

RESEARCH ARTICLE

Molecular diversity analysis, drought related marker-traits association mapping and discovery of excellent alleles for 100-day old plants by EST-SSRs in cassava germplasms (*Manihot esculenta* Cranz)

Bin Wang^{1,2,3}, Xin Guo^{1,2,3}, Pingjuan Zhao^{2,3}, Mengbin Ruan^{2,3}, Xiaoling Yu^{2,3}, Liangping Zou^{2,3}, Yiling Yang^{1,2,3}, Xiao Li^{2,3}, Deli Deng^{2,3}, Jixiang Xiao^{2,3}, Yiwei Xiao^{2,3}, Chunji Hu^{2,3}, Xue Wang^{2,3}, Xiaolin Wang^{2,3}, Wenquan Wang^{2,3}, Ming Peng^{2,3*}

1 College of plant science & technology, Huazhong Agricultural University, Wuhan, Hubei, PR China, **2** Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, Haikou, PR China, **3** Key Laboratory of Biology and Genetic Resources of Tropical Crops, Ministry of Agriculture, Haikou, PR China

* pengming@itbb.org.cn



OPEN ACCESS

Citation: Wang B, Guo X, Zhao P, Ruan M, Yu X, Zou L, et al. (2017) Molecular diversity analysis, drought related marker-traits association mapping and discovery of excellent alleles for 100-day old plants by EST-SSRs in cassava germplasms (*Manihot esculenta* Cranz). PLoS ONE 12(5): e0177456. <https://doi.org/10.1371/journal.pone.0177456>

Editor: Tapan Kumar Mondal, National Bureau of Plant Genetic Resources, INDIA

Received: November 30, 2016

Accepted: April 27, 2017

Published: May 11, 2017

Copyright: © 2017 Wang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was financially supported by the National Natural Science Foundation of China NSFC-CGIAR Project (grant no. 31561143012), the National Science Foundation of China (grant no. 31501378), the national key technology R&D program of China (grant no. 2015BAD15B01) and

Abstract

Cassava is the third largest food crop of the world and has strong ability of drought tolerance. In order to evaluate the molecular diversity and to discover novel alleles for drought tolerance in cassava germplasms, we examined a total of 107 abiotic stress related expressed sequence tags—simple sequence repeat (EST-SSR) markers in 134 cassava genotypes coming from planting regions worldwide and performed drought related marker-traits association mapping. As results, we successfully amplified 98 of 107 markers in 97 polymorphic loci and 279 alleles, with 2.87 alleles per locus, gene diversity of 0.48 and polymorphic information content (PIC) of 0.41 on average. The genetic coefficient between every two lines was 0.37 on average, ranging from 0.21 to 0.82. According to our population structure analysis, these samples could be divided into three sub-populations showing obvious gene flow between them. We also performed water stress experiments using 100-day old cassava plants in two years and calculated the drought tolerance coefficients (DTCs) and used them as phenotypes for marker-trait association mapping. We found that 53 markers were significantly associated with these drought-related traits, with a contribution rate for trait variation of 8.60% on average, ranging between 2.66 and 28.09%. Twenty-four of these 53 associated genes showed differential transcription or protein levels which were confirmed by qRT-PCR under drought stress when compared to the control conditions in cassava. Twelve of twenty-four genes were the same differential expression patterns in omics data and results of qRT-PCR. Out of 33 marker-traits combinations on 24 loci, 34 were positive and 53 negative alleles according to their phenotypic effects and we also obtained the typical materials which carried these elite alleles. We also found 23 positive average allele effects while 10 loci were negative according to their allele effects (AAEs). Our results on molecular diversity, locus association and differential expression under drought can prove

the Hainan province innovative research team foundation (grant no. 2016CXTD013).

Competing interests: The authors have declared that no competing interests exist.

beneficial to select excellent materials through marker assisted selection and for functional genes research in the future.

Introduction

Cassava, an important food and starch crop, has excellent adaptability to multiple environments. However, seasonal drought every year reduces its yield in major planting regions [1, 2]. One approach to solve this problem is to develop new cultivars which are not only drought-tolerant but also have high yield.

Analysis of the genetic diversity of germplasm is a valuable tool for cassava breeding, especially when aiming for drought tolerance [2]. High genetic diversity in germplasm improves the probability to select excellent new germplasm. Genetic diversity is also used as a reference to select materials with outstanding characteristics, such as high yield and drought tolerance, and parent materials with excellent breeding potential and multiple elite alleles of interest.

DNA markers have become a powerful tool for genetic diversity analysis of germplasm and for discovering alleles of interest. Single nucleotide polymorphism (SNP) chips, such as GoldenGate Infinium from Illumina, SNPStream from Beckman Coulter or GeneChip from MegAllel and Affymetrix [3], and high throughput sequencing methods, such as reduced-representation sequencing, restrict site associated DNA (RAD) or exon trapping sequencing, are the two major techniques used in modern molecular marker technology. While GBS strategy become more and more popular for its cheaper and massive information, SSR markers remained to be widely used in MAS and primary mapping researches. At present, SSR markers that detect co-dominance and higher polymorphisms are also valid alternative. EST-SSRs are designed on expressed sequences which involved in the variety of metabolic functions [4]. Gene-based SSR markers are located in target genes, returning clear information the same as EST-SSRs. Thus, both of EST-SSRs and gene-based SSRs had potential use in quantitative trait loci (QTL) mapping and association mapping researches.

Molecular marker technology has also rapidly developed in cassava, with more and more studies using EST-SSRs in combination with other molecular markers such as genomic SSRs [5], amplified fragment length polymorphism (AFLP)[6] or EST-SNPs [7]. This allowed constructing cassava linkage maps in F1 generation segregation population derived from heterozygote parents. Gene-based SSRs have also been used in cassava researches. A previous report studied 846 putative SSRs in 8577 cassava unigene sets and evaluated 124 new, unique polymorphic gene-based SSRs in 25 cassava cultivars and their wild relatives [8]. Moreover, SNP genotyping methods were also used in cassava [9, 10]. For all this, SNP chips in cassava genetic linkage and physical mapping, genotype-by-sequencing (GBS) approaches in re-sequencing cassava cultivars and their wild ancestor, together with the announcement of a database containing cassava genome and SNPs information database derived from 53 cassava cultivars promoted molecular marker development in cassava [11–13].

Linkage QTL mapping and association mapping are used to explore novel alleles in crops such as rice, maize, wheat, cotton, soybean, rape and sugar beet. However, their application in cassava is still rare. The major method to explore novel alleles in cassava is linkage QTL mapping in F1 generation derived from heterozygous parents. Previous studies used this for yield related traits [6, 14], starch pasting viscosity [15], cyanogenic glucoside [14, 16] and cassava mosaic disease (CMD) resistance related traits [17, 18]. Due to the high frequency natural cross-pollination in cassava, the presence of just a few flowers and a high rate of heterozygosis, it is difficult to cultivate homozygote lines and it is impossible to fine map them using high

generation backcross populations as done in other crops. Even so, application of association mapping methods in natural population to discover novel alleles of important agronomic traits in cassava might be a viable option.

Because of the complexity of drought stress responses and the bottle necks of breeding methods based on phenotype selection, researchers need to consider QTL mapping in drought tolerance breeding projects [19]. In order to analyze population structure, association mapping of drought tolerance related traits and excellent alleles for drought tolerance in available cassava germplasms, we genotyped 134 germplasms by using 107 EST-SSR markers. Moreover, we used reverse genetics data to confirm our association mapping results. Finally, we obtained a series of casual associated loci with elite alleles and their carriers.

Materials and methods

Plant materials

The 134 cassava accessions used in our study were obtained from 11 countries and areas as described in [S1 Table](#). They were major cultivars or breeding materials in these countries or areas which could be considered as hybrids and were vegetative reproduction by stems. All of these accessions were used as the basic population to analyze genetic diversity, water depress experiment and marker-traits association mapping. Two cassava cultivars Arg7 and SC124 which had different growing strategies under drought stress [20] were used to perform water depress experiment for qRT-PCR verification.

Water stress experiments, phenotyping and statistical methods

In order to evaluate the drought tolerance of cassava germplasms, we preformed water stress experiments in the year 2014 and 2015 in Haikou, Hainan province, China. Stems of cassava germplasms which were harvested in January in 2014 were cut into equal length fragments and grown in flowerpots which were filled with 1:2 ratio of sand to soil. Water depress experiment was performed after 100 days accessions were planted. Plants with even growth potential were chosen for water stress experiments. Drought treatment group and control group contained six plants respectively and be divided to two repeats. The former was not watered for 12 days while the latter continued to be watered. Three plants of both two repeats with uniform phenotypes were taken as samples from drought treatment group and control group respectively. Above-ground fresh weight (AGFW), storage root fresh weight (SRFW), storage root number (SRN), storage root dry weight (SRDW) and storage root dry matter percentage (SRDMP) per plant were measured. And leaves and root were cut into pieces and mixed to put into the liquid nitrogen and then turn to ultra-low temperature freezer (-80°C). Samples were used to test the physiological traits containing content of proline, soluble reducing sugar (SRS) and malondialdehyde (MDA) per unit fresh weight and activity of superoxide dismutase (SOD), peroxidase (POD) and catalase (CAT) per unit fresh weight. The measure method of proline content is ninhydrin colorimetry, anthrone colorimetry for SRS content, thiobarbituric acid chromatometry (TBA) for MDA content, nitroblue tetrazolium reduced (NBT) method for SOD activity, guaiacol method for POD activity and hydrogen peroxide (H_2O_2) ultraviolet (UV) absorption method for CAT activity. And all of tests for 6 physiological phenotypes were performed by using respective assay kit (Comin biotechnology Co., ltd. Suzhou, China). We used only 95 germplasms in 2015 to water depress experiment due to limited breeding stems and low seedling emergence rate of some germplasms. And seven traits of these cassava genotypes, including AGFW, Proline, MDA, SRS, SOD, POD and CAT, were also detected in 2015. All phenotypes were turned into drought tolerance coefficients (DTC), as described before [21]. And then, maximum, minimum, average, standard deviation and coefficient variation

were analyzed by SPSS18.0 software (<http://www-01.ibm.com/software/analytics/spss/>). Box-plot figures of DTCs were drawn by using GraphPad Prism version 5.0 for Windows, Graph-Pad Software (San Diego, California, USA, www.graphpad.com)

DNA extraction and SSR genotyping

Young leaves of 134 accessions were used for genomic DNA as described before [5]. We chose 107 pairs of SSR markers, containing 55 pairs of EST-SSRs [5] and 52 pairs of gene-SSRs [8], to genotyping cassava germplasm using a previously described PCR program [22]. PCR products were mixed with a fluorescent dye to perform capillary electrophoresis. SSR fragments were assigned to genotypes according to the method reported by Wang et al. [23].

