

RESEARCH

Open Access



# Construction of an ultrahigh-density genetic linkage map for *Manihot esculenta* Crantz and identification of QTL for root quantity traits

Qi Liu<sup>1,2</sup>, Zixuan Li<sup>1,2</sup>, Zihao Wang<sup>1,2</sup>, Yanjie Lu<sup>1,2</sup>, Sirong Jiang<sup>1,2</sup>, Chengcai Xia<sup>1,2</sup>, Pengliang An<sup>1,2</sup>, Long Zhao<sup>1</sup>, Ke Deng<sup>3</sup>, Zhiqiang Xia<sup>1,2\*</sup> and Wenquan Wang<sup>1,2\*</sup>

## Abstract

Cassava, *Manihot esculenta* Crantz, is the main raw material used in starch production in China. However, due to the small planting scale and high demand in China, large-scale imports are needed. To improve cassava yield and to meet China's needs, we examine the agronomic traits of root weight, root number, and root length-to-width ratio per plant. By constructing two semi-sibling genetic maps and using years of data for quantitative trait locus (QTL) localization, we compare two population mapping results to screen co-located 15 QTLs, and transcriptome analysis to explore candidate genes related to these traits. We found *OsWRKY78* in rice to be homologous to candidate gene *Manes.03G051300*, which can regulate rice stem elongation and seed size, and *Manes.18G023500* to be homologous to *MeMYB108*, which can reduce leaf shedding and regulate cassava biomass. Through QTL mapping, we identify key genes related to yield traits that can be used in cassava molecular breeding to improve cassava yield.

**Keywords** *Manihot esculenta*, Cassava root, Linkage map, Quantitative trait locus

## Introduction

Cassava is an important energy plant and one of six major global food crops [1]. It is widely grown in tropical regions of Africa, Asia, and Latin America [2]. Characterized by its adaptability, fast growth and high yield, cassava is an important food crop, especially in the tropics and subtropics, and is of high economic and social value because of its ability to grow in less-fertile soils [3] and

because it is not limited by season or moisture. Cassava starch is suitable for producing high-quality modified starch because of its large particles and high transparency, and it is necessary for modified starch in some industries. Cassava tubers are more valuable because they contain large amounts of starch, and small amounts of protein among other nutrients. These tubers can be processed into a variety of food products, used as animal feed, for brewing beer, and for producing alcohol for bioethanol [4]. For some African countries, approximately 25% of energy acquisition is obtained from cassava [5]. However, despite its importance, the average cassava yield in Africa has not significantly increased since 1961.

Quantitative trait locus (QTL) localization is based on genetic linkage analysis, using linkage relationships between molecular markers and target traits to identify

\*Correspondence:

Zhiqiang Xia

zqxia@hainanu.edu.cn

Wenquan Wang

994341@hainanu.edu.cn

<sup>1</sup> School of Breeding and Multiplication (Sanya Institute of Breeding and Multiplication)/Hainan Seed Industry Laboratory, Hainan University, Sanya, China

<sup>2</sup> College of Tropical Agriculture and Forestry, Hainan University, Haikou, China

<sup>3</sup> Academy of Agriculture and Forestry Sciences, Qinghai University, Xining, China



genomic regions associated with those traits [6]. With development of high-throughput sequencing technology the drawing of fine genetic maps for cassava (*Manihot esculenta* Crantz) has considerably progressed, with multiple genetic linkage maps having been constructed using RFLP (Restriction Fragment Length Polymorphism), SSR (simple sequence repeat), and SNP (single nucleotide polymorphism) markers. These maps cover 18 linkage groups, with an average molecular marker distance ranging 4.51–4.81, with linkage map lengths ranging 1707.9–2412.0 centimorgans (cM) [7–9]. Aided by these maps, significant progress has also been made in identifying Quantitative Trait Loci (QTL) for various cassava traits. Boonchanawiwat et al. searched for QTL that affected the height of cassava plants and the height of the first branch, by performing linkage map analysis in self-pollinated progeny [10]. Nang et al. performed QTL analysis on a reciprocal cross population and reported plant and first branch heights to be controlled by multiple genes [11]. Okogbenin et al., investigating the genetic basis of early root bulking in cassava by genetic analysis and QTL localization in a population of non-self-crossing parents, reported early root bulking to be a complex trait controlled by multiple genes [12]. Akinbo et al. identified several QTLs that affected the protein content of cassava roots through QTL analysis in a backcross population [13]. Zou et al. constructed an amplified-fragment single nucleotide polymorphism and methylation (AFSM) QTL map by crossing two non-inbred parents, KU50 and SC124 (KS population), to obtain 186 cassava populations, and identified 260 candidate QTL genes for cold stress and 301 candidate QTL genes for storing root quality and yield [14, 15]. Okogbenin, and Ewa's studies identified several QTLs associated with cassava plant size and yield traits [16–18]. Overall, progress has been made in the study of QTL localization for root number in cassava, but further research is needed to improve understanding of genetic mechanisms controlling root number, and to use this knowledge to improve cassava varieties.

Hyper-seq is a novel, effective and flexible marker-assisted selection method for genotyping that is based on the polymerase chain reaction (PCR). Compared with commonly used whole genome sequencing methods such as restriction site-associated DNA tag sequencing fragments generated by, for example, type IIB restriction endonucleases (2b-RAD) and AFSM, Hyper-seq is more efficient and less costly. Its greatest advantage is that it can complete genome simplification, barcode addition, and simplified genome sequence amplification in one step. A Hyper-seq library is constructed with specially designed primers and DNA (for PCR amplification) or leaf (for direct PCR amplification) templates, without the

need for restriction enzyme digestion and barcode connection steps [19].

In the present research, two half sibling populations were obtained by crossbreeding cassava cultivar SC205(♀) with HB60(♂) and 18R(♂), respectively. To study the genetic mechanisms and key genes affecting cassava yield traits, in order to improve cassava breeding and increase cassava yield. Three traits related to cassava yield were selected for phenotype analysis, and gene typing was performed using target sequence method. The phenotype and genotype data were integrated for association mapping analysis to identify QTLs closely related to the infected area, which were validated through gene expression analysis.

## Materials and methods

### Mapping populations and DNA extraction

We crossed cassava cultivar SC205 with cultivars HB60 and 18R to produce two F1 populations, with 259 offspring from SC205×18R (A1 population) and 233 offspring from SC205×HB60 (A2 population) (Supplementary Table S1). Populations were planted on two plots of land in Chengmai Meiyu Village, Hainan Province, in April of 2021 and 2022. Each planting was replicated twice, with two rows per replicate, and six plants per row. After 6 months, young leaf samples were collected, from which DNA was extracted using the CTAB method, and evaluated using nanodrop [20]. After 10 months of planting, cassava root samples were collected for RNA extraction for transcriptome sequencing.

### Hyper-seq library preparation and sequencing

The DNA concentration of all samples was homogenized to 150–200 ng /μL, and the reaction volume was increased to 20 μL by adding 1 μL of DNA sample, mix, F' primer, and ddH<sub>2</sub>O for PCR amplification reaction; obtain PCR products, purify PCR products using TIANGEN(DP214) reagent kits, perform double end sequencing using MGI sequencer (DNBSEQ-T7), and upload data to BGI raw data were uploaded to BGI.

### Sequence analyses and genotyping

Raw data quality control was performed using FastQC(fastqc -o outputfile inputfile) software [21]. Original data were divided into each sample using Barcode and fastq-multx(fastq-multx -B barcode.txt -m 1 -b itaq.1.fastq itaq.2.fastq -o %R1.fastq -o %R2.fastq) software (<https://github.com/brwnj/fastq-multx>). The AM5608.1 version genome was selected as a reference and BWA [22] (bwa index -a bwts cassava.fa) was used to establish a reference genome index to generate BAM files. Variant calling was performed using SAMtools and BCFtools [23, 24], and high-quality SNPs were filtered

from VCF files using PLINK [25] software ( $-\text{geno } 0.1$   $-\text{maf } 0.05$   $-\text{hwe } 0.0001$ ). SNP marker density was visualized using the CMplot package in R, and variant sites were annotated using SNPEff [26].

### Transcriptome analyses and genotyping

Raw transcriptome data were analyzed using FastQC, followed by adapter removal, base modification, trimming, and filtering of low-quality sequences using Trimmomatic (trimmomatic PE -phred33 input\_forward.fq.gz input\_reverse.fq.gz) [27]. Hisat2 [28] was used to construct a reference genome index of the AM5608.1 version genome and to align paired-end clean reads to the reference genome. FeatureCounts [29] was used to accurately count reads; reads of all samples were then combined and exported as a single RNA-Seq\_Practice\_countstable file. Expression levels of each gene in each sample were quantified using FPKM and TPM calculation formulae. Gene expression levels were obtained for different samples.

