeline to construct a simulated genome called “Mock,” adapted to the population and built from GBS data. This genome is already being used in genomic studies and indicated that the Mock produces comparable results when it comes to organizing populations, identifying heterotic groups, selecting testers, and predicting genomic characteristics of hybrids compared to standard approaches (SNP-Array and GBS) (MACHADO et al., 2023; SABADIN et al., 2022). This suggests that simulated genomes, can be a good alternative, especially for species where the reference genome is not available. However, no studies have been identified

on the results generated by these different genotyping approaches in Genome-Wide Association Studies (GWAS).

Other studies have compared datasets from different high-throughput genotyping technologies in GWAS. Darrier et al. (2019) using standard platforms, GBS and SNP-Array, demonstrated efficiency in characterizing genetic diversity in barley, although accessing different regions of the genome. Despite capturing different regions, there was a positive correlation between the genetic distance matrices of both approaches, validating the use of either one for the characterization. These authors emphasized that the choice between GBS and SNP-Array genotyping platforms should be based on various factors, including the nature of the research and group preferences. For example, GBS may be preferable for studies requiring broader genomic coverage due to its ability to sequence a large number of genetic markers. Conversely, SNP-Array may be more appropriate for analyses focused on specific genome regions. Group preferences, previous experience, and practical considerations such as cost and resource availability also influence platform choice. In a study with inbred maize lines, Negro et al. (2019) concluded that GBS and SNP-Array were complementary for detecting QTL marking different haplotypes in association studies. Assuming they are complementary, combining these platforms seeks to determine if it will result in greater data accuracy.

To date, there is no study comparing GBS, SNP-Array, and simulated genome for GWAS published yet. The application of studies of this nature is crucial because they provide evidence that the information obtained from various genotyping approaches may be complementary during the genotyping process, thus demonstrating an efficient alternative for identifying polymorphisms. This, in turn, should offer better support to breeding programs that consistently grapple with the necessity of identifying more efficient and tolerant genotypes against various abiotic and biotic factors. In this context, the objectives of this study were: i) to verify if there is a difference in the identification of genes with significant effects among genotyping platforms, SNP-Array, GBS, and simulated genome (“Mock”) in GWAS; ii) once differences are confirmed, to determine if the identified genomic regions are complementary and if they provide better accuracy.

2 MATERIAL AND METHODS

To enhance the comprehensibility of the analyses conducted in this study, we present a workflow in which the experimental and data analysis components are summarized in Figure 1. The subsequent sections provide detailed explanations.

2.1 Genetic material and experimental trials

This study used maize as the model species in a public diversity panel consisting of 360 tropical inbred lines (Yassue *et al.*, 2021). The data to be explored were obtained from eight experiments conducted in 2020 and 2021, as detailed below. This study involves contrasting water supply conditions, well-watered (WW), and water stressed (WS), so a pilot experiment was conducted before to the main experiments. A water retention curve was established through regression to obtain field capacity and determine the amount of water to be provided via irrigation (DE SOUZA SILVEIRA et al., 2024). This pilot experiment involved five randomly selected lines from the panel and five levels of water supply: 100% of water applied (WA), 80% of WA, 70% of WA, 50% of WA, and 40% of WA. As a result, the WW and WS points were determined, with the 80% WA and 40% WA treatments representing these conditions, respectively.

The main experiments were conducted at experimental fields of the Department of Agriculture at UFC, Campus do Pici, Fortaleza-CE, located at 3°44’24.27” S latitude and 38°34’29.93” W longitude. The main experiments were conducted under WW and WS in augmented partially repeated block design (augmented p-rep designs), with two temporally spaced replicates (WILLIAMS et al., 2011). Five common treatments (checks) were used, randomly selected from within the panel and distributed in each block, within the WW and WS conditions (Figure 2).

These experiments were always conducted in the second semester of each year, following the rainy season in the region, a period that resembles the climate of the semi-arid zone. The sowings were carried out in plastic pots with a capacity of 2000 cm^3, containing substrate (easily reproducible) in a ratio of 3:1 (sand: earthworm humus). The use of earthworm humus was chosen due to its easy obtainability and its effectiveness in providing nutrients to the plants. The use of sand is justified by its easy acquisition, availability, and low cost.

Two seeds were sown per pot at an average depth of 3-4 cm. Thinning was performed when the seedlings reached the V2 stage, leaving only one seedling per pot (plot). At this same phenological stage, water deficit was also initiated, which continued until the V6 stage (harvest). Planting and topdressing fertilizations were based on the chemical analysis of the substrate, taking into consideration the crop recommendations, in order to isolate nutritional stress during the experimental conduct.

As the experiment was conducted in an open field, irrigation control for each experiment was carried out manually and daily. Thus, 15 random samples were used to calculate the daily average weight of the pots within each water supply level. Subsequently, the difference between the current weight and the total weight obtained at each water supply level was calculated to replenish the water volume. It is worth noting that, for each vegetative stage, the average plant weight was obtained in order to subtract it along with the current weight, thus not affecting the volume of water to be replenished.

2.2 Phenotypic data

The phenotypic evaluation was conducted when most plants reached the V6 phenological stage. The traits considered in this study were:

- Plant height (PH) - measured from the soil to the insertion of the flag leaf, measured using a graduated ruler (cm);
- Stalk diameter (SD) - average of two measurements above ground level at the second node of the stem obtained using a caliper (mm);
- Chlorophyll content estimation - using SPAD, measuring three leaves per plant to get the average.

Subsequently, the plants were cut off at ground level, placed in paper bags, and placed in forced-air oven at 65°C for 72 hours to obtain:

- Shoot dry matter (SDM)- quantified using an electronic analytical balance (0,005 g).

2.3 Phenotypic analysis

The outliers of the phenotypic data for the traits described in section 2.2 were removed. Then, the remaining data were adjusted for normality using the *bestNormalize*

package (PETERSON, 2017) and the assumptions of normal distribution were checked via the Shapiro test and Q-Q *plots*. Subsequently, equations of mixed linear models were fitted to obtain the BLUP by REML for each trait studied under WW and WS conditions, using the *sommer* package (Covarrubias-Pazaran, 2016).

These analyses were performed using the following model:

$$\boldsymbol{y} = \mathbf{X_1}\boldsymbol{t} + \mathbf{X_2}\boldsymbol{l} + \mathbf{X_3}\boldsymbol{n} + \mathbf{Z_1}\boldsymbol{b} + \mathbf{Z_2}\boldsymbol{g} + \mathbf{Z_3}\boldsymbol{i} + \boldsymbol{\varepsilon} \qquad \text{Eq. 1}$$

where, $\boldsymbol{y}$ is the vector of phenotypic values of the inbred lines panel and checks; $\mathbf{X_1}$, $\mathbf{X_2}$, and $\mathbf{X_3}$ are incidence matrices for ***t, l,*** and ***n*** fixed effects; $\mathbf{Z_1}$, $\mathbf{Z_2}$ and $\mathbf{Z_3}$ are incidence matrices for ***b***, ***g*** e ***i*** random effects; ***t*** is the water supply fixed effect vector (WW and WS conditions); ***l*** is the replicate (season) fixed effect vector within water supply; ***n*** is the number of leaves used as a covariate to correct for differences in plant vigor and development; ***b*** is the block/water supply/season random effect vector, where $b \sim N(0, I\sigma_b^2)$; ***g*** s the genotype random effect vector, where $g \sim N(0, I\sigma_g^2)$; ***i*** is the random effect vector of the genotype–water supply interaction, where $i \sim N(0, I\sigma_i^2)$; $\boldsymbol{\varepsilon}$ is the experimental error, where $\boldsymbol{\varepsilon} \sim N(0, R\sigma_e^2)$, obtained using a structured diagonal matrix to make it possible to estimate two residual variances, one for each water supply level (σ_{eWW}^2 and σ_{eWS}^2). The significance of fixed effects was assessed using the Wald test, and random effects using the likelihood ratio test.

The variance components were used to estimate the heritabilities (h^2) by the following estimator:

$$\boldsymbol{h^2} = \frac{\boldsymbol{\sigma_g^2}}{\boldsymbol{\sigma_g^2 + \frac{\sigma_{ge}^2}{s} + \frac{\left(\sigma_{eWW}^2 + \sigma_{eWS}^2\right)}{rs}}} \qquad \text{Eq. 2}$$

where h^2 refers to the entry-mean heritability; σ_g^2 is the genotypic variance of the inbred lines panel, σ_{ge}^2 is the variance of the genotype–water supply interaction; $\sigma_{eWW\,e}^2\,\sigma_{eWS}^2$ are the environmental variance components in WW and WS; *s* are levels of WW and WS; and *r* is the number of repetitions in each water supply level.

The reliability of selection for each line $[R^2(\widehat{\alpha_i})]$ was obtained by the following expression (GORJANC et al., 2015):

$$R^2(\widehat{\alpha_i}) = 1 - \frac{Var(\alpha_i - \widehat{\alpha_i})}{Var(\alpha_i)} \qquad \text{Eq. 3}$$

where $Var(\alpha_i - \widehat{\alpha_i})$ is the variance of the prediction error (PEV) of line *i* and $Var(\alpha_i)$ is the genotypic variance of the trait.

The de-regressed BLUPs (dBLUPs) were obtained by calculating the ratio between the BLUPs of each inbred line in WW and WS and their respective average reliabilities. After these analyses, 313 lines remained out of the 360 in the panel. The dBLUPs of these lines in WW and WS were used in the GWAS analyses.

2.4 Genotypic data

The lines were genotyped using two SNP genotyping platforms: Affymetrix® Axiom Maize Genotyping Array with 18.413 SNP markers (SNP-Array) and genotyping-by-sequencing (GBS) process following the sequencing protocol established by Poland et al. (2012). In this method, genomic DNA was digested by two restriction enzymes, PstI and MseI, to reduce the genome complexity. Subsequently, specific adapters for sequencing on the Illumina NextSeq 500 platform (Illumina Inc., San Diego, CA, United States) were attached to the digested fragments.

The primary GBS data were employed for two purposes: firstly, to perform SNP calling using the temperate line B73 as the reference genome (RefGen v4). Secondly, to construct a simulated reference genome (mock genome) for SNP calling, following the pipeline proposed by Melo et al. (2016), considering all the lines in the panel (Mock).

Therefore, the SNP data were subjected to three GWAS approaches: 1) SNP-Array; 2) GBS with SNP calling based on the B73 reference genome (GBS-B73); 3) GBS using the simulated genome as the reference (GBS-Mock). The SNPs for the GBS dataset was identified from raw data using the TASSEL 5.0 GBSv2 pipeline (GLAUBITZ et al., 2014), considering both GBS-B73 and GBS-Mock as reference genomes, employing the BWA aligner. Using the BWA aligner (LI; DURBIN, 2009), where the tags were aligned against the reference genome (GBS-B73 and GBS-Mock).

The SNP sets obtained in these scenarios were submitted to quality control parameters as call rate (CR) and Minor Allele Frequency (MAF) procedures, where markers with CR < 90% and MAF lower than 5%, and non-biallelic markers were removed from the datasets(Morosini *et al.*, 2017). Imputation of missing data was performed using the Beagle 5.0 algorithm Browning et al. 2018).

2.5 Population structure and LD decay

In order to minimize potential bias caused by population structure, a PCA was performed based on the additive genomic relationship matrix among the remaining 313 panel lines, following VanRaden (2008) using the *SNPRelate* package (Zheng *et al.*, 2012). FarmCPU automatically incorporated the correction via PCA in the association analysis. Two principal components were used to correct for the population structure effect, and the best fit to the model was determined based on Q-Q *plots*. The most likely number of groups within the panel was determined according to Yassue et al. (2021) as it involved the same diversity panel.

The Linkage Disequilibrium (LD) estimation between each pair of SNP within the chromosomes was calculated by the square of the allele frequency correlation (r^2), among all SNP within a distance less than 1 Mbp, and the r^2 values were plotted against the base pair distance of the SNP pair to obtain the LD decay by chromosome. This procedure was performed with all SNP retained from the quality control procedures.

2.6 Association analysis (GWAS)

GWAS were performed for each trait under WW and WS conditions using the FarmCPU method (LIU et al., 2016). The *FarmCPU.P.Threshold* function was employed to obtain the *p-threshold*, specific for each trait via a simulation process with 100 permutations. Subsequently, the cutoff point was obtained by the ratio between the *p-threshold* and the number of markers used. Subsequently, p-values (significance), MAF, and ASE (Average Effect of Allele Substitution) were obtained for each significantly associated SNP, designated hereafter as a potential candidate gene underlying the target trait. Furthermore, the coefficient of determination for each significant SNP (R^2_{SNP}) was obtained based on ASE and MAF using equations described in Da et al. (2014). Next, multiple linear regressions were established for each trait using the significant SNPs as predictor variables to quantify the influence of the markers on the expression of that trait (R^2_{TOT}). The Manhattan and Q-Q *plots* graphs were generated using the CMplot package (YIN, 2020) and the graphs showing the proportion of phenotypic variance explained by the SNP were generated using the ggplot2 package (Wickham, 2011) in the R software. Venn diagrams based on the common gene functionality for the traits at each water supply level were created using LucidChart (lucidchart.com).

2.7 Correlation among markers of different scenarios

Given the stability and efficiency of SNP-Array technology in accurately genotyping numerous markers, we conducted Pearson correlation analysis (r) among significant markers with known functions identified in GWAS within the GBS-Mock scenario and markers present in the SNP-Array scenario for each trait under both WW and WS conditions. This approach aimed to assess the concordance and potential overlap between markers identified through different genotyping methods and their associations with specific traits. By comparing these markers across scenarios, we sought to elucidate common genetic factors contributing to trait variation and explore the utility of integrating data from diverse genotyping platforms in genomic analyses related to crop improvement and adaptation to environmental stressors.

