Euphytica (2021)217:60
https://doi.org/10.1007/s10681-021-02800-4

# Genome-wide association study of drought tolerance in cassava

**Priscila Patrícia dos Santos Silva · Massaine Bandeira e Sousa · Eder Jorge de Oliveira · Carolina Vianna Morgante · Carlos Roberto Silva de Oliveira · Simone Leal Vieira · Jerônimo Constantino Borel**

Received: 22 January 2019 / Accepted: 1 March 2021


**Abstract** Despite being considered a drought-tolerant species, cassava (*Manihot esculenta* Crantz) exhibits reduced growth and productivity under prolonged water stress. Therefore, the objective of this study was to identify the genomic regions associated with water deficit in cassava using a genome-wide association study (GWAS). A total of 49 cassava genotypes were evaluated under two water conditions: irrigated (control) and water deficit. An additional set of 252 clones grown for several years under water deficit conditions were used to validate the GWAS. The following traits were evaluated: yields of storage roots (RoY), shoots (ShY), and starch (StY); root dry matter content (DMC); drought tolerance index (DTI) and drought tolerance stability index (DTSI). The GWAS was performed using a multiple mixed linear model with a kinship matrix and population structure for each trait and water condition. Broad-sense heritability ($h^2$) estimates were variable depending on the water conditions and characteristics under analysis. Overall, 62 single nucleotide polymorphisms (SNPs) were identified across all 18 cassava chromosomes. Additionally, it was possible to identify

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s10681-021-02800-4.

P. P. dos Santos Silva
Centro de Ciências Agrárias, Ambientais E Biológicas, Universidade Federal do Recôncavo da Bahia, Cruz das Almas, BA 44380-000, Brazil
e-mail: prisilva.bio@gmail.com

M. B. eSousa · E. J. de Oliveira (✉)
Nugene, Embrapa Mandioca E Fruticultura, Cruz das Almas, BA 44380-000, Brazil
e-mail: eder.oliveira@embrapa.br

M. B. eSousa
e-mail: massainebandeira@hotmail.com

C. V. Morgante
Embrapa Semiárido, Rodovia BR-428, Petrolina, PE 56302-970, Brazil
e-mail: carolina.morgante@embrapa.br

C. R. S. de Oliveira · S. L. Vieira · J. C. Borel
Universidade Federal do Vale do São Francisco, Petrolina, PE 56300-000, Brazil
e-mail: carlos-robertoliveira@hotmail.com

S. L. Vieira
e-mail: simoneleal.25@hotmail.com

J. C. Borel
e-mail: jeronimo.borel@univasf.edu.br

Published online: 16 March 2021

specific and stable markers across environments. The identified SNPs comprise approximately 160 transcripts, of which 119 were previously described and 41 have known functional annotations. Some of these transcripts are related to proteins involved in drought tolerance, such as APETALA 2 domain (AP2), photosystem II oxygen-evolving enhancer protein, PR5-like receptor kinase-related, beta-fructofuranosidase/saccharase, leucine zipper, and bZIP transcription factor. There is great potential for applying these SNPs in the marker-assisted selection of new cassava varieties that are tolerant of water deficit.

**Keywords** *Manihot esculenta* · Associative mapping · SNP · Drought tolerance

## Introduction

Cassava (*Manihot esculenta* Crantz) is the third most important source of calories in the tropics after rice and maize, providing food for over 800 million people, mainly in tropical regions in Latin America, Asia, and Africa (Ceballos et al. 2010; Liu et al. 2011; CIAT 2017). Cassava is typically grown by small-scale farmers in tropical countries in the equatorial region (between 30° north and 30° south of the Equator) with annual rainfalls of 500 mm (in semiarid areas) to 2000 mm (humid ecological zones) at altitudes ranging from 0 to 2000 m. This range indicates its wide adaptability to different environments and farming ecosystems (El-Sharkawy 2012; Okogbenin et al. 2013). The phenotypic plasticity of cassava is also reflected in its drought tolerance since it can produce significant yields compared to other annual crops, even under low rainfall and soil fertility conditions (El-Sharkawy 2007; Okogbenin et al. 2013).

Water deficit is one of the most important stressors with a direct effect on agriculture because it directly interferes in plant growth and development (Cattivelli et al. 2008). Cassava cultivation under water deficit conditions, inadequate management strategies, the non-use of pesticides and agricultural inputs, and the use of varieties with low yield potential can reduce crop production (Oliveira et al. 2015). This scenario is observed in semiarid regions of the Brazilian Northeast since the average root yield was 9.5 t $ha^{-1}$ when compared to the 23.6 t $ha^{-1}$ obtained by certain genotypes under experimental water stress conditions (IBGE, 2017; Oliveira et al. 2015). Furthermore, El-Sharkawy (2012) found that cassava is adaptable to prolonged periods of water stress and yields reasonable root development in many situations. For instance, even when cultivated in water deficit and low-temperature environments during winter, some cassava cultivars had root yields of up to 66 t $ha^{-1}$ in the Limpopo River basin in South Africa (Ogola and Mathews 2011). Therefore, important genetic variability in cultivated cassava may facilitate greater yield potential for cultivation in semiarid regions through the improvement and selection of genotypes that are more tolerant of water deficit.

Water deficit tolerance is a complex quantitative trait regulated by several genes that hamper the selection process under field conditions (Okogbenin et al. 2013). Notably, the identification of genomic regions involved in stress response, understanding of genetic control, and the development of marker-assisted selection tools can contribute to improving the phenotypic selection process and thus reduce the time required for developing new cassava varieties with water deficit tolerance.

Efforts have been and continue to be undertaken to identify regions linked to complex cassava traits via quantitative trait loci (QTL) mapping (Masumba et al. 2017; Sedano et al. 2017). Information on the QTLs that regulate water deficit response can elucidate the physiological basis of drought tolerance and aid in the selection of genotypes with higher yields under water stress conditions (Tuberosa and Salvi 2006). However, QTL analysis has certain limitations, such as only a small number of QTLs with the greatest effects being detected in contrast to the polygenic nature of the total genetic variation observed for most quantitative traits (Dekkers 2004). This limitation is largely due to only the allelic diversity of the bi-parental population being detected and the resolution of the genetic mapping being limited due to the reduced number of recombinants produced in the segregating populations.

Next-generation DNA sequencing (NGS) technologies have facilitated large-scale single nucleotide polymorphism (SNP) genotyping of various species using a fast and affordable process. The availability of genomic information on a large scale has enabled the use of linkage disequilibrium (LD) strategies, such as genome-wide association studies (GWAS), to further refine trait mapping at the population level. GWASs

explore historical recombination events that naturally occurred through several generations using QTL mapping (Rosenberg et al. 2010; Korte and Farlow 2013). GWAS overcome some limitations of conventional QTL analysis since they allow the identification of phenotypes of interest, provide insight into the genetic architecture of traits, and suggest potential candidates for mutagenesis and transgenesis. Additionally, they allow the selection of parental lines for QTL analysis and are thus a complementary strategy for the more refined mapping of complex traits.

In maize, molecular markers with significant associations to drought tolerance have been identified, validated, developed, and effectively utilized for selecting tolerant genotypes (Hao et al. 2010; Liu et al. 2013). In other crops, the same approach has been used to understand drought tolerance based on various physiological traits (Wehner et al. 2015) and for selection based on the drought tolerance coefficient (Ma et al. 2016), plant height, and flowering traits (Farfan et al. 2015), and grain yield (Pantalião et al. 2016). Furthermore, GWASs have served as a powerful tool for identifying genomic regions linked to other abiotic stressors (e.g., nitrogen use efficiency) (Morosini et al. 2017).

Since drought tolerance has a strong genotype × environment interaction (G × E), a GWAS based on multi-environment trials provides an additional advantage by facilitating the detection of regions associated with genes expressed in specific conditions and environments, thereby reducing the noise of non-genetic environmental factors and making the results more robust and reliable (Mathews et al. 2008; Malosetti et al. 2013; Farfan et al. 2015; Gutiérrez et al. 2015). However, GWASs of the effects of abiotic stress have not been well explored in cassava crops. Therefore, the objective of the present study was to evaluate a diverse panel of cassava under two hydric conditions to identify the genomic regions associated with cassava drought tolerance through a GWAS.

## Material and methods

### Phenotypic data 1

Forty-nine genotypes, including landraces and improved varieties with a history of drought tolerance (collected in semiarid regions or selected under dry conditions), were evaluated in field conditions (Supplementary Material, Table S1). The cassava genotypes were evaluated in two growing seasons (2012/2013 and 2013/2014) and subjected to one of two hydric conditions: well-watered (WW) or water deficit (WD). In both conditions, we used a complete randomized block design (CRBD) using three replicates with ten plants per plot (two rows with five plants) and a spacing of 0.90 m between rows and 0.80 between plants. The planting was performed using cuttings (16 cm) following standard recommendations and agricultural practices for the crop, according to Souza et al. (2006).

All blocks were irrigated for up to four months after planting (MAP). Water was supplied every two days via inline dripping (4 L $h^{-1}$) according to the plants' evapotranspiration, which was estimated using data provided by a meteorological station close to the experimental area. After this period, irrigation for half of the blocks (WD treatment group) was suspended until harvest for drought assessment in 49 genotypes, while irrigation was maintained in the other blocks (WW treatment group).

The trial was performed at the Bebedouro Experimental Station at Embrapa Semiarid, Petrolina, State of Pernambuco, Brazil (9°22′S, 40°22′W at 376 m altitude). The climate in this region is semiarid (Bsh type) and with low annual rainfall. The soil in this region is classified as red-yellow dystrophic, clay loam texture, with flat relief. The main climatic conditions in the different years of study, e.g., average temperature, humidity, and precipitation is presented in Fig. 1 (Embrapa Semiárido 2013–2020). The growing seasons were marked by meteorological conditions with low precipitation volumes (i.e., drought), primarily from May to October.

Harvests were conducted at 12 MAP, and the following traits were evaluated: shoot yield (ShY in t $ha^{-1}$), determined by weighing the aboveground parts of plants after cutting them 10 cm from the soil surface; fresh root yield (RoY in t $ha^{-1}$), measured by weighing all roots (commercial and non-commercial) from the plots; dry matter content in the roots (DMC%); and starch yield (StY in t $ha^{-1}$), obtained by multiplying the starch content and fresh root yield. Both dry matter and starch content were obtained by specific weight according to Kawano et al. (1987).

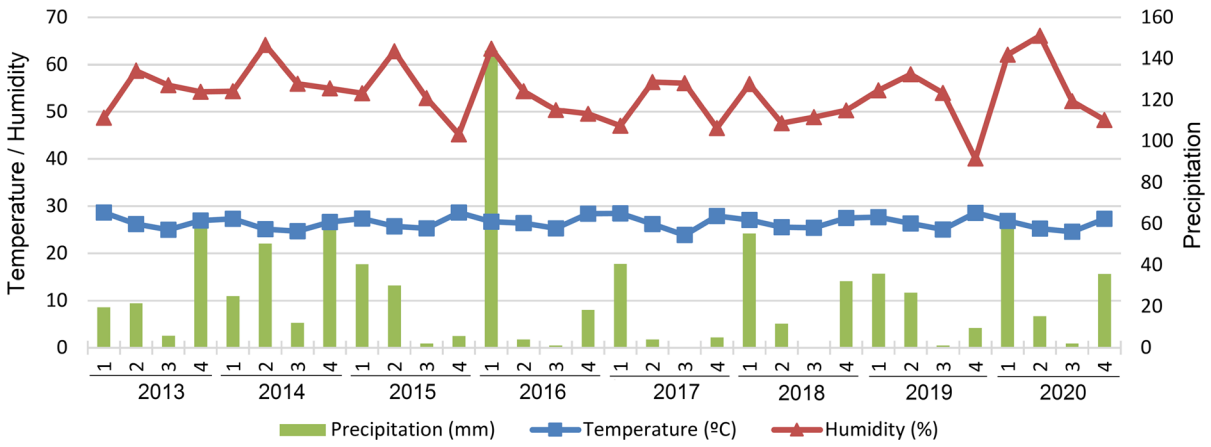



**Fig. 1** Average climatic data of temperature (°C) and humidity (%), as well as the total precipitation (mm) assessed quarterly from 2013 to 2020 in Petrolina (Pernambuco, Brazil)

Phenotypic data 2

Previous studies have demonstrated the possibility of early drought tolerance screening for cassava clones (Vitor et al. 2019). Therefore, we initiated a series of field evaluations to assess a large number of cassava clones at the Agricultural Sciences Campus of the Federal University of Vale do São Francisco, located in the municipality of Petrolina (State of Pernambuco, Brazil) (9°16′10″ S, 40°33′43″ W and an average altitude of 373 m). This study site is located in the region of Vale do Submédio São Francisco and has a flat relief. The soil is classified as Yellow Argisol and the climate is Bsh type (semiarid). From 2013 to 2020, 252 cassava clones (Supplementary Material, Table S2) were evaluated, with the number of evaluations for each clone ranging from 1 to 6 years. The clones were evaluated in a CRBD with four replications, where each plot comprised seven plants. The climatic data from 2013 to 2020 is presented in Fig. 1.

The main difference between these field trials and previous ones is that plants in the current trial were kept under supplementary irrigation until 3 MAP and then submitted to another three months of water stress. Due to the high number of evaluated genotypes and the need for rapid screening for drought tolerance, the clones were not replicated in well-watered trials. Therefore, these clones were only submitted to water deficit conditions; therefore, it was not possible to calculate drought tolerance index and stability. The experimental conditions, crop management, and evaluated characteristics were similar to those of Experiment 1.

This second set of phenotypic data was used for multi-environment analysis and further validation of the genomic regions associated with tolerance to water deficit identified in previous tests. The stages of the data collection from the different field trials are presented as a supplementary material (Figure S1).

Genotypic data

DNA samples were collected from a cassava diversity panel and extracted using the cetyltrimethylammonium bromide (CTAB) protocol described by Doyle and Doyle (1987). Quantification was performed on agarose gel 1.0% (w/v) stained with ethidium bromide (1.0 mg $mL^{-1}$) using a standard series of phage Lambda concentrations (Invitrogen P7589 lambda DNA). The genomic DNA was adjusted to a final concentration of 20 ng $\mu l^{1}$.