Genetic diversity analysis

Number of alleles per locus, major allele frequency, gene diversity and polymorphic information content (PIC) were calculated using Powermarker Version 3.25 software [24]. Genetic coefficient between every two genotypes was calculated using NTSYSPC2.10e software [25]. Analysis of molecular variation and sub-population genetic differences (PhiPT) and gene flow (Nm) was performed using Genalex 6.2 [26].

Population structure analysis

In order to analyze the population genetic structure and deal with its interference for association mapping, STRUCTURE V2.0 software [27, 28] was applied to analyze the genotype data of cassava germplasms. Both of burnin and MCMC were set to 100,000 and K ranged from 1 to 10 with three integrations for every K value. LnP(D) and Variance of LnP(D) which were calculated by STRUCUTRE software were used to estimate the ΔK for every K value as previously described [29]. Two curve graphs were drawn that LnP(D) and ΔK ranged with K value increasing from 1 to 10 respectively. The real K value was that one co-responding the first tunnel of ΔK . These three Q matrices of the real K values, estimated by using STRUCTURE, were combined with CLUMPP 2.0 [30]. The combined Q matrix was used as covariance matrix in association mapping and source data to draw the stacked column chart of Q matrix in excel. Besides, all of these 134 genotypes were divided into three clusters and mixed cluster according to the method reported by Yang et al. [31].

Marker-trait association mapping

DTCs of traits obtained in 2014 and 2015 were used as the phenotypes in association mapping while the SSR fragments information of germplasms as genotypes. Marker-trait association mapping was performed in TASSEL V2.1 [32] by using Q + K+ MLM with MAF > 0.05 filter condition. We obtained significance P values and phenotypes variation R^2 and P values were modified according to the FDR method [33]. Marker-trait associations were considered significant if the FDR of the P values were lower than 0.05. All of significant associated markers were mapped in the *Manihot esculenta* Cranz genome v6.1 (<https://phytozome.jgi.doe.gov/pz/portal.html>)[13] and were used to draw a graph of distributions of the associated markers by MAPChart V2.0 software [34] according to the physical position of 107 EST-SSR markers.

Expression analysis of associated genes

In order to confirm whether associated genes were actual functional genes, gene codes in the cassava genome v6.1 and v4.1 were used to search the transcriptomics under water stress, obtained from genebank and containing the controls and drought data of three genotypes:

SC124, Arg7 and W14, sequenced by Illumina HiSeq 2000 by the Institute of tropical bioscience and biotechnology, Chinese Academy of Tropical Agricultural Sciences, Haikou, PR China (Accession numbers of genebank, SRR2360410, SRR2361999, SRR2388947, SRR2388956, SRR2404199, SRR2404206, SRR2495946, SRR2495947, SRR2495949, SRR2495950, [SRR2496326](#), [SRR2496093](#)) [35] and the iTRAQ-based proteomics database [20]. Log₂ based FPKM change fold values equal to FPKM of associated genes under drought condition based on the transcriptomics reported by Wei et al. [35] were calculated divide them for the controls and used to draw a heat map using the MeV2.0 software [36].

Furthermore, in order to confirm significantly different expressed genes, relative expression level of the 24 genes which were selected from previous omics database were detected in two cassava cultivars Arg7 and SC124 by qRT-PCR method. Samples which were used to qRT-PCR experiment derived from independent drought stress experiment of cassava cultivars. Every pot contained both of the two cultivars plants to keep them in coincident water depression. The conditions of plants growing, methods of drought treatments and tissues collection were the same as described before. Tissues of every cultivar were collected from the mixed materials of five independent plants in both of the drought treatment groups and control groups respectively. Total RNA extraction and cDNA first strand reverse experiments were performed by using RNAPrep Pure Plant Kit (Polysaccharides&Polyphenolics-rich, Tiangen Biotech, Beijing Co. Ltd) and FastQuant RT Kit (with gDnase, Tiangen Biotech, Beijing Co. Ltd) respectively. All of primers used to amplify target genes were shown in [S2 Table](#). House-keeping gene *MeActin* with JGI ID code Manes.12G150500 was used as internal control and three technique repeats were performed for every sample. qRT-PCR performed by using SYBR Premix Ex Taq™ II (perfect real time) kit (Takara Biotechnology, Dalian, China) on StepOne plus system. RQ was estimated by StepOne software v2.1 through $2^{-\Delta\Delta CT}$ method based on C_T values [37]. Significance level of relative expression between drought treatment groups and control groups were calculated by two-tailed *T* test method in Microsoft excel. And differential expression genes were selected with $P < 0.05$ at least in one tissue vs. cultivars.

Discovery of elite alleles

Our association mapping and expression analysis indicated that there were total twenty-four genes significantly associated with drought tolerance. Phenotypic effects of the alleles at these loci were estimated as previously reported [38] and alleles whose phenotypic effects were positive would be considered positive alleles, while alleles whose phenotypic effects were negative values would be considered as negative alleles. The average positive (negative) allele effect of the locus (AAE) was also calculated according to a previously reported method [38]. Loci of which AAE had positive values were considered positive loci of associated traits; otherwise they were counted as negative loci.

Results

Genetic diversity analysis

To evaluate the molecular diversity and genetic structure of cassava germplasm resources, we used 107 EST-SSR markers, 98 of which were located in cassava chromosomes ([S1 Fig.](#)) while others were mapped in scaffolds of cassava genome version 6.1. Ninety-eight genes were aligned and other nine markers were not successfully mapped. Finally, Ninety-eight SSRs primers couples successfully amplified 97 polymorphic loci and 279 alleles. The MAF per locus was 0.62 on average, ranging from 0.29 to 0.99. The number of alleles per locus was 2.87 on average, ranging from 2 to 8. Gene diversity per locus was 0.48 on average, ranging from 0.02–0.80. Average PIC value per locus was 0.41, ranging from 0.02 to 0.77. Genetic coefficient

between every two accessions was 0.37 on average, with a maximum of 0.82 measured between Xinxuan048 and Royang9-1 and a minimum of 0.21 between ECU81 and R7. Genetic coefficients between every two lines were widely ranged, indicating that the genetic relationships between germplasms were very complex.

All of 134 germplasms were divided into four groups according to their area of origin (Asia, South America, Central America and Africa). As was shown in Table 1, result of AMOVA analysis estimated that molecular variations were present mainly within subpopulations. However, the PhiPT was 4.00E-3 with a P value of 0.41, indicating that this was not significant. The gene flow (Nm) was 116.87. These data could indicate that there was no significant sub-population differentiation among the 134 cassava germplasms and that the gene flow was very strong among subpopulations.

Population structure

In order to distinct the genetic structure of the 134 germplasms, we used the Structure software to estimate the number of sub-populations. As shown in Fig 1A and 1B, LnP(D) was the first significant peak at K = 3 while the ΔK was also the first tunnel and peak value at K = 3. Thus, total 134 germplasms were divided into three clusters. Then, we obtained the results of three integrations of K = 3 and combined the Q values of every germplasms into one Q matrix using CLUMPP V2.0 (Fig 1C). We divided the sub-groups using the STRUCTURE software according to previously reported (Yang et al. 2011), obtaining a number of genotypes for the three clusters and mixed cluster of 4, 46, 74 and 10, accounting for 3.00%, 34.33%, 55.22% and 7.46% of the total, respectively. AMOVA for these three clusters, performed by using the Genalex software, showed that PhiPT was 0.03 (P = 0.02) and Nm was 18.73, indicating that these three clusters were significantly different with a strong gene flow. However, we found no significant association between the three clusters and their original countries or areas (Fig 1D).

Summary of description statistics for drought-related traits

In order to evaluate the drought tolerance of cassava germplasms, we preformed water stress experiments in the years 2014 and 2015, as described in the methods section. We tested seven physiological phenotypes of 100-day old plants, including AGFW, Proline, SRS and MDA content, as well as SOD, POD and CAT activity. Storage root-related traits of 100-day old plants were only tested in 2014 (Fig 2). Using the DTC definition reported by [21], we translated these phenotypes into the respective DTCs and analyzed the results using the descriptive statistics function of the SPSS software (S3 Table). All of these traits were qualitative traits with a continuous distribution character. The maximum DTC range was that of POD in leaves in 2015 while the minimum was that of SRS in leaves in 2015. Coefficient DTC variation were very different for different traits. Among them, the biggest CV % value was that of DTC in leaves in 2014 and the smallest CV % value was that of SRS in leaves in 2015. This indicated that DTCs were very distinct in different germplasms.

Marker-traits association mapping

To explore novel allele for drought tolerance in cassava, we performed marker-traits association mapping by using the Q+K+MLM model. As shown in S4 Table, fifty-three pairs of

Table 1. AMOVA of cassava germplasm resources.