2.8 Gene annotation

A candidate gene association mapping was performed for traits with significant SNP. The physical positions of SNP for GBS-Mock were assigned using BLAST (Altschul *et al.*, 1990) to align them with the maize genome assembly for comparison purposes. These positions were used to obtain 41 bp DNA fragments on a single chromosome (Supplementary Table S1). Subsequently, exclusively for GBS-Mock, a BLAST was conducted on MaizeGDB via blastn, utilizing the B73 RefGen_v4 sequence database to locate the chromosome by inserting the DNA fragment. The MaizeGDB database and its functional information associated with each SNP based on B73 RefGen_v4 were utilized for all scenarios. After defining the region to be considered, potential candidate genes flanking each marker were identified. Candidate genes linked to each trait were determined through annotation within a sliding window of 50 kb around each significant SNP, following a conservative approach described by Yassue et al. (2021). All genes within a range of 50 kb downstream and 50 kb upstream were annotated. Subsequently, they were assessed and considered based on two criteria: proximity to the SNP and functional similarity as per databases available on the Maize eFP Browser (2023) and Maize Genomics Resource (2023).

3 RESULTS

3.1 Phenotypic analysis

In general, significant effects were detected for all sources of variation, except for the G x WA interaction, in the studied traits (Table 1). The variance components showed a similar pattern for all traits, with a predominance of genotypic variance over the residual variance of the interaction. Except for the PH trait, there was a higher residual variance for the well-watered environment than the low-water availability. The genotypic variance component ranged from 0.09 to 0.19, and the genotype x environment interaction approached zero for all traits, affecting the estimates of heritabilities and accuracy. Heritabilities ranged from moderate to high magnitude, ranging from 0.58 to 0.73. PH was the trait least influenced by the environment, showing a higher coefficient of genotypic variation, 0.197. The adjusted means fall within the same range observed in other studies.

3.2 Genotypic scenarios: number and distribution of SNP

After the quality control, heterozygous markers were eliminated using the MAF and CR procedures, resulting in 12.704 SNP markers for SNP-Array out of a total of 18.413, 11.153 out of 131.350 for GBS-B73, and 4.935 out of 46.926 for GBS-Mock, which were used in the association analyses (Table 2). Approximately 69% of the marker set remained in the SNP-Array, while 10.5% remained in the GBS-Mock and 8.5% in the GBS-B73 scenario. However, there was a balanced distribution of SNP across the chromosomes in the standard scenarios (SNP-Array and GBS-B73).

3.3 GWAS analysis

Significant SNP were found on five of the ten maize chromosomes for the SNP-Array scenario and four for GBS-B73 for the SPAD trait under the WW condition (Figure 3a; Table 3). The Q-Q *plots* showed data fitted to the model (Figures 3c, 3d, and 3e). The significant marker/trait association threshold ranged from 4.99 to 12.85 (Table 4).

A total of 46, 34, and 31 significant SNP were found for SNP-Array, GBS-B73, and GBS-Mock, respectively (Table 3). Of these, at least one SNP was common among these

scenarios (Figure 4). SPAD had the highest number of significant SNP, totaling 34, followed by PH, SDM, 27, and SD, 23. The SNP array presented more markers for SPAD and PH and GBS-B73 for SD, and there was an equivalence among the three scenarios for SDM. Overall, GBS-B73 and GBS-Mock showed some similarity in the quantity of markers.

3.4 Correlation among SNP in the GBS-Mock and SNP-Array scenarios

Our results revealed 20 significant markers identified in the GBS-Mock that positively correlated with the SNP-Array scenario to traits under different environmental conditions (Table 5). Pearson correlation coefficients (r) were observed, ranging from weak to strong. Specifically, for SDM in WW conditions, correlations ranged from 0.94 to 0.30. Similarly, SPAD values showed moderate to strong correlations with markers, ranging from 0.52 to 0.76 in WW conditions and from 0.40 to 0.76 in WS conditions. For PH, correlations were moderate, with values of 0.36 for WW and 0.51 and 0.53 for WS. Notably, SD exhibited correlations ranging from 0.35 to 0.87 in WW conditions and from 0.30 to 0.87 in WS conditions. Additionally, SDM showed moderate to strong correlations, ranging from 0.46 to 0.94 in WW conditions and 0.47 in WS conditions.

3.5 Candidate genes and functional annotations

Based on the physical location of significant SNP in the B73 reference genome for SNP-Array and GBS-B73 and the reference genome for GBS-Mock, genomic regions and candidate genes related to significant loci were identified (Table 4). In some cases, the same genes and regions were identified for a given trait under both water supply conditions. For example, *Zm00001d042735* and *Zm00001d001852* in the GBS-Mock scenario for SPAD and SD, respectively; *Zm00001d017978* located on chromosome 5 in SNP-Array for PH. Similarly, identical genes and regions were found in different scenarios, for instance, *Zm00001d031759* located on chromosome 1 was detected in SNP-Array and GBS-B73 for SPAD in WW and WS. The same gene was also identified for different traits, such as *Zm00001d005090* for SD and SDM in GBS-B73.

The genomic regions and candidate genes with similar functions were grouped considering each trait under the same water supply level across genotyping scenarios (Figure 4; Supplementary Table S2). For SNP-Array and GBS-B73, regions and genes with the same

functionality on the same chromosome were observed, such as *Zm00001d031445* and *Zm00001d027626*, both on chromosome 1, which are correlated with ethylene biosynthesis for SDM in WW. Conversely, these platforms also identified genomic regions and candidate genes on different chromosomes but with coinciding functions. For example, *Zm00001d026477* on chromosome 10 and *Zm00001d027695* on chromosome 1 are responsible for responses to abiotic stress by reactive oxygen species (ROS), jasmonic acid (JA), and ethylene; *Zm00001d044194* on chromosome 3 and *Zm00001d018127* on chromosome 5 function in the regulation of the circadian cycle for SPAD under WW; *Zm00001d017978* on chromosome 5 and *Zm00001d008952* on chromosome 8 are involved in endoglucanase activity for PH in WW; and *Zm00001d053809* on chromosome 4 and *Zm00001d042481* on chromosome 3 for GBS-B73 are associated with ubiquitin proteins for PH in WS; *Zm00001d016786* on chromosome 5 and *Zm00001d005090* on chromosome 2 act in response to water stress through abscisic acid (ABA) for SDM in WS.

In scenarios involving GBS-B73 and GBS-Mock, genomic regions and candidate genes with similar functions were identified for *Zm00001d021708* on chromosome 7 and *Zm00001d012719* on the single chromosome, related to plant responses to ABA for PH in WW; *Zm00001d014899* on chromosome 5 and *Zm00001d001852* on the single chromosome, associated with the phytohormone gibberellin for SD in WW; *Zm00001d00509* on chromosome 2 and *Zm00001d053262* on the single chromosome, involved in ABA regulation for SD in WS.

3.6 Phenotypic variation explained by SNP in different genotyping scenarios

The proportions of phenotypic variance explained by significant SNP (R^2_{TOT}) for the analyzed traits under both water supply conditions, ideal (WW) and deficit (WS), were less explained in the isolated genotyping scenarios for the studied traits (Figure 5). Regarding to the isolated scenarios, R^2_{TOT} in SNP-Array ranged from 0.18 for SD (WW in WS) to 0.53 for SPAD (WW), GBS-B73 ranged from 0.11 for SD (WS) to 0.48 for SD (WW), and GBS-Mock from 0.11 for PH (WW) to 0.53 for SPAD (WW). Overall, SNP-Array performed better independently for SPAD and PH, except for SD (WW), where GBS-B73 stood out, and SDM, was almost the same among the scenarios. When combined, the value of R^2_{TOT} ranged from 0.26 in SNP-Array + GBS-B73 for SD (WS) to 0.65 in SNP-*Array* + GBS-Mock for SPAD (WW).

The best scenario combination was SNP-Array + GBS-Mock for SPAD (WW) with an increase of 0.12 in accuracy compared to the best isolated scenario. For PH and SDM under

WW condition, SNP-Array + GBS-B73 was superior, increasing accuracy by 0.07 and 0.16, respectively, compared to the best single scenario. For SD, combining GBS-B73 + GBS-Mock increased accuracy by 0.05. Regarding water availability, the ideal water supply condition achieved better overall accuracy, except for PH in isolated SNP-Array and combined with GBS-Mock. In the WS condition, better accuracy was also observed for all traits when combining scenarios.

4 DISCUSSION

Water is one of the most important factors limiting crop growth. Maize requires a large amount of water throughout all stages of development, from seed germination to the reproductive phase. In this context, the significant effect of water supply levels reveals contrasting conditions in WW and WS, indicating that the irrigation treatments used in the present study to generate contrasting environments were sufficient for all traits (Table 1). Moreover, the significance of genotypes suggests that the panel used in this study exhibits genetic variability. Previous studies have also reported genetic diversity for the same tropical maize germplasm panel (YASSUE et al., 2021; DE SOUZA SILVEIRA et al., 2023). Genetic variability is a fundamental factor for any breeding program.

However, the interaction effect shows that the responses were not differentiated for the genotypes across environments; they exhibit similar phenotypic responses to environmental changes. Genotype x environment is important when estimating heritability because it influences a trait's genetic and environmental variation (FALCONER and MACKAY, 1996). The low effect of interaction also maximizes the accuracy (RESENDE et al., 2012), high accuracy estimatives indicate good experimental precision. Heritability was higher for plant height, followed by stem diameter, consistent with Sabiel et al. (2014) results, who reported moderate heritabilities for plant height and stem diameter in maize under water stress.

4.1 SNP in genotyping scenarios

Advances in molecular biology have facilitated the creation of high-throughput, precise, and cost-effective technologies, encompassing the development of platforms and novel genotyping methodologies. Platforms such as SNP-Array and GBS are well-suited for genotyping hundreds to thousands of samples, each containing numerous SNP markers, in a single assay, and at a significantly faster pace (RASHEED et al., 2017). This study had there was a balanced distribution of SNP across chromosomes in the SNP-Array and GBS-B73 genotyping scenarios, perhaps attributed to using the same reference genome (Table 2). The inbred line B73 has been utilized as the reference genome for maize sequencing (SCHNABLE et al., 2009) and an example of a reference genome-based pipeline is TASSEL-GBS.

In the GBS-Mock scenario, a smaller number of SNP markers was observed. In cases where a reference genome is not yet available, a simulated genome can be employed to

perform SNP discovery, serving as a valid alternative, especially for minor crops (MACHADO et al., 2023; SABADIN et al., 2022). Regarding the smaller number of markers observed in GBS-B73 compared to SNP-Array, this may be related to the low genomic coverage of GBS resulting in missing SNP (Wang *et al.*, 2020). However, this issue can be partially addressed by using software employed in imputation, as missing SNP are imputed to fill in the gaps in obtaining intermediate genotype information.

4.2 GWAS and candidate genes

GWAS has emerged as a crucial tool, allowing for a systematic approach to identifying associations between thousands of genomic loci and complex traits. In this study, overall, more SNP were identified in association with the trait under ideal water supply conditions than under water deficit conditions in all genotyping scenarios (Table 3). A similar result was found by De Souza Silveira et al. (2023), who identified more SNPs associated with root traits of tropical maize under ideal water supply conditions than those subjected to water scarcity. Moreover, Yassue et al. (2021; 2023) found more SNP associated with tropical maize traits not evaluated under inoculation by growth-promoting bacteria, such as plant height, stem diameter, and aboveground dry mass. These authors also consider that growth-related traits, such as plant height, stem diameter, and dry mass, are complex and controlled by many genes with small individual effects.

The genes found in the study have small effects (ASE), revealing the polygenic nature of the traits, controlling a relatively small portion of the genotypic variation (Table 4). Complex traits in plants, such as height, diameter, and tolerance to environmental stresses, often have a multifactorial genetic basis involving the interaction of various genes and environmental factors. Thus, knowledge of the genomic regions associated with the traits of interest will provide insight into this genetic basis. Additionally, the study also detected a common marker associated with more than one trait at different water supply levels, indicating a possible pleiotropic effect. Bouchet et al. (2017) reported pleiotropy among phenology-related traits, such as plant height and leaf number, and Zhang et al. (2022) for maize productivity traits. Pleiotropic effects in GWAS studies can increase the complexity of understanding genetic and phenotypic relationships, indicating that phenotypes are more interconnected than initially thought. This complicates the interpretation of study results, as it may need to be clarified which phenotype is directly influenced by the variant and to what extent. In genetic improvement

studies, pleiotropic effects can affect the selection of desirable traits, as a single genetic variant can influence multiple agronomic or desirable traits.

The candidate gene *Zm00001d005090*, associated with SD under both water conditions and SDM under water deficit, possibly indicating a pleiotropic effect regulating the expression of these two traits. This gene is responsible for the clathrin heavy chain, one of the main subunits of clathrin, an essential protein in eukaryotic cells playing a crucial role in the endocytosis process. Hence, endocytosis takes place in many vital processes for the plant development, such as abscisic acid (ABA) responses (Sutter *et al.*, 2007). These authors state that in situations involving ABA, specific proteins in the plasma membrane are negatively regulated through the induction of their endocytosis. It has been demonstrated that ABA and salicylic acid positively regulate a gene encoding a clathrin chain in maize (Zeng *et al.*, 2013). ABA is produced in various parts of plants, including the stem, and it influences gene expression by activating stress-response protein-coding genes and repressing growth-related genes. There is also evidence that clathrin impacts Arabidopsis's stomatal function, gas exchange, and vegetative growth (Larson *et al.*, 2017). Thus, this gene may have a pleiotropic effect, resulting in reduced height, stem diameter growth, and dry mass.