Next, the DNA samples were genotyped at the Genomic Diversity Facility at Cornell University via genotyping by sequencing (GBS). Briefly, the DNA was digested using the *Ape*KI restriction enzyme to prepare the libraries following the protocol described by Elshire et al. (2011). The linkage between the *Ape*KI-cut adapter and the genomic DNA was performed after the digestion of the samples, and the samples were multiplexed for sequencing using a Genome Analyzer 2000 (Illumina, Inc., San Diego, CA). The genomic data were subjected to quality control by removing markers with a call rate of $\geq$ 0.80

and a minor allele frequency (MAF) of $< 0.05$. The data were imputed using Beagle software (Browning and Browning 2009). After the quality control step, the marker matrix consisted of 25,597SNPs.

## Predicted genotypic values and selection indices for the GWAS

Cassava genotypes were initially evaluated in four environments: the 2013/2014 growing season in well-watered (WW13) and water deficit (WD13) conditions and the 2014/2015 growing season in well-watered (WW14) and water deficit (WD14) conditions. In the second stage of this study, an additional 252 cassava clones were evaluated between 2013 and 2020 in seven other trials under water deficit conditions only. For each trait and hydric condition, the best linear unbiased predictions (BLUPs) for the GWAS analysis were obtained via single and multi-environment analyses.

In the single-environment analysis, the phenotypic observations $Y_{ij}$ of genotype $i$ in replicate $j$ were adjusted according to Eq. (1):

$$Y_{ij} = \mu + g_i + r_j + \epsilon_{ij}, \quad (1)$$

where $\mu$ is the general mean, $g_i$ is the random effect vector of genotype $i$, $r_j$ is the vector of fixed effects of replicate $j$, and $\epsilon_{ij}$ is the random residual effect of genotype $i$ for replicate $j$. For the multi-environment model, the phenotypic observations $Y_{ijk}$ of genotype $i$ in replicate $j$ within environment $k$ were modeled by Eq. (2):

$$Y_{ij} = \mu + e_k + g_i + \left({}^{r}/_{e}\right)_{jk} + (g \times e)_{ik} + \epsilon_{ijk}, \quad (2)$$

where $\mu$ is the overall mean, $e_k$ is the fixed effect of environment $k$, $g_i$ is the random effect of genotype $i$, $(r/e)_{jk}$ is the random effect of replicate $j$ nested in environment $k$, $(g * e)_{ik}$ is the random effect of the interaction between genotypes and environments, and $\epsilon_{ijk}$ is the random residual effect of genotype $i$ on replicate $j$ in environment $k$. The BLUPs and variance components were estimated using the restricted maximum likelihood (REML) method and broad-sense heritability ($h^2$) using the *gdata* package in R software version 4.0.1 (R Core Team R 2020).

The BLUPs of all traits were used to estimate the drought tolerance index (DTI) (Fernandez 1992) and drought tolerance stability index (DTSI) (Bouslama and Schapaugh 1984). The indices were calculated from Eqs. (3 and 4)

$$DTI = \frac{Y_s \times Y_p}{\left(\overline{y}_p\right)^2} \quad (3)$$

$$DTSI = \frac{Y_s}{Y_p}, \quad (4)$$

where, $Y_s$ and $Y_p$ are the traits of a given genotype under dry and irrigated conditions, respectively, and $\overline{y}_p$ is the average of all genotypes for a given trait under the well-irrigated condition.

Moreover, $h^2$ was estimated by the environment according to $h^2 = \frac{\sigma_G^2}{\sigma_F^2 + \sigma_E^2}$, where $\sigma_G^2$ is the genotypic variance, $\sigma_F^2$ is the phenotypic variance, and $\sigma_E^2$ is the environmental variance. Considering each hydric condition, $h^2$ was estimated using the expression $h^2 = \frac{\sigma_G^2}{(\sigma_G^2 + \sigma_A^2 + \frac{\sigma_{GxE}^2}{r} + \frac{\sigma_E^2}{rxa})}$, where $\sigma_G^2$ is the genotypic variance, $\sigma_A^2$ is the environmental variance, $\sigma_{GxA}^2$ is the variance of the genotype × environment interaction, $\sigma_{GxA}^2$ is the error variance between plots, $r$ is the number of replicates, and $a$ is the number of environments.

## Linkage disequilibrium, kinship, and population structure

The LD estimates for each chromosome were assessed using a correlation coefficient ($r^2$) between two loci. LD plots were generated using the *LDheatmap* package in R software version 4.0.1 (R Core Team 2020). To investigate LD decline, $r^2$ values were plotted against genetic distance in base pairs (bp) using nonlinear regression to fit according to Weisberg (2005).

A heatmap of the genomic relationships between cassava genotypes was generated using the VanRaden (2008) method, implemented in the *GAPIT* (Genome Association and Prediction Integrated Tool) package (Lipka et al. 2012) in R software version 4.0.1 (R Core Team 2020). The population structure was estimated using fastStructure (v1.0) clustering and stratification software (Raj et al. 2014). The fastStructure algorithm determines the number of groups (K) that best explains the population structure. In this case, multiple K

options were tested (1–10) to determine the ideal number of groups that best fit the population structure.

### GWAS analysis

Three phenotypic inputs were used for each trait in the GWAS analyses to identify QTLs considering specific environments, multi-environments as well as the DTIs and DTSIs. The utilized variables were as follows: (1) BLUPs for each trait and environment (Eq. 1); (2) BLUPs obtained by the multi-environment model (Eq. 2), and (3) BLUPs for each trait considering drought tolerance and DTSIs based on analyzing the WW and WD data simultaneously (Eqs. 3 and 4).

The GWAS analyses were performed using the multiple mixed linear model (MLMM) implemented in the *FarmCPU* (fixed and random model circulating probability unification) package in R software version 4.0.1 (R Core Team 2020). The efficient mixed-model association (EMMA) algorithm implemented in *FarmCPU* reduces computational time in the estimation of variance components for each marker (Kang et al. 2008).

The MLMM model is divided into two steps: a fixed effects model (FEM) and a random effects model (REM), which were used iteratively. Initially, REM estimates the markers associated with the traits as covariates to control false positives and then uses them to obtain the genomic kinship matrix (kinship). These associated markers are considered pseudo-quantitative trait nucleotides (QTNs). Then, the FEM tests all markers individually while using the kinship matrix as a covariate to control false positives and false negatives. At each iteration, the *p*-value of the test markers and associated markers is evaluated. The kinship matrix and population structure were used as covariates by the MLMM model, as proposed by Yu et al. (2006).

Bonferroni correction was used to correct for multiple tests, where the significance of associations between the markers and the phenotype were estimated using the $-\log_{10}$ *p*-value at 1 and 5% significance levels. Additionally, the variance of the significant SNPs was estimated by the equation $\widehat{\sigma^2_{SNP}} = \widehat{a^2}.p(1-p)$, where $\widehat{a}$ is the estimated effect of SNP and $p$ is the MAF (Zhang et al. 2010). The effect of the SNP was expressed in terms of the proportion of the genetic variance explained by the marker.

### In silico annotation of SNPs

The putative biological functions of significant SNPs were determined by aligning the SNP sequences with proteins related to water deficit tolerance using a 20 kb window in the Phytozome database (http://www.phytozome.net) using BLASTX.

## Results

### Variance components and heritability

Heritability ($h^2$) was generally high, especially in the well-watered condition in 2013/2014, in which it ranged from 0.64 (DMC) to 0.80 (RoY and StY). Under the water deficit condition in 2013/2014, the values were lower, ranging from 0.30 (ShY) to 0.60 (DMC) (Table 1).

Under the well-watered condition in 2014/2015, the $h^2$ values of all traits were similar to those from 2013/2014, ranging from 0.50 (DMC) to 0.81 (RoY). Similar to the 2013/2014 season, $h^2$ was lower in the water deficit condition than in the well-watered experiments for all traits except DMC, whose estimates were higher in the latter condition ($h^2 = 0.81$). In the multi-environment analysis of the well-watered condition, $h^2$ values ranged from low (0.20 for ShY) to high (0.69 for RoY), while $h^2$ estimates for the water deficit condition ranged from low (0.26 for ShY) to medium (0.47 for RoY). Except for the ShY trait, $h^2$ estimates were higher in the well-watered condition.

Among all traits, the lowest genetic correlation values (ranging from 0.22 to 0.29) were identified for the water deficit conditions in 2013 and 2014 (WD13 and WD14) (Table 2), while the highest values were identified for the well-watered conditions during the two growing seasons (WW13 and WW14). On the other hand, genetic correlations of medium magnitude were identified for the well-watered and water deficit conditions in 2013 (WW13 × WD13), whose variations were 0.39 (DMC) to 0.58 (ShY). Additionally, estimates of covariance between environments were greater than zero for all traits.

**Table 1** Variance components and broad sense heritability estimates for single and multi-environment analyses of well-watered and water deficit conditions for fresh root yield (RoY), shoot yield (ShY), dry matter content (DMC), and starch yield (StY)

| Components | Well-watered 2013/2014 | | | | Well-watered 2014/2015 | | | |
|---|---|---|---|---|---|---|---|---|
| | RoY | ShY | DMC | StY | RoY | ShY | DMC | StY |
| $\sigma^2_G$ | 204.35 | 38.77 | 6.82 | 15.71 | 95.47 | 71.09 | 3.35 | 6.77 |
| $\sigma^2_\varepsilon$ | 51.36 | 18.91 | 3.89 | 3.86 | 21.79 | 66.42 | 3.32 | 1.75 |
| $h^2$ | 0.80 | 0.67 | 0.64 | 0.80 | 0.81 | 0.52 | 0.50 | 0.79 |
| | Water-deficit 2013/2014 | | | | Water-deficit 2014/2015 | | | |
| $\sigma^2_G$ | 20.23 | 4.34 | 15.59 | 1.29 | 9.07 | 20.14 | 15.84 | 0.32 |
| $\sigma^2_\varepsilon$ | 15.80 | 10.14 | 10.46 | 0.97 | 7.86 | 20.17 | 3.64 | 0.37 |
| $h^2$ | 0.56 | 0.30 | 0.60 | 0.57 | 0.53 | 0.50 | 0.81 | 0.46 |
| | Well-watered—multi-environment | | | | Water-deficit—multi-environment | | | |
| $\sigma^2_G$ | 94.99 | 36.88 | 4.48 | 7.15 | 3.94 | 3.24 | 5.50 | 0.19 |
| $\sigma^2_{GxA}$ | 55.87 | 18.31 | 0.58 | 4.10 | 8.65 | 8.78 | 9.70 | 0.54 |
| $\sigma^2_A$ | 22.39 | 137.49 | 0.78 | 1.82 | 0.18 | 4.25 | 4.61 | 0.04 |
| $\sigma^2_\varepsilon$ | 33.38 | 46.19 | 3.50 | 2.59 | 13.10 | 15.67 | 7.15 | 0.67 |
| $h^2$ | 0.69 | 0.20 | 0.56 | 0.68 | 0.47 | 0.26 | 0.39 | 0.40 |

$\sigma^2_G$—genotypic variance; $\sigma^2_\varepsilon$—error variance; $h^2$—broad sense heritability; $\sigma^2_{GxA}$—genotype by environment interaction; $\sigma^2_A$—environment variance

## Population structure

There was a low degree of relationship between the 49 cassava genotypes (Fig. 2). However, it is possible to observe two distinct genetic groups based on the genetic similarity analysis. The distant relationships between most genotypes may be advantageous for genetic improvement actions due to the possibility that these genotypes present alternative alleles for drought tolerance. However, we observed certain exceptions, with the highest genetic similarity shown between the landraces BGM-0279 versus Engana Ladrão and BGM-0163 versus BGM-0876, and among the improved genotypes BGM-0360 versus 9624-09, GCP-095 versus GCP-009, and GCP-014 versus Mani Branca, whose kinship ranged from 0.51 to 0.57.

The fastStructure algorithm was used to infer the population structure by estimating the proportion of ancestry between the accessions. The number of groups that best explained the population structure was K = 2 (Fig. 2). This result corroborates the genomic kinship matrix and genetic similarity analysis (dendrogram), which structured the genotypes based on their geographic origin and water deficit tolerance. The blue group consisted of most genotypes, including improved varieties and landraces. Most genotypes in the yellow group are derived from crosses between contrasting parents for water deficit tolerance and were selected from semiarid conditions in Northeast Brazil—except the GCP-014 and GCP-179 genotypes, which were included in the blue group.

## Linkage disequilibrium analysis

A total of 25,597 SNPs (mean of 1422 SNPs per chromosome) were used in the LD analysis. The overall mean $r^2$ was 0.047, which ranged from 0.00 to 1.00. LD decay demonstrated differences at the chromosome level, presenting a chromosome mean of 4.76% of SNP pairs with an $r^2$ of $> 0.20$. Chromosome 1 had the highest percentage of SNPs in LD, with an $r^2$ of $> 0.20$ (8.23%). Moreover, chromosomes 17 and 10 showed the lowest percentages of SNP pairs with an $r^2$ of $> 0.20$ (3.58 and 3.59%, respectively). Mean estimates of $r^2$ at the chromosome level ranged from 0.041 (chromosome 10) to 0.055 (chromosome 18) (Supplementary Material, Table S3).