Source	df	SS	MS	Est. Var.	%
Among Pops	3	87.741	29.247	0.115	0.4
Within Pops	126	3377.212	26.803	26.803	99.6
Total	129	3464.954		26.918	100%

<https://doi.org/10.1371/journal.pone.0177456.t001>

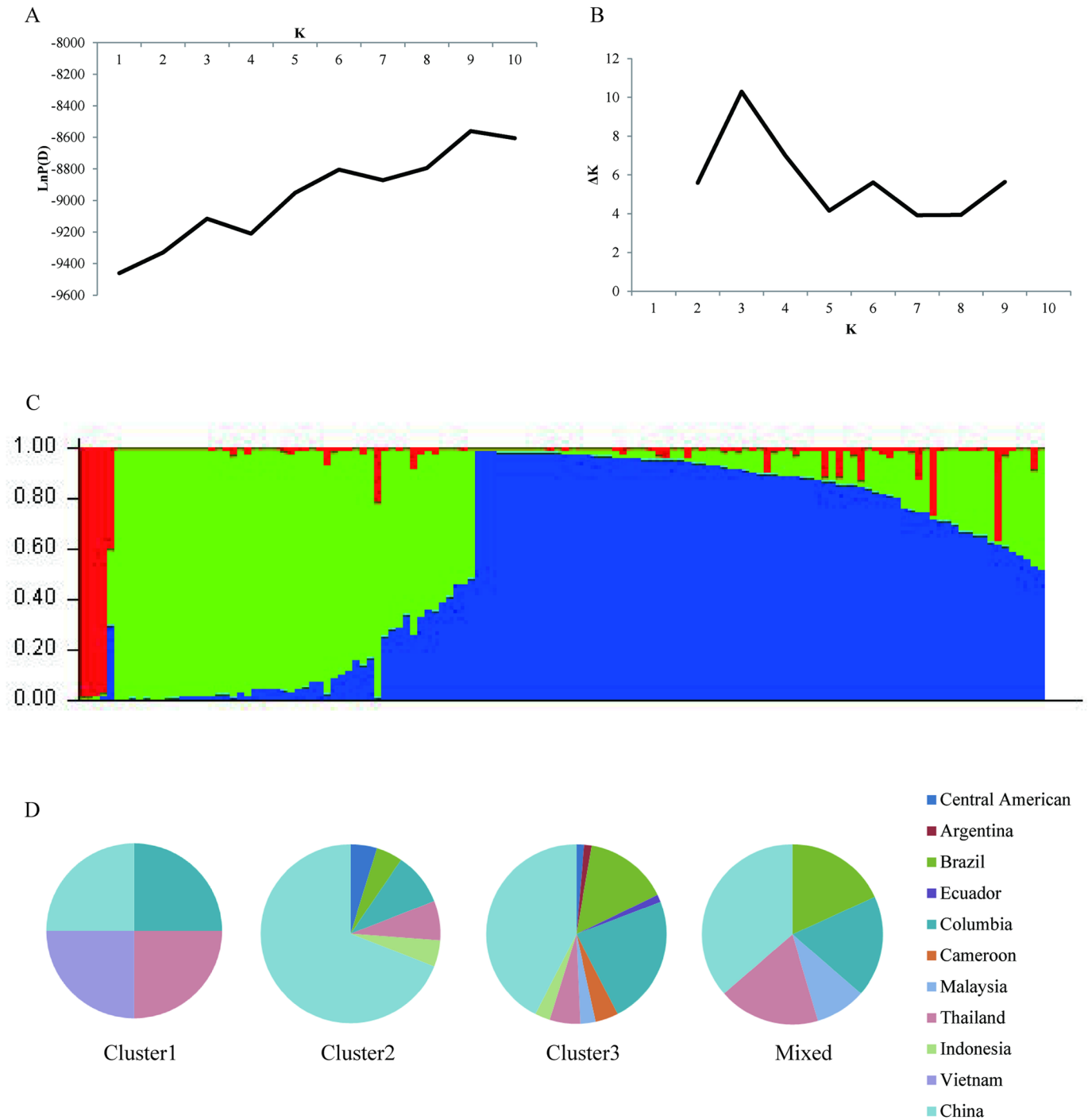


Fig 1. Population structure analysis. A, LnP(D) value graphed against K from 1 to 10. B, ΔK values at K from 1 to 10. C, stacked column chart of Q matrix. D, Origin distributions of the three clusters and mixed cluster.

<https://doi.org/10.1371/journal.pone.0177456.g001>

EST-SSR markers, located in 17 chromosomes of cassava except the 4th chromosome (Fig 3), were associated to DTCs of drought tolerance related traits ($P_{FDR} < 0.05$). Explanation of phenotypic variation (R^2) was 8.60% on average, ranging from 2.66% to 28.09% (S4 Table).

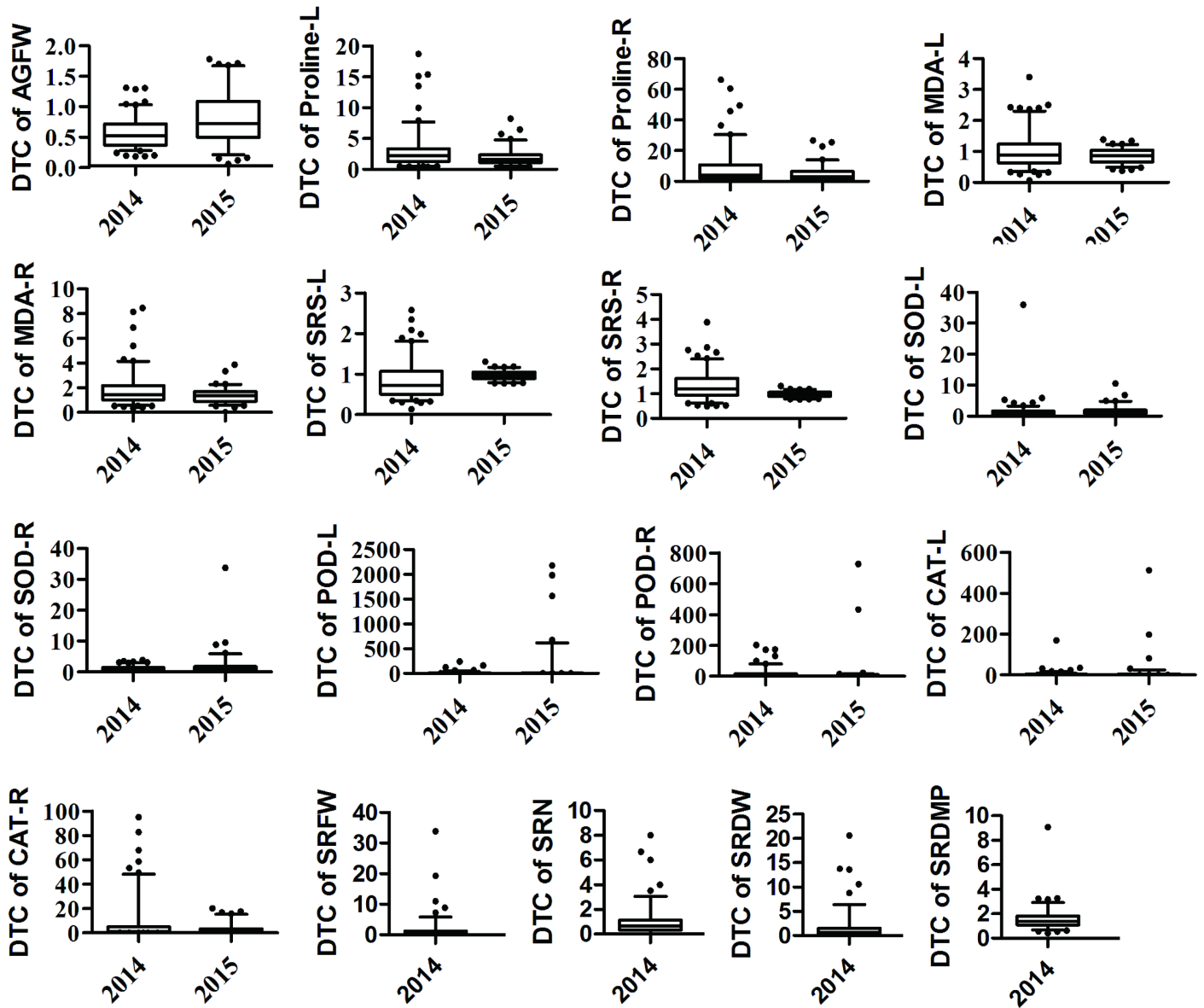


Fig 2. Boxplots of DTCs of cassava germplasm phenotypes in water stress experiments.

<https://doi.org/10.1371/journal.pone.0177456.g002>

Twenty-two marker-trait combinations had an explanation of phenotypic variation of them higher than 10% and accounted for the 22% of total significant marker-trait combinations. In our two-year water stress experiment, 3 SSRs, EME309, MESSR64 and MESSR71, were associated with the same traits and organs. Another marker, EME164, was associated with the same traits but in different organs. In 2014, we detected 47 marker-trait combinations of 33 loci, of which ten associated to multiple phenotypes while 23 others were associated to a single phenotype. In 2015, forty marker-trait combinations of 31 loci were associated, with eight of them that were multiple-association while 23 were associated to one phenotype. Besides, there were 13 markers that were associated not only to the physiological traits but also to biomass and storage root related traits of 100-day old plant.

