



OPEN ACCESS

EDITED BY

William Underwood,
Agricultural Research Service (USDA),
United States

REVIEWED BY

Simon Peter Abah,
National Root Crops Research Institute
(NRCRI), Nigeria
Kenneth Fafa Egbadzor,
Ho Polytechnic, Ghana

*CORRESPONDENCE

Daniel Kwadjo Dzidzienyo
✉ ddzidzienyo@waccci.ug.edu.gh

RECEIVED 06 July 2025

ACCEPTED 21 August 2025

PUBLISHED 10 September 2025

CITATION

Olayinka AF, Dzidzienyo DK, Mbanjo EGN,
Offei SK, Tongoona PB, Danquah EY, Egesi C
and Rabbi IY (2025) Genome-wide
association study identifies novel genes
for plant architecture and yield traits in
cassava (*Manihot esculenta* Crantz).
Front. Plant Sci. 16:1660789.
doi: 10.3389/fpls.2025.1660789

COPYRIGHT

© 2025 Olayinka, Dzidzienyo, Mbanjo, Offei,
Tongoona, Danquah, Egesi and Rabbi. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Genome-wide association study identifies novel genes for plant architecture and yield traits in cassava (*Manihot esculenta* Crantz)

Abiodun Fatai Olayinka^{1,2}, Daniel Kwadjo Dzidzienyo^{1,3*},
Edwige Gaby Nkouaya Mbanjo², Samuel Kwame Offei¹,
Pangirayi Bernard Tongoona¹, Eric Yirenkyi Danquah¹,
Chiedozie Egesi^{4,5} and Ismail Yusuf Rabbi²

¹West Africa Centre for Crop Improvement (WACCI), College of Basic and Applied Sciences, University of Ghana, Accra, Ghana, ²International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria,

³Biotechnology Centre, College of Basic and Applied Sciences, University of Ghana, Accra, Ghana,

⁴Cassava Research Program, National Root Crops Research Institute (NRCRI), Umudike, Nigeria,

⁵Department of Plant Breeding and Genetics, Cornell University, Ithaca, NY, United States

Cassava (*Manihot esculenta* Crantz) cultivars with compact plant types and moderate plant heights are required for mechanical farming to boost productivity. Plant architecture is a complex trait controlled by environmental and genetics factors. However, little is known about the genetic basis of cassava plant architecture. This research sought to bridge the knowledge gap by elucidating the genetic basis of traits related to plant architecture, yield, and productivity in cassava. A panel of 453 cassava clones developed at the International Institute of Tropical Agriculture was genotyped using two distinct genotyping platforms: low-density DArTseq and DArTag. Plant architecture, yield, and productivity-related traits were evaluated at three locations across two growing seasons in Nigeria. Following data filtering, 420 clones, 54,574 DArTSeq, and 2,527 DArTag single-nucleotide polymorphism (SNP) markers were used for genome-wide association studies (GWAS). Of the 16 SNPs identified by GWAS using DArTSeq markers, only one was detected during validation, and the remaining SNPs may be false positives. Sixteen SNPs were found to be significant using DArTag markers. Fifteen of these were associated with 21 putative candidate genes for five plant architecture traits (17 genes) and three yield traits (four genes). Six of the identified candidate genes were novel. The identified candidate genes were associated with various metabolic processes, including plant architecture, adaptation, root development, plant growth, and stress response. The limited number of significant markers identified using DArTSeq markers could be explained by the large gaps and uneven marker distribution observed across the genome with the DArTSeq platform compared to DArTag. The findings of this study provide new insights

into the genetic basis of plant architecture and yield in cassava. Cassava breeders could leverage this knowledge to optimize plant architecture and yield in cassava through marker-assisted selection and targeted manipulation of candidate genes.

KEYWORDS

SNP markers, *Manihot esculenta* Crantz, mechanical farming, DArTseq, DArTag, novel genes

Introduction

Cassava is Africa's second most important food crop. Over 40% of the continent's population relies on it to fulfil their daily caloric and nutritional needs (Alves, 2002; Nweke, 2004; FAO, 2020). Cassava starch is a versatile raw material for food, feed, and industrial applications (Li et al., 2017). Despite the huge prospects of its value chain, cassava production in Nigeria remains low at approximately 10 tons per hectare, while demand is increasing (FAO, 2021). The increasing demand for cassava-based products necessitates the drive to increase crop performance to boost productivity.

Plant architecture plays a crucial role in agriculture. It affects crop growth, yield, and stress resistance (Lauri and Lespinasse, 2001; Rymaszewski et al., 2017; Mansaray et al., 2020). Cassava cultivars with appropriate plant architecture are needed to maximize cassava yield potential, mechanize cassava farming, and boost productivity per unit area (Patiño et al., 2002; FAO, 2013). Plant architecture and yield are complex traits known to be controlled by both environmental and genetic factors (Zhang et al., 2017; Yu et al., 2020). Several studies have attempted to dissect the genetic basis of cassava plant architecture-related traits and their effects on cassava yield using biparental populations, with limited breakthroughs (OkogBenin and Fregene, 2003; Mora Moreno et al., 2016; Srisawad et al., 2023).

Genome-wide association studies (GWAS), which use a diversity panel, are gaining popularity for marker-trait associations (Dash et al., 2021; Uffellmann et al., 2021). GWAS has helped unravel the genetic architecture of important crop plant traits, including yield and quality traits, early bulking and storage root formation, and nitrogen use efficiency in cassava (Abah et al., 2024; Aghogho et al., 2024; Mbe et al., 2024), plant architecture in maize (Lu et al., 2024), and drought tolerance in wheat (Yang et al., 2024). Single-nucleotide polymorphisms (SNPs) have been the marker of choice for GWAS analyses owing to their stable inheritance patterns, low mutation rates, and compatibility with high-throughput genotyping technologies (Morgil et al., 2020; Panahi et al., 2024).

The reliability and effectiveness of GWAS could be influenced by the density and distribution of SNP markers throughout the genome (Spindel et al., 2015; Uffellmann et al., 2021; Aalborg et al.,

2024). Information on the impact of marker density and distribution on GWAS output is limited. Two genotyping platforms, DArTseq and DArTag, were used to conduct GWAS analyses. This study aimed to investigate the genetics of plant architecture and yield-related traits in cassava, as well as how marker density and distribution affect GWAS effectiveness and reliability.

Materials and methods

The cassava genotypes used in this study included 453 genotypes from a preliminary yield trial (PYT) developed at the IITA and five commercial varieties used as checks. The details of the accessions, crosses, and parents are reported in [Supplementary Table S1](#). Descriptions of the commercial check varieties are provided in [Supplementary Table S2](#). The trial was conducted in three locations in Nigeria: Ikenne (Lat 6.8718° N, Long 3.7106° E, humid forest zone), Onne (Lat 4.7363° N, Long 7.1545° E, humid forest zone), and Mokwa (Lat 9.2934° N, Long 5.0493° E, Southern Guinea Savannah zone) across two planting seasons (2020/2021, and 2021/2022). The trial was laid out in an augmented block design with two replications per location for each season. Each plot comprised three rows with three plants planted on ridges per row. A spacing of 1m was maintained between ridges, while a spacing of 0.8m was observed within plants in a row.

Phenotypic data collection

Ten above-ground plant traits: stem diameter (STMDI9), shoot weight (SHTWT), plant height at 9 months after planting (PLTHT9), number of lodged plants per plot (LODG), branching habit at 9 months after planting (BRNHB9), angle of branching (ANGBR9), number of plants per stand (PPSTD9), height at first branch at 9 months after planting (BRNHT9), plant height at 6 months after planting (PLTHT6), top yield (TYLD), height at first branch at 6 months after planting (BRNHT6), and eight yield-related traits: fresh root yield (FYLD), dry yield (DYLD), starch content (SC), dry matter content (DM), number of harvested roots (RTNO), fresh root weight (RTWT), and harvest index (HI), were

measured at specific phenological stages of the crop (3, 6, 9, and 12 months after planting). The phenological stage and method used to assess the traits were based on the parameters described by (Fukuda et al., 2010). The details of the traits evaluated and the evaluation procedures are presented in [Supplementary Table S3](#).