### Linkage map construction

A linkage group refers to a set of genes or markers that are located on the same chromosome and are genetically linked to each other [30]. We constructed a genetic linkage map of common parents from the 259 offspring dataset of Group A1 and the 233 offspring dataset of Group A2, and compared the differential loci between parents to obtain the genotype of parent specific variant loci. Firstly, we used VCFtools [31] to extract the shared specific variant loci between offspring and parents. PLINK [24] ( $-\text{bfile myfile} -\text{geno } 0.05$ ) was then used to filter mutation sites, and PLINK ( $-\text{file myfile} -\text{recodeA}$ ) was used to encode the offspring genotype data as 012. Then import the markers into QTL IciMapping software [32] and use it to construct a linkage map. Initially, markers were grouped into linkage groups based on a LOD score threshold of over 3. Within each group, markers were ordered using the nnTwoOpt algorithm, which is a hybrid approach combining the Nearest Neighbor Algorithm and the Two-opt Algorithm [33], and genetic distances were calculated using the Kosambi mapping function [34]. The map was further refined by selecting a window size of 10 markers and adjusting the marker data based on the sum of adjacent recombination frequencies. This adjustment ensured the shortest possible genetic distances between markers. During the refinement process, marker positions were iteratively optimized to minimize map expansion and distortion, thereby enhancing the accuracy and stability of the final linkage map. The final linkage map, including marker names, genetic positions, and linkage group assignments, was exported from

QTL IciMapping. The map was then visualized using the qtl package in R [35].

### Phenotypic assessment analyses

Phenotypic data were measured in the field 10 months after planting, when three plants with similar growth vitality were selected, RW, RN, and RLW of each was measured. Finally, the phenotype field raw data for 2022 and 2023 were obtained. Then, outlier phenotypic data were removed according to the  $3\sigma$  rule (values outside  $\mu - 3\sigma$  and  $\mu + 3\sigma$  intervals are outliers, where  $\mu$  is the sample data mean and  $\sigma$  is the data standard deviation).

After handling the outliers, we choose a linear mixed model approach to analyze all phenotypic traits. The best linear unbiased estimation (BLUE) of each individual was used to perform QTL mapping. In order to obtain joint localization results of different populations with data between single year and multiple years, Model 1 was used to obtain phenotype data of different populations in 2022 and 2023, and Model 2 was used to obtain merged phenotype data of 22–23.

$$Y_{ijl} = \mu + \beta_i + \gamma_j + \epsilon_l + e_{ijl} \quad (1)$$

$$Y_{ijkl} = \mu + \beta_i + \gamma_j + \delta_k + \epsilon_{jl} + e_{ijkl} \quad (2)$$

where  $Y_{ijl}$  is the phenotypic observation related to line  $i$ , location  $j$ , repetition  $l$ ,  $Y_{ijkl}$  is the phenotypic observation related to line  $i$ , location  $j$ , year  $k$  repetition  $l$ ,  $\mu$  is the global intercept.  $\beta_i$  is the fixed effect of the  $i$ -th variety.  $\gamma_j$  is the random effect at position  $j$ .  $\delta_k$  is the random effect of the  $k$ th year.  $\epsilon_l$  is a random effect of  $l$  repeated combinations.  $\epsilon_{jl}$  is a random effect of  $j$  positions and  $l$  repeated combinations.  $e_{ijl}$  and  $e_{ijkl}$  is error term, representing random errors in observations.

Lme4 is an R package that can generally be used to solve mixed models. It supports linear mixed models, generalized linear mixed models, and nonlinear mixed models. We selected lme4 package based on Model 3 to calculate the heritability of different traits in two populations. Spearman's coefficients were used to evaluate correlations between traits, and Shapiro–Wilk tests were used to evaluate normality, kurtosis, and skewness ( $\alpha = 0.05$ ).

$$h^2 = \frac{V_g}{V_g + V_{GL}/L + V_{GY}/Y + V_{GLY}/L * Y + V_e/Y * L * R} \quad (3)$$

where  $V_g$  is the variance component of cultivar,  $V_{GL}$  is the interaction variance component between cultivar and location,  $V_{GY}$  is the variance component between cultivar and year, and  $V_e$  is the residual variance component.  $V_{GLY}$  is the interaction variance component between cultivar and location and repetition, and  $V_e$  is the residual

variance component, Y, L and R are the the number of year, location, and repetition respectively.

### QTL Detection and candidate gene analysis

MapQTL5 software [36] was used for QTL calculation. Linkage maps and phenotype information from 2022 and 2023 were used for Multiple-QTL Mapping (MQM Mapping) to map QTL and estimate their effects [37]. The LOD score for significant QTL was determined through a test analysis (1000 permutations, total error level 5%). Candidate loci with  $\text{LOD} > 2$  were screened, by collating the loci mapped in the two populations; loci within a 50 Kb interval upstream and downstream of the candidate loci that are present in both populations and in three or more data sets were selected as co-located loci. Collation of the co-located loci to determine the co-located QTL, and estimation of the contribution of each identified QTL to the total phenotypic variance through analysis of variance. Loci and QTL were organized based on population and year information (e.g., 22-A1), and the final QTL were determined. The nomenclature for identified QTL involves a trait abbreviation preceded by “q,” the relevant chromosome name, and a numerical identifier (e.g., qRW—1a). A confidence interval of 1 Mb flanking the QTL identification region was used to screen candidate genes.

Candidate genes were extracted from the QTL interval and functionally annotated using annotation information for the cassava AM560.8.1 genome. Important genes were selected for analysis. Gene numbers were compared with annotation files of the AM560v8.1 reference genome to obtain corresponding GO numbers for each gene. Candidate gene numbers and GO numbers were input into WEGO (<http://wego.genomics.org.cn/>) for GO functional annotation, and gene sequences were extracted using SeqKit subseq [38]. The input file was organized and imported into KEGG (<http://kobas.cbi.pku.edu.cn/genelist/>) for pathway enrichment analysis. An enrichment plot was generated using R.

## Results

### SNP marker density and annotation

Approximately 406 Gb of sequencing data were obtained, with an average sequencing depth of 812 Mb per sample (the cassava AM560.8.1 version genome size is 617 Mb). In total, 1,215,893,952 high-quality reads were obtained, with an average of 2,461,324 reads per sample. Obtained two mutation site vcf files for the (2,417,535) and A2 populations (1,376,544). The A1 population had 161,361 unique variant sites shared with the SC205 parent (the A1 maternal parent, abbreviated A1\_SC205) and the A2 population had 121,556 unique variant sites shared with the SC205 parent (the A2 maternal parent, abbreviated

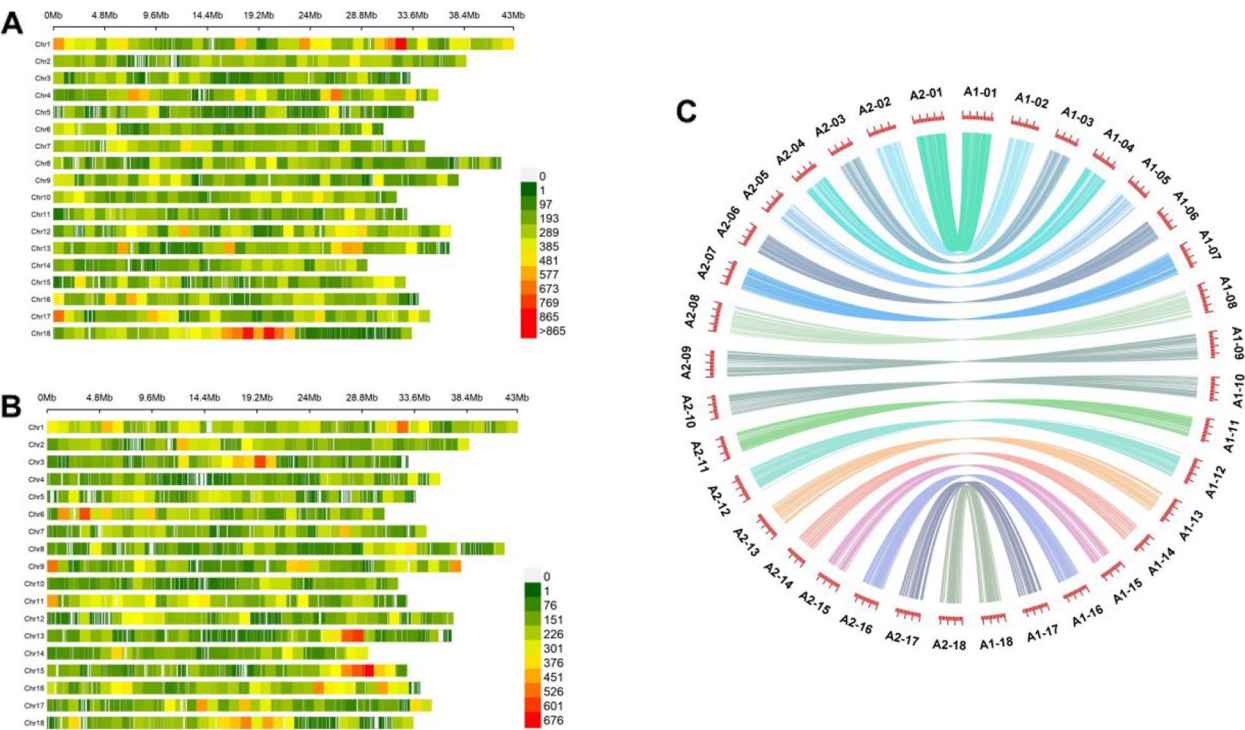
A2\_SC205). An SNP marker density map indicated that variant sites were densely distributed across the genome (Fig. 1A, B). Obtained variant sites were annotated to better understand their position on the genome; annotation information and SNP statistics by type and region are provided in Supplementary Table S2.