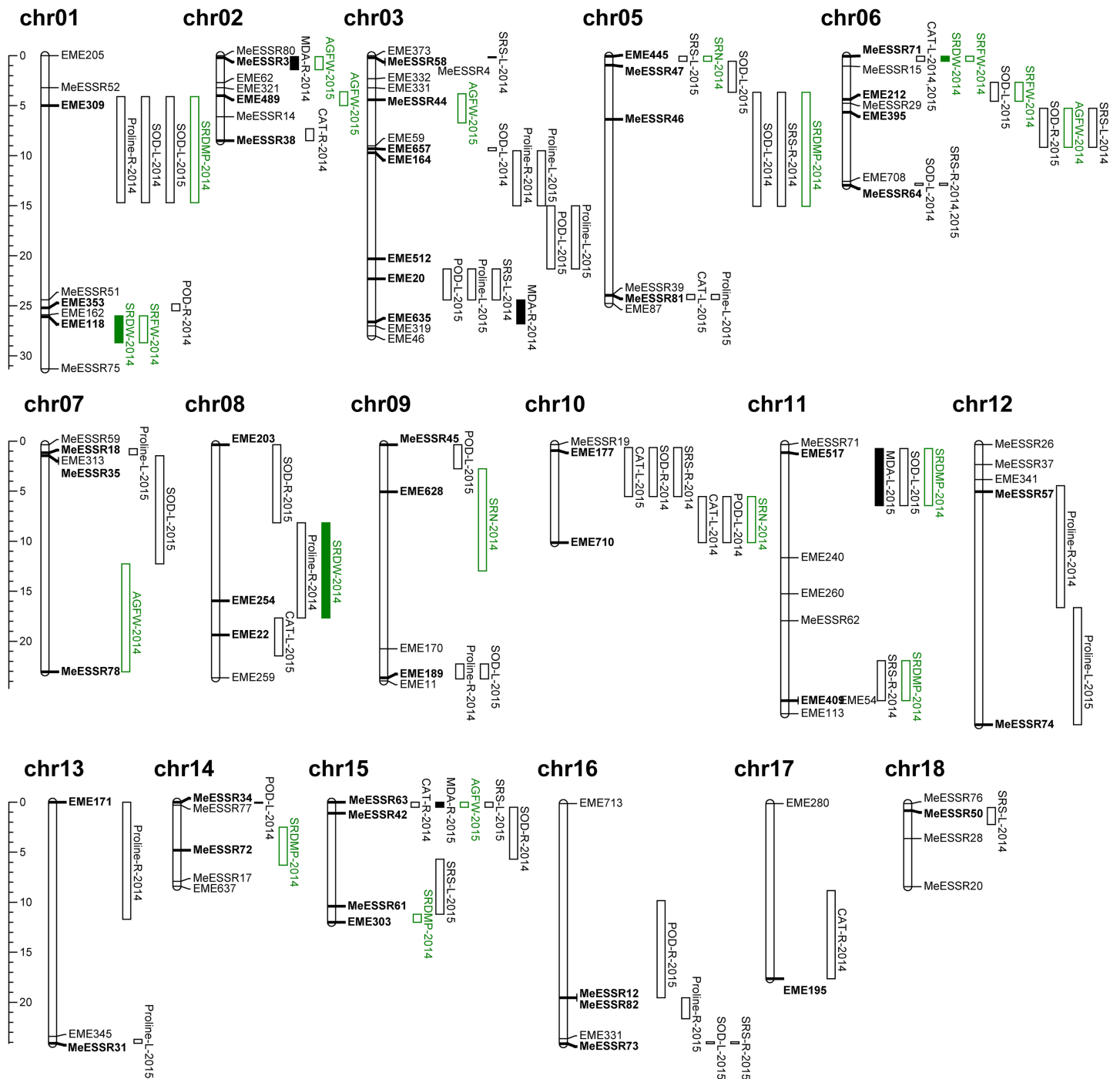


Fig 3. Distribution of associated markers in cassava genome ($P < 0.05$).

<https://doi.org/10.1371/journal.pone.0177456.g003>

Expression pattern analysis

In order to confirm these genes co-localized with associated EST-SSRs and to discover the real functional genes, these 53 associated makers were mapped in the cassava genome. 51 of them were located in coding genes in cassava genome version 4.1 and 6.1 (S4 Table). 44 genes were also present in cassava RNA sequencing data in control and water-stressed cassava plants

derived from genebank. We used Log₂ based on FPKM change fold values under water stress to draw the heat map, with 2.0 and -2.0 as the upper and lower limit respectively. As results, 24 genes that were differently expressed at least in one sample vs. cultivar under drought stress and control condition (Fig 4). And according to the protein database reported by previous study [20], four proteins encoded by four associated genes differential expressed in response to water stress in at least one organ, cultivar or time point (Table 2). According to the transcriptomics and proteomics database of cassava water stress plants, there were four genes that were associated to water stress as for DNA and mRNA expression, as well as protein levels, while 24 genes that were associated at least two or three of them.

All of 24 genes were confirmed by qRT-PCR and the relative expression of every tissue compared drought to control groups were calculated and *T* test was performed to estimate whether they were significantly differential expression ($P < 0.05$) or not (Fig 5). As was shown in Fig 5 and Table 3, there were 9 genes, Manes.02G009300 (co-located with MeSSR36), Manes.09G068800 (co-located with EME628), Manes.02G062400 (co-localized with EME212), Manes.02G124800 (co-localized with MeSSR38), Manes.03G053400 (co-located with MeSSR44), Manes.06G045100 (co-located with MeSSR71), Manes.07G029200 (co-located with MeSSR18), Manes.13G013400 (co-located with EME171) and Manes.10G078800 (co-located with EME710) which were significantly up-regulated expression while other 3 genes, Manes.12G005000 (co-located with EME425), Manes.12G129800 (co-located with MeSSR74) and Manes.13G127100 (co-located with MeSSR31) were significantly down-regulated expression under water stress. They were consistent with previous omics data described above and results of qRT-PCR. In addition, there were 8 genes were partly consensus with omics data while another 4 were opposite (Fig 5). All of the twenty-four differential expression genes revealed by qRT-PCR were meaningful to marker assistant selection in cassava breeding.

Discovery of elite alleles

To estimate the phenotypic effects of alleles and their AAEs, we used 24 loci co-locating with genes were significantly differential expression under water stress confirmed by qRT-PCR as target for elite allele's discovery, as previously reported [38]. As were shown in S5 Table, all of these 24 associated loci contained 33 marker-trait combinations and 87 allele-trait combinations, and these 87 allele-trait combinations also included 34 positive and 53 negative alleles. All of these positive alleles except alleles which were associated with MDA content were elite alleles while negative alleles associated with MDA content were elite alleles because of MDA effects to drought tolerance. We obtained the typical materials that carried elite alleles were obtained (S5 Table). Finally, total of 23 positive and 10 negative AAEs were identified.

Discussion

Drought tolerance is a complex trait and difficult to study for it's more affected. We performed the water stress experiments by using 100-day old plants of cassava germplasms for their suitable size of plants, being convenient to control the water stress condition in the green house and 100-day old also is a key time point for cassava storage root development and enlargement. So, we could not only evaluate the survival rate but also potential of biomass and storage root development and enlargement of cassava under water stress. Besides, it is valid for multiple crops that water stress experiment is performed in their seedling stage [39–45].

Molecular diversity, water depressed experiment and association mapping in our study were based on 134 cassava genotypes. Although 134 genotypes are too little for a large scale GWAS study, it was difficult to water stress experiment of cassava, a big plant crops, in greenhouse and our researches should be considered as a candidate association mapping strategy.

Besides, results of association mapping in our study were valid which has been verified by drought-related transcriptomics and proteomics of cassava and qRT-PCR and many other direct or indirect evidences.

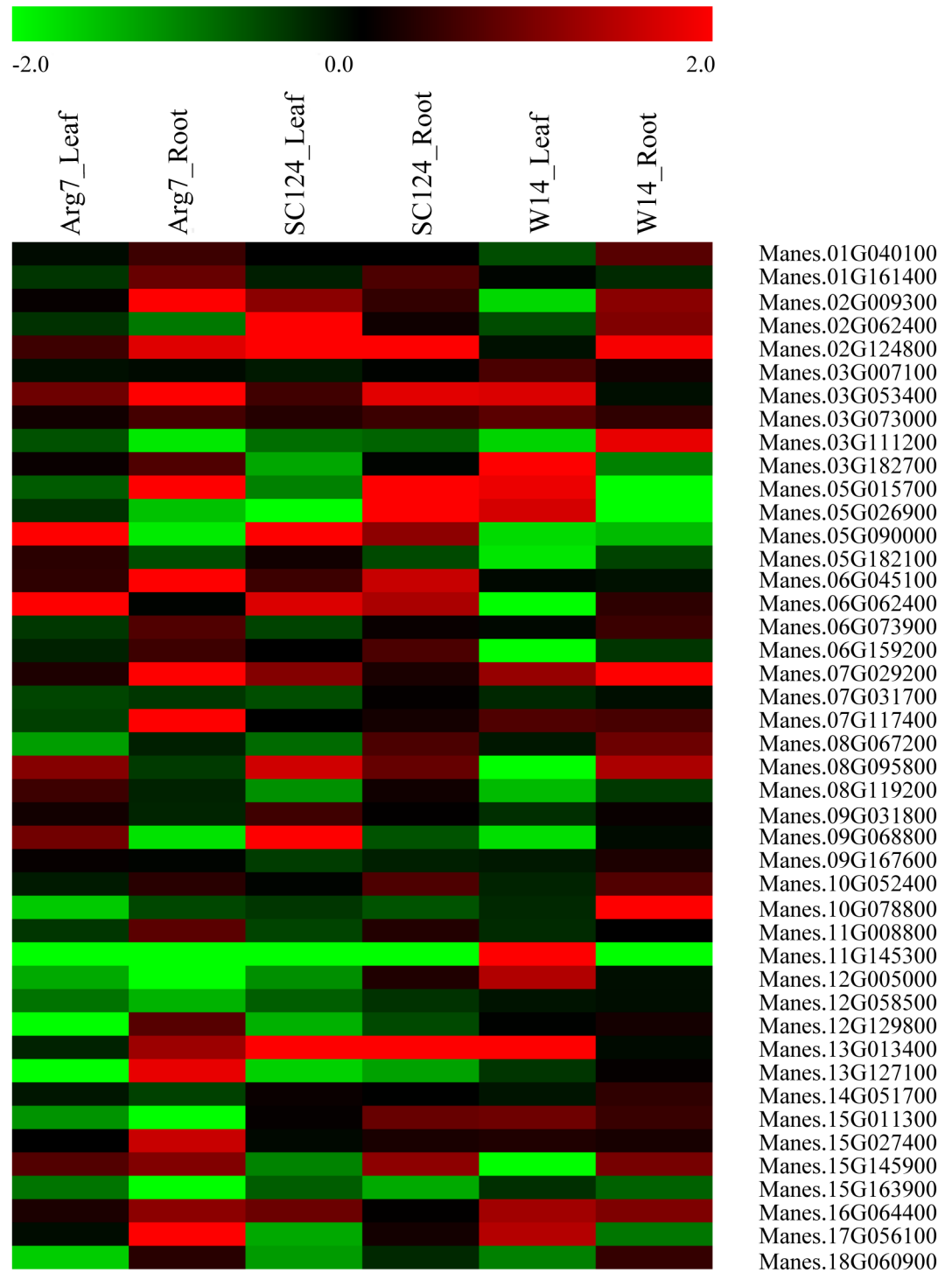


Fig 4. Expression profiles of associated genes in cassava. Log₂ based FPKM change fold values drought stress plants by respective control value were used to create the heat map. The differential expression thresholds of significant up and down regulation were set to 2.0 and -2.0 respectively.

<https://doi.org/10.1371/journal.pone.0177456.g004>

Table 2. Significant differential expression of associated genes in cassava leaves in the iTRAQ-based proteomics database reported by previous (Zhao et al. 2015).