SNP genotyping and marker filtering

Leaf samples were collected from a single plant per cassava clone, 4–6 leaf discs of 6mm in diameter were collected from healthy young leaves. The leaf discs were placed in the corresponding wells of the DNA plate on ice in a sampling bag. The plate was immediately transferred to the laboratory, where it was covered with Parafilm and kept at -80°C in a freezer prior to freeze-drying. The samples were freeze-dried at the IITA Bioscience Centre using a lyophilizer (LABCONCO FreeZone 18 Liter -50°C freeze-dryer) operated at -51°C and 5.0 pa for a minimum of 72 hrs. The freeze-dried leaf samples were shipped to Diversity Arrays Technology in Canberra, Australia, where genotyping was performed using low-density DArTseq and DArTag technology. Upon receipt, the raw SNP data were filtered using PLINK 1.9 software (Purcell et al., 2007; Chang et al., 2015). SNP markers with less than 5% minor allele frequency (MAF) of more than 10% missing call rate were pruned.

Statistical analyses of phenotypic data

Plant architecture and yield-related traits were analyzed using the spatial single-trial model fitted using the R package 'SpATS' (Rodriguez-Alvarez et al., 2018). The mathematical formula for the adopted model is as follows:

$$y_{ijk} = \mu + g_i + r_j + b_k + s(x_{ijk}, y_{ijk}) + e_{ijk}$$

where y_{ijk} is the phenotypic value of the i^{th} genotype in the j^{th} block and k^{th} incomplete block, μ is the overall mean, g_i is the random effect of the i^{th} genotype, r_j is the random effect of the j^{th} block, b_k is the random effect of the k^{th} incomplete block, $s(x_{ijk}, y_{ijk})$ is the smooth bivariate function of the row and column coordinates of the plot, and e_{ijk} is the residual error. The best linear unbiased prediction (BLUP) values derived from this model were used as phenotypic values in the marker-trait association analysis.

The linear mixed model for the analysis of the pooled data (obtained from the three test environments) was fitted as follows using the lmer function from the lme4 package (Bates et al., 2015) of the R software (R Core Team, 2020).

$$\eta_{ijklm} = \mu + \alpha_i + \beta_{ij} + \gamma_k + \delta_{ik} + \varepsilon_{ijl} + e_{ijkl}$$

where η_{ijklm} is the yield value for the m^{th} plot in the l^{th} block of the j^{th} replicate of the i^{th} environment for the k^{th} accession; μ is the overall mean; α_i is the fixed effect of the i^{th} environment; β_{ij} is the fixed effect of the interaction between the i^{th} environment and the j^{th} replicate; γ_k is the random effect of the k^{th} accession; δ_{ik} is the random effect of the interaction between the i^{th} environment and

the k^{th} accession; ε_{ijl} is the random effect of the interaction between the i^{th} environment, the j^{th} replicate, and the l^{th} block; and e_{ijkl} is the random residual error.

The random effects γ_k , δ_{ik} , and ε_{ijl} are assumed to be independent and normally distributed with a mean of 0 and variance components σ^2_a , σ^2_{ea} , and σ^2_b , respectively. The residual error e_{ijkl} is also assumed to be independent and normally distributed with a mean of 0 and variance component σ^2_e .

The best linear unbiased prediction (BLUP) values derived from this model were used as the pooled phenotypic values across environments. The broad-sense heritability (H^2) values were calculated as follows:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{ge}^2}{r} + \frac{\sigma_e^2}{er}}$$

Where H^2 is the broad-sense heritability estimate, σ_g^2 is the genetic variance component of the accession effect, σ_e^2 is the variance component of the residual error, and r is the number of replicates.

Genotypic correlation analysis for the plant architecture and yield-related traits was performed using the statistical software Meta-R Version 6.0 (Alvarado et al., 2020). Variance components were estimated through a mixed linear model fitted using Restricted Maximum Likelihood (REML) to obtain estimates of genotypic variance, variance due to genotype-by-environment interaction, and residual variance prior to the computation of the genotypic correlations among traits.

The formula for the mixed linear model is as follows:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_k + e_{ijk}$$

y_{ijk} is the observation for the i^{th} genotype in the j^{th} block and the k^{th} replication

μ is the overall mean

α_i is the fixed effect of the i^{th} genotype

β_j is the random effect of the j^{th} block

$(\alpha\beta)_{ij}$ is the random interaction effect of the i^{th} genotype and the j^{th} block

γ_k is the random effect of the k^{th} replication within the block

e_{ijk} is the residual error.

The estimates of the genetic correlation coefficient values were obtained using the variance components from the mixed linear model:

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

y is the vector of observations

β is the vector of fixed effects

u is the vector of random effects

e is the vector of residuals

X and Z are design matrices.

The genetic correlation between the two traits was calculated as follows:

$$rg = \frac{\sigma_{g1g2}}{\sqrt{\sigma_{g1}^2 \sigma_{g2}^2}}$$

$\sigma_{g_1g_2}$ is the covariance between the random effects of the two traits. $\sigma_{g_1}^2$ and $\sigma_{g_2}^2$ are the variances of the random effects of the two traits, respectively. Data visualization was performed with the R package ‘corrplot’ version 0.92 (Wei and Simko, 2021). Path Analysis was performed using the R package ‘lavaan’ version 0.6-17 (Rosseel, 2012) and visualized using the R package ‘semPlot’ (Epskamp et al., 2019).

Marker coverage, SNP density, and linkage disequilibrium

Marker coverage and SNP density were visualized using the SRplot platform (Tang et al., 2023). Linkage Disequilibrium (LD) analysis was conducted using PLINK v1.9 (Purcell et al., 2007). Marker pairs exhibiting perfect LD scores ($r^2=1$) were excluded before further analysis. An LD decay plot was created utilizing R software (R Core Team, 2020).

Population structure analysis

Population structure analysis was performed using ADMIXTURE (Alexander et al., 2009), which uses the maximum likelihood estimate to assign genotypes to putative populations (K). Cross-validation was performed to determine the optimal number of clusters. The results of the population structure analysis were visualized using the R package ‘ggplot2’ (Wickham, 2016). Ward’s minimum variance method (ward.D2) was utilized for hierarchical clustering on the Q-matrix produced by ADMIXTURE. The R package ‘dendextend’ (Galili, 2015) was used to plot the dendrogram.

Genetic association analysis

GWAS analysis was performed with the BLUP values obtained from the trials conducted in each of the three test environments as well as the pooled data across all the environments through the Mixed Linear Model (MLM) described in (Zhang et al., 2010), using the R package Genome Association and Prediction Integrated Tools (GAPIT) (Lipka et al., 2012). This model decomposes the observed phenotype (y) into fixed effects ($X\beta$), random genetic effects (Zu), and residual effects (e). The formula is:

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

where X is the design matrix for fixed effects, Z is the design matrix for random effects, β is the vector of fixed effects, u is the vector of random effects, and e is the vector of residuals.

In addition to the MLM model, the Fixed and random model Circulating Probability Unification (FarmCPU) model outlined in (Liu et al., 2016) was used for GWAS carried out using DArTseq markers. FarmCPU uses a modified MLM method, Multiple Loci Mixed Model (MLMM), and incorporates multiple markers simultaneously as covariates in a stepwise MLM to partially

remove the confounding effects between testing markers and kinship. The formula is:

$$y = X\beta + Z_1u_1 + Z_2u_2 + e$$

where X is the design matrix for fixed effects, Z_1 is the design matrix for random effects from kinship, Z_2 is the design matrix for random effects from multiple markers, β is the vector of fixed effects, u_1 is the vector of random effects from kinship, u_2 is the vector of random effects from multiple markers, and e is the vector of residuals.