The LD decay pattern, based on the distance (kb) between SNPs, was investigated considering all 18 cassava chromosomes. The distribution generally showed a rapid LD decay as a function of the increase

**Table 2** Genetic correlation (upper diagonal) and covariance (lower diagonal) between well-watered and water deficit conditions in the growing seasons of 2013–2015 for some agronomic traits in cassava

| Traits | Trial | WW13 | WW14 | WD13 | WD14 |
|---|---|---|---|---|---|
| RoY | WW13 | | 0.64 | 0.57 | 0.36 |
| | WW14 | 83.20 | | 0.41 | 0.37 |
| | WD13 | 26.48 | 13.13 | | 0.24 |
| | WD14 | 13.46 | 9.29 | 2.17 | |
| ShY | WW13 | | 0.56 | 0.58 | 0.38 |
| | WW14 | 24.97 | | 0.56 | 0.49 |
| | WD13 | 4.23 | 5.22 | | 0.22 |
| | WD14 | 8.87 | 14.44 | 1.06 | |
| DMC | WW13 | | 0.74 | 0.39 | 0.51 |
| | WW14 | 2.83 | | 0.43 | 0.56 |
| | WD13 | 3.00 | 2.31 | | 0.29 |
| | WD14 | 4.59 | 3.48 | 3.70 | |
| StY | WW13 | | 0.65 | 0.53 | 0.36 |
| | WW14 | 6.25 | | 0.33 | 0.33 |
| | WD13 | 1.89 | 0.78 | | 0.24 |
| | WD14 | 0.65 | 0.40 | 0.10 | |

WW13 and WW14: well-watered condition in the growing seasons of 2013/2014 and 2014/2015, respectively; WD13 and WD14: water deficit condition in the growing seasons of 2013/2014 and 2014/2015, respectively; RoY—fresh root yield; ShY—shoot yield; DMC—dry matter content; StY—starch yield

in physical distance. We observed a rapid LD decay based on the physical distance between loci, with a range close to 2000 bp (Fig. 3).

## GWAS considering single and multiple environments

In the analyses considering the well-watered and water deficit conditions, 23 SNPs were identified for all traits evaluated in single and multiple environments (Table 3). We identified SNPs that have a significant association with DMC distributed across six different chromosomes (1, 9, 12, 13, 15, and 17) in the well-watered condition (Fig. 4). In the WW13 environment, the most significant association was observed for the SNP S12_25684461 [*p*-value (−log10) = 7.89], with a negative effect −1.64 and low variance (0.288). In the WW14 environment, the most significant association was observed for the SNP S9_21799548 [*p*-value (−log10) = 9.92], with a positive effect of 1.19 and a variance of 0.318 (Table 3).

Besides the markers associated with DMC in the well-watered condition, five other SNPs were associated with this trait under the water deficit condition. Among the five SNPs associated with DMC under water deficit conditions, three are located on chromosomes 4, 7, and 11 and two are located on chromosome 13 (Fig. 4). The SNP S11_4654140 was the most significant [*p*-value (−log10) = 17.57], presenting a high positive effect (20.20) and thus high variance (101.99) (Table 3). GWAS data from multiple trials under water deficit from 2013 to 2020 (i.e., the large panel of cassava clones) indicated that the SNPs S7_20574171 [*p*-value (−log10) = 6.24] and S4_5993266 [*p*-value (−log10) = 5.99] were retained for DMC in relation to the panel of 49 clones and showed a certain level of stability regardless of the genetic background and environmental conditions (Table 4, Fig. 5).

For the water deficit condition in 2013, 8 SNPs associated with RoY and StY were identified. Additionally, two SNPs located on chromosomes 1 and 2 (S1_18594607 and S2_4054519) had a common association. For both traits, the SNP S1_18594607 had a negative effect (−8.49 and −2.70, with a variance of 17.99 and 1.82 for RoY and StY, respectively) and the SNP S2_4054519 had a positive effect (1.27 and 0.52, with a variance of 0.27 and 0.05, for RoY and StY, respectively) (Table 3). The same SNPs were also detected for these two traits based on the analysis of a large panel of cassava clones. In the case of S1_18594607, the effects were −1.71 and −0.47, with a variance of 0.51 and 0.02 for RoY and StY, respectively. The effects of SNP S2_4054519 remained positive (0.27 and 0.22, with a variance of 0.11 and 0.08 for RoY and StY, respectively) (Fig. 5).

Other significant SNPs for RoY were located on chromosomes 3 (S3_5509517), 15 (S15_7659343), and 16 (S16_5925755), while the -log10(*p*-value) of these SNPs ranged from 8.00 to 8.32 (Table 3). However, the two latter SNPs were not identified in the analysis of the second set of field trials; therefore, they possibly suffer a greater influence from the genotype × environment interaction.

For StY, specific SNPs located on chromosomes 3, 4, and 14 were identified, with *p*-values ranging from 6.48 to 13.38 (Table 3 and Fig. 5). Among the most important SNPs for StY, S14_6078279 presented the

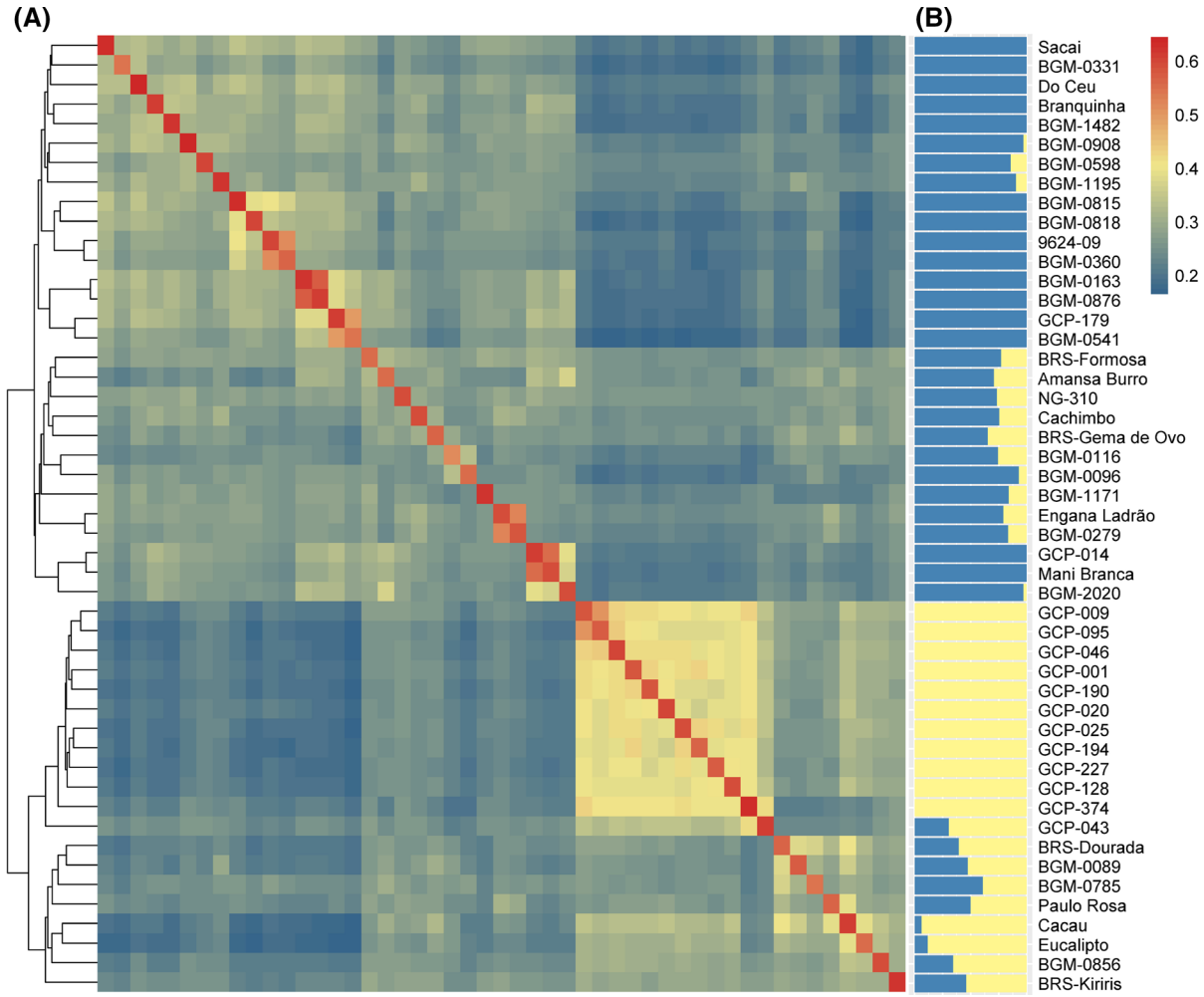



**Fig. 2** **a** Heatmap of the genomic kinship matrix obtained by the VanRaden method (2008) based on 25,597 SNPs; and **b** Population structure estimated by the fastStructure method with K = 2 in the 49 cassava genotypes evaluated for drought tolerance

highest significance [*p*-value (−log10) = 13.38], with a positive effect (1.43) and low variance (0.17). Additionally, an analysis of this large set of cassava clones identified genomic regions associated with water deficit tolerance in four other chromosomes, i.e., 1 (S1_19570959), 2 (S2_5233609), 9 (S9_28881383), and 13 (S13_3375787). The SNP from chromosome 3 (S3_5509478), which was identified in the set subjected to different water stresses, was again associated with the RoY trait in the water deficit assays (Fig. 5). Another significant SNP for RoY was identified on chromosome 1 (S1_18572487).

The SNP S4_21615445, located on chromosome 4, presented a specific and stable association across the different environments for the ShY trait (Fig. 6). In 2014, this SNP was associated with ShY in the water deficit condition and the multi-environment analyses. Additionally, it had a positive effect in both environments [4.44 (2014) and 1.13 (multi-environment)] and greater variance (1.81) in the 2014/2015 growing season (Table 3). Notably, this SNP was also specifically identified under water deficit in the analysis of a larger set of germplasms in the second stage of this study. However, the effect of the reference allele was negative in the multi-environment analysis (−0.73) [*p*-value (−log10) = 6.17] (Table 4). In the same data set, we identified other genomic regions associated with ShY on chromosomes 5 [SNP S5_27432060, with a *p*-value (−log10) of 10.59], 13 [SNPs S13_1372916 and S13_3289247, with *p*-values

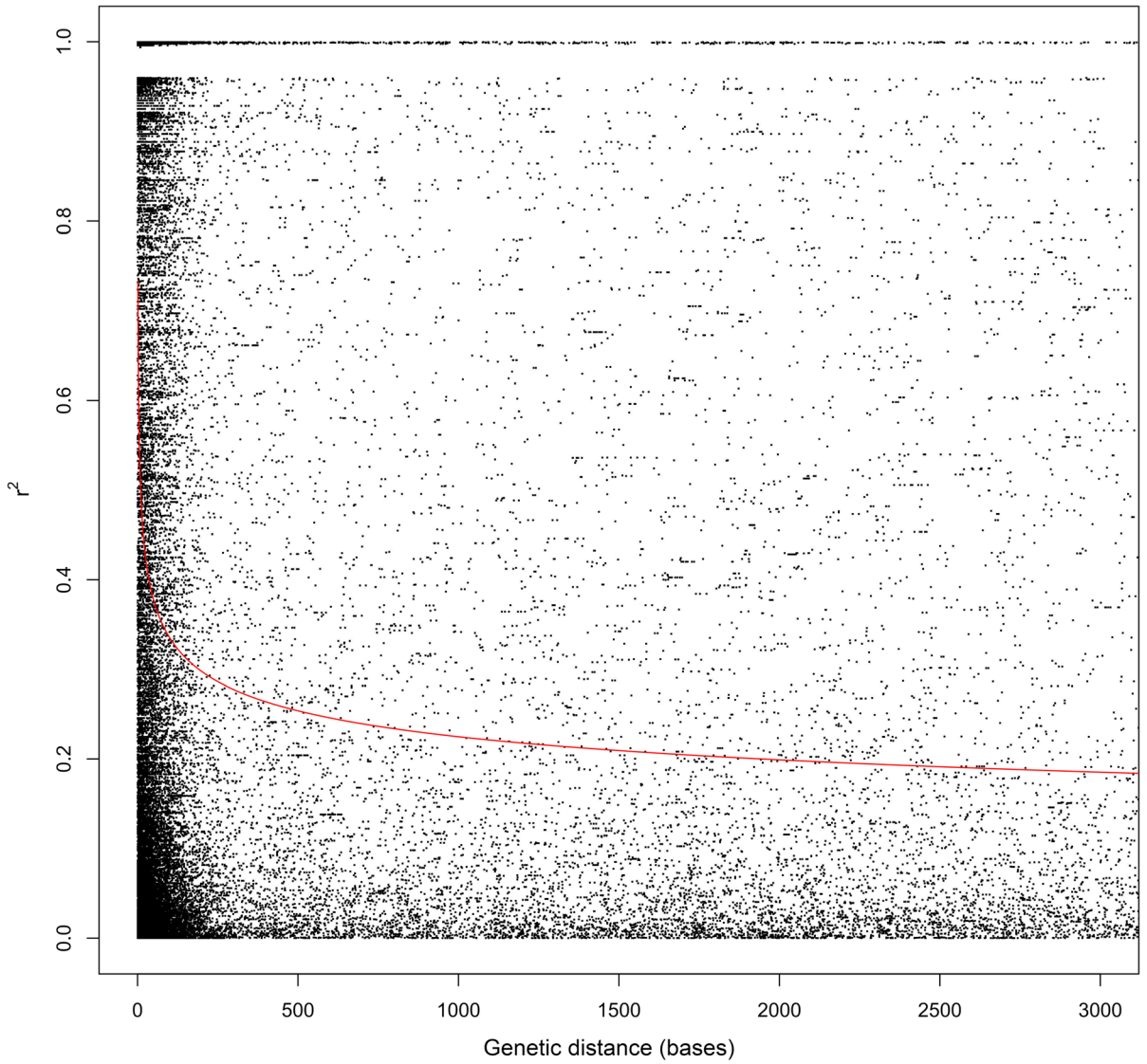



**Fig. 3** Decay pattern of linkage disequilibrium between pairs of single-nucleotide polymorphism (SNP) markers as a function of physical distance in base pairs (bp), based on analysis of all chromosomes

(−log10) of 10, 80 and 6.07, respectively], 15 [SNP S15_16117735, with a *p*-value (−log10) of 6.73], and 16 [SNP S16_23475633, with a *p*-value (−log10) of 6.46] (Fig. 5).

## Genome-wide association studies using selection indices

The selection indices considered the data for the well-watered and water deficit conditions simultaneously (different years). The purpose of using these selection indices related to water stress was to identify genomic regions strictly related to drought tolerance and yield stability.