Marker	Cassava genome version 6.1	Cassava genome version 4.1	Arg-L1/Arg-L0	SC-L1/SC-L0	Associated Traits	Description
EME171	Manes.13G013400	cassava4.1_015024m	0.347±0.13	0.414 ±0.18	CAT	SMALL HEAT-SHOCK PROTEIN HSP20 FAMILY
EME254	Manes.08G095800	cassava4.1_013978m	0.944±0.97	0.221 ±0.05	PRO,TDMP	Ferritin
EME710	Manes.10G078800	cassava4.1_011832m	2.000±4.11	3.180 ±10.31	CAT, POD, TN	Caffeoyl-CoA O-methyltransferase
MeESSR46	Manes.05G090000	cassava4.1_018205m	4.067 ±18.94	7.510 ±57.53	SOD, SRS, TDMP	MLP-LIKE PROTEIN 423-RELATED

Note: Values are expressed as the average fold change in protein abundance between stressed (L1) and control (L0) leaves of Arg7 and SC124 cassava cultivars identified in the three experimental replicates. Significance thresholds were set at 1.67 and 0.6 for up and down regulation, respectively.

<https://doi.org/10.1371/journal.pone.0177456.t002>

Genetic diversity and population structure

All 107 EST-SSR markers used in this study were localized in all 18 cassava chromosomes that had potential to analysis of genetic diversity and population structure. Compared with previous studies on genetic diversity of cassava germplasms, the number of alleles per locus in our study was lower than that previously reported for 42 cassava land races in Brazil [46]. This might be due to the EST-SSR markers used in our study, that is, 55 previously reported EST-SSRs [5] and 52 EST-SSRs which were developed against unique polymorphism [8], were naturally lower abundant polymorphisms SSRs than genomic SSRs. Genetic coefficients between every two accessions were very low on average with a wide range. These indicated that genetic relationships between these germplasms were very complex and should due to the intricacy of their original areas and breeding history. However, AMOVA of germplasms showed that sub-populations, dependent on geographical origin, were not significantly different and had a strong gene flow among them. That might a consequence of the cassava introduction in breeding projects.

Results of population structure analysis in our study showed that all 134 germplasms could be divided into three clusters which we evaluated by AMOVA, finding significant differences and a strong gene flow. However, there were not obviously relationship between clusters and origin area. While the germplasms themselves derived from different areas of the world, most of them came from China, while east-south Asia and Africa, two of major cassava producing regions, and were rare in these accessions. Moreover, cassava land races from China represents mostly outbred offspring derived from parent plants from other countries.

The population differential coefficients were lower than those of the 283 cassava samples calculated by Fregene et al. [47] and that of 36 African cassava land races reported by Raji et al. [48]. This indicates that the 134 cassava germplasms used in our study are suitable for association mapping based on linkage disequilibrium because of that their population differentiation was not very strong as well as abundant genetic and phenotypic diversity.

Marker-trait association mapping

In our study, fifty-three EST-SSRs located on 17 chromosomes and accounted for 54.08% of 98 markers. In order to understand the EST-SSRs used in our study, we obtained the reference EST sequences and genebank EST library (S6 Table). Nine EST libraries contained all 107 EST-SSRs and three of them sequenced by previous reports aiming to identify the different water or abiotic stress tolerance related genes in cassava cultivars [49–51]. 96 EST-SSRs

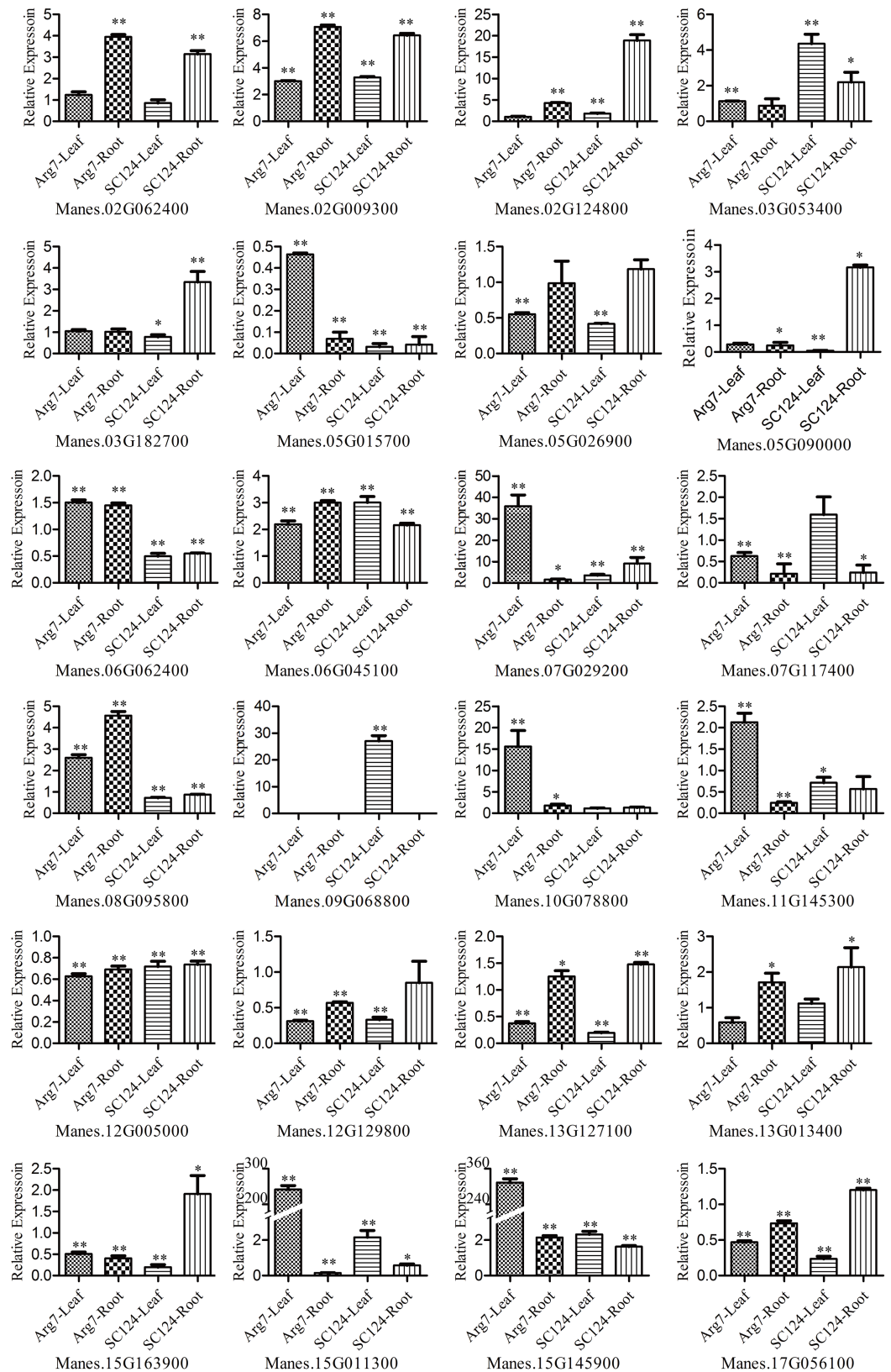


Fig 5. Relative expression verification of 24 selected associated genes by qRT-PCR. All of the 24 genes were selected according to the significantly differential expression from omics data. * and ** standard for

significance of differential expression $P < 0.05$ and $P < 0.01$ respectively which were estimated from relative expression between drought and control treatments by *T* test. Manes.09G068800 was only expressed in the leaf of SC124 under drought and control conditions.

<https://doi.org/10.1371/journal.pone.0177456.g005>

derived from these three libraries, accounting for 89.7% of 107 markers. This allowed us to map functional genes for drought tolerance by association mapping and this might be the reason why so many loci were associated with drought-related traits in our study.

Just as reported in other crops before, we found markers for repeat, pleiotropy markers and one-to-one phenotype-marker association combinations in this study [23, 52–53]. Twenty-two major marker-trait combinations ($R^2 > 10\%$) were explored, accounting for 25.29% of all combinations. Seventeen markers associated with multiple phenotypes in one year. Moreover, we found thirteen pleiotropic loci in two years that were not only associated with biomass and storage root related traits of 100-day old plants, but also with drought tolerance. All of these genes are potential candidate genes for drought tolerance and high yield breeding in cassava.

Due to few related reports in cassava, it was difficult for us to compare our results with previous researches. Instead, we compared our results to cassava drought related transcriptomics and proteinomics and performed qRT-PCR to verify differential expression genes. Markers associated mean that genes they co-localized with the map or their neighboring chromosome regions (within covering \pm LD decay distance) regulated the development of phenotypes they were associated to, meanwhile alleles of actual functional genes had different phenotypic effects. However, it was not to say any EST-SSR and enzymes/molecules co-localized with the map location of identified candidate genes. Expression analysis and functional annotation were two effective methods of the real functional candidate gene discovering for associated region.

Some associated genes or their orthologues or paralogues have similarly functions in cassava or other model plants. MeSSR50 associated with SRS content in leaf in 2014 in our study. According to previous reports [20, 54], 14-3-3 protein family might play roles in cassava storage root enlargement stage and be related to the starch and sugar metabolic and drought responsive. That indicated that our results were supported by previous reports. Another EST-SSR marker, MeSSR18, associated with proline content. The marker which located in the gene annotated as *MeGSTU7*, with gene code ID Manes.07G029200, whose expression rose after drought treatment for 12 days. The expression of one of its paralogous genes improved after by drought stress and recovered to normal after cassava plants were re-watered. A previous report linked *MeGSTU7* to AGFW [55]. Although results of our study and previous reports [55] were not completely consistent, the expression pattern of the two member of the *GSTU7* gene family in cassava were similarly under drought condition. Furthermore, *Arabidopsis* orthologues of associated genes discovered in our study played roles in water stress or abiotic stress resistance, such as H(+)-translocating (pyrophosphate-energized) inorganic pyrophosphatase (AtVHP1)[56], heat shock protein (AtHSP21)[57], homobox protein 16 (HB16)[58], decapping 5 (DCP5)[59]. These supported loci had potential to be considered as candidate genes for further functional verification and tag markers in MAS project for cassava drought tolerance and yield breeding.

Discovery of elite alleles

In order to explore the elite alleles in associated loci, we estimated AEEs of 24 associated loci whose expression changed under drought condition, indicating that they might be functional genes (S5 Table). POD, SOD, CAT, Proline and SRS were positive factors and MDA as negative on the basis of their contribution to drought resistance. We defined elite alleles as alleles

Table 3. Functional annotation of twelve differential expression genes supported by previous omics data and qRT-PCR.