These models deployed different approaches in accounting for the effects of population structure and environmental variation thereby precluding any possibility of false discovery. However, the following criteria were used in calling SNPs that share a significant correlation with the plant architecture and yield traits: Bonferroni correction, the threshold of the P-value, Manhattan plot and the QQ plot.

The Bonferroni-corrected p-value [$\log_{10}(0.05/\text{number of SNPs})$] was used as a cutoff value for identifying significant SNPs. Manhattan plots were generated to visualize the results, while quantile-quantile (QQ) plots showed the distribution of observed p-values against those predicted under the null hypothesis and captured potential inflation or deflation of the test statistic. Additionally, we used the mixed linear model, which excludes (MLMe) and includes (MLMi) candidate markers in the genetic relationship matrix (GRM) via a leave-one chromosome-out analysis implemented in GCTA (Yang et al., 2011; Yang et al., 2014). The exclusion of the tested SNP from the relationship matrix (GRM) used in the random effect reduces proximal contamination or overcorrection, thereby preventing deflating associations on the same chromosome as the tested SNP. This helped verify the authenticity of the significant SNPs identified using the other two models.

Candidate gene analysis

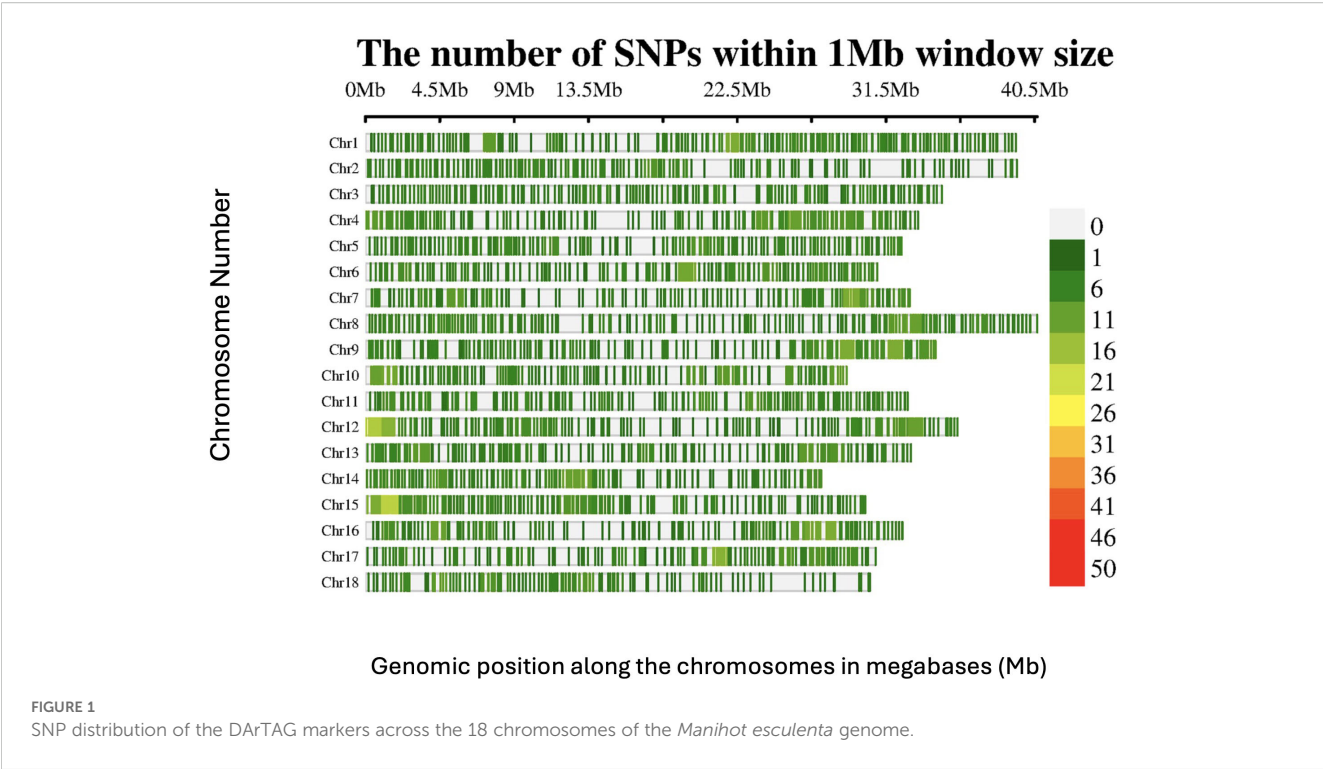
The significant SNPs were mapped onto genes within the 5,000 bp windows using *Manihot esculenta* v7.1 of the Phytozome genome browser (Goodstein et al., 2014). The UniProt Consortium database (Aleksander et al., 2023) was used for gene ontology annotation.

Results

Marker coverage and SNP density

A total of 2,527 DArTag SNP markers, uniformly distributed over the 18 chromosomes, were retained after marker filtering (Figure 1). The average number of SNPs per 1MB window across the genome ranged from 1 to 26 SNPs for most of the genomic regions. Chromosome 8 had the highest number of SNPs (164), whereas chromosome 18 had the lowest (106).

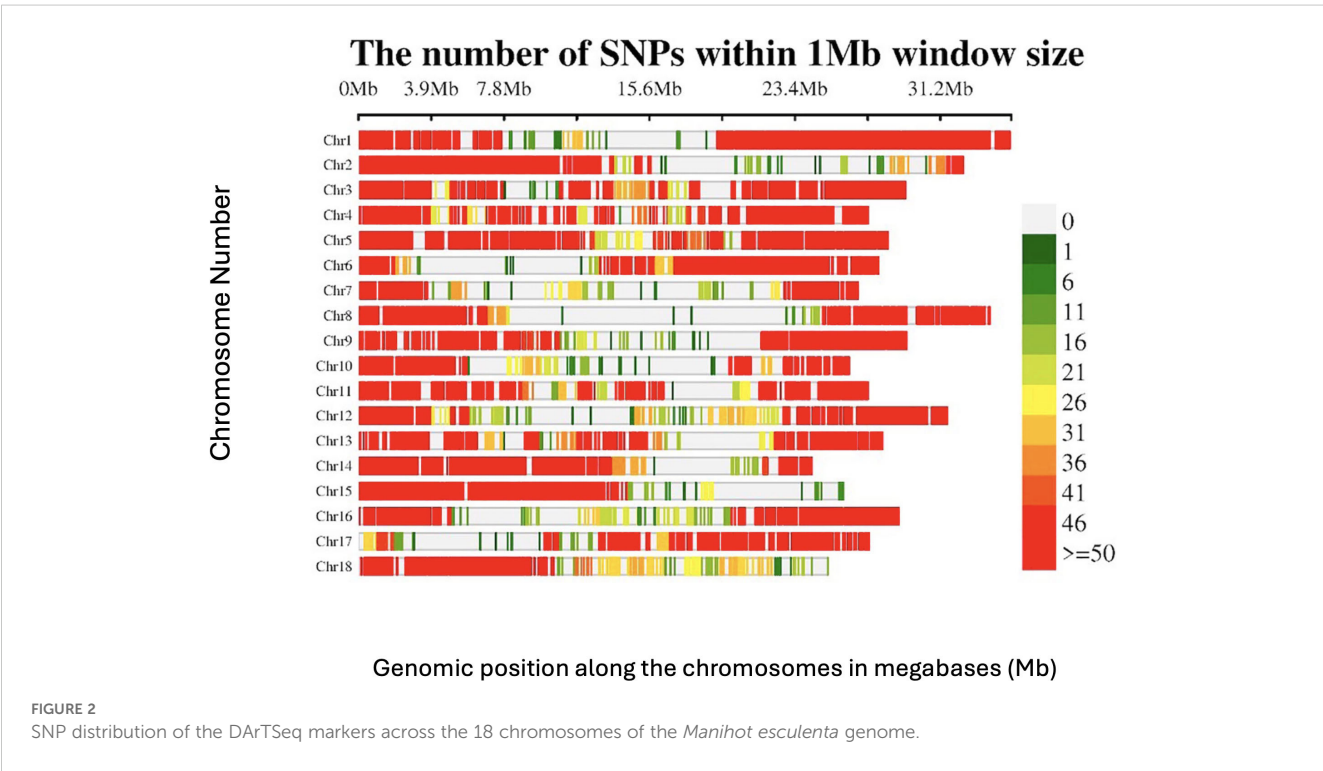
The 54,574 filtered DArTseq SNP markers were not evenly distributed across the 18 chromosomes (Figure 2). Some genomic regions, especially those close to the centromeric and telomeric



regions, showed very low SNP densities, and in some cases, no markers, whereas other regions displayed a dense clustering of SNPs (>50 SNPs per 1 Mb window). Chromosome 1 had the highest number of SNPs (7235), with an average of 207 SNPs/Mb, whereas chromosome 7 had the fewest (1588), with approximately 59 SNPs/Mb.