By filtering and separating sites and redundant sites, 5913 were obtained in population A1 and 6648 in population A2 for construction of high-density genetic maps of population maternal parents, and 18 linkage groups were identified finally (consistent with the number of haploid chromosomes in cassava). The total map distance of the A1\_SC205 genetic map was 2492.62 cM, the average distance between markers was 0.43 cM, the length of 18 linkage groups ranged 82.42–219.53 cM, and the longest linkage group was LG01. The total map distance of the A2\_SC205 genetic map was 2445.17 cM, the average distance between markers was 0.37 cM, the length of linkage groups ranged 72.10–206.95 cM, and the longest linkage group was LG01 (Supplementary Table S3). Collinearity in linkage map markers between populations was observed by counting common loci, of which there were 1792; for A1, the average collinearity rate was 30.37%, with a maximum of 38.01% (LG16); for A2 the average collinear rate was 26.69%, with a maximum of 39.40% (LG18) (Supplementary Table S3). TBtools was used to draw chromosome collinearity for each population (Fig. 1C). The collinearity relationship between linkage maps of each population (Supplementary Fig. 1) revealed a certain degree of collinearity on each chromosome, laying a foundation for subsequent co-localization.

### Statistical analysis and correlation of phenotypic data

Summary data for parental population phenotypic mean values, statistical data for different years, and combined Best Linear Unbiased Estimators (BLUE) values for correlated traits between populations, are presented in Supplementary Table S4. A combined analysis of populations revealed a coefficient of variation (CV) for RW to range 0.413–0.656, and heritability values ranged 0.16–0.23. For RN, the CV ranged 0.244–0.342, and heritability values ranged 0.48–0.5. For RLW, the CV ranged 0.207–0.283, and heritability values ranged 0.48–0.59 (Table 1). In both populations, RW was most variable and had relatively lower heritability, while RN and RLW had smaller trait variabilities and higher stability and heritability. This suggests that data stability may influence the magnitude of trait heritability. Additionally, the similarity in heritability between populations indicates a relatively stable heritability.

Overall skewness was close to 0 and kurtosis to 3, suggesting approximate data normality and suitability for subsequent analysis. Data also approximately follow a



**Fig. 1** **A** A1 population SNP density map, with an average of 8610.2 mutation sites per chromosome; **B** A2 population SNP density map, with an average of 6478.9 mutation sites per chromosome; **C** Collinearity analysis of two group linkage maps (A1 with an average collinearity of 30.37%, and peak (38.01%) collinearity in LG16; A2 with an average collinearity of 26.69% and peak (39.40%) collinearity in LG18)

**Table 1** Statistical analysis of population traits of hybrid progenies

population	Trait	Year	Mean	standarddeviation	Range	Skew	Kurtosis	CV	Heritability
A1	RN	2022	6.589	1.901	2.510–12.510	0.194	2.472	0.288	0.48
		2023	5.353	1.824	0.720–10.720	0.538	3.165	0.341	
		22–23	5.738	1.620	1.57–11.57	0.079	3.064	0.282	
	RW	2022	2.614	1.560	0.120–9.310	0.936	4.158	0.597	0.16
		2023	2.643	1.734	0.230–9.120	1.297	4.635	0.656	
		22–23	2.513	1.474	0.08–9.103	0.934	4.357	0.587	
	RLW	2022	6.788	1.556	3.290–12.330	0.635	3.516	0.229	0.58
		2023	8.695	1.956	1.340–13.860	−0.385	3.754	0.225	
		22–23	7.608	1.572	0.935–12.812	−0.156	4.123	0.207	
A2	RN	2022	6.585	2.204	1.00–14.160	0.419	3.318	0.335	0.53
		2023	4.684	1.600	0.720–9.720	0.438	3.753	0.342	
		22–23	5.599	1.364	2.237–10.73	0.211	3.726	0.244	
	RW	2022	2.161	1.219	0.140–8.670	1.394	6.670	0.564	0.23
		2023	1.563	0.791	0.060–5.090	1.010	4.917	0.506	
		22–23	1.789	0.740	0.081–4.327	0.074	2.866	0.413	
	RLW	2022	6.818	1.761	3.450–12.460	0.750	3.746	0.258	0.59
		2023	7.432	2.106	0.830–14.660	0.406	4.172	0.283	
		22–23	7.100	1.537	2.876–11.604	0.225	2.835	0.217	

normal distribution (Fig. 2A). Violin plots and trait error distribution plots (Fig. 2B) reveal minimal overall variation in phenotype data for each population over the two years, indicating low data error and high stability.

R software was used to compute trait correlation coefficients and correlations between populations, and to generate a correlation heatmap (Fig. 2C, D). All three traits correlated positively in both populations, with RN strongly and positively correlated with RW. Boxplots revealed similar overall trends in the data for RN and RW.

QTL localization

QTL mapping results reveal 188 QTLs to be located with RLW in different years between populations, along with 157 QTLs associated with RN and 213 QTLs related to RW. Quantities of QTLs mapped in different populations and years are summarized in Table 2, and detailed in Supplementary Table S5. To validate QTL effectiveness, positional information of mapped loci in populations was compared; 3 co-located RLW-related QTLs, 3 RN-related QTLs, and 9 RW-related QTLs were identified (Table 3).

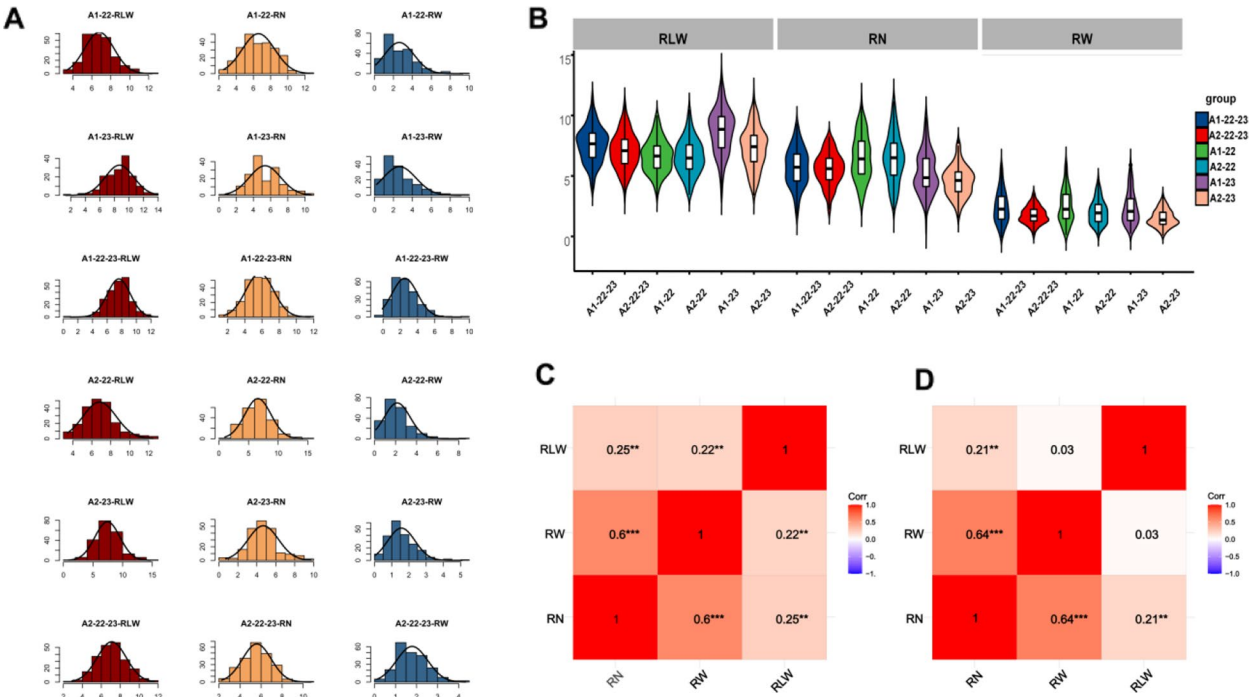
The 3 QTLs associated with RLW are positioned on chromosomes 1, 7, and 9. Notably, the QTL designated as qRLW-7 exhibits a LOD score exceeding 3. These QTLs contributes an average phenotypic effect of 10.53% and has been consistently identified in both A1 and A2 populations. The 3 QTLs of RN are located on chromosomes 3, 16, and 17. Two of them have a mapping interval of

**Table 2** Number of loci for different traits in different populations and years

Year	RLW	RN	RW	Total
A1-22	20	9	17	46
A1-23	38	41	31	110
A1-22+23	32	13	19	64
A2-22	38	41	74	153
A2-23	35	21	14	70
A2-22+23	25	32	58	115
Total	188	157	213	558

less 100 bp. The 9 QTLs of RW occur mainly on chromosomes 7, 9, 17, and 18, with two QTLs located on chromosome 18 and three QTL positioning intervals less 150 bp. At the same time, QTLs related to RLW and RN also occur on chromosomes 7 and 9. QRW-18a, located around 2 Mb on chromosome 18, occurred in 5 sets of data from populations A1 and A2, with an LOD value of 3.34.