SNP were found to be associated with the trait simultaneously in both water availability levels, such as the gene *Zm00001d017978* identified in association with the PH trait in the SNP-Array scenario and the gene *Zm00001d001852* in association with the SD trait in the GBS-Mock scenario. *Zm00001d017978* has a putative function in the endoglucanase enzyme, a subgroup of a larger enzyme family called cellulase. Cellulases are part of a superfamily of enzymes called hydrolases that use water to break down molecules. All cellulases are essential to degrading cellulose, a structural polysaccharide found in plant cell walls (Rahman *et al.*, 2018). The cell wall plays a crucial role in plants' support and mechanical support, allowing them to grow by providing rigidity and resistance. Therefore, any alteration in cellulose degradation, caused by overexpression or underexpression of enzymes can affect structural integrity and, consequently, plant height. The applied water deficit may have negatively affected stem elongation, contributing to plant height, as at the V6 stage, the stem initiates the accelerated elongation phase. The gene *Zm00001d001852* has a putative function as Gibberellin-regulated protein 2 (GRP) with expression positively regulated by gibberellin. The plant hormone gibberellin regulates major aspects of plant growth and development (YAMAGUCHI, 2008), stimulating cell division and growth. The effect of gibberellin on stem diameter may be related to cell division and radial expansion of cells, increasing the number of

cell layers. Additionally, there is evidence that biotic stresses impact gibberellin and GRP levels, as it has been reported that a slight increase in temperature can raise endogenous gibberellin concentration (Camut *et al.*, 2019).

The genes *Zm00001d042735* and *Zm00001d031759* were also identified at both water supply levels and are associated with the SPAD trait. The first one was identified in the GBS-Mock scenario, while the other one was identified in both the SNP-Array and GBS-B73 scenarios, and both belongs to the zinc finger family. Zinc finger proteins are named for their three-dimensional structure resembling a finger, binding to zinc ions through amino acids in the peptide sequence and are widely distributed in eukaryotic organisms (Han *et al.*, 2020). They bind to specific genetic sequences, interact with various proteins, participate in signal transduction, and regulate gene expression, playing an essential role in growth, development, and environmental adaptation. *Zm00001d042735* was described as a RING-type E3 ubiquitin transferase. Ubiquitin is a protein that acts as a molecular marker, signaling various cellular functions such as protein degradation, cell cycle regulation, cellular stress response, and intracellular signaling (Lee; Kim, 2011). E3 ubiquitin proteins respond to water stress by regulating ABA biosynthesis and signal transduction, modifying and degrading stress-related proteins (Han *et al.*, 2022). An example is ZmAIRP4 involved in ABA signaling in maize and the overexpression of this gene increased water stress tolerance in *Arabidopsis* (Yang *et al.*, 2018). Changes in water content induced by water stress can directly affect the SPAD index and chlorophyll content, as ABA concentration increases, causing stomatal closure to reduce water loss, which may affect the expression of genes related to stress response.

The gene Zm00001d031759, also belonging to the zinc finger protein family, has a putative function in the Protein shoot gravitropism 5 group, acting in the morphogenesis of aerial organs and responses to gravitropism. Some genes from the shoot gravitropism family have been identified and are involved in the perception and signal transduction for gravity associated with the branching angle (YAMAUCHI et al., 1997). It has also been found that loss of functionality of the shoot gravitropism 5 gene (SGR5) resulted in decreased starch accumulation in aerial tissues and consequently reduced gravity sensitivity (TANIMOTO; TREMBLAY; COLASANTI, 2008). Gravity is an important regulator of plant architecture, allowing plants to optimize their position relative to the soil for nutrient absorption and to light for photosynthesis. Furthermore, some genes and regions manifest for the expression of the trait independently of the water supply level, probably unrelated to water stress.

4.2.1 Genes associated with phytohormone signaling pathway

Genes and regions shared among the genotyping scenarios were identified based on their function for the same trait (Supplementary Table S2). For example, genes *Zm00001d026477* in SNP-Array and *Zm00001d027695* in GBS-B73 are related to jasmonic acid (JA) response, associated with SPAD in WW traits. Jasmonate ZIM domain proteins, known as JAZ proteins, play a crucial role in pathogen responses (Ishiga *et al.*, 2013) and are important signaling molecules in the JA pathway (Liu *et al.*, 2017). Glutaredoxins are associated with water-induced stress response in maize, also participating in the abiotic stress response mediated by JA and ethylene through their interaction with transcription factors (Ding *et al.*, 2019). As JA is involved in various signaling pathways regulating physiological and molecular processes in plants, in defense against biotic and abiotic stresses, such as drought (Rehman *et al.*, 2023), signaling pathways induce stomatal closure, activating potassium efflux in guard cell protoplasts (Evans, 2003) enhancing the plants' ability to cope with environmental stresses.

Regarding ABA regulation, *Zm00001d016786* was associated with SDM in WS in SNP-Array, and *Zm00001d005090* in GBS-B73. *Zm00001d021708* was found in GBS-B73, and *Zm00001d012719* in GBS-Mock for PH under WW conditions. *Zm00001d005090* and *Zm00001d053262* were also identified in GBS-B73 and GBS-Mock, respectively, for SD under WS conditions. Protein disulfide-isomerase (PDI) is a member of the thioredoxin superfamily of redox proteins with multiple physiological functions (Khan; Mutus, 2014), playing a crucial role in abiotic stress tolerance. Thioredoxin (TRXo1) is involved in ABA perception through redox regulation of specific receptors (De Brasi-Velasco *et al.*, 2023). In maize, genes related to PDI were highly responsive to ABA and water stress (LIU et al., 2009). Additionally, a PDI-like protein strongly associated with aboveground biomass and leaf size was identified (Kang *et al.*, 2015). According to Tanz et al. (2012), PDI is a family proteins affect chlorophyll biosynthesis in Arabidopsis seedlings.

The PPR (pentatricopeptide repeat) proteins are located in mitochondria or chloroplasts. In contrast, the BZIP (basic leucine zipper) proteins constitute a family of transcription factors (TFs) associated with plant growth, development, and stress responses. A typical PPR protein is targeted to mitochondria or chloroplasts, binds to one or several organellar transcripts, and influences their expression by altering RNA sequence, turnover, processing, or translation (Barkan; Small, 2014). It has been found that the PPR96 protein,

located in mitochondria, altered the transcription levels of various stress-responsive genes under ABA treatments (Liu *et al.*, 2016). BZIP proteins are involved in various stress responses, primarily through the ABA signaling pathway (Uno *et al.*, 2000). Changes in the transcription levels of maize BZIP TFs were observed in response to ABA treatments (Cao *et al.*, 2019).

As mentioned earlier in SDM and SD, the Clathrin heavy chain indicates possible pleiotropy. Calcium-dependent lipid-binding protein acts in response to abiotic stress, such as drought. The expression of sANN3, a calcium-dependent lipid-protein, increased in response to water stress in rice, inducing various genes in the ABA signaling pathway and promoting root growth to enhance water absorption and stomatal closure to reduce water loss (LI et al., 2019). Therefore, these proteins and the biosynthesis pathways in ABA regulation may influence photosynthesis and plant development and growth.

The genes *Zm00001d014899* in GBS-B73 and *Zm00001d001852* in GBS-Mock are associated with the trait SD under WW conditions, involved with the phytohormone gibberellin. The first encodes a protein from the tetratricopeptide repeat (TPR)-like superfamily. Proteins containing tetratricopeptide repeats play an important role in protein-protein interaction and regulating various cellular functions (Rosado *et al.*, 2006). They serve different crucial roles in plants, including their involvement in phytohormone signaling, such as gibberellin (JACOBSEN; OLSZEWSKI, 1993; SILVERSTONE et al., 2007). Therefore, TPR-repeat-containing proteins are pivotal in signaling phytohormones and regulating various physiological processes, including growth, development, and environmental response. Gibberellin-regulated protein 2 (GRP) was mentioned earlier, occurring at both levels of water availability for SD.

Genes associated with SDM under WW conditions were found on the same chromosome, *Zm00001d031445* in the SNP-Array and *Zm00001d027626* in the GBS-B73, both involved in ethylene biosynthesis. The ethylene-insensitive3-like/ethylene-insensitive3 (EIL/EIN3) is one of the major regulatory families in ethylene signaling, also serving as a hub for ethylene connections with various plant responses to different environmental conditions (Wu; Yang, 2019). Ethylene is a crucial regulator in stress signaling, and its interaction with a receptor complex triggers the inactivation of kinase response, resulting in the initial dephosphorylation of EIN2, followed by the cleavage of the C-terminal of EIN2. Subsequently, EIN2 translocate to the nucleus, regulating the activation of EIN3/EIL1. These proteins, in turn, exert control over ethylene response factors (Yoshida *et al.*, 2011).

S-adenosyl-L-methionine synthetase, known as SAM, is a donor of methyl groups in the biosynthesis of nucleic acids, proteins, lipids, polysaccharides, and secondary compounds (Heidari *et al.*, 2020). SAM is involved in many important biological processes such as the biosynthesis of ethylene. Yu et al. (2012) found that alterations in the expression level of SAM affected protein synthesis, phytohormones (JA and ethylene), and genes related to stress defense response. Ethylene is a volatile compound produced endogenously by the plant for growth regulation - roots, stems, leaves, and flowers (Shilev, 2020). Plants increase the synthesis of this hormone when subjected to stressful situations, whether biotic or abiotic. Water deficit, in particular, is one of the main factors related to its increase (Apelbaum; Yang, 1981). Thus, in response, the plant alters its growth rates, decreases biomass, and reduces development (Glick, 2014).

4.2.2 Genes associated with the circadian clock

Zm00001d044194 was identified in the SNP-Array, and *Zm00001d018127* in the GBS-B73 under WW condition associated with the SPAD trait acting in the circadian clock. The MYB proteins constitute one of the most extensive families of transcription factors found in plants, playing an important role in growth and development, with widespread expression in the development of corn and soybeans in stress responses, and are closely correlated with the circadian rhythm (Du *et al.*, 2013). MYB-related genes can act as repressors and activators associated with the circadian clock (KAMIOKA et al., 2016; HSU; DEVISETTY; HARMER, 2013; HU et al., 2024; SCHAFFER et al., 1998).

The SNW/Ski domain protein is involved in the post-transcriptional regulation of circadian clock genes. SkipP interacts with the serine/arginine-rich spliceosomal protein 45 (SR45) and controls the circadian cycle through alternative splicing of circadian clock genes under biotic stress conditions (Wang *et al.*, 2012). The circadian clock in plants refers to an internal timing system on a cycle of approximately 24 hours that regulates the behavioral and physiological processes of plants, including photosynthesis (Niwa; Yamashino; Mizuno, 2009). Likely, each guard cell maintains its circadian rhythm, and the involvement of a clock controlling stomatal opening seems to be advantageous for the plant, helping prevent unnecessary water loss through transpiration (DODD et al., 2005; GORTON et al., 1993). Thus, besides the environmental and internal factors that influence stomatal function, the circadian

pattern in regulating stomatal movements is advantageous as it can enhance both photosynthetic efficiency and water use efficiency.

4.2.3 Genes associated with ubiquitination regulation

The genes associated with the PH trait under WS conditions were *Zm00001d053809* in SNP-Array and *Zm00001d042481* in GBS-B73, which are related to the regulation of protein ubiquitination. Culins neddylation modulates the ubiquitin ligase activity of the complex, leading to increased ubiquitination and degradation of target proteins by the proteasome (BISWAS et al., 2007; MOHANTY; CHATTERJEE; DAS, 2021; PAN et al., 2004). Neddylation is the post-translational protein modification most closely related to the regulation of protein ubiquitination (Rabut; Peter, 2008).

Ubiquitin thioesterases play a fundamental role in regulating the degradation of proteins marked with ubiquitin in plants. The ubiquitin system regulates virtually all aspects of cellular function (Ernst *et al.*, 2013), playing an important role in controlling abiotic stress and processes that affect agronomic traits. For example, the ubiquitin-proteasome system is an essential pathway for protein degradation in plant growth and development (Linden; Callis, 2020). In the regulation of transcription responsive to ABA, the ubiquitin-proteasome system is involved, allowing plants to respond to abiotic stresses such as drought (Dreher; Callis, 2007). Thus, ubiquitination affects gene expression or protein abundance to determine agronomic traits and stress control, enabling dynamic adjustments in physiological and biochemical responses contributing to plant survival and adaptation under adverse conditions.

4.2.4 Coincident genes among genotyping scenarios

Concerning the SNP-Array and GBS-B73 genotyping scenarios, these platforms are based on the same reference genome (B73) and are physically fixed, making it possible to determine the physical position of the marker in the genome. The coincidence between genes and regions on the same chromosome occurred only for *Zm00001d031445* in the SNP-Array and *Zm00001d027626* in the GBS-B73, both on chromosome 1. However, it was observed that, even though there was no coincidence regarding the physical position of the markers and chromosomes, there was still similarity regarding the gene functions.

Considering the three scenarios, when considering the identification of the gene and region, it was observed that there was coincidence only for one marker, in the SPAD trait under both irrigation conditions. However, when deeper analyses were conducted regarding the gene function, it highlighted possible coincidences. Negro et al. (2019) concluded that GBS and SNP-Array were complementary for detecting QTLs in maize, marking different haplotypes. In a study performed in barley by Darrier et al. (2019), GBS and SNP-Array were shown to be efficient in accessing diversity. Still, they accessed different regions of the genome. However, even though they captured different regions, there was a positive correlation between the similarity matrices of both approaches. Thus, even when accessing different genome regions, these platforms demonstrate that they can be complementary. In the study, there was also a coincidence for the simulated genome, GBS-Mock, validating the complementarity for this scenario as well.