Linkage disequilibrium

The extent of LD decay was compared between the two marker types used (Figures 3, 4). A rapid LD decay was observed using DArTag markers. The LD between two SNPs fell to 0.2 (the standard threshold that indicates the end of LD) at a relatively



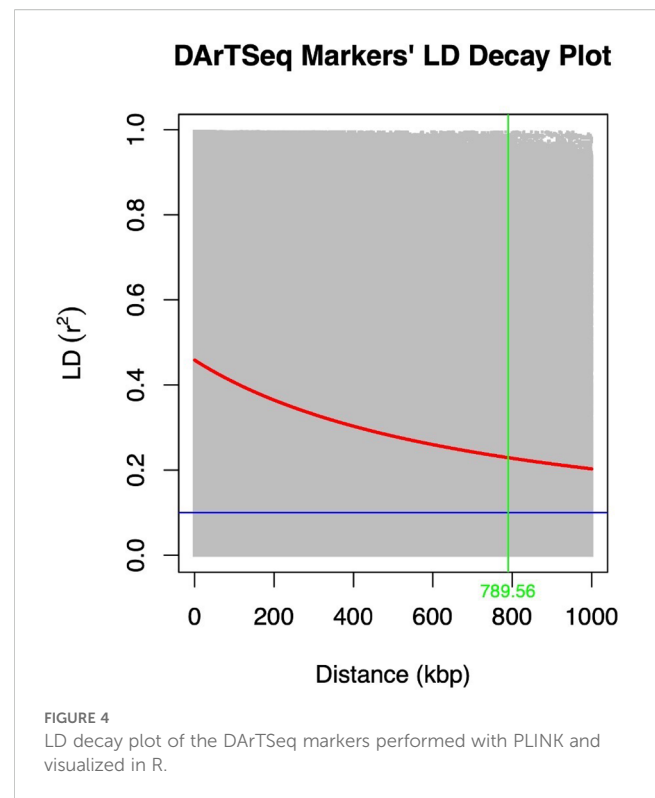
more rapid rate, approximately 229 kbp (Figure 3), contrary to the output obtained with the DArTSeq markers (Figure 4), where the LD between two SNPs fell to 0.2 at a longer distance of approximately 790 kbp.

Population structure analysis

Population structure analyses revealed five subgroups (Figure 5). The cassava accessions were developed from crosses made using parents with diverse genetic backgrounds (Supplementary Table S1). The plot also shows that there is a balance between the extent of admixture and homogeneity within the population. This shows that there exists enough genetic diversity within the population to capture the variation in the plant architecture and yield traits.

Variance Components and broad-sense heritability of plant architecture and yield-related traits

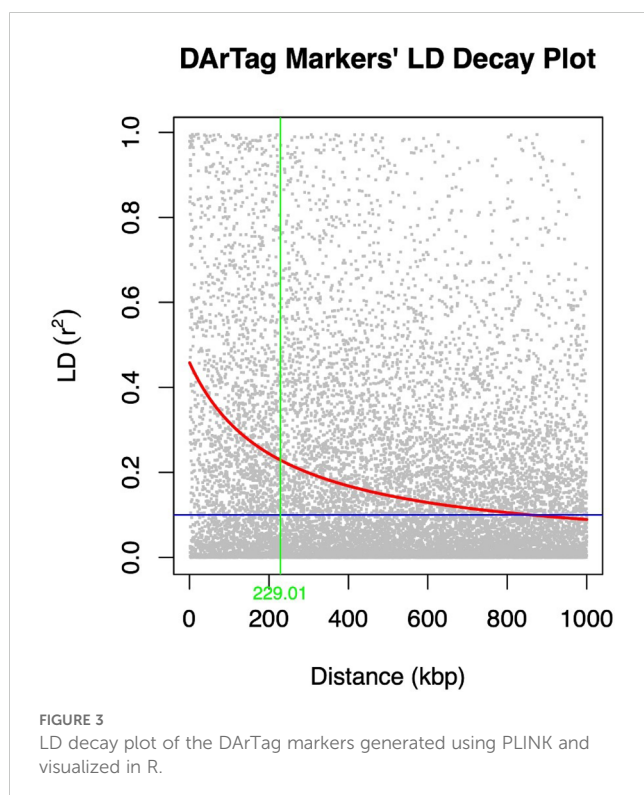
All 18 traits exhibited greater genotypic variance than both the error variance and the variance due to the interaction between genotype and environment (Table 1). Both genetic and interaction factors had a strong influence on the performance of the accessions for all measured traits ($p < 0.001$). The broad-sense heritability varied from moderately high (0.62; SHTWT) to high (0.88; BRNHT9). The broad-sense heritability for key plant architecture



traits, viz., ANGBR9 (0.78), STMDI9 (0.73), PLTHT9 (0.80), PLTHT6 (0.83), and BRNHT6 (0.86) were found to be high, while the following yield-related traits, including DM (0.68), SC (0.70), FYLD (0.67), and HI (0.86) had relatively high broad-sense heritability values (Table 1). The phenotypic and genotypic coefficients of variation ranged from 8.30% and 6.84% (DM), respectively, to 78.28% and 65.03% (LODG) (Table 1).

The exploratory analysis revealed the interrelationships between plant architecture and yield-related traits. The genetic correlation coefficient matrix revealed positive and significant relationships between FYLD and DYLD ($r = 0.70 - 0.89$, $p < 0.001$), PLTHT6 and PLTHT9 ($r = 0.93 - 1.00$, $p < 0.001$), STMDI9 and PLTHT6 ($r = 0.49 - 0.63$, $p < 0.001$), and STMDI9 and PLTHT9 ($r = 0.51 - 0.68$, $p < 0.001$) (Figures 6–8). Negative and significant relationships were observed between BRNHT6 and BRNLEV9 ($r = -0.03 - (-0.83)$, $p < 0.001$) and BRNHT9 and BRNLEV9 ($r = -0.81 - (-0.82)$, $p < 0.001$) (Figures 6–8). In addition to these, significant negative relationships were observed between LODG and FYLD ($r = -0.18 - (-0.65)$, $p < 0.001$), RTNO ($r = -0.19 - (-0.57)$, $p < 0.001$), PPSTD9 ($r = -0.04 - (-0.30)$, $p < 0.001$) and HI ($r = -0.40 - (-0.73)$, $p < 0.001$) respectively. Similarly, there was a strong negative and significant correlation between HI and some crucial plant architecture traits, viz., STMDI9 ($r = -0.38 - (-0.61)$, $p < 0.001$), PLTHT6 ($r = -0.36 - (-0.52)$, $p < 0.001$), and PLTHT9 ($r = -0.33 - (-0.42)$, $p < 0.001$) (Figures 6–8).