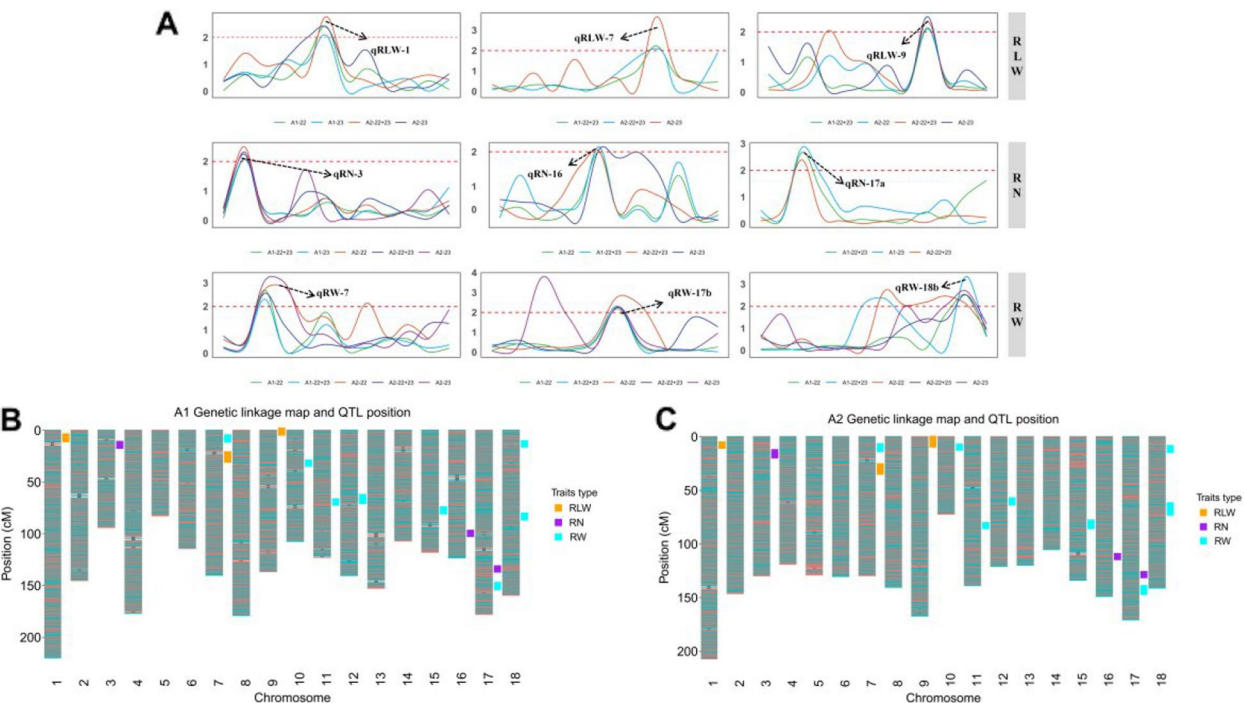
To understand variation in LOD scores of QTLs mapped for different traits in different years within populations, three co-located QTLs for each trait were selected, and QTL interval and nearby LOD waveforms were plotted (Fig. 3A). There was a consistent overall fluctuation in LOD scores within mapped QTL



**Fig. 2** A Violin plots of trait distribution; B Error plots of trait distribution; and Trait correlation coefficients for C) Group A1 and D) Group A2

**Table 3** Colocation QTLs mapped for root quantity traits in Cassava

QTL_Name	Group	Trait	Position	Population	LOD	%Expl
RLW-1	1	RLW	446,386–526,632	A1-22,A1-22 + 23,A2-23,A2-22 + 23	2.77	8
RLW-7	7	RLW	7,270,622–7,306,733	A1-22 + 23,A2-23,A2-22 + 23	3.65	14.8
RLW-9	9	RLW	56,068–75,706	A1-22 + 23,A2-22,A2-23,A2-22 + 23	2.5	8.8
RN-3	3	RN	4,862,573–4,922,415	A1-23,A1-22 + 23,A2-22,A2-23,A2-22 + 23	2.5	25.9
RN-16	16	RN	26,309,903–26,309,939	A1-22,A1-22 + 23,A2-23,A2-22 + 23	2.14	11.5
RN-17	17	RN	27,016,633–27,016,679	A1-23,A1-22 + 23,A2-22 + 23	2.85	31.2
RW-7	7	RW	3,472,827–3,497,170	A1-22,A1-22 + 23,A2-22,A2-23,A2-22 + 23	3.08	12.8
RW-9	9	RW	140,164–148,432	A1-22 + 23,A2-22,A2-23,A2-22 + 23	7.6	12.8
RW-10	10	RW	4,262,539–4,262,551	A1-22 + 23,A2-22,A2-23,A2-22 + 23	3.11	13.6
RW-11	11	RW	16,003,135–16,003,215	A1-23,A1-22 + 23,A2-22,A2-22 + 23	3.52	15.4
RW-12	12	RW	21,852,432–21,935,379	A1-23,A1-22 + 23,A2-22,A2-23	2.53	8.4
RW-15	15	RW	23,949,858–23,961,834	A1-22,A1-22 + 23,A2-22,A2-23,A2-22 + 23	4.14	10.2
RW-17	17	RW	31,297,009–31,368,543	A1-22,A1-22 + 23,A2-22,A2-23,A2-22 + 23	3.78	9.8
RW-18	18	RW	16,818,273–16,893,445	A1-22,A1-22 + 23,A2-23,A2-22 + 23	3.08	16.7
RW-18	18	RW	2,317,804–2,317,946	A1-22,A1-22 + 23,A2-22,A2-23,A2-22 + 23	3.34	30.5



**Fig. 3** **A** LOD value fluctuation chart: for each trait, select the main three QTLs for localization, and draw a LOD waveform of the localization population based on physical location. The trend of LOD values in the QTL interval is consistent, and LOD peaks in the QTL interval; **B** A1 linkage map; **C** A2 linkage map. QTL labels (**B**, **C**): yellow, RLW related QTLs; purple, RW related QTLs; blue, RN related QTLs

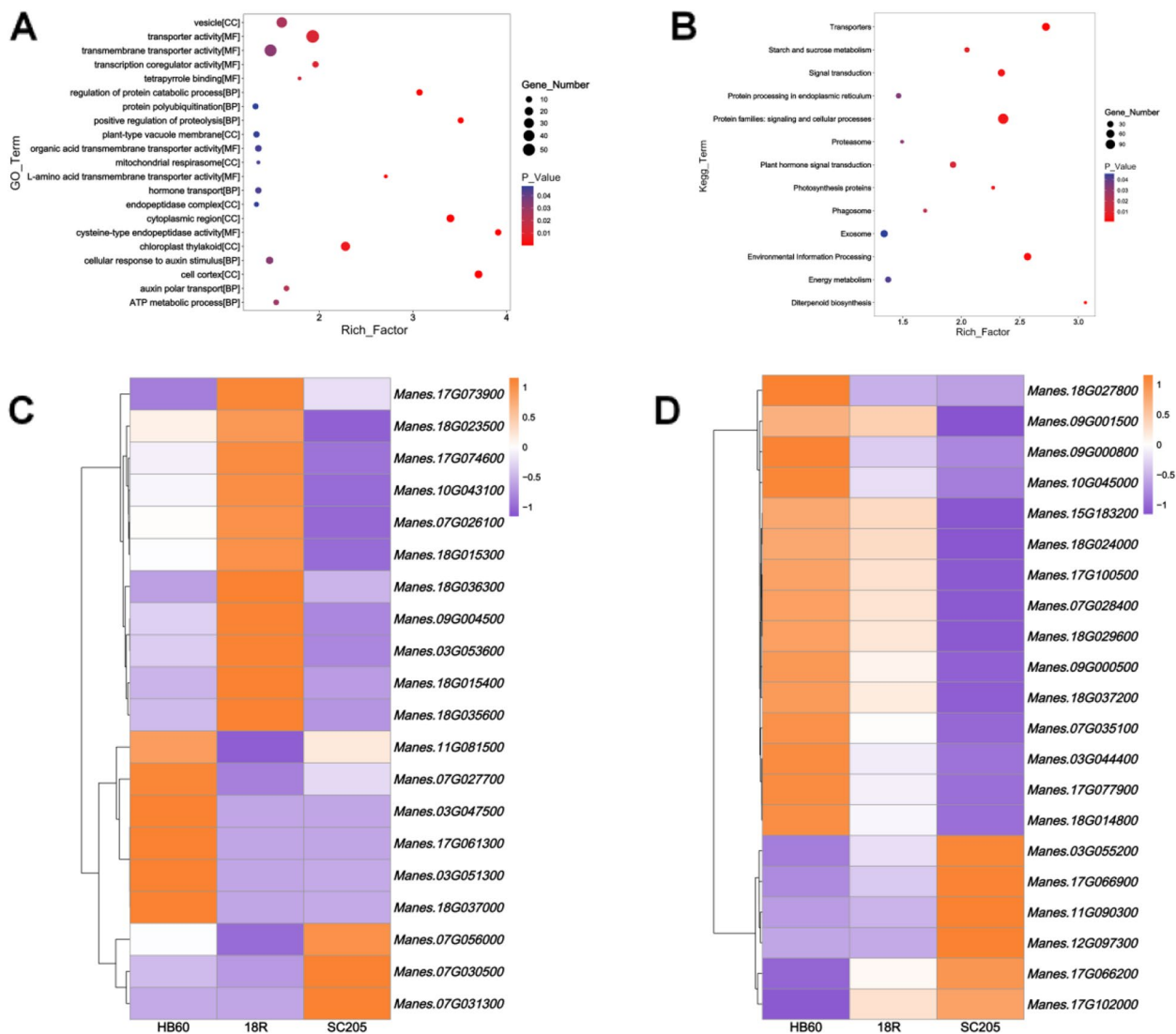
intervals and their vicinity across datasets. Two linkage maps were drawn using genetic distance and markers, and QTL localization information was labeled based on subsequent localization results (Fig. 3B, C).