4.3 Association of markers in genotyping scenarios

The correlation between the markers in the SNP-Array and GBS-Mock scenarios provides information about the location of the markers on the chromosomes. Identifying a marker highly correlated with the GBS-Mock suggests that this marker is likely on a specific chromosome. The strength of the correlation between two markers is related to their physical proximity; the closer the markers are, the stronger the linkage disequilibrium (LD) (Myles *et al.*, 2009). When markers are closer, there is a higher likelihood that they will be inherited together, leading to a stronger correlation between them. This is because when two markers are very close, they have fewer opportunities for recombination during meiosis, the process of gamete formation, which maintains stable combinations of adjacent alleles across generations. This information can be useful for guiding research, providing an initial direction for investigating the specific position of the marker in the genome.

However, according to the study results, the markers are located throughout the genome and not necessarily physically close. In other words, despite the relationship between the strength of the correlation and the physical proximity of the markers, the results showed that the markers are distributed across the entire genome. This suggests that other factors, besides physical proximity, may influence the correlation between the markers, such as genetic inheritance patterns, recombination rate, and genomic structure, highlighting the importance of considering these aspects.

4.4 Combining genotyping scenarios

The combination of genotyping scenarios can be a valid alternative for GWAS studies, providing higher resolution results than those obtained in isolated scenarios. In the approach involving Array and GBS, it was noticed that one tool complements the other, regardless of how GBS data are explored, whether with the referenced genome or *in-silico*, as there was little difference between SNP-Array + GBS-B73 and SNP-Array + GBS-Mock. Using multiple genotyping platforms, it is possible to capture a broader range of genetic markers in linkage disequilibrium with the loci of interest, which can increase the ability to detect significant associations between genetic variants and phenotypes in GWAS studies.

With regard to the use of simulated genomes, Machado et al. (2023) and Sabadin et al. (2022), assert that it is an excellent strategy for studies on diversity, population structure, heterotic group definition, tester selection, and genomic prediction for minor crops. Another caveat is that using temperate germplasm as a reference genome may introduce a significant bias when analyzing tropical germplasm (Xu *et al.*, 2017). As a result, favorable alleles hidden in tropical maize, in specific tropical genomic regions, may be lost (Rasheed *et al.*, 2017). With GBS, marker discovery and genotyping occur simultaneously, mitigating this bias and enabling the identification of markers in the analyzed diversity panel (Heslot *et al.*, 2013). Furthermore, combining information obtained via conventional approaches with a reference genome obtained from the simulated genome should improve accuracy in association studies and impact the advancement of genetic research and the development of breeding strategies.

4.5 Applicability

Negro et al. (2019) and Darrier et al. (2019) highlighted the complementarity between standard genotyping platforms for GWAS, demonstrating that both SNP-Array and GBS can identify markers strongly linked to genes influencing key phenotypic traits. However, adopting different genotyping platforms may incur substantial costs due to their distinct methodologies. Conversely, GBS genotyping offers the flexibility to utilize both the reference genome and in-silico genome, thereby avoiding additional expenses associated with combining these scenarios. In our study, combining GBS-B73 and GBS-Mock datasets resulted in a notable increase in accuracy for several traits compared to the highest accuracy achieved by

GBS alone. Specifically, we observed accuracy gains of 0.06, 0.03, 0.05, and 0.15 for SPAD, PH, SD, and SDM, respectively. This integration of datasets allows for more comprehensive analyses, capturing a broader range of SNPs and providing enhanced resolution in explaining phenotypic variation. Ultimately, leveraging a single genotyping method enables more informative and efficient exploration of data, facilitating a deeper understanding of the genetic basis of traits and informing crop improvement strategies.

Indeed, when a study aims to uncover greater genetic polymorphism within a species, and SNP-Array technology is unavailable, leveraging GBS approaches becomes a viable alternative. By conducting GWAS using GBS methods, researchers can effectively identify additional polymorphisms, thereby increasing the resolution and depth of the study. This strategy proves particularly beneficial for minor or orphan crops that possess a genome reference but lack access to SNP-Array technology. In such cases, GBS offers a cost-effective and accessible means to explore the genetic diversity present within these crops, facilitating a more comprehensive understanding of their genetic architecture and potential avenues for crop improvement. By harnessing the power of GBS-based GWAS, researchers can unlock valuable insights into the genetic factors underlying traits of interest, ultimately contributing to the development of improved varieties tailored to the specific needs of these crops.

5 CONCLUSIONS

The conclusions drawn from the study emphasize the importance and effectiveness of combining multiple genotyping scenarios (SNP-Array, GBS-B73, and GBS-Mock) for association mapping in crops, in specific case under varying water supply conditions. Despite differences in genotyping methods, genes and regions associated with specific traits were consistently identified, indicating the reliability and robustness of the approaches.

Furthermore, the study highlights that certain candidate genes shared functional similarities across genotyping scenarios, even when their physical positions on chromosomes did not align. This suggests that functional similarity, rather than physical proximity, plays a crucial role in influencing traits of interest.

In terms of accuracy, combining GBS-B73 and GBS-Mock was found to improve the accuracy of trait associations across all traits and water supply conditions. This suggests that this combined approach not only enhances accuracy but also provides better resolution in GWAS. Importantly, this strategy offers a cost-effective solution for genotyping, making it more accessible for research and breeding programs.

Overall, the findings underscore the importance of considering multiple genotyping scenarios and highlight the value of combining GBS-B73 and GBS-Mock approaches for association mapping studies in crops, particularly for traits related to drought tolerance. This approach not only improves accuracy but also reduces costs, making it a practical and viable option for crop improvement programs.

REFERENCES

ALTSCHUL, S. F. *et al.* Basic local alignment search tool. **Journal of Molecular Biology**, [*s. l.*], v. 215, n. 3, p. 403–410, 1990.

APELBAUM, A.; YANG, S. F. Biosynthesis of Stress Ethylene Induced by Water Deficit. **Plant Physiology**, [*s. l.*], v. 68, n. 3, p. 594–596, 1981.

BARKAN, A.; SMALL, I. Pentatricopeptide repeat proteins in plants. **Annual Review of Plant Biology**, [*s. l.*], v. 65, p. 415–442, 2014.

BISWAS, K. K. *et al.* Genetic Characterization of Mutants Resistant to the Antiauxin p -Chlorophenoxyisobutyric Acid Reveals That AAR3 , a Gene Encoding a DCN1-Like Protein, Regulates Responses to the Synthetic Auxin 2,4-Dichlorophenoxyacetic Acid in Arabidopsis Roots. **Plant Physiology**, [*s. l.*], v. 145, n. 3, p. 773–785, 2007.

BOUCHET, S. *et al.* Association mapping for phenology and plant architecture in maize shows higher power for developmental traits compared with growth influenced traits. **Heredity**, [*s. l.*], v. 118, n. 3, p. 249–259, 2017.

BROWNING, B. L.; BROWNING, S. R. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. **American Journal of Human Genetics**, [*s. l.*], v. 84, n. 2, p. 210–223, 2008.

CAMUT, L. *et al.* Root-derived GA12 contributes to temperature-induced shoot growth in Arabidopsis. **Nature Plants**, [*s. l.*], v. 5, n. 12, p. 1216–1221, 2019.

CAO, L. *et al.* Systematic analysis of differentially expressed maize ZmbZIP genes between drought and rewatering transcriptome reveals bZIP family members involved in abiotic stress responses. **International Journal of Molecular Sciences**, [*s. l.*], v. 20, n. 17, 2019.

COVARRUBIAS-PAZARAN, G. Genome-Assisted prediction of quantitative traits using the r package sommer. **PLoS ONE**, [*s. l.*], v. 11, n. 6, p. 1–15, 2016.

DA, Y. *et al.* Mixed model methods for genomic prediction and variance component estimation of additive and dominance effects using SNP markers. **PLoS ONE**, [*s. l.*], v. 9, n. 1, 2014.

DARRIER, B. *et al.* A comparison of mainstream genotyping platforms for the evaluation and use of barley genetic resources. **Frontiers in Plant Science**, [*s. l.*], v. 10, n. April, p. 1–14, 2019.

DAVEY, J. W. *et al.* Genome-wide genetic marker discovery and genotyping using next-generation sequencing. **Nature Reviews Genetics**, [*s. l.*], v. 12, n. 7, p. 499–510, 2011.

DE BRASI-VELASCO, S. *et al.* Thioredoxin TRXo1 is involved in ABA perception via

PYR1 redox regulation. **Redox Biology**, [*s. l.*], v. 63, n. March, 2023.

DE SOUZA SILVEIRA, M. V. *et al.* Association mapping for image-based root traits in tropical maize under water stress in semi-arid regions. **Agronomy Journal**, [*s. l.*], 2024.

DING, S. *et al.* Identification of Maize CC-Type Glutaredoxins That. **Genes**, [*s. l.*], v. 10(8):610, 2019.

DODD, A. N. *et al.* Cell biology: Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. **Science**, [*s. l.*], v. 309, n. 5734, p. 630–633, 2005.

DREHER, K.; CALLIS, J. Ubiquitin, hormones and biotic stress in plants. **Annals of Botany**, [*s. l.*], v. 99, n. 5, p. 787–822, 2007.

DU, H. *et al.* Genome-wide identification and evolutionary and expression analyses of MYB-related genes in land plants. **DNA Research**, [*s. l.*], v. 20, n. 5, p. 437–448, 2013.

ELBASYONI, I. S. *et al.* A comparison between genotyping-by-sequencing and array-based scoring of SNPs for genomic prediction accuracy in winter wheat. **Plant Science**, [*s. l.*], v. 270, n. August, p. 123–130, 2018.

ERNST, A. *et al.* A Strategy for Modulation of Enzymes in the Ubiquitin System. **Science**, [*s. l.*], v. 339, n. 6119, p. 590–595, 2013.

EVANS, N. H. Modulation of guard cell plasma membrane potassium currents by methyl jasmonate. **Plant Physiology**, [*s. l.*], v. 131, n. 1, p. 8–11, 2003.

FALCONER, D. S.; MACKAY, T. F. C. **Introduction to quantitative genetics**. 4.eded. London: Longman: [*s. n.*], 1996.

FRENSCH, J.; HSIAO, T. C. Transient Responses of Cell Turgor and Growth of Maize Roots as Affected by Changes in Water Potential. **Plant Physiology**, [*s. l.*], v. 104, n. 1, p. 247–254, 1994.

GLICK, B. R. Bacteria with ACC deaminase can promote plant growth and help to feed the world. **Microbiological Research**, [*s. l.*], v. 169, n. 1, p. 30–39, 2014.

GORJANC, G.; BIJMA, P.; HICKEY, J. M. Reliability of pedigree-based and genomic evaluations in selected populations. **Genetics Selection Evolution**, [*s. l.*], v. 47, n. 1, 2015.

GORTON, H. L. *et al.* Linked references are available on JSTOR for this article : Circadian Rhythms in Stomatal Responsiveness to Red and Blue Light ’. [*s. l.*], v. 103, n. 2, p. 399–406, 1993.

HAN, G. *et al.* C2H2 Zinc Finger Proteins: Master Regulators of Abiotic Stress Responses in Plants. **Frontiers in Plant Science**, [*s. l.*], v. 11, n. February, p. 1–13, 2020.

HAN, G. *et al.* RING Zinc Finger Proteins in Plant Abiotic Stress Tolerance. **Frontiers in Plant Science**, [*s. l.*], v. 13, n. April, p. 1–16, 2022.

HEIDARI, P. *et al.* Insights into the SAM synthetase gene family and and hormone treatments. **Plants**, [*s. l.*], v. 9, n. 5, p. 586, 2020.

HESLOT, N. *et al.* Impact of Marker Ascertainment Bias on Genomic Selection Accuracy and Estimates of Genetic Diversity. **PLoS ONE**, [*s. l.*], v. 8, n. 9, 2013.

HSIAO, T. C. Plant Responses to Water Stress. **Annual Review of Plant Physiology**, [*s. l.*], v. 24, p. 519–570, 1973.

HSU, P. Y.; DEVISETTY, U. K.; HARMER, S. L. Accurate timekeeping is controlled by a cycling activator in Arabidopsis. **eLife**, [*s. l.*], v. 2013, n. 2, p. 1–20, 2013.

HU, Z. *et al.* Differential Response of MYB Transcription Factor Gene Transcripts to Circadian Rhythm in Tea Plants (*Camellia sinensis*). **International journal of molecular sciences**, [*s. l.*], v. 25, n. 1, 2024.

INGVARSSON, P. K.; STREET, N. R. Association genetics of complex traits in plants. **New Phytologist**, [*s. l.*], v. 189, n. 4, p. 909–922, 2011.

ISHIGA, Y. *et al.* Jasmonate ZIM-Domain (JAZ) Protein Regulates Host and Nonhost Pathogen-Induced Cell Death in Tomato and Nicotiana benthamiana. **PLoS ONE**, [*s. l.*], v. 8, n. 9, 2013.

JACOBSEN, S. E.; OLSZEWSKI, N. E. Mutations at the SPINDLY locus of *Arabidopsis* alter gibberellin signal transduction. **The Plant Cell**, [*s. l.*], v. 5, n. 8, p. 887–896, 1993.

KAMIOKA, M. *et al.* Direct repression of evening genes by CIRCADIAN CLOCK-ASSOCIATED1 in the *Arabidopsis* circadian clock. **Plant Cell**, [*s. l.*], v. 28, n. 3, p. 696–711, 2016.

KANG, Y. *et al.* Genome-wide association of drought-related and biomass traits with HapMap SNPs in *Medicago truncatula*. **Plant Cell and Environment**, [*s. l.*], v. 38, n. 10, p. 1997–2011, 2015.