Path coefficient analysis demonstrated that plant architectural traits had both direct and indirect effects on fresh root yield and fresh root weight in cassava. For the trials conducted in all three test environments (Figures 9–11), RTWT had an absolute (1.0) positive contribution to FYLD. However, HI had a strong positive



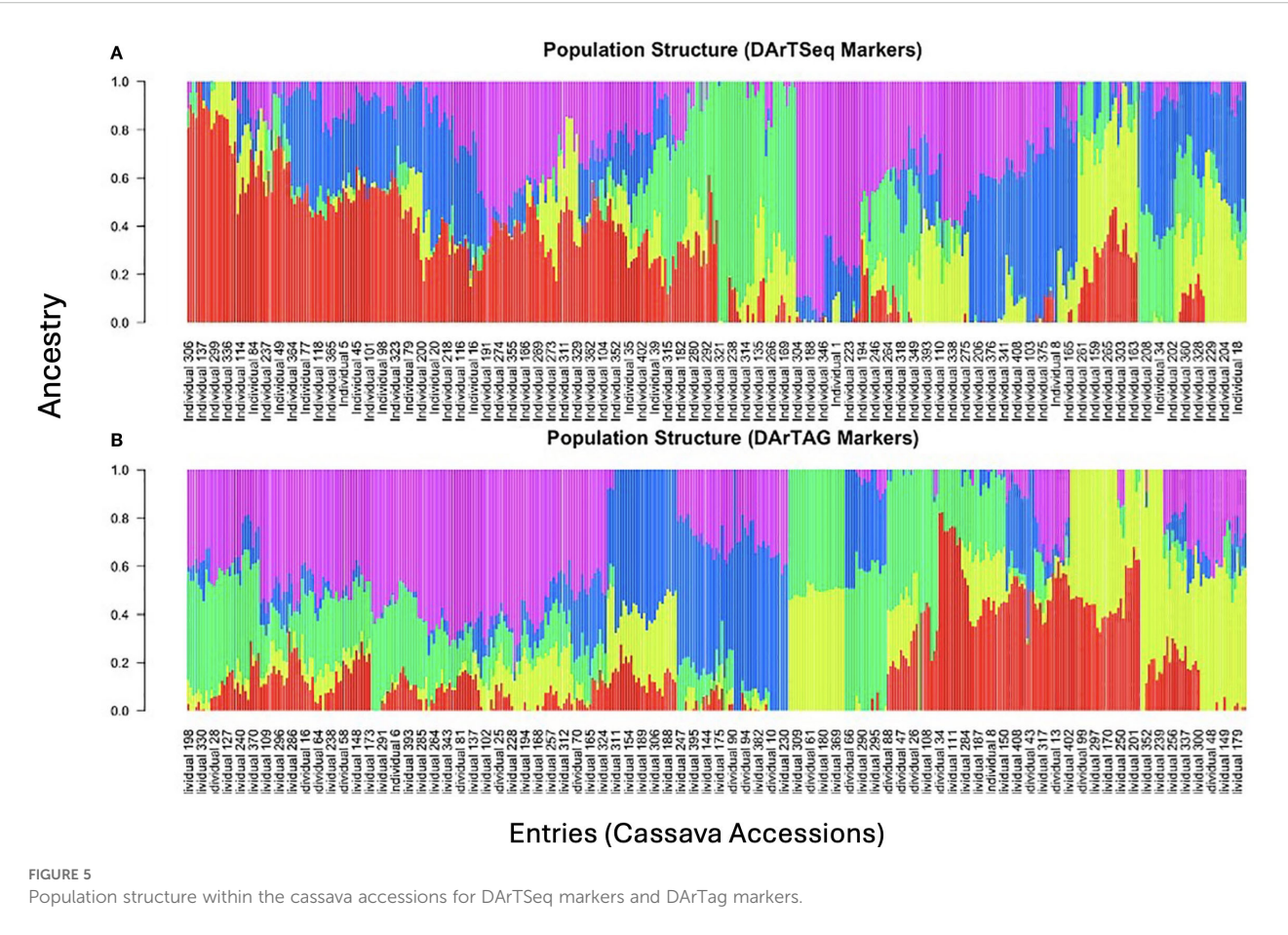


TABLE 1 Estimates of variance components and broad-sense heritability for plant architecture and yield-related traits in 438 cassava accessions evaluated across three locations in Nigeria.

Trait	Mean	σ^2_g	$\sigma^2_{G \times E}$	σ^2_e	H^2	PCV(%)	GCV(%)
STMDI9	24.00	5.56	0.81	1.28	0.73	11.52	9.82
FYLD	23.00	37.20	9.00	9.32	0.67	32.40	26.52
DYLD	7.80	3.70	0.97	1.08	0.64	30.74	24.66
SC	19.80	5.84	1.71	0.85	0.70	14.63	12.21
DM	32.60	4.97	1.53	0.80	0.68	8.29	6.84
RTWT	19.20	25.50	6.67	6.43	0.66	32.36	26.30
SHTWT	21.00	21.80	4.73	8.50	0.62	28.19	22.23
PLTHT9	176.90	507.80	40.83	84.67	0.80	14.23	12.74
HI	0.47	0.01	0.00	0.00	0.86	22.98	21.28
BRNHB9	2.60	0.17	0.04	0.03	0.71	18.78	15.86
ANGBR9	81.00	60.70	6.00	10.93	0.78	10.88	9.62
RTNO	44.60	131.40	29.87	41.58	0.65	31.93	25.70
PPSTD9	2.00	0.10	0.01	0.05	0.63	19.90	15.81
BRNHT9	83.80	580.00	33.63	48.00	0.88	30.69	28.74
PLTHT6	158.00	459.00	30.57	64.67	0.83	14.90	13.56
TYLD	25.90	31.50	6.40	13.17	0.62	27.59	21.67
BRNHT6	74.00	477.60	36.00	44.93	0.86	31.94	29.53
BRLEV9	2.30	0.34	0.06	0.03	0.78	28.62	25.35

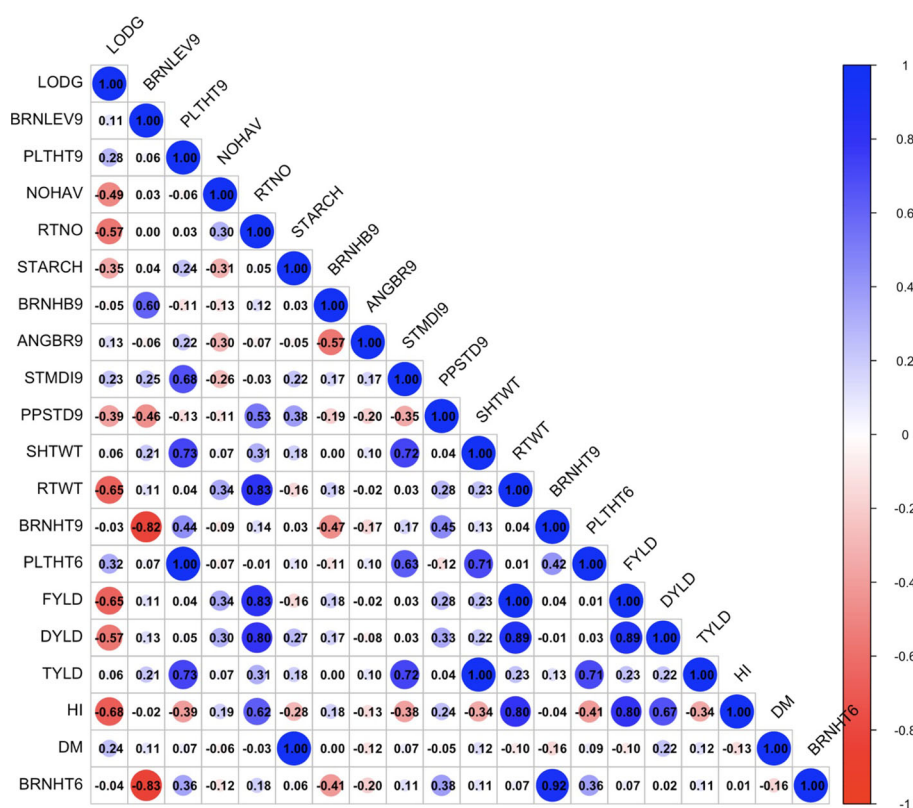


FIGURE 6

Genotypic correlation coefficient plot of plant architecture and yield traits in Ikenne trial. NOHAV, number of harvested plants per plot; RTNO, number of harvested roots per plot; SHTWT, shoot weight; SC, starch content; FYLD, fresh root yield; DYLD, dry yield; HI, harvest index; LODG, number of lodged plants per plot; PLTHT6, plant height at 6 months after planting; BRNHT6, height at first branch at 6 months after planting; PLTHT9, plant height at 9 months after planting; BRNHT9, height at first branch at 9 months after planting; BRNLEV9, level of branching at 9 months after planting; BRNHB9, branching habit at 9 months after planting; ANGBR9, angle of branching; PPSTD9, number of plants per stand, and STMDI9, stem diameter.