Marker	Gene ID#	Tissue ^a	Associated Traits	R ²	AAE ^b	Description	GO annotation	Enzyme	Pathway
MeESSR36	Manes.02G009300	Leaf, Root	AGFW-2015 MDA-R-2014	7.90 3.84	positive negative	NAD DEPENDENT EPIMERASE/DEHYDRATASE	F:UDP-glucose 4-epimerase activity P:galactose metabolic process	EC:5.1.3.2	map00052 Galactose metabolism; map00520. Amino sugar and nucleotide sugar metabolism
EME628	Manes.09G068800	Leaf	SRN-2014	5.28	positive	AQUAPORIN PIP2-1-RELATED	C:integral component of plasma membrane P:water transport C:plasmodesma F:glycerol channel activity P:cellular water homeostasis P:glycerol transport P:ion transmembrane transport F:water channel activity C:vacuole		
EME710	Manes.10G078800	Leaf, Root	CAT-L-2014 POD-L-2014 SRN-2014	16.75 5.23 5.55	positive positive positive	Caffeoyl-CoA O-methyltransferase	P:coumarin biosynthetic process F:caffeoyl-CoA O-methyltransferase activity F:metal ion binding P:lignin biosynthetic process P:methylation	EC:2.1.1.104	map00940. Phenylpropanoid biosynthesis: map00360, Phenylalanine metabolism; map00945, Stilbenoid, diarylheptanoid and gingerol biosynthesis: map00941, Flavonoid biosynthesis
EME425	Manes.12G005000	Leaf, Root	AGFW-2015 SOD-R-2015	16.64 12.11	positive positive	PROTEIN PHOSPHATASE 2C 33-RELATED	F:protein serine/threonine phosphatase activity P:protein dephosphorylation		
MeESSR74	Manes.12G129800	Leaf, Root	Proline-L-2015	9.13	positive	HOMEBOX-LEUCINE ZIPPER PROTEIN ATHB-16-RELATED	C:nucleus F:transcription factor activity sequence-specific DNA binding F:transaminase activity P:regulation of transcription DNA-templated F:sequence-specific DNA binding		
MeESSR31	Manes.13G127100	Leaf, Root	Proline-L-2014	8.31	positive	CYCLIN-DEPENDENT KINASE INHIBITOR 6	P:cell cycle arrest C:nucleus F:cyclin-dependent protein serine/threonine kinase inhibitor activity P:negative regulation of protein serine/threonine kinase activity		
MeSSR38	Manes.02G124800	Leaf, Root	CAT-R-2014	3.37	negative	17.6 KDA CLASS I HEAT SHOCK PROTEIN 1-RELATED	P:response to oxidative stress, response to heat C:cytoplasm F:protein binding, protein self-association		map04141:Protein processing in endoplasmic reticulum
MeSSR44	Manes.03G053400	Leaf, Root	AGFW-2015	6.95	positive	Cation transport ATPase	P:copper ion transport, metal ion transport C:plasma membrane, chloroplast thylakoid membrane, integral component of membrane F:integral component of membrane F:nucleotide binding, copper-exporting ATPase activity, copper ion transmembrane transporter activity, ATP binding, ATPase activity, coupled to transmembrane movement of ions, phosphorylative mechanism, cation-transporting ATPase activity, metal ion binding	EC:3.6.3.4	
MeESSR18	Manes.07G029200	Leaf, Root	Proline-L-2015	5.28	negative	GLUTATHIONE S-TRANSFERASE, GST, SUPERFAMILY, GST DOMAIN CONTAINING	P:glutathione metabolic process, toxin catabolic process, response to toxic substance, response to salicylic acid C:cytoplasm, cytosol F:glutathione transferase activity	EC:2.5.1.18	map00480:Glutathione metabolism

(Continued)

Table 3. (Continued)

Marker	Gene ID#	Tissue ^a	Associated Traits	R ²	AAE ^b	Description	GO annotation	Enzyme	Pathway
EME171	Manes.13G013400	Root	CAT-R-2015	5.34	negative	SMALL HEAT-SHOCK PROTEIN HSP20 FAMILY	P: response to heat, response to high light intensity, response to hydrogen peroxide C: chloroplast		map04141:Protein processing in endoplasmic reticulum
EME212	Manes.02G062400	Root	SOD-L-2015 SRFW-2014	15.2 6.21	positive positive	Domain of unknown function			
MeESSR71	Manes.06G045100	Leaf, Root	CAT-L-2014 CAT-L-2015 SRDW-2014 SRFW-2014	5.77 10.47 6.91 10.99	positive positive positive positive	Inorganic diphosphatase / Pyrophosphate phosphohydrolase	P: response to water deprivation, response to salt stress, auxin polar transport, establishment or maintenance of transmembrane electrochemical gradient, proton transport, leaf development C: Plant-type vacuole, mitochondrion, vacuole, vacuolar membrane, golgi apparatus, cytosol, plasma membrane, chloroplast, plant-type vacuole membrane, chloroplast envelope, endosome membrane, membrane, integral component of membrane F: inorganic diphosphatase activity, hydrogen-translocating pyrophosphatase activity, metal ion binding	EC:3.6.1.1	map00190:Oxidative phosphorylation

^a was the differential expressed tissue under drought stress.

^b was the AAEs of EST-SSR loci of associated genes for traits which they were associated to.

<https://doi.org/10.1371/journal.pone.0177456.t003>

positively influencing drought tolerance and high yield breeding, therefore loci associated with POD, SOD, CAT, Proline, AGFW and storage root-related traits and a negative association with MDA. We found three elite alleles as for MDA content and two associated loci were negative effect according to their AAE values. All phenotypes and AAEs of these could be useful for molecular assisted selection and pyramiding breeding in the future.

These alleles, whose phenotypes comprising multiple traits were consistent with comprehensive breeding targets for drought tolerance and high yield breeding in pleiotropic loci, might be valuable for future research. For example, EME710 associated with SRN, CAT and POD activity, with positive AAEs for three traits. Moreover, the allele B in EME710 was an elite allele according to our definition. There were another two loci like EME710 of all 9 casual pleiotropic loci. Another 4 loci did not show elite allele characteristics for all associated phenotypes. Among them, MeSSR36 was associated AGFW and MDA content. As mentioned above, MDA content is negative factor for drought tolerance. Therefore, allele B in MeSSR36 was the allele to consider for drought tolerance and high yield breeding because of its positive effect on AGFW content and negative on MDA content. All of these pleiotropic variations have a potential application for comprehensive breeding.

Cross-talk about associated mapping and expression analysis

Compared to associated genes and omics data and results of qRT-PCR by using independent drought stress experiment, 24 differential expression genes were obtained under water stress. Twelve of them selected for their consensus in multiple independent experiments would be actual functional genes of cassava drought resistance. Furthermore, amino acid sequences of them were used to functional annotation. And we tried to explain part of cassava drought resistance mechanisms revealed by these functional genes.

As was shown in Table 3, Manes.02G009300 (co-localized with MeSSR36) was annotated to UGE5 protein (EC: 5.1.3.2) and might regulate the galactose metabolism (map00052) and amino sugar and nucleotide sugar metabolism (map00520). Homolog genes of it in *Arabidopsis thaliana* were reported to regulate cell wall biosynthesis. *AtUGE2* and *AtUGE4* were synergistic to regulate galactose content of cell wall which was correlated to shoot growth. *AtUGE5* was slightly cooperated with *AtUGE4* to regulate the root growth and galactose content which might be induced by stress [60]. In our study, *MeUGE5* was associated with DTCs of AGFW and MDA and was significantly up-regulated expression. So, we indicated that *MeUGE5* might promote shoot growth and enhance the cell membrane stability. Besides, *MeUGE5* was a positive effect locus for DTC of AGFW but negative one for DTC of MDA content according to AAE. These might support our speculation.

Manes.10G07880 (co-localized with EME710) which was annotated to Caffeoyl-CoA O-methyltransferase (EC: 2.1.1.104) might regulate multiple metabolism pathways and flavonoid biosynthesis pathway was one of them. Flavonoid biosynthesis pathway was considered as one of important active oxygen elimination mechanisms under stress condition [61]. In our study, it was associated with DTCs of CAT, POD and SRN with positive effects and up-regulated under drought stress. All of the above indicated that Manes.10G078800 might participate in the active oxygen elimination during cassava drought resistance process.

The third up-regulated gene Manes.09G068800 (co-localized with EME628) was annotated to AQUAPORIN PIP2-1-RELATED which was intrinsic protein of cell membrane and had glycerol channel activity and water channel activity. Its homologous gene in *Arabidopsis* played role in water absorption of root. And in our study, its expression patterns under water stress agreed with its possible functions. It was indicated that up-regulation of *MePIP2-1* might enhance the water up-take ability of cassava root and have positive effects on DTC of SRN and

cassava drought resistance. But it was still unknown for us how *MePIP2-1* regulates DTCs of SRN of cassava under drought stress.

Besides the three genes, another two genes, Manes.03G053400 and Manes.06G045100 which were annotated as positive regulators [62, 63] were up-regulated under drought stress with positive AAEs (Table 3). However, there were 3 genes, Manes.02G124800, Manes.07G029200 and Manes.13G013400, which were annotated as positive regulators [55, 64–65] for drought resistance were up-regulated under water deprived condition with negative AAEs for their associated traits (Table 3). That was different from our knowledge and might due to functional alleles and complex relationships of their associated traits and pathways these genes directly participated in.

There were three down-regulated expression genes under water stress, Manes.12G005000 (co-localized with EME425), Manes.12G129800 (co-localized with MeSSR74) and Manes.13G127100 (co-localized with MeSSR31), that were annotated to *MePP2C33*, *MeHB16* and *MeKRP6* respectively. PP2C proteins were negative regulators of ABA pathway and stress-induced MAPK signaling pathway and they also involved in regulation the cell cycle [66, 67]. *MeHB16* was a member of the HD-ZIP super-family and its homologous gene *AtHB16* acted as a negative regulator of cell expansion, leaf growth and photoperiod of *Arabidopsis* [68]. As was shown in Fig 5, *MePP2C33* and *MeHB16* were down-regulated in Arg7 root under drought stress while not significant changed in SC124. Besides, *MePP2C33* was associated with DTC of AGFW with positive effect of AAE. That might account for the continuous growth of Arg7 under drought condition reported previously [20]. As results of GO annotations, *MeKRP6* might participate in cell cycle arrest and negative regulation of protein serine/threonine kinase activity process. The energy sensor *AtSnRK1* plays a cardinal role in the control of cell proliferation in *A. thaliana* plants through the inhibition of *AtKRP6* biological function by phosphorylation [69]. However, *MeKRP6* was down-regulated in leaf tissue of SC124 under drought stress in our results which could be explained for its function and expression pattern but not for the difference between cassava cultivars. All of these three negative regulators were down-regulated under water depressing condition which indicated that these genes might regulate the continuous growing ability of cassava plant under drought stress. Differential expression patterns between *MeKRP6* and the other two genes in cassava cultivars might be related to the different strategies of cultivars for drought resistance. As was shown in Table 3, AAEs of these negative regulators were positive. That might be for their expression patterns under drought stress or loss of function in natural alleles of these negative genes. However, it was difficult to explain the mechanisms based on evidences at present that these down-regulated genes were associated with phenotypes under drought stress but it was no doubt that they were actual functional genes under drought stress.