**Candidate gene analysis**

Based on localization results, interval and upstream and downstream genes were identified as candidate genes. GO and KEGG analyses were performed to better understand gene functions. GO enrichment

analysis revealed these genes to be mainly enriched in “polar transport of auxin,” “auxin transport,” “cytokinin metabolic process,” “ATP biosynthetic process,” “auxin-activated signaling pathway,” and “transcription coregulator activity” (Fig. 4A). KEGG analysis revealed candidate genes to be mainly enriched in pathways such as “protein photosynthesis proteins,” “starch and sucrose metabolism,” and “plant hormone signal transduction” (Fig. 4B). Through screening and identification, a total of 39 genes were obtained (Table 4), including 5 candidate genes associated with RLW, 11 candidate genes associated with RN, 23 candidate genes

associated with RW, and 7 genes containing 5 classes of transcription factors (WARY, BES 1, bZIP, ARF, MYB). Remaining candidate genes included auxin-responsive factor, zinc finger (C3HC4-type RING finger) family protein, ethylene sensor, SNARE-like superfamily protein, plant glycogenin-like starch initiation protein, and glucose-6-phosphate dehydrogenase 6. Expression levels of candidate genes were calculated based on QTL localization candidate genes and parental transcriptome data (Supplementary Table S6). Depending on whether expression level changes were consistent with trends in parental traits, the candidate genes were



**Fig. 4** **A** GO function enrichment (gene function mainly enriched in cellular response to auxin stimuli, auxin polar transport, and other biological processes); **B** KEGG pathway: Plant hormone signal transduction, Starch and cross metabolism pathways, etc.; **C** Gene expression heatmap inconsistent with trends in change of parental traits, where expression changes of candidate genes in different parents are not significantly related to changes in parental traits; **D** Gene expression heatmap consistent with trends in change of parental traits, where expression levels of candidate genes increase or decrease synchronously with the size of parental traits in different parents

**Table 4** Candidate genes for root quantity traits in Cassava

Trait	Chromosome	Gene_ID	Physical position(bp)	Gene Function
RN	3	<i>Manes.03G044400</i>	3,875,045–3,881,657	Ankyrin repeat family protein
RN	3	<i>Manes.03G047500</i>	4,297,715–4,299,007	Winged-helix DNA-binding transcription factor family protein
RN	3	<i>Manes.03G051300</i>	4,782,760–4,784,024	WRKY
RN	3	<i>Manes.03G053600</i>	5,181,953–5,190,433	Zn-dependent exopeptidases superfamily protein
RN	3	<i>Manes.03G055200</i>	5,383,408–5,388,267	Ankyrin repeat family protein
RLW	7	<i>Manes.07G056000</i>	6,355,527–6,365,241	Tetratricopeptide repeat (TPR)-like superfamily protein
RW	7	<i>Manes.07G026100</i>	2,819,441–2,822,633	Smg-4/UPF3 family protein
RW	7	<i>Manes.07G027700</i>	3,050,129–3,053,172	SNARE-like superfamily protein
RW	7	<i>Manes.07G028400</i>	3,120,266–3,131,581	Protein phosphatase 2C family protein
RW	7	<i>Manes.07G030500</i>	3,349,025–3,352,522	Zinc finger (C3HC4-type RING finger) family protein
RW	7	<i>Manes.07G031300</i>	3,431,995–3,432,570	Dehydroascorbate reductase 2
RLW	9	<i>Manes.09G001500</i>	149,564–163,612	DHHC-type zinc finger family protein
RLW	9	<i>Manes.09G000800</i>	433,877–435,511	DNA/RNA helicase protein
RLW	9	<i>Manes.09G000500</i>	459,723–477,394	Ribosomal protein L32e
RLW	9	<i>Manes.09G004500</i>	1,145,489–1,158,661	bZIP
RW	9	<i>Manes.09G001500</i>	149,564–163,612	DHHC-type zinc finger family protein
RW	10	<i>Manes.10G043100</i>	4,441,036–4,451,646	Plant-specific GATA-type zinc finger transcription factor family protein
RW	10	<i>Manes.10G045000</i>	4,799,921–4,809,695	Clast3-related
RW	11	<i>Manes.11G081500</i>	15,096,310–15,098,673	Adaptin family protein
RW	11	<i>Manes.11G090300</i>	16,982,368–16,982,661	R-protein L3 B
RW	12	<i>Manes.12G097300</i>	21,899,393–21,904,241	Signal transduction histidine kinase, hybrid-type, ethylene sensor
RW	15	<i>Manes.15G183200</i>	24,954,254–24,957,013	Triosephosphate isomerase
RN	17	<i>Manes.17G061300</i>	26,043,602–26,053,323	Plant glycogenin-like starch initiation protein 3
RN	17	<i>Manes.17G066200</i>	26,623,698–26,625,728	WRKY
RN	17	<i>Manes.17G066900</i>	26,690,489–26,691,693	P-loop containing nucleoside triphosphate hydrolases superfamily protein
RN	17	<i>Manes.17G073900</i>	27,398,788–27,404,693	Auxin efflux carrier family protein
RN	17	<i>Manes.17G074600</i>	27,527,652–27,529,759	Glucose-6-phosphate dehydrogenase 6
RN	17	<i>Manes.17G077900</i>	27,772,133–27,779,912	With no lysine (K) kinase 4
RW	17	<i>Manes.17G100500</i>	30,715,977–30,722,346	Transcriptional factor B3 family protein / auxin-responsive factor AUX/IAA-related
RW	17	<i>Manes.17G102000</i>	30,954,184–30,956,884	WRKY
RW	18	<i>Manes.18G014800</i>	1,564,349–1,568,553	Lysine histidine transporter 1
RW	18	<i>Manes.18G015300</i>	1,619,465–1,621,243	Ubiquitin-like superfamily protein
RW	18	<i>Manes.18G015400</i>	1,638,932–1,645,840	ARF
RW	18	<i>Manes.18G023500</i>	2,256,111–2,258,325	MYB
RW	18	<i>Manes.18G027800</i>	2,392,755–2,401,894	Nodulin MtN3 family protein
RW	18	<i>Manes.18G029600</i>	2,509,358–2,512,562	BES1
RW	18	<i>Manes.18G035600</i>	3,171,581–3,174,946	Raffinose synthase family protein
RW	18	<i>Manes.18G036300</i>	3,195,620–3,197,448	Octicosapeptide/Phox/Bem1p family protein

classified into two categories, and expression level heat maps of localized genes were plotted (Fig. 4C, D).

## Discussion

### Genetic map and QTL mapping precision

High-quality SNP markers were obtained using hyper-seq simplified sequencing technology, and genetic maps were constructed for two populations. Multiple cassava maps have been previously constructed using markers such as

RFLP and SSR, ranging 100–510 markers, map lengths of 845.2–1420.3 cM, and with spacing of 5.6–17.92 cM/marker [27, 39–42]. We increase marker density, and by producing a final genetic map spacing of 0.37–0.43 cM/marker, we provide a higher precision cassava genetic map. A total of 15 QTLs related to traits were identified. There are two of the QTLs in RN have a mapping interval of less than 100 bp, which has high accuracy and helps to accurately locate candidate genes related to this trait.