contribution to RTWT (0.69) in the trials conducted at Ikenne and Mokwa (Figures 9, 10), whereas it had no contribution to the output of this trait in the trial conducted at Onne. The contributions of PLTHT9 to FYLD through RTWT were also positive across the three test environments, where a near absolute contribution (0.99) was recorded in the trial conducted at Onne (Figure 11). In contrast, irregular relationships were observed across the test environments for LODG and PLTHT6 with respect to their contributions to RTWT. For LODG, a negative contribution with a low magnitude (-0.14) was reported in two test environments (Figures 9, 10), whereas the contribution was positive (0.13) in the third test environment (Figure 11). Similar results were obtained for PLTHT6, which had a positive contribution (0.23) to RTWT in the Ikenne trial (Figure 9), but negative effects on this trait in the Mokwa (-0.08) and Onne (-0.35) trials (Figures 10, 11).

Marker-trait analysis

Four hundred and twenty (420) accessions, 2,527 DArTag, and 54,574 DArTSeq SNP markers were used in GWAS analyses. The GWAS conducted using the DArTag markers and the mixed linear

model revealed 16 significant marker-trait hits for eight plant architecture traits (ANGBR9, BRNHB9, PLTHT6, PLTHT9, BRNLEV3, BRNLEV9, SHTWT, and TYLD) and three yield traits (RTWT, FYLD, and DYLD) (Table 2; Figure 12). Some of the significant SNP hits were location specific while some were detected across all the locations (Table 3). Significant SNPs were found on chromosomes 1, 2, 6, 8, 11, 12, 14, and 15. ANGBR9 had significant SNPs on chromosomes 12 and 14, while three chromosomes had significant marker-trait associations (MTAs) for more than one trait: chromosome 8 for PLTHT9, FYLD, and RTWT; chromosome 6 for SHTWT and TYLD; and chromosome 12 for ANGBR9 and BRNHB9. The remaining traits, DYLD, BRNLEV3, BRNLEV9, and BRNHB9, only had significant MTAs on chromosomes 2, 1, 15, and 12, respectively (Table 2; Figure 12). Further analyses of the 16 significant SNP hits led to the identification of 21 putative candidate genes for nine traits (Table 4). These included ANGBR9 (7), PLTHT9 (2), BRNLEV3 (1), BRNLEV9 (6), FYLD (1), RTWT (1), SHTWT (1), TYLD (1), and DYLD (3) (Table 4).

Only eight traits had significant SNP hits when DArTseq markers and the FarmCPU model were used (Table 5; Figure 13). These included six plant architecture traits (ANGBR9, PT, VR, BRNLEV6, BRNLEV9, and PPSTD9) and two yield traits (HI and

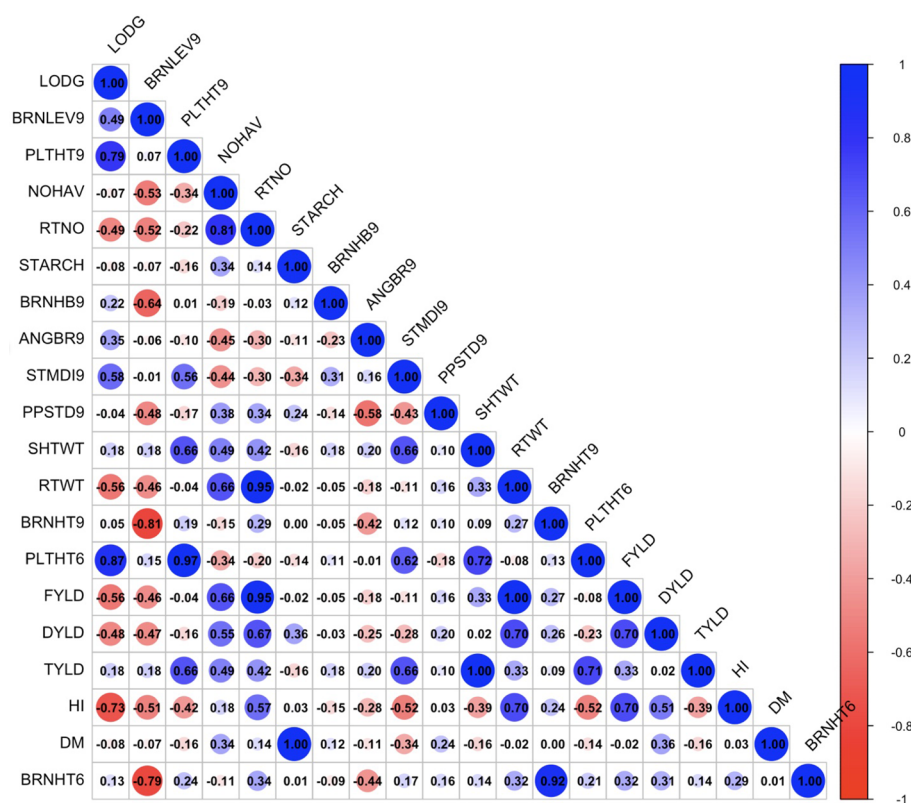


FIGURE 7

Genotypic correlation coefficient plot of plant architecture and yield traits in Mokwa trial. NOHAV, number of harvested plants per plot; RTNO, number of harvested roots per plot; SHTWT, shoot weight; SC, starch content; FYLD, fresh root yield; DYLD, dry yield; HI, harvest index; LODG, number of lodged plants per plot; PLHT6, plant height at 6 months after planting; BRNHT6, height at first branch at 6 months after planting; PLHT9, plant height at 9 months after planting; BRNHT9, height at first branch at 9 months after planting; BRNLEV9, level of branching at 9 months after planting; BRNHB9, branching habit at 9 months after planting; ANGBR9, angle of branching; PPSTD9, number of plants per stand; STMDI9, stem diameter.

DM). The 17 significant SNPs recorded for these traits were found on chromosomes 1, 3, 4, 5, 6, 8, 9, 11, 12, 15, and 16. A total of 19 putative candidate genes were identified near the significant SNPs (Table 6). Significant SNP were validated using MLMe analysis. The validation procedure confirmed the genuineness of each significant DArTag marker as a result of the mixed linear model performed in GAPIT, and the validation analyses performed in GCTA were identical. In contrast, only one of the significant SNPs on chromosome 6 detected using DArTSeq markers and the FarmCPU model (S6_21336567) was detected by the MLMe analysis conducted in GCTA. The Q-Q plots revealed a huge deflection from the observed to expected variation when DArTSeq markers and the FarmCPU model were used, whereas little or no deflection was observed when using DArTag markers and the mixed linear model. This suggests that both the SNPs and candidate genes discovered through the GWAS conducted using the DArTSeq markers and the FarmCPU model were probably artifacts.

Discussion

Plant height and stem diameter are primary plant architectural traits that significantly influence the position and alignment of

other plant architectural traits (Laurans et al., 2024). The negative correlation between FYLD, the primary predictor of yield, and PLTH is consistent with the conclusions of (Edet et al., 2015), who also reported a negative relationship between the two traits. This implies that sustainable improvement in FYLD productivity in cassava could be achieved by attaining a balance in the expression of these inversely related traits.