We identified 14 SNPs that have a significant association with the DTI (Fig. 7 and Table 5). These markers were located on 12 chromosomes, thereby proving the polygenic nature of drought tolerance in cassava. Among these markers, four were associated with StY (with a variance between 0.75 and 1.86 and a positive effect, except for the SNP S10_20405553) and one was associated with ShY (the same marker

**Table 3** Single nucleotide polymorphism (SNP) markers associated with dry matter content, fresh root yield, starch and shoot yield in well-watered and water deficit conditions during the growing seasons of 2013/2014 and 2014/2015, identified through genome-wide association studies (GWAS)

| Environment | SNP | Ch[1] | Position (pb) | MAF[2] | *P*-valor (−log10)[3] | Effect | Var (SNP) |
|---|---|---|---|---|---|---|---|
| *Dry matter content* | | | | | | | |
| Well-watered (2013) | S1_28193620 | 1 | 28,193,620 | 0.47 | 6.08* | 1.18 | 0.344 |
| | S12_25684461 | 12 | 25,684,461 | 0.12 | 7.89** | −1.64 | 0.288 |
| | S17_3366403 | 17 | 3,366,403 | 0.34 | 6.08* | −1.14 | 0.290 |
| Well-watered (2014) | S9_21799548 | 9 | 21,799,548 | 0.34 | 9.92** | 1.19 | 0.318 |
| | S13_26038260 | 13 | 26,038,260 | 0.42 | 6.46** | −0.62 | 0.092 |
| | S15_9075007 | 15 | 9,075,007 | 0.47 | 6.00* | −0.71 | 0.125 |
| Water-deficit (2014) | S4_5993266 | 4 | 5,993,266 | 0.41 | 7.82** | −1.03 | 0.254 |
| | S7_20574171 | 7 | 20,574,171 | 0.47 | 7.56** | 2.70 | 1.819 |
| | S11_4654140 | 11 | 4,654,140 | 0.49 | 17.57** | 20.20 | 101.992 |
| | S13_22348129 | 13 | 22,348,129 | 0.06 | 6.28** | −1.48 | 0.126 |
| | S13_23968102 | 13 | 23,968,102 | 0.16 | 6.72** | 2.11 | 0.609 |
| *Fresh root yield* | | | | | | | |
| Water-deficit (2013) | S1_18594607 | 1 | 18,594,607 | 0.49 | 8.57** | −8.49 | 17.993 |
| | S2_4054519 | 2 | 4,054,519 | 0.21 | 7.26** | 1.27 | 0.274 |
| | S3_5509517 | 3 | 5,509,517 | 0.11 | 8.01** | 2.61 | 0.679 |
| | S15_7659343 | 15 | 7,659,343 | 0.19 | 8.00** | 2.29 | 0.818 |
| | S16_5925755 | 16 | 5,925,755 | 0.14 | 8.32** | −1.88 | 0.431 |
| *Starch yield* | | | | | | | |
| Water-deficit (2013) | S1_18594607 | 1 | 18,594,607 | 0.49 | 9.50** | −2.70 | 1.817 |
| | S2_4054519 | 2 | 4,054,519 | 0.21 | 10.46** | 0.52 | 0.046 |
| | S3_3056452 | 3 | 3,056,452 | 0.36 | 6.48** | −0.42 | 0.040 |
| | S4_3558714 | 4 | 3,558,714 | 0.49 | 7.76** | 0.38 | 0.036 |
| | S14_6078279 | 14 | 6,078,279 | 0.09 | 13.38** | 1.43 | 0.170 |
| *Shoot yield* | | | | | | | |
| Water-deficit (2014) | S4_21615445 | 4 | 21,615,445 | 0.10 | 8.16** | 4.44 | 1.809 |
| Water-deficit (Multi) | S4_21615445 | 4 | 21,615,445 | 0.10 | 7.49** | 1.13 | 0.117 |

[1]Chromosome; [2]minor allele frequency; [3]SNPs with significant association at *5% and **1% by Bonferroni test

identified in single and multi-environments, with a variance of 0.333 and a positive effect of 1.99). Nine SNPs indicated a significant association between DTI and DM. In single environments, all five SNPs had positive effects, while the other four SNPs had negative effects and low variance (0.001–0.003) in the multi-environment analysis.

For the DTSI, 17 SNPs with a significant association were identified for all traits. Two SNPs were identified for RoY, which were located on chromosomes 10 and 15; four were identified for ShY, of which two were located on chromosome 5 and the other two on chromosomes 6 and 12; eight were identified for DMC, located on chromosomes 3, 6, 7, 11, 12, and 17; and three SNPs associated with StY were located on chromosomes 6, 10, and 16 (Fig. 8 and Table 6). Despite the high *p*-values (−log10 = 22.50) observed, the variance of these SNPs was null or low (varied between 0.00 and 0.09); consequently, the SNPs effects were also low, ranging from −0.13 (S12_552437) to 0.60 (S11_4654140) (Table 6).

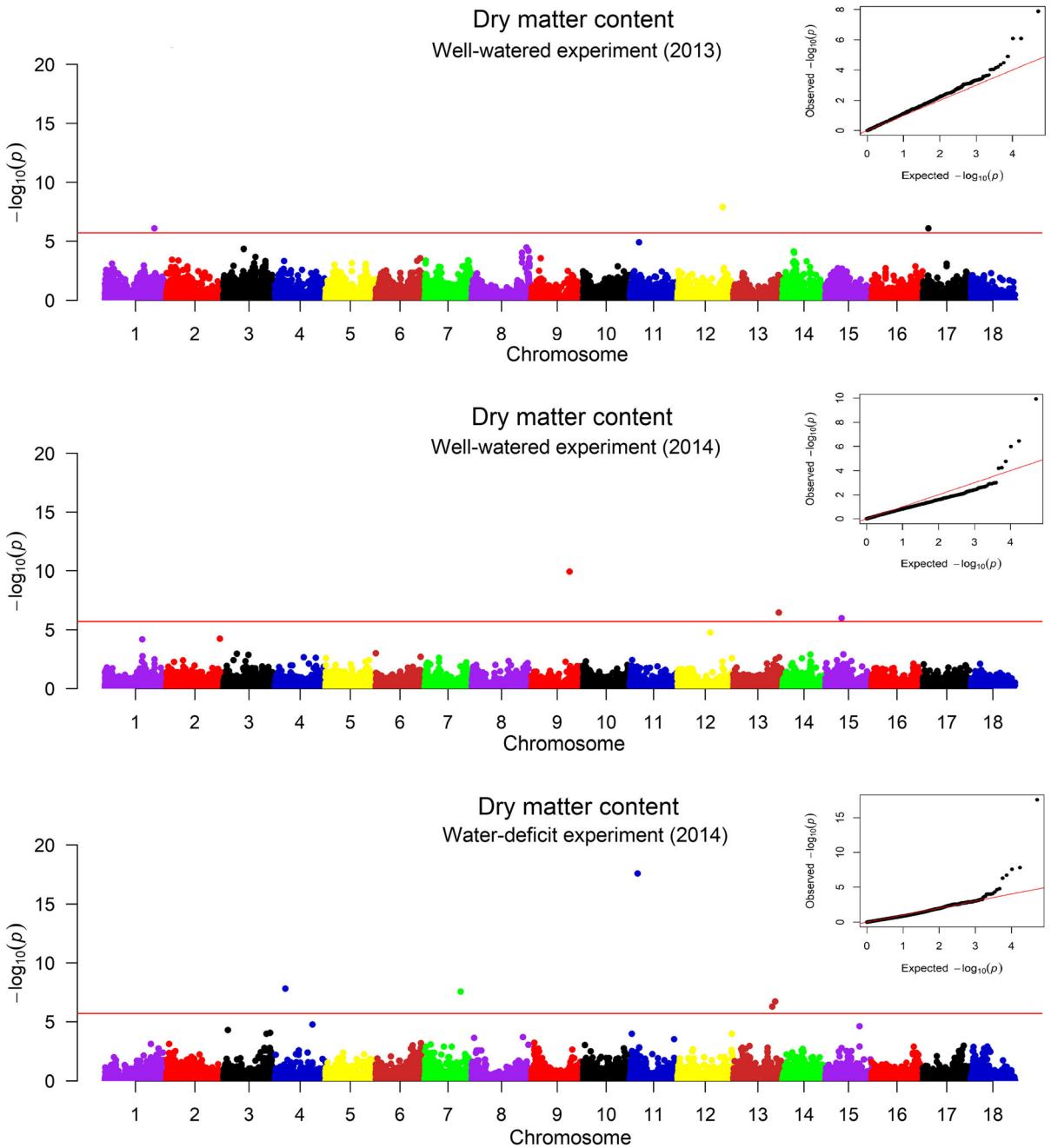



**Fig. 4** Manhattan plot indicating the single nucleotide polymorphisms (SNPs) associated with the dry matter content in 49 cassava genotypes evaluated under well-watered and water deficit conditions in single environments (growing seasons of 2013/2014 and 2014/2015). The SNP locations on each chromosome and the association test ($-\log_{10}$(p)) are represented on the x and y axis, respectively. The red line indicates the Bonferroni correction ($p < 0.05$). The plot above refers to the quantile–quantile (QQ) of the *p*-values observed and expected from the genome-wide association study

**Table 4** Single nucleotide polymorphism (SNP) markers associated with dry matter content, fresh root yield, starch and shoot yield in water deficit conditions in the growing seasons of 2013, 2014, 2016, 2017, 2018, 2019 and 2020 in Petrolina (PE, Brazil), harvested at six months after planting, identified through genome-wide association studies (GWAS)

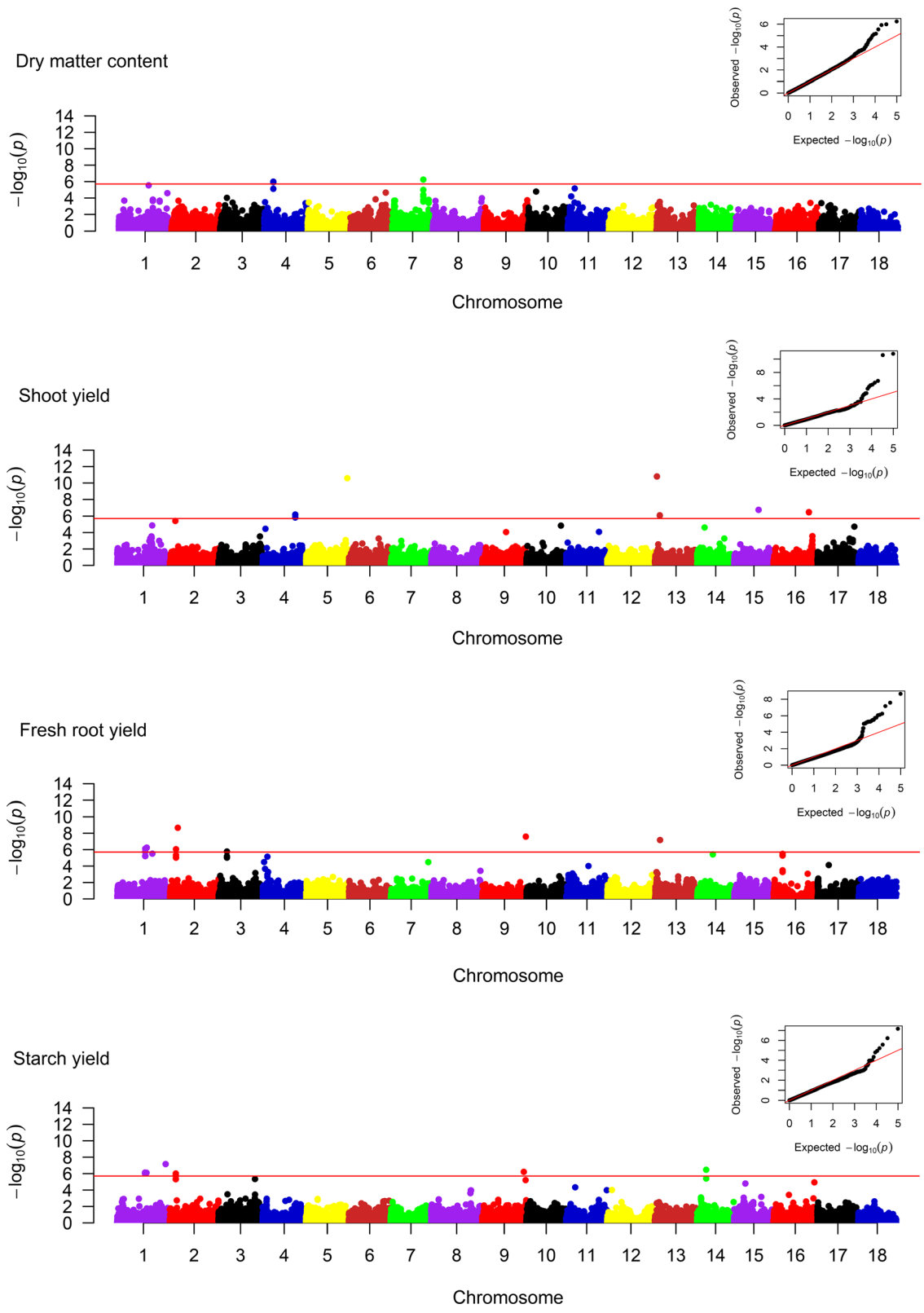
| Trait | SNP | Ch[1] | Position (pb) | MAF[2] | *P*-valor (−log10)[3] | Effect | Var (SNP) |
|---|---|---|---|---|---|---|---|
| Dry matter content | S7_20574171 | 7 | 20,574,171 | 0.55 | 6.24** | 5.73 | 0.82 |
| | S4_5993266 | 4 | 5,993,266 | 0.54 | 5.99* | 2.89 | 0.30 |
| | S4_5993255 | 4 | 5,993,255 | 0.43 | 5.92* | 3.62 | 0.44 |
| Fresh root yield | S2_5233609 | 2 | 5,233,609 | 0.57 | 8.66** | −1.99 | 0.58 |
| | S9_28881383 | 9 | 28,881,383 | 0.50 | 7.59** | −1.38 | 0.30 |
| | S13_3375787 | 13 | 3,375,787 | 0.58 | 7.17** | −1.00 | 0.36 |
| | S1_19570959 | 1 | 19,570,959 | 0.26 | 6.25** | −0.95 | 0.32 |
| | S1_18594607 | 1 | 18,594,608 | 0.25 | 6.12** | −1.71 | 0.51 |
| | S2_4054519 | 2 | 4,054,519 | 0.36 | 6.06** | 0.27 | 0.11 |
| | S3_5509478 | 3 | 5,509,478 | 0.28 | 5.77* | 1.28 | 0.36 |
| | S1_18572487 | 1 | 18,572,487 | 0.51 | 5.73* | −1.16 | 0.29 |
| Shoot yield | S13_1372916 | 13 | 1,372,916 | 0.28 | 10.80** | −2.35 | 0.59 |
| | S5_27432060 | 5 | 27,432,060 | 0.29 | 10.59** | 4.36 | 1.02 |
| | S15_16117735 | 15 | 16,117,735 | 0.42 | 6.73** | 2.05 | 0.89 |
| | S16_23475633 | 16 | 23,475,633 | 0.31 | 6.46** | 3.88 | 1.20 |
| | S4_21615445 | 4 | 21,615,445 | 0.44 | 6.17** | −0.73 | 0.26 |
| | S13_3289247 | 13 | 3,289,247 | 0.36 | 6.07** | −3.22 | 1.02 |
| | S4_21615485 | 4 | 21,615,485 | 0.44 | 5.83* | −1.79 | 0.59 |
| Starch yield | S1_32277166 | 1 | 32,277,166 | 0.56 | 7.16** | −0.75 | 0.21 |
| | S14_6078279 | 14 | 6,078,279 | 0.49 | 6.47** | −0.42 | 0.12 |
| | S9_27685650 | 9 | 27,685,650 | 0.44 | 6.22** | 0.54 | 0.23 |
| | S1_18601155 | 1 | 18,601,155 | 0.46 | 6.13** | −0.79 | 0.28 |
| | S1_19570959 | 1 | 19,570,959 | 0.26 | 6.10** | 0.10 | 0.05 |
| | S1_18572487 | 1 | 18,572,487 | 0.51 | 6.09** | −0.62 | 0.21 |
| | S1_18594607 | 1 | 18,594,608 | 0.25 | 6.09** | −0.47 | 0.02 |
| | S1_18572475 | 1 | 18,572,475 | 0.42 | 6.03** | 0.23 | 0.06 |
| | S2_4054519 | 2 | 4,054,519 | 0.36 | 6.02** | 0.22 | 0.08 |
| | S2_4054559 | 2 | 4,054,559 | 0.57 | 5.77* | −0.62 | 0.15 |