KHAN, H. A.; MUTUS, B. Protein disulfide isomerase a multifunctional protein with multiple physiological roles. **Frontiers in Chemistry**, [*s. l.*], v. 2, n. AUG, p. 1–9, 2014.

LARCHER, W. **Physiological Plant Ecology: Ecophysiology and Stress Physiology of Functional Groups.** Berlin: Springer Nature.: [*s. n.*], 2003.

LARSON, E. R. *et al.* Clathrin heavy chain subunits coordinate endo- and exocytic traffic and affect stomatal movement. **Plant Physiology**, [*s. l.*], v. 175, n. 2, p. 708–720, 2017.

LEE, J. H.; KIM, W. T. Regulation of abiotic stress signal transduction by E3 ubiquitin ligases in *Arabidopsis*. **Molecules and Cells**, [*s. l.*], v. 31, n. 3, p. 201–208, 2011.

LI, X.; ZHANG, Q.; YANG, X.; HAN, J.; ZHU, Z. OsANN3, a calcium-dependent lipid binding annexin is a positive regulator of ABA-dependent stress tolerance in rice. **Plant Science**, [*s. l.*], v. 284, p. 212–220, 2019.

LINDEN, K. J.; CALLIS, J. The ubiquitin system affects agronomic plant traits. **Journal of Biological Chemistry**, [*s. l.*], v. 295, n. 40, p. 13940–13955, 2020.

LIU, X; HUANG, M; FAN, B; BUCKLER, E. S; ZHANG, Z. Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient GenomeWide Association Studies. **PLoS genetics**, [*s. l.*], v. 12, n. e1005767, 2016.

LIU, Y.; WANG, X.; SHI, Y.; HUANG, Y.; SONG, Y.; WANG, T. Expression and characterization of a protein disulfide isomerases in maize. **Journal of Biochemistry and Molecular Biology**, [*s. l.*], v. 25, p. 229–234, 2009.

LIU, F. *et al.* Plant jasmonate ZIM domain genes: Shedding light on structure and expression patterns of JAZ gene family in sugarcane. **BMC Genomics**, [*s. l.*], v. 18, n. 1, p. 1–17, 2017.

LIU, J. M. *et al.* The E-subgroup pentatricopeptide repeat protein family in *Arabidopsis thaliana* and confirmation of the responsiveness PPR96 to abiotic stresses. **Frontiers in Plant Science**, [*s. l.*], v. 7, n. DECEMBER2016, p. 1–14, 2016.

MACHADO, I. P. *et al.* On the usefulness of mock genomes to define heterotic pools, testers, and hybrid predictions in orphan crops. **Frontiers in Plant Science**, [*s. l.*], v. 14, n. June, p. 1–15, 2023.

MELO, A. T. O.; BARTAULA, R.; HALE, I. GBS-SNP-CROP: A reference-optional pipeline for SNP discovery and plant germplasm characterization using variable length, paired-end genotyping-by-sequencing data. **BMC Bioinformatics**, [*s. l.*], v. 17, n. 1, p. 1–15, 2016.

MOHANTY, P.; CHATTERJEE, K. S.; DAS, R. NEDD8 Deamidation Inhibits Cullin RING Ligase Dynamics. **Frontiers in Immunology**, [*s. l.*], v. 12, n. August, p. 1–12, 2021.

MOROSINI, J. S. *et al.* Association mapping for traits related to nitrogen use efficiency in tropical maize lines under field conditions. **Plant and Soil**, [*s. l.*], v. 421, n. 1–2, p. 453–463, 2017.

MYLES, S. *et al.* Association mapping: Critical considerations shift from genotyping to experimental design. **Plant Cell**, [*s. l.*], v. 21, n. 8, p. 2194–2202, 2009.

NEGRO, S. S. *et al.* Genotyping-by-sequencing and SNP-arrays are complementary for detecting quantitative trait loci by tagging different haplotypes in association studies. **BMC Plant Biology**, [*s. l.*], v. 19, n. 1, p. 1–22, 2019.

NIWA, Y.; YAMASHINO, T.; MIZUNO, T. The Circadian Clock Regulates the Photoperiodic Response of Hypocotyl Elongation through a Coincidence Mechanism in *Arabidopsis thaliana*. **Plant and Cell Physiology**, [*s. l.*], v. 50, n. 4, p. 838–854, 2009.

PAN, Z.-Q. *et al.* Nedd8 on cullin: building an expressway to protein destruction. **Oncogene**, [*s. l.*], v. 23, n. 11, p. 1985–1997, 2004.

PETERSON, R. A. **Estimating normalization transformations with bestNormalize**. [*S. l.: s. n.*], 2017.

POLAND, J. A. *et al.* Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. **PLoS ONE**, [*s. l.*], v. 7, n. 2, 2012.

RABUT, G.; PETER, M. Function and regulation of protein neddylation. "Protein modifications: beyond the usual suspects" review series. **EMBO Reports**, [*s. l.*], v. 9, n. 10, p. 969–976, 2008.

RAHMAN, M. S. *et al.* Endoglucanase (EG) Activity Assays Chapter 13. **Methods in molecular biology**, [*s. l.*], n. February 2019, p. 169–183, 2018.

RASHEED, A. *et al.* Crop Breeding Chips and Genotyping Platforms: Progress, Challenges, and Perspectives. **Molecular Plant**, [*s. l.*], v. 10, n. 8, p. 1047–1064, 2017.

RAZA, A. *et al.* Impact of climate change on crops adaptation and strategies to tackle its outcome: A review. **Plants**, [*s. l.*], v. 8, n. 2, 2019.

REHMAN, M. *et al.* The Multifaceted Role of Jasmonic Acid in Plant Stress Mitigation: An Overview. **Plants**, [*s. l.*], v. 12, n. 23, 2023.

RESENDE, M. D. V. *et al.* **Seleção Genômica Ampla (GWS) via Modelos Mistos (REML/BLUP), Inferência Bayesiana (MCMC), Regressão Aleatória (RR) e Estatística Espacial**. Viçosa : Universidade Federal de Viçosa, Departamento de Estatística: [*s. n.*], 2012.

ROSADO, A. *et al.* The *Arabidopsis* Tetratricopeptide Repeat-Containing Protein TTL1 Is Required for Osmotic Stress Responses and Abscisic Acid Sensitivity. **Plant Physiology**, [*s. l.*], v. 142, n. 3, p. 1113–1126, 2006.

SABADIN, F. *et al.* Population-tailored mock genome enables genomic studies in species without a reference genome. **Molecular Genetics and Genomics**, [*s. l.*], v. 297, n. 1, p. 33–46, 2022.

SABIEL, S. A. I. *et al.* Genetic Variation of Plant Height and Stem Diameter Traits in Maize (*Zea mays* L.) under Drought Stress at Different Growth Stages Yield improvement of rapeseed in rice-rapeseed relay cropping View project breeding faba bean for drought and heat tolerance. **Journal of Natural Sciences Research www.iiste.org ISSN**, [*s. l.*], v. 4, n. 23, 2014.

SCHAFFER, R. *et al.* The late elongated hypocotyl mutation of *Arabidopsis* disrupts circadian rhythms and the photoperiodic control of flowering. **Cell**, [*s. l.*], v. 93, n. 7, p. 1219–1229, 1998.

SCHNABLE, P. S. *et al.* The B73 maize genome: Complexity, diversity, and dynamics. **Science**, [*s. l.*], v. 326, n. 5956, p. 1112–1115, 2009.

SHILEV, S. **Plant-growth-promoting bacteria mitigating soil salinity stress in plants**. [*S. l.: s. n.*], 2020.

SHINJIRO YAMAGUCHI. Gibberellin metabolism and its regulation. **Annual Review of Plant Biology**, [*s. l.*], v. 59, p. 225–251, 2008.

SILVERSTONE, A. L. *et al.* Functional Analysis of SPINDLY in Gibberellin Signaling in *Arabidopsis*. **Plant Physiology**, [*s. l.*], v. 143, n. 2, p. 987–1000, 2007.

SUTTER, J. U. *et al.* Abscisic Acid Triggers the Endocytosis of the *Arabidopsis* KAT1 K+ Channel and Its Recycling to the Plasma Membrane. **Current Biology**, [*s. l.*], v. 17, n. 16, p. 1396–1402, 2007.

TAIZ, L. ZEIGER, E. MOLLER, I.M. MURPHY, A. **Plant Physiology and Development**. 6. ed. Sinauer Associates Incorporated; Sunderland, Massachusetts, USA.: [*s. n.*], 2015.

TANZ, S. K. *et al.* The SCO2 protein disulphide isomerase is required for thylakoid biogenesis and interacts with LCHB1 chlorophyll a/b binding proteins which affects chlorophyll biosynthesis in *Arabidopsis* seedlings. **Plant Journal**, [*s. l.*], v. 69, n. 5, p. 743–754, 2012.

TURNER, N. C.; JONES, M. M. Turgor maintenance by osmotic adjustment : A Review and evaluation. **Adaptation of Plants to Water and High Temperatures Stress**, [*s. l.*], n. January, p. 87–103, 1980.

UNO, Y. *et al. Arabidopsis* basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. **Proceedings of the National Academy of Sciences of the United States of America**, [*s. l.*], v. 97, n. 21, p. 11632–11637, 2000.

VANRADEN, P. M. Efficient methods to compute genomic predictions. **Journal of Dairy Science**, [*s. l.*], v. 91, n. 11, p. 4414–4423, 2008.

WANG, N. *et al.* Applications of genotyping-by-sequencing (GBS) in maize genetics and breeding. **Scientific Reports**, [*s. l.*], v. 10, n. 1, p. 1–12, 2020.

WANG, X. *et al.* SKIP is a component of the spliceosome linking alternative splicing and the circadian clock in *Arabidopsis*. **Plant Cell**, [*s. l.*], v. 24, n. 8, p. 3278–3295, 2012.

WICKHAM, H. ggplot2. **WIREs Computational Statistics**, [*s. l.*], v. 3, n. 2, p. 180–185, 2011.

WILLIAMS, E.; PIEPHO, H. P.; WHITAKER, D. Augmented p-rep designs. **Biometrical Journal**, [*s. l.*], v. 53, n. 1, p. 19–27, 2011.

WU, Y. S.; YANG, C. Y. Ethylene-mediated signaling confers thermotolerance and regulates transcript levels of heat shock factors in rice seedlings under heat stress. **Botanical Studies**, [*s. l.*], v. 60, n. 1, 2019.

XU, Y. *et al.* Enhancing genetic gain in the era of molecular breeding. **Journal of Experimental Botany**, [*s. l.*], v. 68, n. 11, p. 2641–2666, 2017.

YANG, L. *et al.* Overexpression of the maize E3 ubiquitin ligase gene ZmAIRP4 enhances drought stress tolerance in *Arabidopsis*. **Plant Physiology and Biochemistry**, [*s. l.*], v. 123, p. 34–42, 2018.

YASSUE, R. M.; GALLI, G. .; CHUN-PENG, J. C. .; FRITSCHE-NETO, R.; GOTA, M. Genome-wide association analysis of hyperspectral reflectance data to dissect the genetic architecture of growth-related traits in maize under plant growth-promoting bacteria inoculation. **Plant Direct**, [*s. l.*], v. 7, n. 4, 2023.

YASSUE, R. M. *et al.* On the genetic architecture in a public tropical maize panel of the symbiosis between corn and plant growth-promoting bacteria aiming to improve plant resilience. **Molecular Breeding**, [*s. l.*], v. 41, n. 10, 2021.

YIN, L. CMplot: circle manhattan plot. **CMplot: Circle Manhattan Plot**, [*s. l.*], 2020.

YOSHIDA, T. *et al. Arabidopsis* HsfA1 transcription factors function as the main positive regulators in heat shock-responsive gene expression. **Molecular Genetics and Genomics**, [*s. l.*], v. 286, n. 5–6, p. 321–332, 2011.

YU, J. G.; LEE, G. H.; PARK, Y. D. Physiological role of endogenous S-adenosyl-L-methionine synthetase in Chinese cabbage. **Horticulture Environment and Biotechnology**, [*s. l.*], v. 53, n. 3, p. 247–255, 2012.

ZENG, M. H. *et al.* Characterization of a gene encoding clathrin heavy chain in maize up-regulated by salicylic acid, abscisic acid and high boron supply. **International Journal of Molecular Sciences**, [*s. l.*], v. 14, n. 7, p. 15179–15198, 2013.

ZHANG, X. *et al.* Genetic architecture of maize yield traits dissected by QTL mapping and GWAS in maize. **Crop Journal**, [*s. l.*], v. 10, n. 2, p. 436–446, 2022.

ZHENG, X. *et al.* A high-performance computing toolset for relatedness and principal component analysis of SNP data. **Bioinformatics**, [*s. l.*], v. 28, n. 24, p. 3326–3328, 2012.

APPENDIX A - LIST OF TABLES AND FIGURES

Table 1. Wald test of fixed effects, likelihood-ratio test (LRT) of random effects, variance components, heritability, accuracy, and adjusted average for SPAD, plant height (PH), stalk diameter (SD), and shoot dry matter (SDM) of the inbred lines evaluated in WW (well-water) and WS (water stress) conditions water supply.