Heritability is an essential parameter in plant breeding, as it indicates the potential for the genetic improvement of a trait through selection. A couple of plant architecture and yield-related traits were reported to have moderately high to high broad-sense heritability, including FYLD (0.67), DM (0.68), HI (0.86), BRNHT6 (0.86), BRNHT9 (0.88), PLHT9 (0.80), and SC (0.70). Similar results were reported (Santos et al., 2023) for FYLD and for DM. High SC content is one of the selling points of cassava varieties targeted for industrial use. The high value of broad-sense heritability recorded for SC in this study corroborated the findings of (Namakula and Nuwamanya, 2022), who reported a heritability value of 0.76. Given that PLHT9, PLHT6, STMDI9, HI, FYLD, SC, and LODG have high values of broad sense heritability (0.80), (0.83), (0.73), (0.86), (0.67), (0.7), and (0.69), respectively, there is significant potential for genetic improvement of these traits in cassava through selection, thereby making it

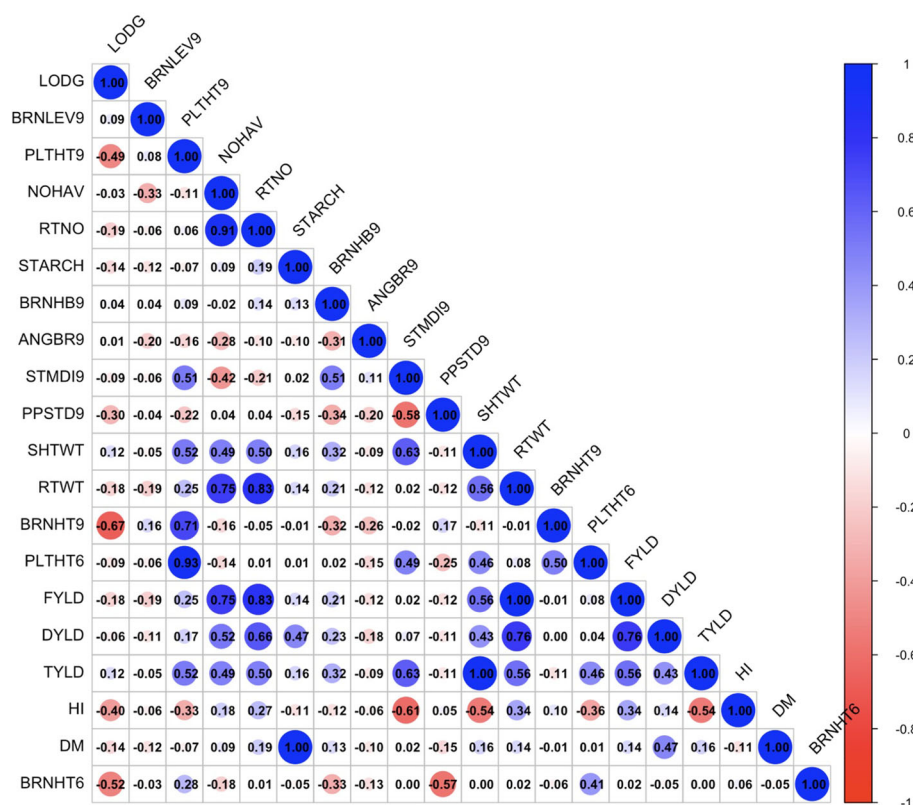


FIGURE 8

Genotypic correlation coefficient plot of plant architecture and yield traits in Onne trial. NOHAV, number of harvested plants per plot; RTNO, number of harvested roots per plot; SHTWT, shoot weight; SC, starch content; FYLD, fresh root yield; DYLD, dry yield; HI, harvest index; LODG, number of lodged plants per plot; PLTHT6, plant height at 6 months after planting; BRNHT6, height at first branch at 6 months after planting; PLTHT9, plant height at 9 months after planting; BRNHT9, height at first branch at 9 months after planting; BRNLEV9, level of branching at 9 months after planting; BRNHB9, branching habit at 9 months after planting; ANGBR9, angle of branching; PPSTD9, number of plants per stand; STMDI9, stem diameter.

feasible to modify these traits to develop cassava varieties with optimal architecture and increased yield.

Marker coverage and SNP density

The DArTseq marker system uses a genome complexity reduction approach, whereby restriction enzymes randomly cut at regions with high frequencies of cutting sites, resulting in an uneven SNP distribution across the genome (Kilian et al., 2012).

This results in the omission and underrepresentation of certain genomic regions, such as repetitive sequences, heterochromatic regions, and areas lacking restriction enzyme recognition sites, leading to observed gaps in genome coverage (Wenzl et al., 2004; Kilian et al., 2012; Sánchez-Sevilla et al., 2015). In contrast, DArTag employs a targeted genotyping approach, where pre-selected SNPs are evenly distributed across the genome, ensuring uniform coverage and minimizing gaps (Hardenbol et al., 2003; Hardigan

et al., 2023). Large gaps, as observed with DArTseq markers, have a significant impact on GWAS results. Regions with sparse SNP coverage may not capture causal variants or nearby loci in linkage disequilibrium (LD), leading to missed associations with traits (Kilian et al., 2012; Sánchez-Sevilla et al., 2015).

Marker-trait association analysis

Subsequent to the output of the significant SNPs validation, the discussion on the GWAS would be limited to the results obtained through the GWAS conducted using the DArTag markers and the mixed linear model. GWAS is a crucial method for investigating the genetic underlining of complex plant traits (Zhu et al., 2008; Bentley et al., 2022; Susmitha et al., 2023). Of the 11 traits with significant SNPs, only ANGBR9 had significant SNPs on multiple chromosomes (chromosomes 12 and 14). This suggests that this trait is complex and influenced by multiple genes. A similar result was recorded for the

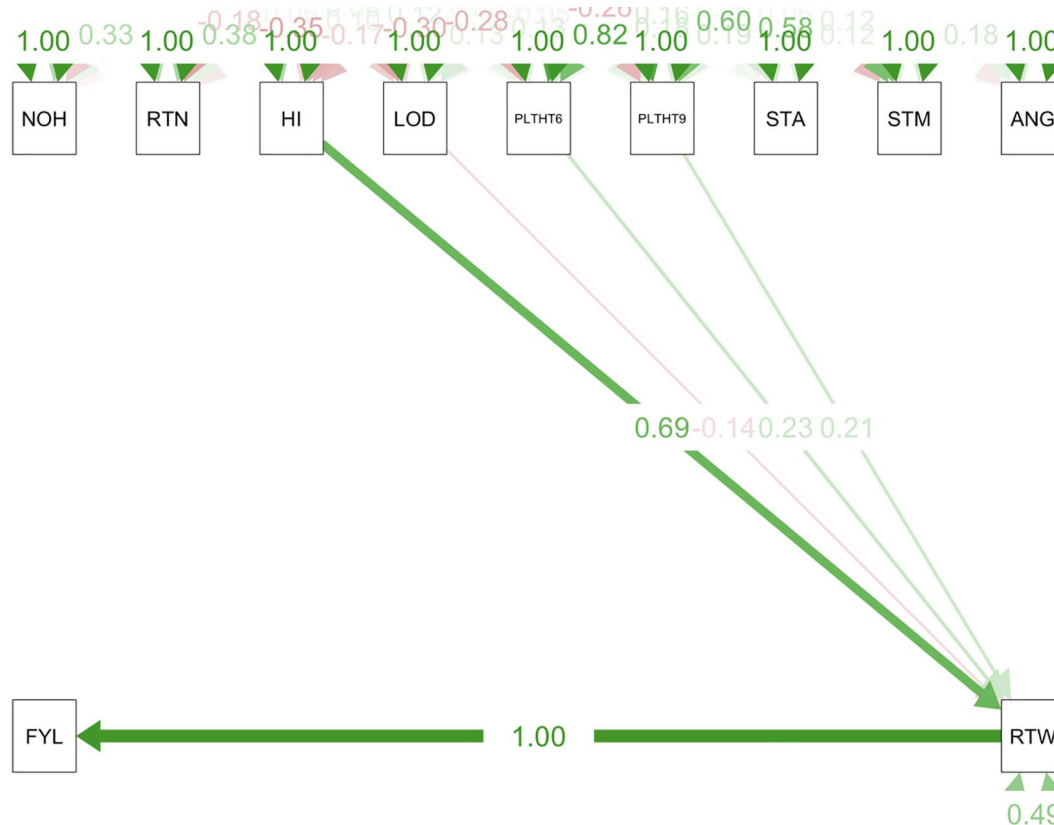


FIGURE 9

Path coefficient analysis plot of plant architecture and yield traits in Ikenne trial. NOH, number of harvested plants per plot; RTN, number of harvested roots per plot; RTW, root weight; SC, starch content; FYL, fresh root yield; HI, harvest index; LOD, number of lodged plants per plot; PLTH6, plant height at 6 months after planting; PLTH9, plant height at 9 months after planting; ANG, angle of branching; STM, stem diameter.