There are three QTL in RW positioning intervals less than 150 bp, indicating high overall positioning accuracy, and the average LOD value of QTLs exceeds 3, indicating a high possibility of the existence of the positioned QTLs. Meantime, Shengkui [43] used GWAS to locate multiple significant loci for cassava agronomic traits, and positioning results of qRN-17a and qRW-18a are consistent. Partly consistent with our results, Ewa [16] reported QTLs related to RW traits to be located on chromosomes 3, 4, and 7, and those related to RN traits to be located on chromosomes 3 and 7, while QTLs related to RN traits were located on chromosomes 3 and 7, consistent with some of the results of this study.

### Functional analysis of candidate genes

We selected three traits related to cassava yield (RN, RW, and RLW) for QTL localization analysis based on a high-precision cassava genetic map. There is a high positive correlation between these traits; RN · RW and RLW can effectively evaluate crop yield and tuber shape, which is also an important factor to affect yield. We also report *Manes.09G001500* to be associated with RN and RW traits; this gene belongs to the DHHC type zinc finger family protein which affects plant growth and development. *OsDHHC01* increases the number of tillers and grain yield in rice and oilseed rape (*Brassica napus*) [44, 45], and *OsDHHC13* positively regulates the oxidative stress response [44]. DHHC type zinc finger protein (DHHC) and S-acylated protein mainly serve as substrates for S-acyltransferase. S-acylation, also known as palmitoylation, is an important protein lipid modification that plays an important role in plant growth, development, and stress response [46]. *OsDHHC30* participates in regulation of salt tolerance in rice through S-acylation modification of OsCBL2/3 [47]. We speculate that this gene may affect cassava growth and development by regulating its stress resistance, thereby affecting the weight and quantity of root tubers. Further research is needed to identify how this gene affects root tuber traits.

*Manes.03G051300*, *Manes.17G066200*, and *Manes.17G102000* belong to the WRKY family, in which *Manes.03G051300* and *Manes.17G066200* are associated with RN traits and *Manes.17G102000* is associated with RW. This family is an important regulator of growth and development in higher plants. The WRKY transcription factor OsWRKY78 regulates rice stem elongation and seed size [48], and SIWRKY23 expressed mainly in tomato roots is associated with seed size [49]. In cassava, the MeWRKY20 gene promotes ABA accumulation [50], and high ABA levels can promote radial expansion of cortical cells and reduce root penetration ability [51]. *Manes.18G015400* and *Manes.17G100500* belong to the AFR transcription factor family,

and *Manes.17G073900* belongs to the Auxin efflux carrier family protein, auxin, Aux) Indole-3-acetic acid, IAA is almost involved in all processes of plant growth and development; most of these processes are regulated by gene expression [52]. Auxin response factor (ARF) is a family of transcription factors that regulates expression of auxin responsive genes [53, 54]. ARF can specifically interact with auxin response elements in the promoter region of auxin responsive genes, and AuxRE binds to “TGTCTC” to activate or inhibit gene expression [55]. The auxin signaling pathway mediated by ARF transcription affects plant growth and development by regulating cell division, elongation, and differentiation [56, 57]. miRNA160 and miRNA167 are involved in *Arabidopsis* AtARF6, regulating adventitious root formation; *Arabidopsis* AtARF17 is also involved in this process [58, 59]. The SlARF2 gene in tomato regulates lateral root formation [60]. DnARF11 is specifically expressed in the roots of *Dendrobium officinale*, indicating its importance in root system establishment [61]. We speculate that these genes affect RN and RW by regulating root tuber development.

The candidate gene *Manes.18G023500* belongs to the MYB transcription factor family, one of the largest transcription factor families in plants with important roles in physiological and biochemical processes in growth and development. MeMYB108 can reduce leaf shedding and regulate cassava biomass [62]. In *Arabidopsis*, *AtMYB3R1* and *AtMYB3R4* (*AtMYB3R1/4*) act as transcriptional activation factors, expressed in proliferating tissue, regulating the cell cycle process, and affecting circadian rhythms [63–65]. Ectopic expression of *AtMYB56* can inhibit root growth, leading to failed root regeneration after stem cell damage [66]. In soybeans, *GmMYB81* regulates development of soybean tissues and embryos and has a significant effect on abiotic stress [67]. We speculate that the *MeMYB67* and *MeMYB61* genes may affect cassava yield by influencing leaf shedding and regulating the cell cycle process.

RLW not only effectively evaluates the shape of cassava, but also relates to yield traits [68, 69]. We identify *Manes.09G004500* (in the bZIP family) to be related to RLW. This family is an important regulatory factor in higher plants and participates in plant growth and development [70]. BZIP is involved in regulating certain signaling pathways related to the ABA pathway. During germination and flowering of *Arabidopsis* seeds, bZIP transcription factors ABI5 (ABA sensitive 5) and ABFs (ABRE binding factors) are key factors regulating ABA signaling Hossain. *Arabidopsis* *abi5* mutant is less sensitive to ABA. Because of ABA-inhibiting seed germination, ABI5 can activate expression of specific genes in seeds through ABA-mediated signaling pathways, thereby promoting seed germination [71] Additionally, *Arabidopsis* *AtbZIP29* is specifically

expressed in proliferative tissues, regulates the number of cells in leaf and root meristem tissues, and promotes tissue regeneration by specifically binding to cell cycle regulatory factors and cell wall regeneration related genes. During the cell division cycle, BZIP29 can interact with bZIP69 to form dimers, which play roles in the root primordia, root crown, and meristem of lateral roots, thereby affecting root shape [72]. We speculate that this gene may affect RLW through similar molecular mechanisms.

## Conclusions

To identify genes related to cassava yield, and to improve cassava yield, we selected three important agronomic traits. By constructing genetic maps of two half-sibling populations sharing the same parent, QTL mapping was performed on two populations, co-located loci were extracted to obtain 15 QTLs for important quantitative traits, and potential candidate genes were identified that affected these traits. We also performed transcriptome analysis to analyze expression levels of different genes. Finally, 5 candidate genes related to RLW, 11 candidate genes related to RN, and 22 candidate genes related to RW were obtained. We have found that previous studies have shown that candidate genes such as *Manes. 09G001500*, *Manes. 17G066200*, and *Manes. 17C073900* belong to gene families that are associated with their traits. Localization of these candidate genes improves our understanding of the genetic basis of cassava yield traits, and identifies potential gene targets to improve cassava yield and to provide reference for improving cassava yield through molecular breeding methods.

## Abbreviations

RFLP	Restriction Fragment Length Polymorphism
SSR	Simple Sequence Repeat
SNP	Single Nucleotide Polymorphism
QTL	Quantitative Trait Loci
PCR	Polymerase chain reaction
WGS	Whole genome sequencing
AFSM	Amplified Fragment Single-nucleotide polymorphisms and Methylation
BLUE	Best Linear Unbiased Estimation
RW	Single plant Root Weight
RN	Root Number
RLW	Root Length to Width ratio
LOD	Logarithm of the Odds
cM	Centimorgan
CV	Coefficient of Variation
KEGG	Kyoto Encyclopedia of Genes and Genomes
GO	Gene Ontology
DHHC	DHHC type zinc finger protein

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-025-06278-3>.

Supplementary Material 1: Fig. S1. Collinear diagram of chromosomal spiders. The figure shows the linkage maps of parents from two different populations. The collinear loci between the two sets of linkage maps are connected and highlighted in red. At the same time, QTLs co located between the two populations are randomly marked with colors in the linkage maps, visually demonstrating the collinearity between the parental linkage maps of different populations.

Supplementary Material 2: Table S1. population sample name, Table S2. Variants information, Table S3. Information for linkage map, Table S4. Phenotypic data, Table S5. QTL information, Table S6. Candidate gene expression levels.

## Acknowledgements

Thank you to all authors for their participation and contribution to this article.

## Authors' contributions

Q.L., Z.X. and W.W. conceived and managed the experiments. L.Z., Z.W., Y.L., P.A., C.X., and S.J. conducted the experiments work. Q.L., Z.X. and Y.L. analyzed the experimental results. Q.L. wrote the manuscript. Z.X., and W.W. reviewed and contributed to improve it and revised the last version of the manuscript. All authors read and approved the final manuscript.

## Funding

This study was funded by the Biological Breeding-National Science and Technology Major Project (2023ZD04073) and supported by the Hainan Yazhou Bay Seed Lab (B23YQ0002), National Natural Science Foundation of China NSFC-CG Joint Fund Key Project, Gene Network and Molecular Design Breeding of Cassava Yield and Root Quality (3181101517), and High-performance Computing Platform of YZBSTCACC.

## Data availability

No datasets were generated or analysed during the current study.

## Declarations

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Written informed consent for publication was obtained from all participants.

## Competing interests

The authors declare no competing interests.

Received: 10 July 2024 Accepted: 19 February 2025

Published online: 25 April 2025

## References

1. Mann C. Reseeding the green revolution [J]. *Science*, 1997;277:209–220. <https://doi.org/10.1126/science.277.5329.1038>.
2. Ukwu N.U, Olanmi B. Crossability among Five Cassava (*Manihot esculenta* Crantz) Varieties. *Modern Concepts & Developments in Agronomy*. 2018;2(4):MCDA.000543. <https://api.semanticscholar.org/CorpusID:92846744>.
3. Sanchez T, Ceballos H, Dufour D, et al. Prediction of carotenoids, cyanide and dry matter contents in fresh cassava root using NIRS and Hunter color techniques [J]. *Food Chem*. 2014;15:444–51. <https://doi.org/10.1016/j.foodchem.2013.11.081>.