Source of variation	SPAD	PH	SD	SDM
	Wald statistic			
Water supply (WA)	1098.27***	4353.03***	4549.80***	5393.32***
Replicates/WA	1076.83***	2542.48***	1842.71***	2462.96***
	Likelihood-ratio test (LRT)			
Block/WA/Season	89.35***	40.76***	46.21***	212.23***
Genotypes (G)	102.59***	201.98***	100.60***	96.35***
G x WA	1.11^{NS}	1.08^{NS}	0.72^{NS}	2.43^{NS}
	Variance components			
σ_g^2	0.192	0.197	0.119	0.099
$\sigma_{g\times e}^2$	0.015	0.007	0.007	0.012
σ_{eWW}^2	0.525	0.266	0.352	0.299
σ_{eWS}^2	0.511	0.279	0.275	0.240
	Heritability and accuracy			
h^2	0.58	0.73	0.59	0.58
	Adjusted means			
	31.13	9.44	7.56	2.32

***: significant at the 0.001 probability level (by Wald test or LRT), respectively. ^{NS} non-significant.

Table 2. Number of markers scored (*raw data*) and the final number of markers (*clean data*) total and per chromosome (Chr) after quality control for all genotyping scenarios used to assess inbred lines evaluated in WW and WS conditions water supply.

	Genotyping scenarios[a]		
	SNP-Array	**GBS-B73**	**GBS-Mock**
Raw data	18,413	131,350	46,926
Clean data	12,704	11,153	4,935
Chrm 1	1,977 (15.6%)	1,651 (15.0%)	Unique chrm
Chrm 2	1,643 (12.9%)	1,411 (12,7%)	
Chrm 3	1,430 (11.3%)	1,269 (11.4%)	
Chrm 4	1,412 (11.1%)	1,177 (10.6%)	
Chrm 5	1,373 (10.8%)	1,336 (12.0%)	
Chrm 6	1,018 (8.0%)	859 (7.7%)	
Chrm 7	957 (7.5%)	831 (7.5%)	
Chrm 8	1,116 (8.8%)	964 (8.6%)	
Chrm 9	973 (7.7%)	855 (7.7%)	
Chrm 10	805 (6.3%)	780 (7.0%)	

[a] SNP-Array: Affymetrix® Axiom Maize Genotyping array; GBS-B73: genotyping-by-sequence with SNP calling using B73 as reference genome; GBS-Mock: genotyping-by-sequence with SNP calling using the mock reference built with all parental lines.

Table 3. Number, average and standard deviation (SD) of significant SNPs per trait in WW and WS conditions water supply and genotyping scenario.

Water Supply	SPAD			PH			SD			SDM		
	SNP-Array	**GBS-B73**	**GBS-Mock**	**SNP-Array**	**GBS-B73**	**GBS-Mock**	**SNP-Array**	**GBS-B73**	**GBS-Mock**	**SNP-Array**	**GBS-B73**	**GBS-Mock**
WW	8	6	5	6	4	2	4	7	4	6	5	4
WS	7	3	5	9	3	3	3	2	3	3	4	5
Overall	15	9	10	15	7	5	7	9	7	9	9	9
Average	7.5	4.5	5	7.5	3.5	2.5	3.5	4.5	3.5	4.5	4.5	4.5
SD	0.5	1.5	0	1.5	0.5	0.5	0.5	2.5	0.5	1.5	0.5	0.5

Table 4. Marker, chromosome (Chr), physical position (pos), -log$_{10}$ of the p-value, minor allele frequency (MAF), allele substitution effect (ASE), the proportion of phenotypic variance explained by the SNP (R^2_{SNP}), and annotation of candidate genes detected by GWAS analysis for traits in three genotyping scenarios under WW and WS conditions water supply.

Trait	Scenario	Marker	Chrm	Posi (bp)	p-value	MAF	ASE	R^2_{SNP}	Putative annotation
SPAD in WW	SNP-Array	*Zm00001d031759*	1	200951693	10.19	0.33	-0.101	9e-04	Protein shoot gravitropism 5
		Zm00001d044194	3	221628210	7.22	0.47	0.067	1e-04	MYB-related-transcription factor 97
		Zm00001d049717	4	41320760	5.72	0.39	0.061	1e-04	Loricrin-related
		Zm00001d018076	5	213811740	6.78	0.30	0.076	1e-04	FLZ-type domain-containing protein
		Zm00001d036010	6	65518917	6.28	0.15	0.100	4e-04	Oligopeptide transporter 7
		Zm00001d036833	6	103276663	6.15	0.31	0.071	1e-04	Putative pentatricopeptide repeat-containing protein
		Zm00001d047696	9	139410201	7.20	0.07	-0.136	1e-04	Senescence associated gene 20
		Zm00001d026477	10	146706556	5.86	0.29	0.069	1e-04	Jasmonate ZIM-domain proten
	GBS-B73	*Zm00001d027695*	1	11300621	6.20	0.08	0.119	1e-04	Glutaredoxin-C13
		Zm00001d028901	1	50207452	7.36	0.11	0.145	9e-04	High chlorophyll fluorescence3
		Zm00001d031759	1	200951623	7.31	0.31	-0.090	4e-04	Protein shoot gravitropism 5
		Zm00001d003603	2	49593983	7.26	0.38	0.082	4e-04	Major facilitator superfamily protein
		Zm00001d018127	5	214805832	6.03	0.05	0.204	9e-04	SNW/SKI-interacting protein
		Zm00001d047574	9	136093835	5.21	0.36	-0.068	1e-04	WRKY domain-containing protein
	GBS-Mock	*Zm00001d036175*	Unique	76554426	5.60	0.09	0.114	1e-04	GDSL esterase/lipase APG
		No hits found	Unique	103179336	8.11	0.15	0.123	9e-04	none
		Zm00001d042735	Unique	201318938	5.26	0.06	0.181	9e-04	RING-type E3 ubiquitin transferase
		Zm00001d042755	Unique	230114762	5.15	0.05	0.137	1e-04	none
		Zm00001d024497	Unique	302570660	6.70	0.38	-0.081	4e-04	none
SPAD in WS	SNP-Array	*Zm00001d030556*	1	143751621	7.08	0.36	0.073	1e-04	Organic cation/carnitine transporter 7
		Zm00001d031759	1	200951693	7.25	0.33	-0.082	4e-04	Protein shoot gravitropism 5
		Zm00001d003871	2	64226257	5.40	0.40	0.061	1e-04	AP2-EREBP-transcription factor 91
		Zm00001d007742	2	238555551	6.82	0.07	-0.130	1e-04	none
		Zm00001d017043	5	183519054	6.75	0.10	-0.113	1e-04	OSJNBb0022F23.8-like protein
		Zm00001d019062	7	14881267	7.12	0.24	-0.084	4e-04	membrane H(+)-ATPase3
		Zm00001d019150	7	19277460	5.67	0.09	-0.112	1e-04	Pathogenesis-related thaumatin superfamily protein
	GBS-B73	*Zm00001d031759*	1	200951623	10.12	0.31	-0.106	2e-03	Protein shoot gravitropism 5
		Zm00001d052519	4	191970203	5.21	0.07	0.155	4e-04	Putative 2-carboxy-D-arabinitol-1-phosphatase

To be continued

		Zm00001d021440	7	152284943	5.77	0.38	0.073	4e-04	HMA domain-containing protein
	GBS-Mock	*No hits found*	Unique	103179336	8.13	0.15	0.15	9e-04	none
		Zm00001d008500	Unique	200328622	5.20	0.07	0.07	1e-04	Histidine-rich calcium-binding protein
		Zm00001d042735	Unique	201318938	10.76	0.06	0.06	5e-03	RING-type E3 ubiquitin transferase
		Zm00001d006357	Unique	306796326	5.56	0.30	0.30	1e-04	Protein GRIP
		Zm00001d029023	Unique	537851163	5.15	0.23	0.23	4e-04	Hexosyltransferase
PH in WW	SNP-Array	*Zm00001d002163*	2	6961288	5.44	0.31	-0.077	4e-04	Diacylglycerol kinase
		Zm00001eb080540	2	43019177	6.67	0.42	-0.081	4e-04	Floury endosperm1
		No hits found	4	181112611	5.32	0.27	-0.078	1e-04	none
		Zm00001d052912	4	204485263	6.92	0.42	0.082	4e-04	Serine/threonine receptor-like kinase NFP
		Zm00001d017978	5	211170831	6.58	0.35	0.086	4e-04	Endoglucanase
		Zm00001d046890	9	109050043	6.74	0.29	0.091	4e-04	Putative ubiquitin-conjugating enzyme family
	GBS-B73	*Zm00001d051672*	4	166302733	5.91	0.11	0.105	1e-04	GPI-anchored protein
		Zm00001d021708	7	160911849	6.78	0.06	0.179	4e-04	Pentatricopeptide repeat-containing protein chloroplastic
		Zm00001d021859	7	164642302	6.91	0.45	-0.089	4e-04	ATP-dependent DNA helicase
		Zm00001d008952	8	26903022	8.45	0.29	-0.102	4e-04	Endoglucanase
	GBS-Mock	*No hits found*	Unique	291972610	5.64	0.26	0.089	4e-04	none
		Zm00001d012719	Unique	305928723	4.99	0.12	-0.108	1e-04	BZIP-transcription factor
PH in WS	SNP-Array	*Zm00001d028464*	1	35732896	6.64	0.08	0.143	4e-04	Rho GTPase activation protein with PH domain
		Zm00001d028660	1	42126116	7.55	0.08	0.149	4e-04	Acyl-CoA N- with type zinc finger domain
		Zm00001eb080540	2	43019177	7.21	0.42	-0.086	4e-04	Floury endosperm1
		No hits found	3	106448106	6.40	0.39	-0.078	4e-04	none
		Zm00001d049192	4	19864256	5.27	0.16	0.080	1e-04	Chaperone protein dnaJ-related
		Zm00001d052912	4	204485263	5.26	0.42	0.066	1e-04	Serine/threonine receptor-like kinase NFP
		Zm00001d053809	4	241454721	5.52	0.43	0.067	1e-04	Defective in cullin neddylation protein
		Zm00001d017978	5	211170831	8.73	0.35	0.097	4e-04	Endoglucanase
		Zm00001d045427	9	21787790	5.30	0.46	0.066	1e-04	Barren stalk fastigiate 1
	GBS-B73	*Zm00001d042481*	3	169056323	6.61	0.42	-0.078	4e-04	Ubiquitin thioesterase OTU
		Zm00001d051672	4	166302733	5.54	0.11	0.103	1e-04	GPI-anchored protein
		Zm00001d008954	8	26903022	10.26	0.29	-0.117	2e-03	Ectonucleotide pyrophosphatase
	GBS-Mock	*Zm00001d034057*	Unique	399525514	5.23	0.34	0.079	4e-04	Protein shoot gravitropism 5
		No hits found	Unique	504442276	5.24	0.36	-0.153	5e-03	none

Continuation

		Zm00001d018703	Unique	922415122	6.27	0.09	0.146	4e-04	F-box domain-containing protein
SD in WW	SNP-Array	*Zm00001d027902*	1	17165877	5.66	0.32	-0.059	1e-04	Protein TIFY
		Zm00001d031832	1	203133117	5.98	0.46	0.059	4e-04	Ubiquitin carboxyl-terminal hydrolase 17
		Zm00001d033038	1	248268614	9.75	0.30	-0.092	3e-03	Stress enhanced protein 1 chloroplastic
		Zm00001d037612	6	131758145	5.96	0.37	-0.061	4e-04	Myb_DNA-bind_3 domain-containing protein
	GBS-B73	*Zm00001d031287*	1	185122781	6.74	0.46	-0.060	4e-04	Oligopeptide transporter 5
		Zm00001d032225	1	217881804	5.75	0.11	-0.095	4e-04	Formin-like protein
		Zm00001d034681	1	299645762	7.71	0.13	-0.106	9e-04	Pentatricopeptide repeat-containing protein
		Zm00001d005090	2	157770180	8.51	0.33	0.077	9e-04	Clathrin heavy chain
		Zm00001d005558	2	178434079	5.79	0.37	-0.055	1e-04	leucine--tRNA ligase
		Zm00001d014899	5	67342178	7.50	0.09	-0.112	4e-04	Tetratricopeptide repeat (TPR)-like superfamily protein
		Zm00001d038199	6	151389473	7.12	0.23	0.073	4e-04	MLO-like protein
	GBS-Mock	*Zm00001d026300*	Unique	205301556	5.03	0.05	-0.109	1e-04	Argonaute 2-like
		No hits found	Unique	277995367	5.37	0.22	0.097	9e-04	none
		Zm00001d001852	Unique	583680097	7.65	0.34	0.066	4e-04	Gibberellin-regulated protein 2
		Zm00001d047956	Unique	831219654	5.16	0.14	0.070	1e-04	Helicase-like transcription factor CHR27
SD in WS	SNP-Array	*Zm00001d028699*	1	43578111	5.80	0.38	-0.052	1e-04	Ypt/Rab-GAP domain of gyp1p superfamily protein
		Zm00001d033038	1	248268614	6.94	0.30	-0.076	4e-04	Stress enhanced protein 1 chloroplastic
		Zm00001d020444	7	114906347	7.16	0.18	-0.092	4e-04	Putative zinc finger motif protein
	GBS-B73	*Zm00001d005090*	2	157770180	12.85	0.33	0.095	2e-03	Clathrin heavy chain
		Zm00001d038199	6	151389473	6.17	0.23	0.065	1e-04	MLO-like protein
	GBS-Mock	*Zm00001d001852*	Unique	583680097	5.41	0.34	0.057	1e-04	Gibberellin-regulated protein 2
		Zm00001d053262	Unique	751364673	6.01	0.48	0.116	6e-03	Calcium-dependent lipid-binding family protein
		Zm00001d047956	Unique	831219654	6.87	0.14	0.089	4e-04	Helicase-like transcription factor CHR27
SDM in WW	SNP-Array	*Zm00001d028912*	1	50880303	7.61	0.07	0.138	9e-04	UPF0481 protein
		Zm00001d031445	1	190284851	6.13	0.35	0.060	9e-04	Ethylene insensitive 3-like 3 protein
		Zm00001d033038	1	248268614	5.60	0.30	-0.062	9e-04	Stress enhanced protein 1 chloroplastic
		No hits found	2	137425632	5.96	0.37	0.057	9e-04	none
		Zm00001d009468	8	65780194	6.51	0.24	-0.067	9e-04	Ereb49 - AP2-EREBP-transcription factor 49
		Zm00001d023272	10	1811784	6.48	0.29	0.059	1e-04	ENTH/VHS family protein
	GBS-B73	*Zm00001d027626*	1	9405820	7.38	0.06	0.120	9e-04	S-adenosyl-L-methionine-dependent methyltransferase superfamily protein