[1]Chromosome; [2]minor allele frequency; [3]SNPs with significant association at *5% and **1% by Bonferroni test

## In silico annotation of SNPs

The 62 SNPs significantly associated with environments with and without water stress and with the tolerance and stability DTIs comprise close to 160 transcripts that were previously identified and available in the Phytozome database (http://www.phytozome.net) (Supplementary Material, Tables S4, S5, S6, S7, and S8). Among these transcripts, four are inserted within the gene regions of the Manes.01G182700.1, Manes.17G012100.1,

**Fig. 5** Manhattan plot indicating the single nucleotide polymorphisms (SNPs) associated with fresh root yield, shoot yield, starch yield and dry matter content in 252 cassava genotypes evaluated under water deficit conditions in multi-trial (growing seasons of 2013, 2014, 2016, 2017, 2018, 2019 and 2020). The SNP locations on each chromosome and the association test ($-\log_{10}$(p)) are represented on the x and y axis, respectively. The red line indicates the Bonferroni correction ($p < 0.05$). The plot above refers to the quantile–quantile (QQ) of the *p*-values observed and expected from the genome-wide association study ►


Dry matter content
Shoot yield
Fresh root yield
Starch yield
−log10(p)
Chromosome
Observed −log10(p)
Expected −log10(p)

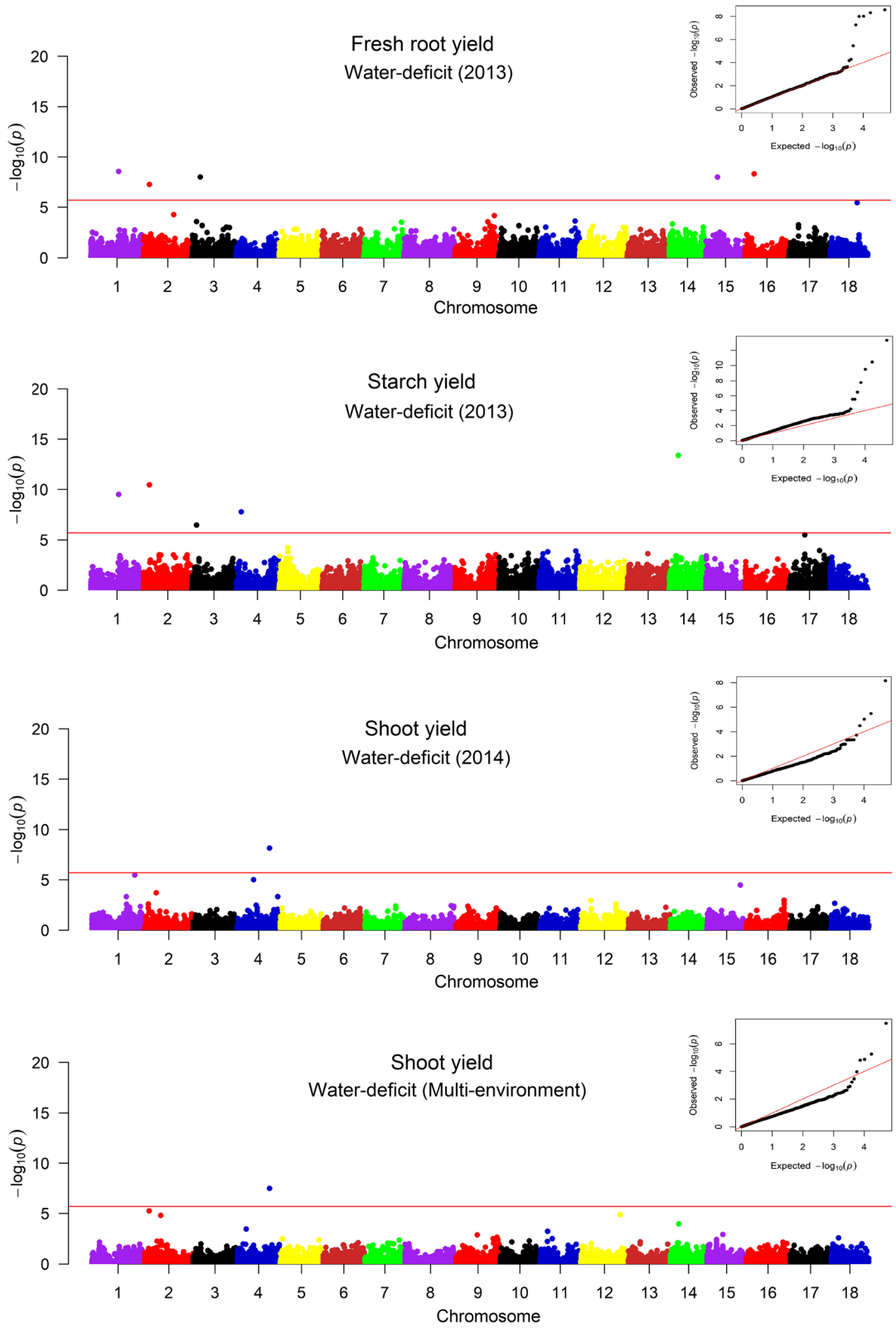
Manes.13G132400.1 and Manes.15G120300.1 transcripts. The transcripts with a known functional annotation are related to the protein kinases associated with leucine-rich repeats, Scarecrow-like 32, oxygen-enhancing proteins of photosystem II, and proteins with functions in the transcriptional regulation of RNA polymerase II, besides the BTB/POZ domain associated with GTPases Rop (Supplementary Material S4).

In the small panel of cassava clones evaluated under a water deficit condition, 31 transcripts related to 14 SNPs were recorded. Of these SNPs, seven were inserted into the coding regions of the Manes.02G053200.1, Manes.03G058400.1, Manes.15G102800.1, Manes.14G074700.1, Manes.07G087500.1, Manes.11G048600.1 and Manes.04G077900.1 transcript (Supplementary Material S5). Among the transcripts, nine have known functional annotations, such as defense against pathogens, proteins involved in the degradation and transport of certain cell elements, proteins capable of making multiple contacts along DNA, ribosomal and histone proteins, transcription factors and regulators, proteins that play a key role in multiple cellular processes, and nuclear pore complex proteins (Supplementary Material, Table S5). An analysis of the second panel of cassava clones in water deficit conditions revealed a greater number of transcripts (68) present in 24 SNPs. Of the transcripts associated with different agronomic traits, 25 were common to the first panel of evaluated germplasm, while 13 transcripts have an unknown function. Seventeen of these SNPs are inserted within genetic regions of the transcripts Manes.01G063300.1, Manes.01G063600.1, Manes.01G244500.1, Manes.02G053200.1, Manes.02G070200.1, Manes.03G058400.1, Manes.04G077900.1, Manes.05G200900.1, Manes.07G087500.1, Manes.09G163200.1, Manes.09G180700.1, Manes.13G014000.1, Manes.13G035800.1, Manes.13G036600.1, Manes.14G074700.1, Manes.15G176100.1, and Manes.16G078600.1 (Supplementary Material, Table S8). The main functional annotations of these transcripts refer to: (i) the glycosaminoglycan component, which is important to connective tissues including chondroitin sulfate, dermatan sulfate, heparan sulfate, heparin, and hyaluronan; (ii) the rabconnectin-3 complex, which is involved in regulating notch signaling as a single-pass transmembrane receptor protein; (iii) proteins modulating plant responses to drought stress; (iv) response to salt stress and water deprivation; (v) plant-specific proteins required for growth and development; (vi) proteins involved in the vindoline biosynthesis pathway; (vii) proteins involved in DNA damage, repair, neurogenesis, transcription, and transcription regulation; (viii) proteins involved in the synthesis of phytochelatins and homophytochelatins, which are the heavy-metal-binding peptides of plants; (ix) proteins involved in the bidirectional long-distance transport of aliphatic compounds and the removal of glucosinolates from the xylem in roots; (x) the transcription factor far upstream element-binding protein (Supplementary Material, Table S8).

**Fig. 6** Manhattan plot indicating the single nucleotide polymorphisms (SNPs) associated with fresh root yield, starch yield and shoot yield in 49 cassava genotypes evaluated under well-watered and water deficit conditions in single environments (growing seasons of 2013/2014 and 2014/2015). The SNP locations on each chromosome and the association test ($-\log_{10}(p)$) are represented on the x and y axis, respectively. The red line indicates the Bonferroni correction ($p < 0.05$). The plot above refers to the quantile–quantile (QQ) of the *p*-values observed and expected from the genome-wide association study

For the DTI, nine significant SNPs were identified within transcripts with known functions, while another 22 SNPs were identified close to other transcripts (distances ranging from 103 to 12,762 bp). Three of these transcripts have known functional annotations related to the formation of protein–protein interactions (leucine-rich repeats), switching many cellular functions, exchanging components between the nucleus and cytoplasm, serving as transcriptional regulators, dehydration-responsive protein, and proteins with a RING finger domain, which serves a key role in the ubiquitination pathway and the domains of several distinct nucleotide-binding protein folds (Supplementary Material S6).

Of the 15 SNPs associated with the DTSI, 12 are within coding regions. Additionally, 26 significant SNPs were identified close to other transcripts (distances ranging from 60 to 13,628 bp). These transcripts generally have functional annotations related to leucine-rich repeats, proteins from the transcriptional apparatus, proteins involved in intracellular transport and motility, plant stress resistance (cadmium, salt,


Fresh root yield
Water-deficit (2013)
Starch yield
Water-deficit (2013)
Shoot yield
Water-deficit (2014)
Shoot yield
Water-deficit (Multi-environment)
−log10(p)
Chromosome
Observed −log10(p)
Expected −log10(p)

and antifungal response), and DNA repair protein (Supplementary Material S7).

## Discussion

### Importance of phenotypic traits for drought tolerance analysis in cassava

Drought stress induced in the 2013 and 2014 trials was severe enough to cause significant changes in the components of variance and $h^2$ in the single and multi-environment analyses (Table 1). Notably, $h^2$ values in the well-watered condition were of medium to high magnitude; however, in the water deficit condition, these values were reduced for most agronomic traits in both the single-environment and the multi-environment analyses. Decreases in $h^2$ estimates were related to the water deficit environment and to the genetic material used since some varieties are susceptible to this abiotic stress (Farfan et al. 2015; Oliveira et al. 2017). Similar results were observed in experiments with different water conditions, in which $h^2$ estimates were lower for traits evaluated under water deficit conditions in crops such as common beans (Hinkossa et al. 2013), potatoes (Cabello et al. 2014), and maize (Beyene et al. 2015).

The genetic correlation was higher for the well-watered treatment groups than for the water deficit groups. Higher genetic correlations are expected for groups evaluated in similar environmental conditions since there is a tendency toward the induction of related responses in the genotypes, resulting in strong genetic correlations (Malosetti et al. 2013). The low genetic correlation observed for drought stress environments may be related to differences in climatic conditions between the two growing seasons due to the higher average annual rainfall in 2013 (347.8 mm) than in 2014 (216.3 mm) as well as interannual differences in mean temperature and relative humidity.

### Genomic association between SNPs and responses to water deficit in cassava

With their reduced sequencing costs, GWASs have been routinely conducted to explore the allelic variation associated with agronomic traits of interest.