Conclusion

In conclusion, all of these 134 cassava accessions showed abundant genetic and phenotypic diversity and a slight sub-population differentiation. The results of association mapping were supported by reverse genetics evidences as well as previous study, allowing selecting elite alleles of the associated loci. These results will make possible to choose cassava resources for drought tolerance and to provide plant materials, candidate genes and reliable experimental support for further research in the mechanism behind cassava drought tolerance.

Supporting information

S1 Fig. Mapping of the 98 EST-SSRs used in this study on cassava chromosomes.
(TIF)

S1 Table. The pedigree of all the 134 cassava accessions.

(XLSX)

S2 Table. Information of primers used in qRT-PCR.

(XLSX)

S3 Table. Summary statistics of drought tolerance coefficient of phenotypes in water stress experiments.

(XLSX)

S4 Table. Results of marker-trait association mapping in cassava germplasm resources ($P_{FDR} < 0.05$).

(XLSX)

S5 Table. Phenotypic effect and their typical materials of alleles and AAE of locus in 24 reliable associated loci.

(XLSX)

S6 Table. Reference EST sequences and genebank EST libraries 106 cassava EST-SSRs derived except one marker.

(XLSX)

Acknowledgments

This work was financially supported by National Natural Science Foundation of China NSFC-CGIAR Project (grant no. 31561143012), the National Science Foundation of China (grant No. 31501378), the national key technology R&D program of China (grant no. 2015BAD15B01) and the Hainan province innovative research team foundation (grant no. 2016CXTD013).

Author Contributions

Conceptualization: MP.

Data curation: WQW.

Formal analysis: BW.

Funding acquisition: MP MBR XLY.

Investigation: BW XG PJZ MBR XLY LPZ YLY XL DLD JXX YWX CJH XW XLW.

Methodology: BW.

Project administration: MP.

Resources: WQW.

Software: BW.

Supervision: MP.

Validation: BW XG.

Visualization: BW.

Writing – original draft: BW MP MBR.

Writing – review & editing: BW MP MBR.

References

1. Howeler RN, Lutaladio, Thomas G. Save and Grow: Cassava, Food and Agriculture Organization of the United Nations, Rome, 2013
2. Okogbenin E, Setter TL, Ferguson M, Mutegi R, Ceballos H, Olasanmi B et al. Phenotypic approaches to drought in cassava: review. *Front Phys*. 2013; 4:93
3. Gupta PK, Rustgi S, Mir RR. Array-based high-throughput DNA markers for crop improvement. *Heredity*. 2008; 101(1):5–18 <https://doi.org/10.1038/hdy.2008.35> PMID: 18461083
4. Haq SU, Kumar P, Singh RK, Verma KS, Bhatt R, Sharma M et al. Assessment of functional EST-SSR markers (Sugarcane) in cross-species transferability, genetic diversity among poaceae plants, and bulk segregation analysis. *Genet Res Int*. 2016
5. Kunkeaw S, Sraphet TYS, Boonchanawiwat A, Lightfoot OBDA, Triwitayakorn K, Tangphatsornruang S. Construction of a genetic linkage map using simple sequence repeat markers from expressed sequence tags for cassava (*Manihot esculenta* Crantz). *Mol Breeding*. 2011; 27:67–75
6. Chen X, Fu YH, Xia ZQ, Jie L, Wang HY, Lu C et al. Analysis of QTL for yield-related traits in cassava using an F1 population from non-inbred parents. *Euphytica*. 2012; 187(2):227–234
7. Rabbi IY, Kulembeka HP, Masumba E, Marri PR, Ferguson M. An EST-derived SNP and SSR genetic linkage map of cassava (*Manihot esculenta* Crantz). *Thero Appl Genet*. 2012; 125(2):329–342
8. Raji AAJ, Anderson JV, Kolade OA, Ugwu CD, Dixon AGO, Ingelbrecht IL. Gene-based microsatellites for cassava (*Manihot esculenta* Crantz): prevalence, polymorphisms and cross-taxa utility. *BMC Plant Bio*. 2009a; 9:118
9. Ferguson ME, Hearne SJ, Close TJ, Wanamaker S, Moskal WA, Town CD et al. Identification, validation and high-throughput genotyping of transcribed gene SNPs in cassava. *Thero App Genet*, 2012; 124:685–695
10. Sakurai T, Mochida K, Yoshida T, Akiyama K, Ishitani M, Seki M et al. Genome-wide discovery and information resource development of DNA polymorphisms in cassava. *Plos One*. 2013; 8(9):e74056 <https://doi.org/10.1371/journal.pone.0074056> PMID: 24040164
11. Soto JC, Ortiz JF, Perlaza-Jiménez L, Vásquez AX, Lopez-Lavalle LAB, Mathew B et al. A genetic map of cassava (*Manihot esculenta* Crantz) with integrated physical mapping of immunity related genes. *BMC Genomics*. 2015; 16:190 <https://doi.org/10.1186/s12864-015-1397-4> PMID: 25887443
12. Pootakham W, Shearman JR, Ruang-areerate P, Sonthirod C, Sangsrakru D, Jomchai N et al. Large-scale SNP discovery through RNA sequencing and SNP genotyping by targeted enrichment sequencing in Cassava (*Manihot esculenta* Crantz). *Plos One*. 2014; 9(12):e116028 <https://doi.org/10.1371/journal.pone.0116028> PMID: 25551642
13. Bredson JV, Lyons JB, Prochnik SE, Wu GA, Ha CM, Edsinger-Gonzales E et al. Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nat Biotech*. 2016; 34:562–570
14. Kizito EB, Rönnberg-Wästljung, Egwang T, Gullberg U, Fregene M, Westerbergh A. Quantitative trait loci controlling cyanogenic glucoside and dry matter content in cassava (*Manihot esculenta* Crantz) roots. *Hereditas*. 2007; 144:129–136 <https://doi.org/10.1111/j.2007.0018-0661.01975.x> PMID: 17850597
15. Thanyasiriwat T, Sraphet S, Whankaew S, Boonseng O, Bao J, Lightfoot DA et al. Quantitative trait loci and candidate genes associated with starch pasting viscosity characteristics in cassava (*Manihot esculenta* Crantz). *Plant Bio*. 2014; 16:197–207
16. Whankaew S, Poopear S, Kanjanawattanawong S, Tangphatsornruang S, Boonseng O, Lightfoot DA et al. A genome scan for quantitative trait loci affecting cyanogenic potential of cassava root in an outbred population. *BMC Genomics*. 2011; 12:266 <https://doi.org/10.1186/1471-2164-12-266> PMID: 21609492
17. Akano AO, Dixon AGO, Barrera CME, Fregene M. Genetic mapping of a dominant gene conferring resistance to cassava mosaic disease. *Theor Appl Genet*. 2002; 105:521–525 <https://doi.org/10.1007/s00122-002-0891-7> PMID: 12582500
18. López C, Quesada-Ocampo LM, Bohorquez A, Duque MC, Vargas J, Tohme J et al. Mapping EST-derived SSRs and ESTs involved in resistance to bacterial blight in *Manihot esculenta*. *Genome*. 2007; 50(12):1078–1088 <https://doi.org/10.1139/G07-087> PMID: 18059536
19. Ashraf M, Athar HR, Harris PJC, Kwon TR. Some prospective strategies for improving crop salt tolerance. *Adv Agron*. 2008; 97:45–110