4. Howeler, Reinhardt H., N. B. Litaladio and Graeme Thomas. Save and grow : cassava : a guide to sustainable production intensification. 2013. <https://api.semanticscholar.org/CorpusID:126510126>.
5. De Souza AP, Massenbourg LN, Jaiswal D, Cheng S, Shekar R, Long SP. Rooting for cassava: insights into photosynthesis and associated physiology as a route to improve yield potential. *New Phytol.* 2016;213:50–65. <https://doi.org/10.1111/nph.14250>.
6. Legarra A, Croiseau P, Sanchez MP, Teyssèdre S, Sallé G, Allais S, Fritz S, Moreno CR, Ricard A, Elsen JM. A comparison of methods for whole-genome QTL mapping using dense markers in four livestock species. *Genet Sel Evol.* 2015Feb 12;47(1):6. <https://doi.org/10.1186/s12711-015-0087-7>.
7. Chen X, Xia Z, Fu Y, et al. Constructing a genetic linkage map using an F1 population of non-inbred parents in cassava (*Manihot esculenta* Crantz) [J]. *Plant Mol Biol Report.* 2010;28:676–83. <https://doi.org/10.1007/s11105-010-0198-9>.
8. Sraphet S, Athipong B, Thanwanit et al. The SSR and EST-SSR-based genetic linkage map of cassava (*Manihot esculenta* Crantz) [J]. *Theoretical and Applied Genetics.* 2011, 122:1161–1170. <https://doi.org/10.1007/s00122-010-1502-5>.
9. International Cassava Genetic Map Consortium (ICGMC). High-resolution linkage map and chromosome-scale genome assembly for cassava (*Manihot esculenta* Crantz) from 10 populations. G3 (Bethesda). 2014 Dec 11;5(1):133–44. <https://doi.org/10.1534/g3.114.015008>.
10. Athipong Boonchanawit, Sraphet Supajit, Boonseng Opas, et al. QTL underlying plant and first branch height in cassava (*Manihot esculenta* Crantz)[J]. *Field crops research*, 2011, 121(3): 343–349. <https://doi.org/10.56899/151.02.11>.
11. Nang Hmwe et al. Identification of QTL Underlying Plant Height and First Branch Height of Cassava[J]. *Philippine Journal of Science*, 2022, (2)(151): 683–693. <https://doi.org/10.56899/151.02.11>.
12. Okogbenin E, Marin J, Fregene M. An SSR-based molecular genetic map of cassava [J]. *Euphytica.* 2006;147:433–40. <https://doi.org/10.1007/s00122-010-1520-5>.
13. Akinbo O, Labuschagne M-T, Marín J, et al. QTL Analysis for Root Protein in a Backcross Family of Cassava Derived from *Manihot esculenta* ssp *flabellifolia*[J]. *Tropical Plant Biology.* 2012;5(2):161–72. <https://doi.org/10.1007/s12042-012-9095-8>.
14. Xia Z, Zou M, Zhang S, Feng B, Wang W. AFSM sequencing approach: a simple and rapid method for genome-wide SNP and methylation site discovery and genetic mapping[J]. *Sci Rep.* 2014Dec;3(4):7300. <https://doi.org/10.1038/srep07300>.
15. Meiling Zou Lu, Cheng ZS, et al. Epigenetic map and genetic map basis of complex traits in cassava population[J]. *Sci Rep.* 2017Jan;25(7):41232. <https://doi.org/10.1038/srep41232>.
16. Okogbenin E, Fregene M. Genetic mapping of QTLs affecting productivity and plant architecture in a full-sib cross from non-inbred parents in Cassava (*Manihot esculenta* Crantz)[J]. *Tag.theoretical & Applied Genetics. theoretische Und Angewandte Genetik*, 2003, 107(8): 1452–1462. <https://doi.org/10.1007/s00122-003-1383-0>.
17. Ewa F, Asiwe JNA, Okogbenin E, Ogonna AC, Egesi C. KASPar SNP genetic map of cassava for QTL discovery of productivity traits in moderate drought stress environment in Africa. *Sci Rep.* 2021May 28;11(1):11268. <https://doi.org/10.1038/s41598-021-90131-8>.
18. Okogbenin E, Fregene M. Genetic analysis and QTL mapping of early root bulking in an F1 population of non-inbred parents in cassava (*Manihot esculenta* Crantz). *Theoretical and Applied Genetics.* 2002 Dec;106(1):58–66. <https://link.springer.com/article/10.1007/s00122-002-1068-0>.
19. Zou M, Hyper-seq XZ. A novel, effective, and flexible marker-assisted selection and genotyping approach[J]. *Innovation.* 2022;3(4):2. <https://doi.org/10.1016/j.xinn.2022.100254>.
20. Murray MG, Thompson WF. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 1980Oct 10;8(19):4321–5. <https://doi.org/10.1093/nar/8.19.4321>.
21. De Sena Brandine G, Smith AD. Falco: high-speed FastQC emulation for quality control of sequencing data. *F1000Res.* 2019 Nov 7;8:1874. <https://doi.org/10.12688/f1000research.21142.2>.
22. Jung Y, Han D. BWA-MEME: BWA-MEM emulated with a machine learning approach. *Bioinformatics.* 2022Apr 28;38(9):2404–13. <https://doi.org/10.1093/bioinformatics/btac137>.
23. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009 Aug 15;25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352>.
24. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. Twelve years of SAMtools and BCFtools. *Gigascience.* 2021 Feb 16;10(2). <https://doi.org/10.1093/gigascience/giab008>.
25. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal Of Human Genetics.* 2007 Sep;81(3):559–75. <https://pubmed.ncbi.nlm.nih.gov/17701901>.
26. Ingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin).* 2012 Apr-Jun;6(2):80–92. <https://doi.org/10.4161/fly.19695>.
27. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014 Aug 1;30(15):2114–20. <https://pubmed.ncbi.nlm.nih.gov/24695404>.
28. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* 2019Aug;37(8):907–15. <https://doi.org/10.1038/s41587-019-0201-4>.
29. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 2014Apr 1;30(7):923–30. <https://doi.org/10.1093/bioinformatics/btt656>.
30. Lander, E. S., Green, P., Abrahamson, J., Barlow, A., Daly, M. J., Lincoln, S. E., & Newburg, L. (1987). MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics*, 1(2), 174–181. <https://pubmed.ncbi.nlm.nih.gov/3692487/>.
31. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R; 1000 Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics.* 2011 Aug 1;27(15):2156–8. <https://doi.org/10.1093/bioinformatics/btr330>.
32. Meng L, Li H, Zhang L, Wang J. QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *The Crop Journal.* 1973;2015(3):269–83. <https://doi.org/10.1016/j.cj.2015.01.001> Lin S, & Kernighan, B.W.
33. Lin S, Kernighan BW. An Effective Heuristic Algorithm for the Traveling-Salesman Problem. *Oper Res.* 1973;21(2):498–516. <https://doi.org/10.1287/opre.21.2.498>.
34. Kosambi DD. The estimation of map distances from recombination values. *Ann Eugen.* 1943;12(1):172–5. <https://doi.org/10.1111/j.1469-1809.1943.tb02321.x>.
35. Broman, K. W., Wu, H., Sen, S., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, 19(7), 889–890. <https://pubmed.ncbi.nlm.nih.gov/1272430>.
36. Ooijen J. MapQTL 5, Software for the mapping of quantitative trait loci in experimental populations. , 2004.
37. Powder KE. Quantitative Trait Loci (QTL) Mapping. *Methods Mol Biol.* 2020;2082:211–29. [https://doi.org/10.1007/978-1-0716-0026-9\\_15](https://doi.org/10.1007/978-1-0716-0026-9_15).
38. Shen W, Le S, Li Y, Hu F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS ONE.* 2016Oct 5;11(10):e0163962. <https://doi.org/10.1371/journal.pone.0163962>.
39. Fregene M, Angel F, Gomez R et al. A molecular genetic map of cassava (*Manihot esculenta* Crantz) [J]. *Theoretical and Applied Genetics.* 1997;95:431–441. <https://doi.org/10.1007/s001220050580>.
40. Kunkeaw S, Yoocha T, Sraphet S, et al. 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. <https://doi.org/10.1007/s11032-010-9414-4>.
41. Kunkeaw S, Tangphatsornruang S, Smith DR et al. Genetic linkage map of cassava (*Manihot esculenta* Crantz) based on AFLP and SSR markers[J]. *Plant Breeding.* 2009.1439–0523.01623. <https://doi.org/10.1111/J.1439-0523.2009.01623.X>.