**Fig. 7** Manhattan plot indicating the single nucleotide polymorphisms (SNPs) associated with the drought tolerance index for starch yield, shoot yield, and dry matter content in 49 cassava genotypes evaluated under well-watered and water deficit conditions in single environments (growing seasons of 2013/2014 and 2014/2015) and multi-environments. The SNP locations on each chromosome and the association test ($-\log_{10}(p)$) are represented on the x and y axis, respectively. The red line indicates the Bonferroni correction ($p < 0.05$). The plot above refers to the quantile–quantile (QQ) of the *p*-values observed and expected from the genome-wide association study

Applying this method allows the identification of regions in haplotype blocks that indicate precise associations between molecular markers and important phenotypic traits (Han and Huang 2013). In rice, 13 SNPs were associated with yield under drought conditions. The genomic regions of these SNPs include close to 30 genes with functional annotations; among these, 10 have functional annotations related to drought and/or tolerance to abiotic stress. Among the identified SNPs, two presented the potential for TaqMan assays to be used in routine PCR (Pantalião et al. 2016). In maize, 42 SNPs were found to be associated with 33 genes, of which three were co-located in drought-related QTL regions (Xue et al. 2013).

Using the GWAS method, these associations can be reliably identified since possible spurious associations due to the kinship between genotypes are reduced by including the genomic kinship matrix and population structure (Flint-Garcia et al. 2005). As indicated by fastStructure and the kinship matrix, there was concordance between the population structures of the two groups in the present work. Therefore, it was possible to structure the genotypes based on their geographic origin, water deficit tolerance, and specific classification in terms of breeding since most genotypes derived from crosses between contrasting parental populations for drought tolerance remained in the same group. Furthermore, in a GWAS of resistance to cassava root rot, 263 accessions were grouped into four groups with no relation to their agronomic traits or geographic origin due to historical and recent hybridizations related to the genotypes included in the different groups (Brito et al. 2017).

The extent of LD is another determinant of GWAS efficiency. LD is evaluated based on the non-random associations between pairs of SNPs with significant

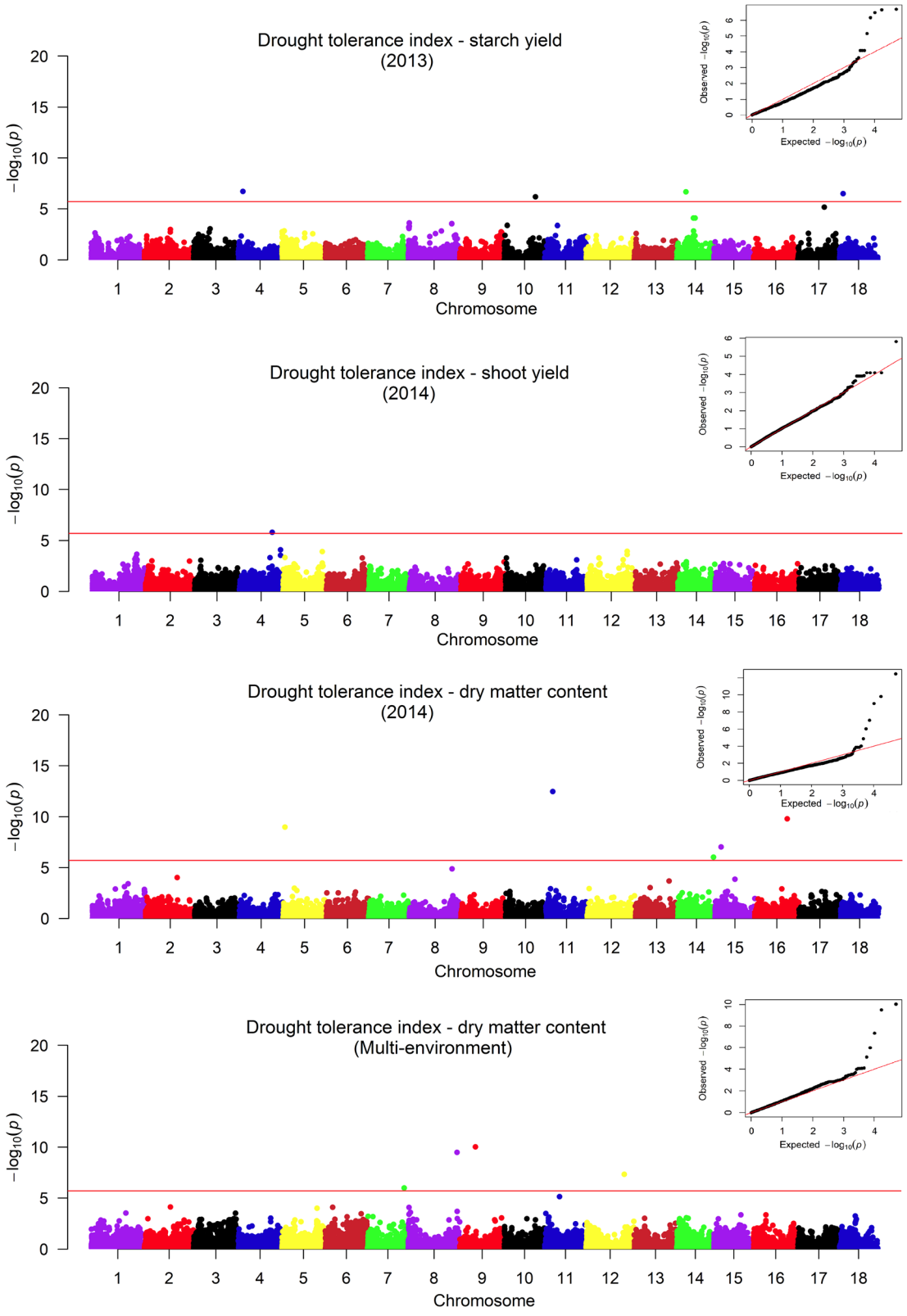

Drought tolerance index - starch yield
(2013)
Drought tolerance index - shoot yield
(2014)
Drought tolerance index - dry matter content
(2014)
Drought tolerance index - dry matter content
(Multi-environment)
−log10(p)
Chromosome
Observed −log10(p)
Expected −log10(p)

**Table 5** Single nucleotide polymorphism (SNP) markers associated with the drought tolerance index for starch yield, shoot yield, and dry matter content during the growing seasons of 2013/2014 and 2014/2015, identified through genome-wide association studies

| Environment | SNP | Ch[1] | Position (pb) | MAF[2] | *P*-valor (−log10)[3] | Effect | Var (SNP) |
|---|---|---|---|---|---|---|---|
| 2013 | *Starch yield* | | | | | | |
| | S4_3015131 | 4 | 3,015,131 | 0.05 | 6.71** | 6.19 | 1.858 |
| | S10_20405553 | 10 | 20,405,553 | 0.11 | 6.18* | −2.74 | 0.749 |
| | S14_6039650 | 14 | 6,039,650 | 0.11 | 6.67** | 4.00 | 1.592 |
| | S18_2494095 | 18 | 2,494,095 | 0.07 | 6.49** | 4.01 | 1.065 |
| 2014 | *Shoot yield* | | | | | | |
| | S4_21615445 | 4 | 21,615,445 | 0.1 | 5.81* | 1.99 | 0.363 |
| | *Dry matter content* | | | | | | |
| | S5_1282566 | 5 | 1,282,566 | 0.05 | 8.99** | 0.26 | 0.003 |
| | S11_4654140 | 11 | 4,654,140 | 0.49 | 12.46** | 1.05 | 0.274 |
| | S14_23407800 | 14 | 23,407,800 | 0.31 | 6.03* | 0.10 | 0.002 |
| | S15_3918053 | 15 | 3,918,053 | 0.21 | 7.03** | 0.10 | 0.002 |
| | S16_21272865 | 16 | 21,272,865 | 0.44 | 9.80** | 0.11 | 0.003 |
| Multi-environment | *Dry matter content* | | | | | | |
| | S7_24121759 | 7 | 24,121,759 | 0.36 | 5.99* | −0.05 | 0.001 |
| | S8_32094464 | 8 | 32,094,464 | 0.42 | 9.49* | −0.08 | 0.002 |
| | S9_10268397 | 9 | 10,268,397 | 0.42 | 10.03* | −0.09 | 0.002 |
| | S12_25167807 | 12 | 25,167,807 | 0.16 | 7.34* | −0.08 | 0.001 |

[1]Chromosome; [2]minor allele frequency; [3]SNPs with significant association at *5% and **1% by Bonferroni test

LD. The $r^2$ parameter has been one of the most common metrics used to evaluate LD (Mangin et al. 2011; Flint-Garcia et al. 2003; Lipka et al. 2015). LD decay in the 49 cassava genotypes occurred rapidly with an increasing physical distance between the loci in the genome (LD with $r^2 = 0.2$, close to 2000 bp). As observed in previous studies of cassava and potato and another species of vegetative propagation, there was significant LD decay in the present study (Esuma et al. 2016; Brito et al. 2017; Stich et al. 2013). To identify significant associations between markers and phenotypes of interest, it is necessary to use a high marker density. However, although a low LD ($r^2 = 0.20$, close to 1320 bp) was identified, significant associations for carotenoid content have been reported in cassava (Esuma et al. 2016).

Determining the magnitude of LD is important since this information establishes the number of SNPs necessary for several studies, including association studies and molecular marker-assisted selection (Grenier et al. 2015; Esuma et al. 2016). Additionally, LD may differ between populations of the same species and is influenced by genome size and complexity, genome recombination patterns, population structure, reproductive system, domestication, and breeding processes (Ching et al. 2002; Mather et al. 2007; Würschum et al. 2013). For example, breeding populations present LD values with a range of 100–500 kb in endogenous maize lines (Ching et al. 2002). In cultivated rice species (*O. japonica*), the LD is higher (> 500 kb) in temperate regions when compared to varieties from tropical regions (150 kb)

**Fig. 8** Manhattan plot indicating the single nucleotide polymorphisms (SNPs) associated with the drought tolerance stability index for fresh root yield, starch yield, shoot yield, and dry matter content in 49 cassava genotypes evaluated in well-watered and water deficit conditions in single environments (growing seasons of 2013/2014 and 2014/2015), and multi-environments. The SNP locations on each chromosome and the association test ($-\log_{10}(p)$) are represented on the x and y axis, respectively. The red line indicates the Bonferroni correction ($p < 0.05$). The plot above refers to the quantile–quantile (QQ) of the *p*-values observed and expected from the genome-wide association study

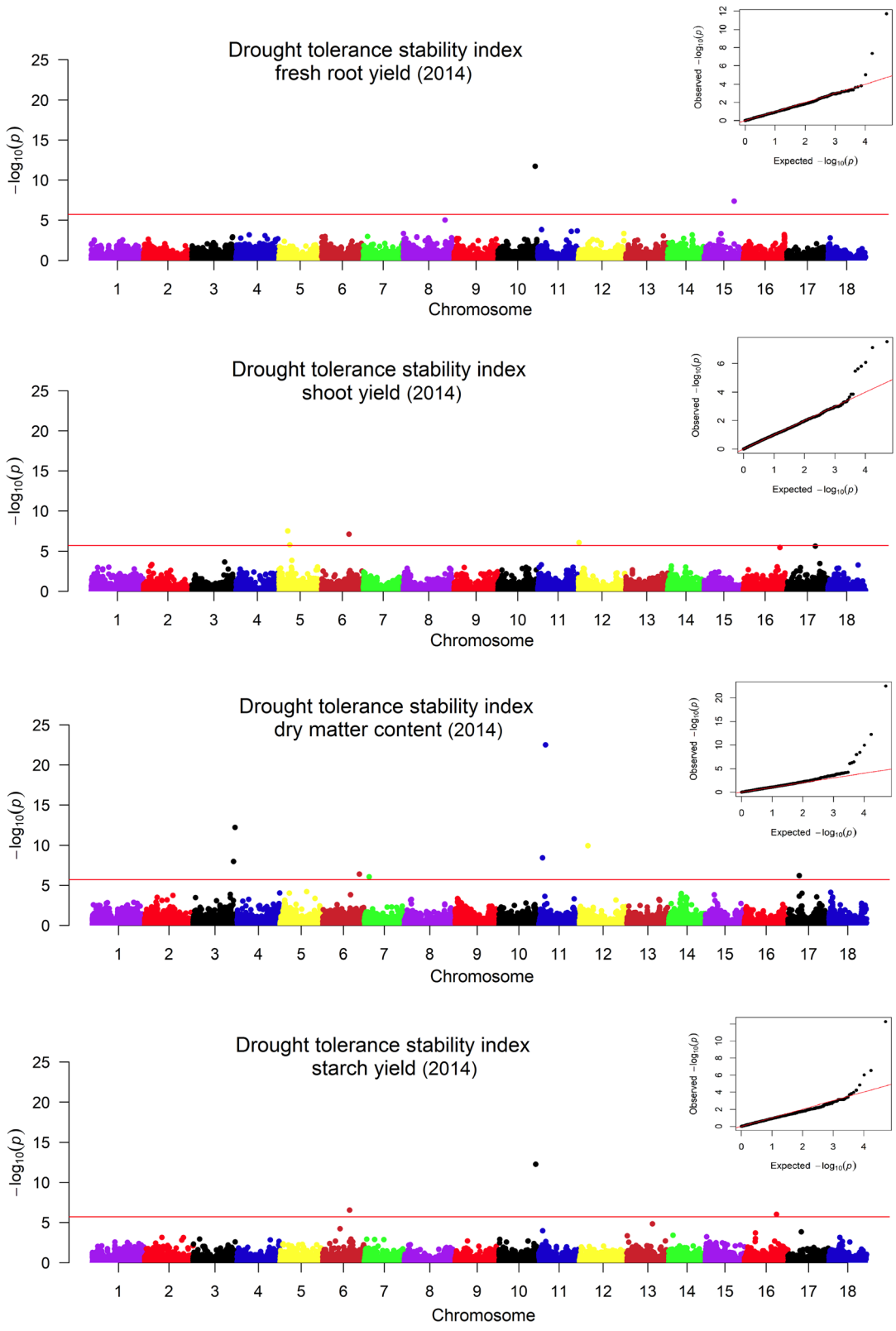

Drought tolerance stability index
fresh root yield (2014)
Drought tolerance stability index
shoot yield (2014)
Drought tolerance stability index
dry matter content (2014)
Drought tolerance stability index
starch yield (2014)
−log10(p)
Chromosome
Observed −log10(p)
Expected −log10(p)