20. Zhao PJ, Liu P, Shao JF, Li CQ, Wang B, Guo X et al. Analysis of different strategies adapted by two cassava cultivars in response to drought stress: ensuring survival or continuing growth. *J Exp Bot.* 2015;
21. Bousslama M, Schapaugh WT Jr. Stress tolerance in soybeans. I. Evaluation of three screening techniques for heat and drought tolerance. *Crop Sci.* 1984; 24(5): 933–937
22. Lin ZX, He DH, Zhang XL. Linkage map construction and mapping QTL for cotton fiber quality using SRAP, SSR and RAPD. *Plant Breeding.* 2005; 124:180–187
23. Wang B, Nie YC, Lin ZX, Zhang XL, Liu JJ, Bai J. Molecular diversity, genomic constitution, and QTL mapping of fiber quality by mapped SSRs in introgression lines derived from *Gossypium hirsutum* × *G. darwinii* Watt. *Thero Appl Genet.* 2012; 125:1263–1274
24. Liu K, Muse SV. Powermarker: integrated analysis environment for genetic marker data. *Bioinformatics.* 2005; 21:2128–2129 <https://doi.org/10.1093/bioinformatics/bti282> PMID: 15705655
25. Adams DJ, Rolf FJ. Ecological character displacement in *Plethodon*: biomechanical differences found from a geometric study. *Proc Natl Acad Sci USA.* 2000; 97:4106–4111 PMID: 10760280
26. Peakall R, Smouse PE. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes.* 2006; 6:288–295
27. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000a; 155:945–959
28. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. Association mapping in structured populations. *Am J Hum Genet.* 2000b; 67:170–181
29. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Eco.* 2005; 14:2611–2620
30. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics.* 2007; 23(14):1801–1806 <https://doi.org/10.1093/bioinformatics/btm233> PMID: 17485429
31. Yang XH, Gao SB, Xu ST, Zhang ZX, Prasanna BM, Li L, Li JS, Yan JB. Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. *Mol Breeding.* 2011; 28: 511–526
32. Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet.* 2006; 38:203–208 <https://doi.org/10.1038/ng1702> PMID: 16380716
33. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser C Appl Stat.* 1995; 57(1):289–300
34. Voorrips RE. MapChart: Software for the graphical presentation of linkage maps and QTLs. *J Heredity.* 2002; 93(1):77–78
35. Wei YX, Shi HT, Xia ZQ, Tie WW, Ding ZH, Yan Y et al. Genome-wide identification and expression analysis of the WRKY gene family in cassava. *Front Plant Sci.* 2016; 7:25 <https://doi.org/10.3389/fpls.2016.00025> PMID: 26904033
36. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N et al. TM4: a free, open-source system for microarray data management and analysis. *Biotech.* 2003; 34(2):374–8
37. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta$ CT method. *Methods.* 2001; 25:402–408 <https://doi.org/10.1006/meth.2001.1262> PMID: 11846609
38. Wen ZX, Zhao TJ, Zheng YZ, Liu SH, Wang CE, Wang F et al. Association analysis of agronomic and quality traits with SSR markers in Glycine max and Glycine soja in China: II. Exploration of elite alleles. *Acta Agro Sinica.* 2008; 34(8):1339–1349
39. Qie LF, Jia GQ, Zhang WY, Schnable J, Shang ZL, Li W et al. Mapping of quantitative trait locus (QTLs) that contribute to germination and early seedling drought tolerance in the interspecific cross *Setaria italica* × *Setaria viridis*. *Plos One.* 2014; 9(7):e101868 <https://doi.org/10.1371/journal.pone.0101868> PMID: 25033201
40. Gavia JL, Engelbrecht BMJ. Effects of drought, pest pressure and light availability on seedling establishment and growth: their role for distribution of tree species across a tropical rainfall gradient. *Plos One.* 2015; 10(11):e0143955 <https://doi.org/10.1371/journal.pone.0143955> PMID: 26619138
41. Zhou Y, Yang P, Cui FL, Zhang FT, Luo XD, Xie JK. Transcriptome analysis of salt stress responsiveness in the seedlings of Dongxiang wild rics (*Oryza rufipogon* Griff.). *Plos One.* 2016; 11(1):e0146242 <https://doi.org/10.1371/journal.pone.0146242> PMID: 26752408
42. Tomar RSS, Tiwari S, Vinod, Naik BK, Chand S, Deshmukh R et al. Molecular and morpho-agronomical characterization of root architecture at seedling and reproductive stages for drought tolerance in wheat. *Plos One.* 2016; 11(6):e0156528 <https://doi.org/10.1371/journal.pone.0156528> PMID: 27280445

43. Min HW, Chen CX, Wei SW, Shang XL, Sun MY, Xia R et al. Identification of drought tolerant mechanisms in maize seedlings based on transcriptome analysis of recombination inbred lines. *Front Plant Sci.* 2016; 7:1080 <https://doi.org/10.3389/fpls.2016.01080> PMID: 27507977
44. Wang XL, Wang HW, Liu SX, Ferjani A, Li JS, Yan JB et al. Genetic variation in ZmVPP1 contributes to drought tolerance in maize seedlings. *Nat Genet.* 2016; 48:1233–1241 <https://doi.org/10.1038/ng.3636> PMID: 27526320
45. Wang XB, Wang LF, Shangguan ZP. Leaf gas exchange and fluorescence of two winter wheat varieties in response to drought stress and nitrogen supply. *Plos One.* 2016; 11(11):e0165733 <https://doi.org/10.1371/journal.pone.0165733> PMID: 27802318
46. Siqueira MVB, Queiroz-Silva JR, Bressan EA, Borges A, Pereira KJC, Pinto JG et al. Genetic characterization of cassava (*Manihot esculenta*) landraces in Brazil assessed with simple sequence repeats. *Genet Mol Bio.* 2009; 32(1):104–110
47. Fregene MA, Suarez M, Mkumbira J, Kulembeka H, Ndedya E, Kulaya A et al. Simple sequence repeat marker diversity in cassava landraces: genetic diversity and differentiation in an asexually propagated crop. *Thero App Genet.* 2003; 107:1083–1093
48. Raji AAJ, Fawole I, Gedil M, Dixon AGO. Genetic differentiation analysis of African cassava (*Manihot esculenta*) landraces and elite germplasm using amplified fragment length polymorphism and simple sequence repeat markers. *Ann Appl Biol.* 2009b; 155:187–199
49. Bonaldo MF, Lennon G, Soares MB. Normalization and subtraction: two approaches to facilitate gene discovery. *Genome Res.* 1996; 6(9): 791–806 PMID: 8889548
50. Lokko Y, Anderson JV, Rudd S, Raji A, Horvath D, Mikel MA et al. Characterization of an 18,166 EST dataset for cassava (*Manihot esculenta* Crantz) enriched for drought-responsive genes. *Plant Cell Rep.* 2001; 26:1605–1618
51. Sakurai T, Plata G, Rodríguez-Zapata, Seki M, Salcedo A, Toyoda A et al. Sequencing analysis of 20000 full-length cDNA clones from cassava reveals lineage specific expansions in gene families related to stress response. *BMC Plant Bio.* 2007; 7:66
52. Nie XH, Huang C, You CY, Li W, Zhao WX, Shen C et al. Genome-wide SSR-based association mapping for fiber quality in nation-wide upland cotton inbred cultivars in China. *BMC Genomics.* 2016; 17:352 <https://doi.org/10.1186/s12864-016-2662-x> PMID: 27177443
53. Nie XH, Tu JL, Wang B, Zhou XF, Lin ZX. A BIL population derived from *G. hirsutum* and *G. barbadense* provides a resource for cotton genetics and breeding. *Plos One.* 2015; 10(10):e0141064 <https://doi.org/10.1371/journal.pone.0141064> PMID: 26517274
54. Wang XC, Chang LL, Tong Z, Wang DY, Yin Q, Wang D et al. Proteomics profiling reveals carbohydrate metabolic enzymes and 14-3-3 proteins play important roles for starch accumulation. *Sci Rep.* 2016; 6:19643 <https://doi.org/10.1038/srep19643> PMID: 26791570
55. Deng DL, Wang B, Zeng CY, Guo X, Peng M. Natural variation analysis of *MeGSTU7*, one drought responsive gene of Cassava (*Manihot esculenta* Cranz). *Chinese J Trop Crops.* 2015; 36(11):1986–1993
56. Hernández A, Herrera-Palau R, Madroñal JM, Albi T, López-Lluch G, Perez-Castiñeira JR et al. Vacuolar H⁺-pyrophosphatase AVP1 is involved in amine fungicide tolerance in *Arabidopsis thaliana* and provides tridemorph resistance in yeast. *Front Plant Sci.* 2016; 7:85 <https://doi.org/10.3389/fpls.2016.00085> PMID: 26904057
57. Zhong LL, Zhou W, Wang HJ, Ding SH, Lu QT, Wen XG et al. Chloroplast small heat shock protein HSP21 interacts with plastid nucleoid protein pTAC5 and is essential for chloroplast development in *Arabidopsis* under heat stress. *Plant Cell.* 2013; 25:2925–2943 <https://doi.org/10.1105/tpc.113.111229> PMID: 23922206
58. Ding Y, Liu N, Virilouvet L, Riethoven JJ, Fromm M, Avramova Z. Four distinct types of dehydration stress memory genes in *Arabidopsis thaliana*. *BMC Plant Bio.* 2013; 13:229
59. Xu J, Chua MH. Dehydration stress activates *Arabidopsis* MPK6 to signal DCP1 phosphorylation. *EMBO J.* 2012; 31:1975–1984 <https://doi.org/10.1038/emboj.2012.56> PMID: 22407295
60. Rösti J, Barton CJ, Albrecht S, Dupree P, Pauly M, Findlay K et al. UDP-Glucose 4-Epimerase isoforms UGE2 and UGE4 cooperate in providing UDP-Galactose for cell wall biosynthesis and growth of *Arabidopsis thaliana*. *Plant Cell.* 2007; 19:1565–1579 <https://doi.org/10.1105/tpc.106.049619> PMID: 17496119
61. Liu M, Li X, Liu Y, Cao B. Regulation of flavanone 3-hydroxylase gene involved in the flavonoid biosynthesis pathway in response to UV-B radiation and drought stress in the desert plant, *Reaumuria soongorica*. *Plant Physio Biotech.* 2013; 73(6):161–167
62. Khadiikar AS, Yadav UP, Salazar C, Shulaev V, Paez-Valencia J, Pizzio GA et al. Constitutive and companion cell-specific overexpression of AVP1, encoding a proton-pumping pyrophosphatase, enhances biomass accumulation, phloem loading, and long-distance transport. *Plant Physio.* 2016; 170:401–414

63. Sautron E, Mayerhofer H, Glustini C, Pro D, Crouzy S, Ravaud S et al. HMA6 and HMA8 are two chloroplast Cu⁺-ATPases with different enzymatic properties. *Biosci Rep*. 2015; 35:e00201 <https://doi.org/10.1042/BSR20150065> PMID: 26182363
64. Sun WN, Montagu MV, Verbruggen N. Small heat shock proteins and stress tolerance in plants. *Bioch Biophys Acta*. 2002; 1577:1–9
65. Kim DH, Xu ZY, Hwang I. *ATHSP17.8* overexpression in transgenic lettuce gives rise to dehydration and salt stress resistance phenotypes through modulation of ABA-mediated signaling. *Plant Cell Rep*. 2013; 32:1953–1963 <https://doi.org/10.1007/s00299-013-1506-2> PMID: 24081610
66. Schweighofer A, Hirt H, Meskiene I. Plant PP2C phosphatases: emerging functions in stress signaling. *Trends Plant Sci*. 2004; 9(5):236–243 <https://doi.org/10.1016/j.tplants.2004.03.007> PMID: 15130549
67. Rodriguez PL. Protein phosphatase 2C (PP2C) function in higher plants. *Plant Mol Bio*. 1998; 38:919–927
68. Wang Y, Henriksson E, Söderman E, Henriksson KN, Sundberg E, Engström P. The *Arabidopsis* homeobox gene, *ATHB16*, regulates leaf development and the sensitivity to photoperiod in *Arabidopsis*. *Dev Biol*. 2003; 264(1):228–239 PMID: 14623244
69. Guérinier T, Millan L, Crozet P, Qury C, Rey F, Valot B et al. Phosphorylation of p27 (KIP1) homologs KRP6 and 7 by SNF1-related protein kinase01 link plant energy homeostasis and cell proliferation. *Plant J*. 2013; 75 (3):515–525 <https://doi.org/10.1111/tpj.12218> PMID: 23617622