Continuation

		Zm00001d028039	1	21295780	5.47	0.21	-0.060	1e-04	S-acyltransferase
		Zm00001d002541	2	14991268	5.34	0.14	0.070	1e-04	RING-type E3 ubiquitin transferase
		Zm00001d043706	3	207925584	5.24	0.26	0.059	1e-04	Transcription factor
		Zm00001d008944	8	26440815	7.07	0.48	0.057	9e-04	Uncharacterized protein
	GBS-Mock	*No hits found*	Unique	277995367	5.67	0.22	0.091	2e-03	none
		No hits found	Unique	300613355	8.07	0.31	-0.105	5e-03	none
		Zm00001d046354	Unique	320987424	5.56	0.19	0.068	1e-04	GATA transcription factor 20
		Zm00001d008954	Unique	511458983	5.63	0.28	-0.064	9e-04	Ectonucleotide pyrophosphatase/phosphodiesterase family member 3
SDM in WS	SNP-Array	*Zm00001d029438*	1	70502769	7.77	0.41	-0.066	9e-04	RING-type E3 ubiquitin transferase
		Zm00001d016786	5	175971996	5.68	0.17	-0.076	9e-04	PDIL5-3-Zea mays protein disulfide isomerase or PDI-like 5-2
		No hits found	7	82368315	5.52	0.19	0.061	1e-04	none
	GBS-B73	*Zm00001d005090*	2	157770180	5.43	0.33	0.051	1e-04	Clathrin heavy chain
		Zm00001d040705	3	59361883	5.31	0.22	-0.052	1e-04	Peroxidase
		Zm00001d008944	8	26440815	5.67	0.48	0.050	1e-04	Uncharacterized protein
		Zm00001d026638	10	149275033	6.76	0.09	-0.094	1e-04	Calcium-binding EF hand family protein
	GBS-Mock	*No hits found*	Unique	277995367	6.61	0.22	0.100	2e-03	none
		No hits found	Unique	300613355	7.29	0.31	-0.096	3e-03	none
		Zm00001d018001	Unique	440728659	5.21	0.13	0.063	1e-04	Xanthine/uracil permease family protein

Conclusion.

Table 5. Pearson correlation among significant markers from the GBS-Mock scenario with known functions and markers from the SNP-Array scenario for traits under WW and WS conditions water supply.

Trait	GBS-Mock		SNP-Array		*r*
	Marker	Position	Chrm	Position	
SPAD in WW	Zm00001d036175	76554426	6	66088433	0.71
			6	68443154	
	Zm00001d042735	201318938	6	76875108	0.76
			6	77806057	
	Zm00001d042755	230114762	1	99951335	0.41
	Zm00001d024497	302570660	10	22670857	0.52
SPAD in WS	Zm00001d008500	200328622	3	179265803	0.40
			4	214338748	
	Zm00001d042735	201318938	6	76875108	0.76
			6	77806057	
	Zm00001d006357	306796326	2	206059268	0.49
	Zm00001d029023	537851163	5	199240894	0.41
PH in WW	Zm00001d012719	305928723	9	1840217	0.36
PH in WS	Zm00001d034057	399525514	1	279982555	0.51
	Zm00001d018703	922415122	8	132436479	0.53
SD in WW	Zm00001d026300	205301556	1	62995980	0.38
	Zm00001d001852	583680097	7	119716831	0.35
	Zm00001d047956	831219654	10	144416974	0.87
SD in WS	Zm00001d001852	583680097	7	119716831	0.35
	Zm00001d053262	751364673	1	33596576	0.30
			1	156697884	
			2	2509863	
			2	176643239	
			2	226012848	
			3	200304796	
			4	30519074	
			4	74777100	
			4	194400472	
			4	195155904	
			4	235373522	
			5	153728179	
			5	219174423	
			7	119716831	
			8	26683716	
			8	136183604	
	Zm00001d047956	831219654	10	144416974	0.87
SDM in WW	Zm00001d046354	320987424	9	84921243	0.94
	Zm00001d008954	511458983	10	94908425	0.46
SDM in WS	Zm00001d018001	440728659	5	211640184	0.47

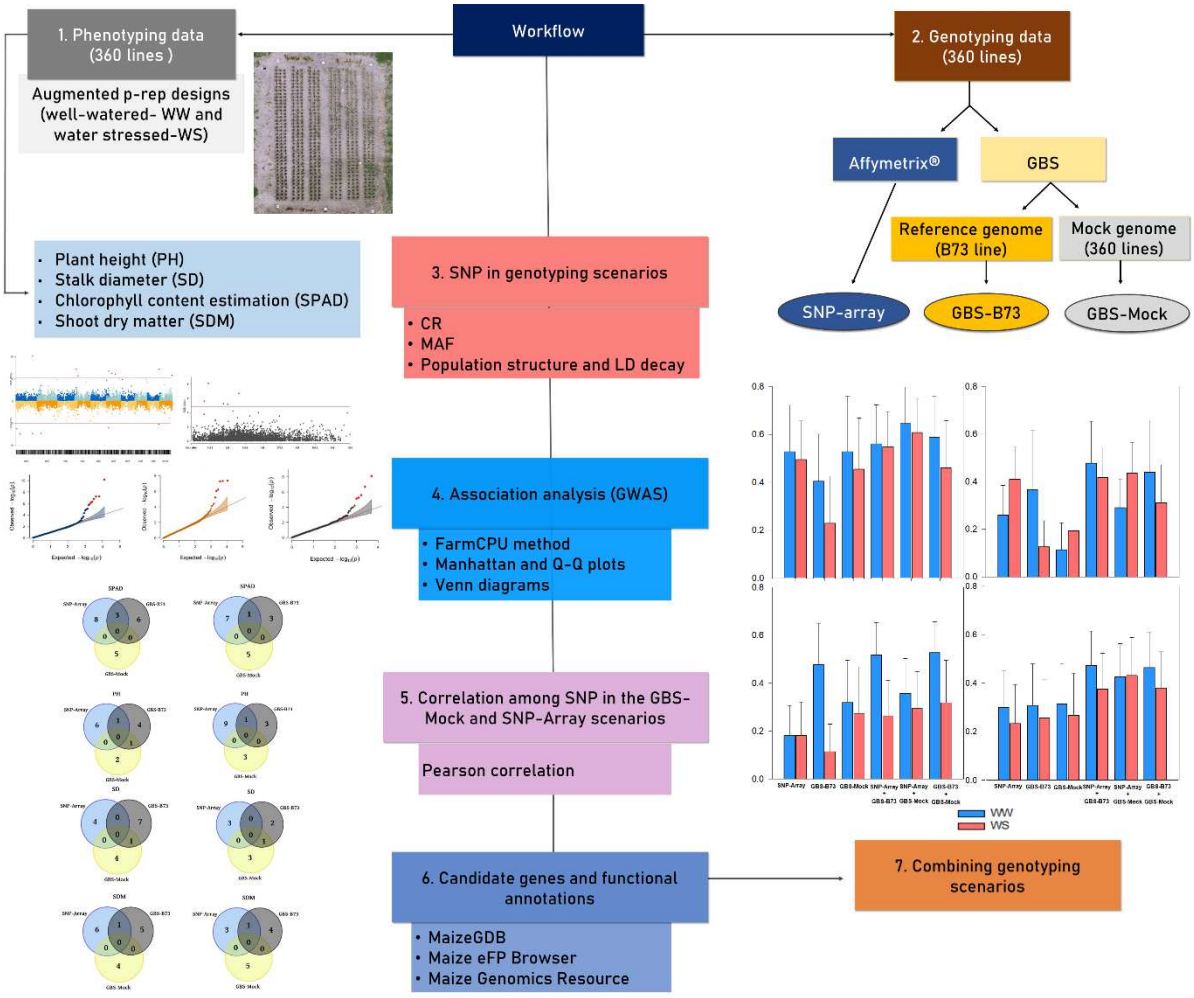

Figure 1. The workflow employed in the study. Different colors are used to represent distinct phases of the analysis.

Figure 2. Aerial image showing an overview of the experimental area. Block on the right shows WW condition and the left WS condition, blocks are 1.5m apart.

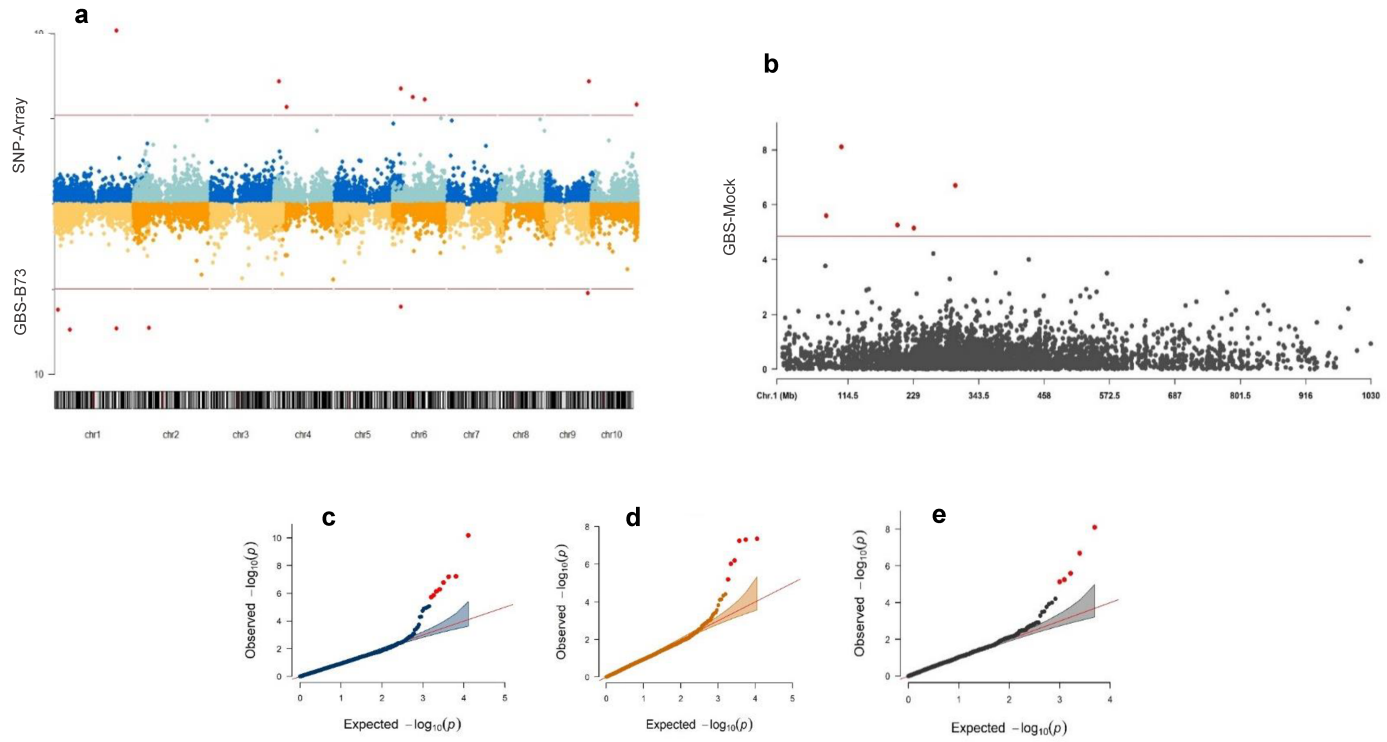

Figure 3. Manhattan plot and quantile-quantile (Q-Q plots) for Genome-Wide Association Study (GWAS) comparison genotyping platforms for tropical maize to water stress. **a** Manhattan plot of GWAS based on SNP-Array and GBS-B73 datasets; **b** Manhattan plot of GWAS based on GBS-Mock datasets. The x-axis shows the numbers chromosome, and the y-axis shows the -log10 of P-values for statistical significance. The horizontal red lines represent the standard genome-wide suggestive significance(threshold). The dots above the red line represent the SNP significance. **c** Q-Q plots of GWAS based on SNP-Array datasets; **d** Q-Q plots of GWAS based on GBS-B73 datasets; **e** Q-Q plots of GWAS based on GBS-Mock datasets. The x-axis shows the -log10 of expected P-values of the association from the chi-square distribution, and the y-axis shows the -log10 of P-values from the observed chi-square distribution. The dots represent the observed data with the top hit SNP being and the red line is the expectation under the null hypothesis of no association