**Table 6** Single nucleotide polymorphism (SNP) markers associated with drought tolerance stability index for fresh root yield, shoot yield, starch yield, and dry matter content in the growing seasons of 2013/2014 and 2014/2015, identified through genome-wide association studies

| SNP | Ch[1] | Position (pb) | MAF[2] | *P*-valor (−log10)[3] | Effect | Var (SNP) |
|---|---|---|---|---|---|---|
| *Total root yield* | | | | | | |
| S10_24780638 | 10 | 24,780,638 | 0.11 | 11.72** | 0.49 | 0.024 |
| S15_19883928 | 15 | 19,883,928 | 0.05 | 7.36** | 0.53 | 0.013 |
| *Shoot yield* | | | | | | |
| S5_5857490 | 5 | 5,857,490 | 0.23 | 7.55** | −0.09 | 0.001 |
| S5_7261808 | 5 | 7,261,808 | 0.37 | 5.81* | −0.07 | 0.001 |
| S6_18511652 | 6 | 18,511,652 | 0.05 | 7.12** | 0.16 | 0.001 |
| S12_552437 | 12 | 552,437 | 0.06 | 6.08* | −0.13 | 0.001 |
| *Dry matter content* | | | | | | |
| S3_27980858 | 3 | 27,980,858 | 0.09 | 12.24** | 0.06 | 0.000 |
| S3_26997330 | 3 | 26,997,330 | 0.05 | 7.97** | −0.06 | 0.000 |
| S6_24608698 | 6 | 24,608,698 | 0.31 | 6.41** | 0.02 | 0.000 |
| S7_3279854 | 7 | 3,279,854 | 0.07 | 6.07* | 0.03 | 0.000 |
| S11_4654140 | 11 | 4,654,140 | 0.49 | 22.50** | 0.60 | 0.089 |
| S11_2699771 | 11 | 2,699,771 | 0.49 | 8.43** | −0.03 | 0.000 |
| S12_5681456 | 12 | 5,681,456 | 0.44 | 9.95** | −0.05 | 0.001 |
| S17_7612220 | 17 | 7,612,220 | 0.30 | 6.21* | 0.03 | 0.000 |
| *Starch yield* | | | | | | |
| S6_18122549 | 6 | 18,122,549 | 0.18 | 6.54** | −0.11 | 0.002 |
| S10_24426293 | 10 | 24,426,293 | 0.13 | 12.28** | 0.29 | 0.010 |
| S16_21413646 | 16 | 21,413,646 | 0.42 | 6.01* | −0.08 | 0.002 |

[1]Chromosome; [2]minor allele frequency; [3]SNPs with significant association at *5% and **1% by Bonferroni test

and *O. indica* (75 kb) (Mather et al. 2007). These results indicate that breeding and domestication processes can strongly interfere with LD behavior.

In allogamous plants such as maize, LD largely depends on the type of population studied. However, LD generally tends to be low due to the larger size, high recombination rates, and higher mobility of the genome of this species (due to transposons and retrotransposons) (Gupta et al. 2005). Thus, the extent of LD in maize is considered low, with rapid decay ranging between 1.0 and 10.0 kb in most populations (Lu et al. 2012; Truntzler et al. 2012). Additionally, the LD decay is faster in hotspot regions prone to recombination, as observed by Würschum et al. (2013) in wheat, where differences in LD measurements were verified throughout the genome.

Marker density is another crucial factor that affects the power of GWASs (Lipka et al. 2015). The extent of LD determines the marker density required for an efficient mapping resolution. If LD decays at a short distance, the mapping resolution tends to be high; however, a high density of widely distributed molecular markers is required to improve the mapping resolution. If the LD extends over a long distance, the mapping resolution tends to be low, but a relatively small number of markers will be required (Grady et al. 2011; Zhu et al. 2008). In the present study, the mean marker density (25,597 SNPs) was used and included a broad representation of the cassava genome, with a distribution of 1 SNP per 22.1 Mb.

The captured genetic effects associated with agronomic traits also depend on the effective size of the mapping population. Small population size may lead to false positives (Brachi et al. 2011), especially for a complex trait such as drought tolerance. Although the cassava evaluated populations were small, the studied genotypes represent a broad panel of genetic diversity for water deficit tolerance, comprising varieties with different levels of drought tolerance and different origins. In other crops, such as maize and *Arabidopsis*, wide genetic diversity in the evaluated panel was efficient for the detection of genomic regions associated with nitrogen use efficiency and flowering, respectively (Atwell et al. 2010; Morosini et al. 2017). Some of these regions have been mentioned in the literature, evidencing the effectiveness of this analysis and the importance of these regions for the genetic control of these traits (Atwell et al. 2010).

In the first panel of 49 cassava clones, it was possible to identify 54 marker-phenotype associations, with 48 SNPs distributed across all cassava chromosomes. Fourteen SNPs are associated with phenotypes in water deficit conditions, with most having a positive effect on the phenotype. Furthermore, 14 are associated with drought tolerance and 17 are associated with drought tolerance stability.

Evaluations of the second panel of 252 cassava clones in multi-trials under water deficit conditions facilitated the identification of two SNPs associated with DMC (S7_20574171 and S4_5993266), three SNPs associated with FRY (S1_18594607, S2_4054519, and S3_5509478), one SNP associated with ShY (S4_21615445), and three SNPs associated with StY (S1_18594607, S2_4054519, and S14_6078279) that were also associated with these same traits in the analysis of the two water regimes (for the first set of cassava clones). These results indicate that the aforementioned genomic regions are conserved under water deficit conditions. Regarding the two sets of germplasms, 62 SNPs were identified that were significantly associated with one of the four traits analyzed and distributed among all cassava chromosomes.

In the well-watered condition, significant SNPs were only detected for DMC; however, there was consistency in the results since common SNPs were obtained in the two evaluated growing seasons. For the well-watered condition groups, there was a high genetic correlation (0.74) between the years of agronomic evaluation as well as a greater similarity in genotype ranking. However, the water deficit groups presented a low genetic correlation between the years of evaluation, ranging from 0.22 (ShY) to 0.29 (DMC), and less similarity was observed in genotype ranking. These facts, which are associated with a strong environmental influence, can help to explain the occurrence of markers specific to each environment in the water deficit condition. However, several SNPs presented a specific and stable effect among the water deficit environments. ShY exhibited a low variance in genotype × environment interaction and a weak influence of the year of evaluation in water deficit trials. As a result, GWASs considering contrasting water stress conditions have identified allelic effects in favorable and stressed environments (Millet et al. 2016). In maize, a low consistency of QTLs among different environments was reported for aflatoxin resistance (Farfan et al. 2015).

## SNPs versus genic regions for drought tolerance in cassava

The identification of genomic regions associated with drought tolerance will contribute to the understanding of genetic factors related to abiotic stress tolerance in cassava. Moreover, it will contribute to deepening knowledge of the genomic regions and related proteins. Among the transcripts related to the significant SNPs, some are directly related to proteins involved in drought tolerance, such as photosystem II oxygen enhancers, PR5-like receptor kinase-related, Beta-fructofuranosidase/Saccharase, basic leucine zipper domain (bZIP domain) transcription factor, and Apetala 2 domain (AP2), which encodes a transcription factor active during both reproductive and vegetative development (Licausi et al. 2013; Pantalião et al. 2016; Ge et al. 2012). In rice, AP2 is involved in responses to environmental stimuli (Licausi et al. 2013; Pantalião et al. 2016). In *Arabidopsis thaliana,* the role of the PR5-like receptor kinase subfamily is involved in plant drought stress signaling. AtPR5K2 physically interacts with proteins involved in initiating abscisic acid (ABA) signaling; therefore, these proteins participate in ABA-dependent drought stress signaling through the phosphorylation of ABA-insensitive proteins (Baek et al. 2019). Moreover, in *Medicago Sativa* L., the genes encoding fructokinase were upregulated in drought-tolerant varieties (Ma et al. 2020). These authors mentioned that fructokinase serves an important role in plant tolerance to abiotic stresses by promoting the accumulation of osmoprotectants.

In wheat, a growing accumulation of the oxygen-evolving enhancer protein 2 of photosystem II during grain development was identified in some genotypes submitted to water deficit conditions. Therefore, this protein may be related to greater drought resistance in plants (Ge et al. 2012). Similar observations were also identified in cassava, where this protein was expressed in genotypes grown under water stress (Lokko et al. 2007).

Notably, the leucine zipper protein is involved in plant growth regulation and drought tolerance response. This protein makes up the subunits of rictor-mTOR (TORC2) protein complexes, which

control cell growth and proliferation in several plants (Jacinto et al. 2006). The Lzipper-MIP protein belongs to the bZIP family (basic leucine zippers), which includes a series of genes involved in drought response in plants such as rice and *Arabidopsis* (Uno et al. 2000; Xiang et al. 2008). In rice, OsbZIP16 and OsbZIP23 are the main genes responsible for conferring drought tolerance (Xiang et al. 2008; Chen et al. 2012). In *Arabidopsis*, it was reported that the homeodomain-leucine zipper proteins constitute a family of transcription factors regulated at the transcriptional level by the availability of water and abscisic acid (Dezar et al. 2005). In cassava, a transcriptome analysis using three genotypes revealed that many MebZIP genes were activated due to water stress, thus highlighting the involvement of these genes in the drought tolerance process (Hu et al. 2016). Therefore, there is a high probability that SNPs close to the transcripts related to these proteins are associated with protein domains and induce drought tolerance.

Some SNPs have also been related to transcripts involved in other stresses, such as the protein kinases belonging to an important family of proteins with multiple functions (e.g., responses to biotic and abiotic stresses). In cassava, these proteins are associated with leucine-rich repeats that confer a greater tolerance to diseases (Louis and Rey 2015). The Gnk2 domain, often found in association with kinase domains, is related to saline stress responses (Miyakawa et al. 2009), while the chloroplast RNA splicing and ribosome maturation (CRM) domain is involved in growth and responses to abiotic stress in plants (Lee et al. 2014). Moreover, the LURP1 protein is associated with defense against pathogens, as reported in *Arabidopsis* (Knoth and Eulgem 2008). Therefore, the identification of transcripts involved in biological processes related to biotic and abiotic stresses—specifically those that confer greater drought tolerance—is an important contribution since these transcripts can guide future research on the functional characterization of genes involved in this important abiotic stress response.

In summary, this is the first study on the use of a GWAS to understand drought tolerance in cassava. SNPs associated with the yields of fresh roots, starch, and shoots as well as dry matter content, DTI, and DTSI were identified. Some of these SNPs are located close to the genomic regions reported as being related to proteins involved in drought tolerance in other crops. Advances have been made toward understanding the causal variants in candidate genes that may influence the levels of water deficit tolerance. However, there is a need for complementary studies aimed at the re-sequencing of candidate genes in QTL mapping, the analysis of a cassava panel germplasm with a greater number of genotypes, and the validation of candidate gene expression.

**Acknowledgements** The authors thank the Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the financial assistance and scholarship support.

## References

Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT, Jiang R, Muliyati NW, Zhang X, Amer MA, Baxter I, Brachi B, Chory J, Dean C, Debieu M, de Meaux J, Ecker JR, Faure N, Kniskern JM, Jones JD, Michael T, Nemri A, Roux F, Salt DE, Tang C, Todesco M, Traw MB, Weigel D, Marjoram P, Borevitz JO, Bergelson J, Nordborg M (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. Nature 465:627–631

Baek D, Kim M, Kumar D, Park B, Cheong MS, Choi W, Park H, Chun H, Park HJ, Lee S, Bressan R, Kim J-Y, Yun D (2019) AtPR5K2, a PR5-Like receptor kinase, modulates plant responses to drought stress by Phosphorylating Protein Phosphatase 2Cs. Front Plant Sci 10:1146

Beyene Y, Semagn K, Mugo S, Tarekegne A, Babu R, Meisel B, Sehabiague P, Makumbi D, Magorokosho C, Oikeh S, Gakunga J, Vargas M, Olsen M, Prasanna BM, Banziger M, Crossa J (2015) Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress. Crop Sci 55:154–163

Bouslama M, Schapaugh WT (1984) Stress tolerance in soybeans. Part 1: evaluation of three screening techniques for heat and drought tolerance. Crop Sci 24:933–937

Brachi B, Morris GP, Borevitz JO (2011) Genome-wide association studies in plants: the missing heritability is in the field. Genome Biol 12:232

Brito AC, Oliveira SAS, Oliveira EJ (2017) Genome-wide association study for resistance to cassava root rot. J Agric Sci 155:1424–1441

Browning BL, Browning SR (2009) A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am J Hum Genet 84:210–223

Cabello R, Monneveux P, Bonierbale M, Khan MA (2014) Heritability of yield components under irrigated and drought conditions in andigenum potatoes. Am J Potato Res 91:492–499

Cattivelli L, Rizza F, Badeck FW, Mazzucotelli E, Mastrangelo AM, Francia E, Marè C, Tondelli A, Stanca AM (2008)

Drought tolerance improvement in crop plants: an integrated view from breeding to genomics. Field Crops Res 105:1–14

Ceballos H, Okogbenin E, Pérez JC, López-Valle LAB, Debouck D (2010) Cassava. In: Bradshaw JE (ed) Hand book of plant breeding: root and tuber crops. Springer, New York, pp 53–96

Chen H, Chen W, Zhou J, He H, Chen L, Chen H, Deng XW (2012) Basic leucine zipper transcription factor osbzip16 positively regulates drought resistance in rice. Plant Sci 193:8–17

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet 8:19–33

CIAT. International Center for Tropical Agriculture (2017) http://ciat.cgiar.org/what-we-do/breeding-better-crops/rooting-for-cassava/. Accessed 15 Dec 2017.

Dekkers JCM (2004) Commercial application of marker-and gene-assisted selection in livestock: strategies and lessons. J Anim Sci 82:313–328

Dezar CA, Gago GM, González DH, Chan RL (2005) Hahb-4, a sunflower homeobox-leucine zipper gene, is a developmental regulator and confers drought tolerance to *Arabidopsis thaliana* plants. Transgenic Res 14:429–440

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small amounts of fresh leaf tissue. Phytochem 19:11–15

El-Sharkawy MA (2007) Physiological characteristics of cassava tolerance to prolonged drought in the tropics: implications for breeding cultivars adapted to seasonally dry and semiarid environments. J Plant Physiol 19:257–286

El-Sharkawy MA (2012) Stress-tolerant cassava: the role of integrative ecophysiology-breeding research in crop improvement. Open J Soil Sci 2:162–186

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:19379

Embrapa Semiárido. Centro de Pesquisa Agropecuária do Trópico Semiárido (2017) Dados meteorológicos de 2013. http://www.cpatsa.embrapa.br:8080/servicos/dadosmet/ceb-anual.html. Accessed 8 Dec 2017.