42. E Okogbenin, Marin J, Fregene M. QTL analysis for early yield in a pseudo F2 population of cassava[J]. *African Crop Science Journal*, 2008,14:275–285. <https://api.semanticscholar.org/CorpusID:55352414>.
43. Zhang Shengkui. Comprehensive evaluation of cassava germplasm resources and genome-wide association analysis of major agronomic traits [D]. Huazhong Agricultural University [2024–12–09]. CNKI:CDMD:1.1018.206875.
44. Peng D, Tan X, Lin Z, Yuan D, Lin J, Liu X, et al. Increasing branch and seed yield through heterologous expression of the novel rice S-acyl transferase gene OsPAT15 in *Brassica napus* L. *Cell Death Dis.* 2018;9:1238. <https://doi.org/10.1270/jsbbs.17126>.
45. Zhou B, Lin JZ, Peng D, Yang YZ, Guo M, Tang DY, et al. Plant architecture and grain yield are regulated by the novel DHHC-type zinc finger protein genes in rice (*Oryza sativa* L.). *Plant Sci.* 2017;254:12–21. <https://doi.org/10.1016/j.plantsci.2016.08.015>.
46. Liu, X.M., Wang, W.W., Yang, Y.Z., Xia, M.L., Li, X., Zuo, D.L. et al. Preliminary study on a rice OsDHHC13 gene involving in the response to oxidative stress. *Journal of Hunan University (Natural Sciences)*, 2016, 43, 110–117. <https://api.semanticscholar.org/CorpusID:116222654>.
47. Tian Y, Zeng H, Wu J, Huang J, Gao Q, Tang D, Cai L, Liao Z, Wang Y, Liu X, Lin J. Screening DHHCs of S-acylated proteins using an OsDHHC cDNA library and bimolecular fluorescence complementation in rice. *Plant J.* 2022Jun;110(6):1763–80. <https://doi.org/10.1111/tpj.15769>.
48. Zhang CQ, Xu Y, Lu Y, et al. The WRKY transcription factor OsWRKY78 regulates stem elongation and seed development in rice. *Planta.* 2011;234(3):541–54. <https://doi.org/10.1007/s00425-011-1423-y>.
49. Singh D, Debnath P, Roohi, et al. Expression of the tomato WRKY gene, SWRKY23, alters root sensitivity to ethylene, auxin and JA and affects aerial architecture in transgenic *Arabidopsis* [J]. *Physiology and Molecular Biology of Plants*, 2020, 26 (6):1187–1199. <https://doi.org/10.1007/s12298-020-00820-3>.
50. Wei YX, Liu W, Hu W, et al. The chaperone MeHSP90 recruits MeWRKY20 and MeCatalase1 to regulate drought stress resistance in cassava [J]. *New Phytol.* 2020;226(2):476–91. <https://doi.org/10.1111/nph.16346>.
51. Huang G, Kilic A, et al. Ethylene inhibits rice root elongation in compacted soil via ABA- and auxin-mediated mechanisms. *PNAS.* 2022Jul 26;119(30):e2201072119. <https://doi.org/10.1073/pnas.2201072119>.
52. Roosjen M, Paque S, Weijers D. Auxin Response Factors: output control in auxin biology. *J Exp Bot.* 2018Jan 4;69(2):179–88. <https://doi.org/10.1093/jxb/erx237>.
53. Ulmasov T, Hagen G, Guilfoyle TJ. ARF1, a transcription factor that binds to auxin response elements. *Science.* 1997;276(5320):1864–8. <https://doi.org/10.1126/science.276.5320.1865>.
54. Tiwari SB, Hagen G, Guilfoyle T. The roles of auxin response factor domains in auxin-responsive transcription. *Plant Cell.* 2003;15:533–43. <https://doi.org/10.1105/tpc.008417>.
55. Guilfoyle TJ, Hagen G. Auxin response factors. *Curr Opin Plant Biol.* 2007;10:453–60. <https://doi.org/10.1016/j.pbi.2007.08.014>.
56. Kato H, Kouno M, Takeda M, Suzuki H, Ishizaki K, Nishihama R, Kohchi T. The roles of the sole activator-type auxin response factor in pattern formation of *Marchantia polymorpha*. *Plant Cell Physiol.* 2017Oct 1;58(10):1642–51. <https://doi.org/10.1093/pcp/pcx095>.
57. Yanlin Li, Zhihong G, Juan S, et al. Plant Auxin Response Factor ARF and Growth and Development [J]. *J Plant Physiol.* 2017;53(10):17. <https://doi.org/CNKI:SUN:ZWSL.0.2017-10-004>.
58. Gutierrez L, Bussell JD, Pacurac DL, Schwambach J, Pacurac M, Bellini C. Phenotypic plasticity of adventitious rooting in *Arabidopsis* is controlled by complex regulation of AUXIN RESPONSE FACTOR transcripts and microRNA abundance. *Plant Cell.* 2009;21:3119–32. <https://doi.org/10.1105/tpc.108.064758>.
59. Mallory AC, Bartel DP, Bartel B. MicroRNA-directed regulation of *Arabidopsis* AUXIN RESPONSE FACTOR17 is essential for proper development and modulates expression of early auxin response genes. *Plant Cell.* 2005;17:1360–75. <https://doi.org/10.1105/tpc.105.031716>.
60. Ren ZX, Liu RY, Gu WT, Dong XC. The *Solanum lycopersicum* auxin response factor SlARF2 participates in regulating lateral root formation and flower organ senescence. *Plant Sci.* 2017;256:103–11. <https://doi.org/10.1016/j.plantsci.2016.12.008>.
61. Carey NS, Krogan NT. The role of AUXIN RESPONSE FACTORS in the development and differential growth of inflorescence stems. *Plant Signal Behav.* 2017;12(4):e1307492. <https://doi.org/10.1080/15592324.2017.1307492>.
62. Wang B, Li S, Zou L, et al. Natural variation MeMYB108 associated with tolerance to stress-induced leaf abscission linked to enhanced protection against reactive oxygen species in cassava [J]. *Plant Cell Rep.* 2022;41(7):1573–87. <https://doi.org/10.1007/s00299-022-02879-6>.
63. Kobayashi K, Suzuki T, et al. Transcriptional repression by MYB3R proteins regulates plant organ growth. *EMBO JOURNAL.* 2015 Aug 4;34(15):1992–2007. <https://doi.org/10.15252/emboj.201490899>.
64. Haga N, Kato K, Murase M, Araki S, Kubo M, Demura T, Suzuki K, Müller I, Voss U, Jürgens G, Ito M. R1R2R3-Myb proteins positively regulate cytokinesis through activation of KNOLLE transcription in *Arabidopsis thaliana*. *Development.* 2007Mar;134(6):1101–10. <https://doi.org/10.1242/dev.02801>.
65. Hanano S, Stracke R, Jakoby M, Merkle T, Domagalska MA, Weisshaar B, Davis SJ. A systematic survey in *Arabidopsis thaliana* of transcription factors that modulate circadian parameters. *BMC Genomics.* 2008Apr;21(9):182. <https://doi.org/10.1186/1471-2164-9-182>.
66. Vilarrasa-Blasi J, González-García MP, Frigola D, Fàbregas N, Alexiou KG, López-Bigas N, Rivas S, Jauneau A, Lohmann JU, Benfey PN, Ibañes M, Caño-Delgado AI. Regulation of plant stem cell quiescence by a brassinosteroid signaling module. *Dev Cell.* 2014Jul 14;30(1):36–47. <https://doi.org/10.1016/j.devcel.2014.05.020>.
67. Ohme-Takagi M, Shinshi H. Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element [J]. *Plant Cell.* 1995;7(2):173–82. <https://doi.org/10.1105/tpc.7.2.173>.
68. Adjepong-Danquah J, Gracen VE, Offei SK, Asante IK, Manu-Aduening J. Genetic variability in storage root bulking of cassava genotypes under irrigation and no irrigation. *Agriculture & Food Security.* 2016;5(9):1–12. <https://doi.org/10.1186/s40066-016-0055-7>.
69. El-Sharkawy MA, Hernandez AD, Hershey C. 1992. Yield stability of cassava during prolonged mid-season water-stress. *Experimental Agriculture.* 1992;28(2):165–174. <https://doi.org/10.1017/S0014479700019608>.
70. Ali Z, Sarwat SS, Karim I, et al. Functions of plant's bZIP transcription factors [J]. *Pakistan Journal of Agricultural Sciences*, 2016, 53(2): 303–314. <https://doi.org/10.21162/pakjas/16.2043>.
71. Hossain MA, Cho JI, Han M, et al. The ABRE-binding bZIP transcription factor OsABF2 is a positive regulator of abiotic stress and ABA signaling in rice [J]. *J Plant Physiol.* 2010;167(17):1512–20. <https://doi.org/10.1016/j.jplph.2010.05.008>.
72. Leene JV, Blomme J, Kulkarni SR, et al. Functional characterization of the *Arabidopsis* transcription factor bZIP29 reveals its role in leaf and root development [J]. *J Exp Bot.* 2016;67(19):5825–40. <https://doi.org/10.1093/jxb/erw347>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.