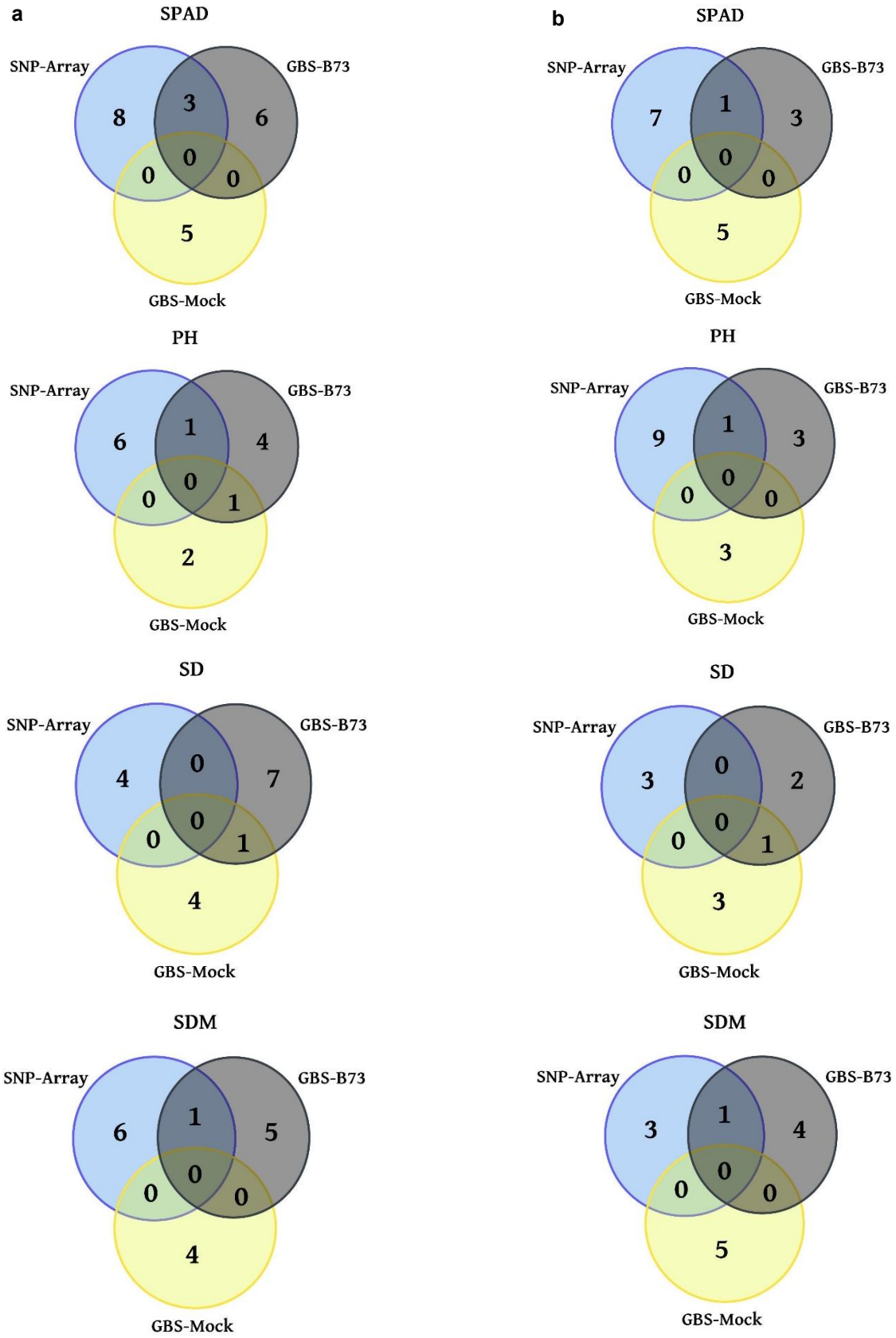

Figure 4. Venn diagrams with the number of significant SNPs for traits in three genotyping scenarios. **a** WW (well-watered) water supply condition column; **b** WS (water-stressed) water supply condition column. SPAD, PH (plant height), SD (stalk diameter), and SDM (shoot dry matter).

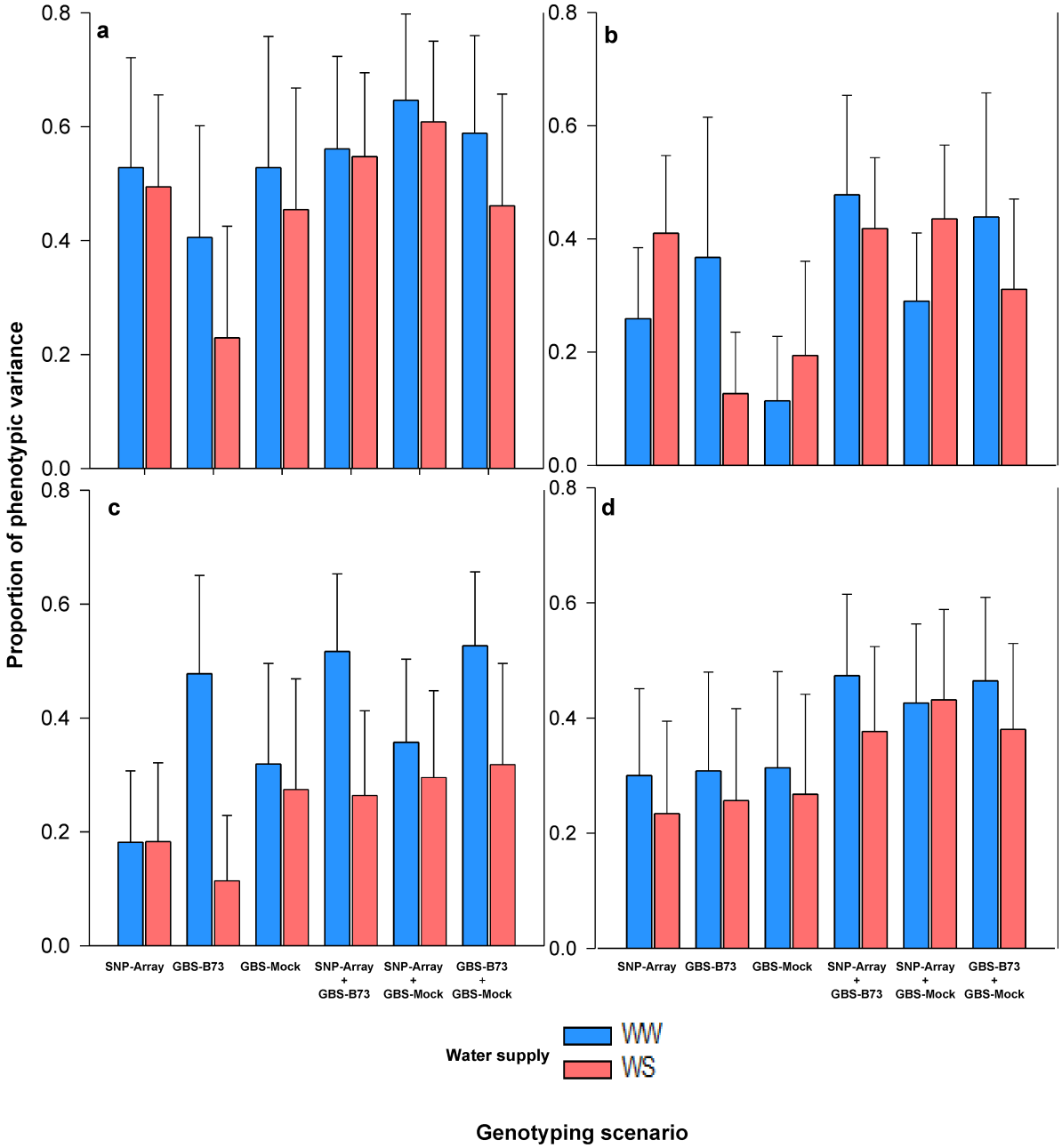

Figure 5. Proportion of phenotypic variance explained by the SNP (R^2_{TOT}) per trait in WW and WS conditions water supply and genotyping scenario. **a** SPAD, **b** PH (plant height), **c** SD (stalk diameter) and **d** SDM (shoot dry matter).

APPENDIX B - SUPPLEMENTARY MATERIAL

Table S1. DNA fragments obtained via BLAST in GBS-Mock for each character under in WW (well-watered) and WS (water stressed) in water supply conditions. SPAD, PH (plant height), SD (stalk diameter) and SDM (shoot dry matter).

SPAD in WW	
DNA fragments	GCAGATCTGCTCACATGTTTCCGGCTCCATCGAAACAGAAC TGAAGCATCCAGCTTCACATACTAGTAGACCACCTTATGGA GAGTGCAATAAGATCGAAGGTCACGTCCAAAAAAAAAAAAAA ACGAGGTCTTAGTTCCAGAAGACAGATTCAAAAAAAAAAAAA TTCCCATTTCCCTCCTTCCTACAAACTGCTTCACAAAAAAA

SPAD in WS	
DNA fragments	TGAAGCATCCAGCTTCACATACTAGTAGACCACCTTATGGA CATGGCCTGCAGCTAGACAAAGTCCGGCACCAACGAACGGC GAGTGCAATAAGATCGAAGGTCACGTCCAAAAAAAAAAAAAA ACTCCCTGCTGCAGCAAAAAATGGTTGTTTGTAATAGTCAC CTGCCTGGCGTGCGTGAGGAGTCTGGCGTCTGACAGATCGG

PH in WW	
DNA fragments	TCCCAGGAAAAGAAAAAAAAAACTAAATTACAGATCGAAGA CTGCAGCCCACGAGCGAGGTGGTGGAGGAGGACGCCCTGCT

PH in WS	
DNA fragments	CTTCGCCTTCTGCAGGTCGCGCGCACCATCCTAGAAGACAA ACAAGTAGTTGTGAAGTTACCACAGCATTCATTGCTATTTG ACCTGCAACACGCGAATGACCGTACCTGCTGTTCCATTGTT

SD in WW	
DNA fragments	AGCAATCCTGTAGAACAGACGGCAGAGAGCTGCATAACGTC CGTCCAGCACGAATAGCCTTGTCCATGGCTTTATCGCGTTG TAGAGCTGGACCGGCCGGGAGGGTTTTTACAGATCGGAAAA AAAAAAAAAAAAAAACCTAAGAAGAAGGCCAGAGTGAGGTA

SD in WS	
DNA fragments	AGAGCTGGACCGGCCGGGAGGGTTTTTACAGATCGGAAAA CATTGAACAGGATGGTCCAAACCCCTCTTCGCCAAACGAT AAAAAAAAAAAAAACCTAAGAAGAAGGCCAGAGTGAGGTA

SDM in WW	
DNA fragments	CGTCCAGCACGAATAGCCTTGTCCATGGCTTTATCGCGTTG TTGTTGCGCGCAATTACTTTGGATGAGGCAAACCCTTAGGG CCCTTGTGGAGGAGTGGTCCATGCGGCCCCAAAGGTATCCT CAGCTGTTTGGGCTGCAGCAAGGGGCGGACCATGTGTGTGC

SDM in WS	
DNA fragments	CGTCCAGCACGAATAGCCTTGTCCATGGCTTTATCGCGTTG TTGTTGCGCGCAATTACTTTGGATGAGGCAAACCCTTAGGG GTGCAAGAGCAAATACAATAATTGGCATATAGCCAAACAAA

Table S2. Marker, chromosome (Chr), physical position (pos), annotation of candidate genes and common function detected by GWAS analysis for traits in three genotyping scenarios under WW (well-watered) and WS (water stressed) conditions water supply. SPAD, PH (plant height), SD (stalk diameter) and SDM (shoot dry matter).

Trait	Scenario	Marker	Chrm	Posi (bp)	Putative annotation	Common function
SPAD in WW	SNP-*Array*	*Zm00001d031759*	1	200951693	Protein shoot gravitr. 5	Zinc finger proteins-gravitropism
	GBS-B73	*Zm00001d031759*	1	200951623	Protein shoot gravitr. 5	
	SNP-*Array*	*Zm00001d026477*	10	146706556	Jasmonate ZIM-domain proten	Jasmonic acid
	GBS-B73	*Zm00001d027695*	1	11300621	Glutaredoxin-C13	
	SNP-*Array*	*Zm00001d044194*	3	221628210	MYB-related-transcription factor 97	Circadian clock
	GBS-B73	*Zm00001d018127*	5	214805832	SNW/SKI-interacting protein	
SPAD in WS	SNP-*Array*	*Zm00001d031759*	1	200951693	Protein shoot gravitr. 5	Zinc finger proteins-gravitropism
	GBS-B73	*Zm00001d031759*	1	200951623	Protein shoot gravitr. 5	
PH in WW	SNP-*Array*	*Zm00001d017978*	5	211170831	Endoglucanase	Cellulose catabolic process
	GBS-B73	*Zm00001d008952*	8	26903022	Endoglucanase	
	GBS-B73	*Zm00001d021708*	7	160911849	Pentatricopeptide repeat-containing p. chloroplastic	ABA
	GBS-Mock	*Zm00001d012719*	Unique	305928723	BZIP-transcription factor	
PH in WS	SNP-*Array*	*Zm00001d053809*	4	241454721	Defective in cullin neddylation protein	Ubiquitination
	GBS-B73	*Zm00001d042481*	3	169056323	Ubiquitin thioesterase OTU	
SD in WW	GBS-B73	*Zm00001d014899*	5	67342178	Tetratricopeptide repeat (TPR)-like superfamily	Gibberellin
	GBS-Mock	*Zm00001d001852*	Unique	583680097	Gibberellin-regulated protein 2	
SD in WS	GBS-B73	*Zm00001d005090*	2	157770180	Clathrin heavy chain	ABA
	GBS-Mock	*Zm00001d053262*	Unique	831219654	Calcium-dependent lipid-binding family protein	
SDM in WW	SNP-*Array*	*Zm00001d031445*	1	190284851	Ethylene insensitive 3-like 3 protein	Ethylene
	GBS-B73	*Zm00001d027626*	1	9405820	S-adenosyl-L-methionine-dep. methy.superf. protein	
SDM in WS	SNP-*Array*	*Zm00001d016786*	5	175971996	PDI-like 5-2	ABA
	GBS-B73	*Zm00001d005090*	2	157770180	Clathrin heavy chain	