Embrapa Semiárido. Centro de Pesquisa Agropecuária do Trópico Semiárido (2017) Dados meteorológicos de 2014. http://www.cpatsa.embrapa.br:8080/servicos/dadosmet/ceb-anual.html. Accessed 8 Dec 2017.

Esuma W, Herselman L, Labuschagne MT, Ramu P, Lu F, Baguma B, Buckler ES, Kawuki RS (2016) Genome-wide association mapping of provitamin A carotenoid content in cassava. Euphytica 212:97–110

Farfan IDB, La Fuente GN, Murray SC, Isakeit T, Huang PC, Warburton M, Williams P, Windham GL, Kolomiets M (2015) Genome wide association study for drought, aflatoxin resistance, and important agronomic traits of maize hybrids in the sub-tropics. PLoS ONE 10:0117737

Fernandez GCJ (1992) Effective selectioncriteria for assessing stress tolerance. In: Kuo CG (ed) Proceedings of the international symposium on adaptation of vegetables and other food crops intemperature and water stress. AVRDC Publication, Tainan, p 257–270

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357–374

Flint-Garcia SA, Thuillet AC, Yu J, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM, Buckler ES (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J 44:1054–1064

Ge P, Ma C, Wang S, Gao L, Li X, Guo G, Ma W, Yan Y (2012) Comparative proteomic analysis of grain development in two spring wheat varieties under drought stress. Anal Bioanal Chem 402:1297–1313

Grady BJ, Torstenson ES, Ritchie MD (2011) The effects of linkage disequilibrium in large scale SNP datasets for MDR. BioData Min 4:11

Grenier C, Cao TV, Ospina Y, Quintero C, Châtel MH, Tohme J, Courtois B, Ahmadi N (2015) Accuracy of genomic selection in a rice synthetic population developed for recurrent selection breeding. PLoS ONE 10:0136594

Gupta PK, Rustgi S, Kulwal PL (2005) Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Mol Biol 57:461–485

Gutiérrez L, Germán S, Pereyra S, Hayes PM, Pérez CA, Capettini F, Locatelli A, Berberian NM, Falconi EE, Estrada R, Fros D, Gonza V, Altamirano H, Huerta-Espino J, Neyra E, Orjeda G, Sandoval-Islas S, Singh R, Turkington K, Castro AJ (2015) Multi-environment multi-QTL association mapping identifies disease resistance QTL in barley germplasm from Latin America. Theor Appl Genet 128:501–516

Han B, Huang X (2013) Sequencing-based genome-wide association study in rice. Curr Opin Plant Biol 16:133–138

Hao Z, Li X, Liu X, Xie C, Li M, Zhang D, Zhang S (2010) Meta-analysis of constitutive and adaptive QTL for drought tolerance in maize. Euphytica 174:165–177

Hinkossa A, Gebeyehu S, Zeleke H (2013) Generation mean analysis and heritability of drought resistance in common bean (*Phaseolus vulgaris* L.). Afr J Agric Res 8:1319–1329

Hu W, Yang H, Yan Y, Wei Y, Tie W, Ding Z, Zuo J, Peng M, Li K (2016) Genome-wide characterization and analysis of Bzip transcription factor gene family related to abiotic stress in cassava. Sci Rep 6:1–12

IBGE. Instituto Brasileiro de Geografia e Estatística. Produção Agrícola (2017) ftp://ftp.ibge.gov.br/Producao_Agricola/Fasciculo_Indicadores_IBGE/estProdAgr_201704.pdf. Accessed 6 Jun 2018.

Jacinto E, Facchinetti V, Liu D, Soto N, Wei S, Jung SY, Huang Q, Qin J, Su B (2006) SIN1/MIP1 maintains rictor-mTOR complex integrity and regulates Akt phosphorylation and substrate specificity. Cell 127:125–137

Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E (2008) Efficient control of population structure in model organism association mapping. Genetics 178:1709–1723

Kawano K, Fukuda WMG, Cenpukdee U (1987) Genetic and environmental effects on dry matter content of cassava root 1. Crop Sci 27:69–74

Knoth C, Eulgem T (2008) The oomycete response gene lurp1 is required for defense against *Hyaloperonospora parasitica* in *Arabidopsis thaliana*. Plant J 55:53–64

Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9:29–37

Lee K, Lee HJ, Kim DH, Jeon Y, Pai HS, Kang H (2014) A nuclear-encoded chloroplast protein harboring a single CRM domain plays an important role in the Arabidopsis growth and stress response. BMC Plant Biol 14:98–109

Licausi F, Ohme-Takagi M, Perata P (2013) APETALA 2/Ethylene Responsive Factor (AP 2/ERF) transcription factors: mediators of stress responses and developmental programs. New Phytol 199:639–649

Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z (2012) GAPIT: genome association and prediction integrated tool. Bioinform 28:2397–2399

Lipka AE, Kandianis CB, Hudson ME, Yu J, Drnevich J, Bradbury PJ, Gore MA (2015) From association to prediction: statistical methods for the dissection and selection of complex traits in plants. Curr Opin Plant Biol 24:110–118

Lokko Y, Anderson JV, Rudd S, Raji A, Horvath D, Mikel MA, Kim R, Liu L, Hernandez A, Dixon AG, Ingelbrecht IL (2007) Characterization of an 18,166 EST dataset for cassava (*Manihot esculenta* Crantz) enriched for drought-responsive genes. Plant Cell Rep 26:1605–1618

Louis B, Rey C (2015) Resistance gene analogs involved in tolerant cassava-geminivirus interaction that shows a recovery phenotype. Virus Genes 51:393–407

Liu J, Zheng Q, Ma Q, Gadidasu KK, Zhang P (2011) Cassava genetic transformation and its application in breeding. J Integr Plant Biol 53:552–569

Liu S, Wang X, Wang H, Xin H, Yang X, Yan J, Li J, Tran LSP, Shinozaki K, Yamaguchi-Shinozaki K, Qin F (2013) Genome-wide analysis of *ZmDREB* genes and their association with natural variation in drought tolerance at seedlings stage of *Zea mays* L. PLoS Genet 9:1003790

Lu Y, Xu J, Yuan Z, Hao Z, Xie C, Li X, Shah T, Lan H, Zhang S, Rong T, Xu Y (2012) Comparative LD mapping using single SNPs and haplotypes identifies QTL for plant height and biomass as secondary traits of drought tolerance in maize. Mol Breed 30:407–418

Ma Q, Xu X, Xie Y, Wang W, Zhao L, Ma D (2020) Comparative analysis of alfalfa (*Medicago sativa* L.) seedling transcriptomes reveals genotypespecic drought tolerance mechanisms. BMC Plant Biol. doi:https://doi.org/10.21203/rs.3.rs-75482/v1

Ma X, Feng F, Wei H, Mei H, Xu K, Chen S, Li T, Liang X, Liu H, Luo L (2016) Genome-wide association study for plant height and grain yield in rice under contrasting moisture regimes. Front Plant Sci 7:1801–1814

Malosetti M, Ribaut J-M, Van Eeuwijk FA (2013) The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. Front Physiol 4:44–61

Mangin B, Siberchicot A, Nicolas S, Doligez A, This P, Cierco-Ayrolles C (2011) Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. Heredity 108:285–291

Masumba EA, Kapinga F, Mkamilo G, Salum K, Kulembeka H, Rounsley S, Bredeson JV, Lyons JB, Rokhsar DS, Kanju E, Katari MS, Myburg AA, van der Merwe NA, Ferguson ME (2017) QTL associated with resistance to cassava brown streak and cassava mosaic diseases in a bi-parental cross of two Tanzanian farmer varieties, Namikonga and Albert. Theor Appl Genet 130:2069–2090

Mather A, Caicedo AL, Polato NR, Olsen KM, Mccouch S, Purugganan MD (2007) The extent of linkage disequilibrium in rice (*Oryza sativa* L.). Genetics 177:2223–2232

Mathews KL, Malosetti M, Chapman S, Mcintyre L, Reynolds M, Shorter R, Van Eeuwijk F (2008) Multi-environment QTL mixed models for drought stress adaptation in wheat. Theor Appl Genet 117:1077–1091

Millet E, Welcker C, Kruijer W, Negro S, Nicolas S, Praud S, Draye X, Negro S, Coupel-Ledru A, Nicolas SD, Laborde J, Bauland C, Praud S, Ranc N, Presterl T, Tuberosa R, Bedo Z, Draye X, Usadel B, Charcosset A, Van Eeuwijk F, Tardieu F (2016) Genome-wide analysis of yield in Europe: allelic effects as functions of drought and heat scenarios. Plant Physiol 172:749–764

Miyakawa T, Miyazono KI, Sawano Y, Hatano KI, Tanokura M (2009) Crystal structure of ginkbilobin-2 with homology to the extracellular domain of plant cysteine rich receptor-like kinases. Proteins: Struct, Funct, Bioinf 77:247–251

Morosini JS, Mendonça LF, Lyra DH, Galli G, Vidotti MS, Fritsche-Neto R (2017) Association mapping for traits related to nitrogen use efficiency in tropical maize lines under field conditions. Plant Soil 421:453–463

Okogbenin E, Setter TL, Ferguson M, Mutegi R, Ceballos H, Olasanmi B, Fregene M (2013) Phenotypic approaches to drought in cassava: review. Front Physiol 4:1–15

Oliveira EJ, Aidar ST, Morgante CV, Chaves ARM, Cruz JL, Coelho Filho MA (2015) Genetic parameters for drought-tolerance in cassava. Pesqui Agropecu Bras 50:233–241

Oliveira EJ, Morgante CV, Aidar ST, Chaves ARM, Antonio RP, Cruz JL, Coelho Filho MA (2017) Evaluation of cassava germplasm for drought tolerance under field conditions. Euphytica 213:188–208

Ogola JBO, Mathews C (2011) Adaptation of cassava (*Manihot esculenta*) to the dry environments of Limpopo, South Africa: growth, yield and yield components. Afr J Agric Res 6:6082–6088

Pantalião GF, Narciso M, Guimarães C, Castro A, Colombari JM, Breseghello F, Rodrigues L, Vianello RP, Borba TO, Brondani C (2016) Genome wide association study (GWAS) for grain yield in rice cultivated under water deficit. Genetica 144:651664

R Core Team. R: a language and environment for statistical computing (2020) R foundation for statistical computing. https://www.R-project.org/. Accessed 8 Aug 2020

Raj A, Stephens M, Pritchard JK (2014) fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics 197:573–589

Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M (2010) Genome-wide association studies in diverse populations. Nat Rev Genet 11:356–366

Sedano JCS, Moreno REM, Mathew B, Léon J, Cano FAG, Ballvora A, Carrascal CEL (2017) Major novel QTL for resistance to cassava bacterial blight identified through a multi-environmental analysis. Front Plant Sci 8:1169

Stich B, Urbany C, Hoffmann P, Gebhardt C (2013) Population structure and linkage disequilibrium in diploid and tetraploid potato revealed by genome-wide high-density

genotyping using the SolCAP SNP array. Plant Breed 132:718–724

Souza LS, Farias A, Mattos P, Fukuda WMG (2006) Aspectos socioeconômicos e agronômicos da mandioca, Embrapa Informação Tecnológica; Cruz das Almas: Embrapa Mandioca e Fruticultura

Uno Y, Furihata T, Abe H, Yoshida R, Shinozaki K, Yamaguchi-Shinozaki K (2000) *Arabidopsis* basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. Proc Natl Acad Sci 97:11632–11637

Vanraden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91:4414–4423

Vitor AB, Diniz RP, Morgante CV, Antônio RP, Oliveira EJ (2019) Early prediction models for cassava root yield in different water regimes. Field Crops Res 239:149–158

Truntzler M, Ranc N, Sawkins MC, Nicolas S, Manicacci D, Lespinasse D, Ribière V, Galaup P, Servant F, Muller C, Madur D, Betran J, Charcosset A, Moreau L (2012) Diversity and linkage disequilibrium features in a composite public/private dent maize panel: consequences for association genetics as evaluated from a case study using flowering time. Theor Appl Genet 125:731–747

Tuberosa R, Salvi S (2006) Genomics-based approaches to improve drought tolerance of crops. Trends Plant Sci 11:405–412

Xiang Y, Tang N, Du H, Ye H, Xiong L (2008) Characterization of OsbZIP23 as a key player of the basic leucine zipper transcription factor family for conferring abscisic acid sensitivity and salinity and drought tolerance in rice. Plant Physiol 148:1938–1952

Xue Y, Warburton ML, Sawkins M, Zhang X, Setter T, Xu Y, Gruldoyma P, Gethi J, Ribaut JM, Li W, Zhang X, Zheng Y, Yan J (2013) Genome-wide association analysis for nine agronomic traits in maize under well-watered and water-stressed conditions. Theor Appl Genet 126:2587–2596

Wehner GG, Balko CC, Enders MM, Humbeck KK, Ordon FF (2015) Identification of genomic regions involved in tolerance to drought stress and drought stress induced leaf senescence in juvenile barley. BMC Plant Biol 15:1

Weisberg S (2005) Applied linear regression, vol 528. Hoboken, NJ: Wiley.

Würschum T, Langer SM, Longin CFH, Korzun V, Akhunov E, Ebmeyer E, Schachschneider R, Schacht J, Kazman E, Reif JC (2013) Population structure genetic diversity and linkage disequilibrium in elite winter wheat assessed with SNP and SSR markers. Theor Appl Genet 126:1477–1486

Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203

Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, Buckler E, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, Buckler ES (2010) Mixed linear model approach adapted for genome-wide association studies. Nat Genet 42:355–362

Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. Plant Genome 1:5–20