angle of branching in a GWAS experiment conducted on cotton across three test environments (Shao et al., 2022) and for fresh root yield in cassava (Hohenfeld et al., 2022).

We found evidence of pleiotropy with some significant SNPs associated with multiple traits (e.g., CassV7_chr12_674448), which were recorded on chromosome 12 for both ANGBR9 and BRNHB9. This implies that the genes controlling these traits could co-segregate and be inherited together (Abah et al., 2024). reported a similar outcome for cassava, where a significant SNP (S4_8840623) was associated with the productivity of dry yield and bulking index. He also reported that these SNPs (S10_2319500, S2_1937678, and S3_3324735) independently influenced both the starch and dry matter contents.

Similarly, significant MTAs for PLTH9, FYLD, and RTWT were found on chromosome 8, validating the results of the path coefficient analysis, where PLTH9 had a significant positive contribution to FYLD through RTWT across the test environments with an almost absolute contribution (0.99) to FYLD in the trial at Onne (Figure 11) (Rabbi et al., 2017). also reported the colocalization of SNPs that control chromameter value, colour chart, and dry matter content in fresh cassava roots, while (Villwock et al., 2025) reported similar results for total carotenoid and dry matter contents.

Plant height is an essential component of plant architecture that influences crop productivity through its effects on planting density,

amount of insolation received, and resistance to lodging (Fei et al., 2022). The efficiency of mechanical harvesting can be influenced by plant height (Yan et al., 2019). The crux of this research is based on identifying genomic regions that could be exploited to develop cassava varieties with ideal plant architecture and improved yield. The identification of these important genomic regions on chromosome 8 that control plant height (CassV7_chr08_17260363) and fresh root yield (CassV7_chr08_6710616 and CassV7_chr08_7101298) constitutes a huge leap towards the realization of the core objective of this research, aimed at unraveling the genetics of plant architecture and its effects on yield in cassava, as well as suitability for mechanized farming.

Putative candidate genes linked to marker loci for traits associated with plant architecture and yield

Of the 21 putative candidate genes identified, six were found to promote yield, plant growth, and development. These include two genes identified on the angle of branching: Manes.12G003900, which encodes a legume lectin domain-containing protein (Aleksander et al., 2023) that promotes plant growth and development and enhances plant defence mechanisms against pathogens (Katoch and Tripathi,

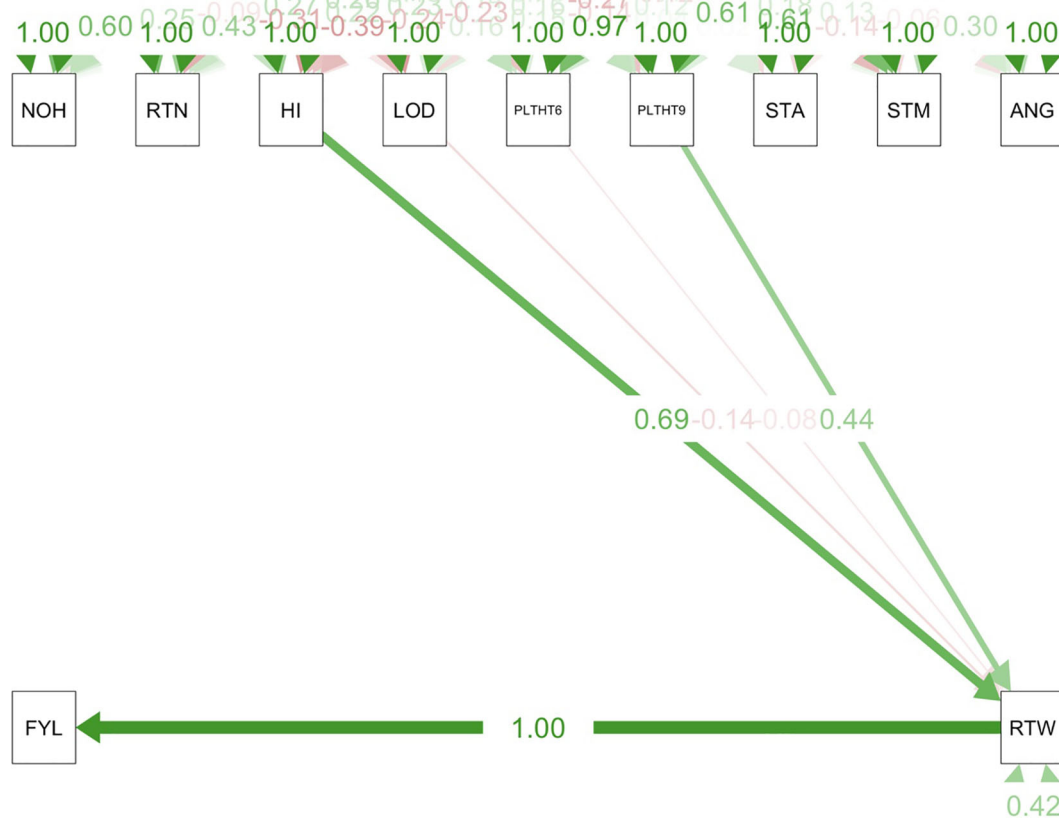


FIGURE 10

Path coefficient analysis plot of plant architecture and yield traits in Mokwa trial. NOH, number of harvested plants per plot; RTN, number of harvested roots per plot; RTW, root weight; SC, starch content; FYL, fresh root yield; HI, harvest index; LOD, number of lodged plants per plot; PLTHT6, plant height at 6 months after planting; PLTHT9, plant height at 9 months after planting; ANG, angle of branching, and STM, stem diameter.

2021), and Manes.12G007000, which encodes a glutaredoxin domain-containing protein (Aleksander et al., 2023) that regulates gene expression and signal transduction and contributes to plant growth and development (Rouhier et al., 2006). Similarly, for plant height, we identified Manes.08G075800, which encodes an oligopeptide transporter 1-like protein (Aleksander et al., 2023). This gene is involved in plant development and adaptation to stress (Liu et al., 2012; Zhai et al., 2014). The level of branching, a measure of flowering events that the plant has undergone, is a fundamental aspect of plant architecture that influences the plant's ability to capture light, distribute nutrients, and overall biomass production and yield potential (Pineda et al., 2020). The candidate gene associated with the level of branching, Manes.01G130800, encodes a PH (Pleckstrin Homology) domain-containing protein (Aleksander et al., 2023), which influences root development in plants through its interactions with phosphoinositides and phosphatidic acid (Lemmon, 2010). Similar to the study by (Baguma et al., 2024), on chromosome 15, we identified a genomic region linked with this trait that was approximately 3,783,500 bp apart from the one reported by (Baguma et al., 2024). The candidate gene Manes.15G171966, identified in this region in this study, is a novel candidate gene, while those reported by (Baguma et al., 2024) have

well-documented functions. This novel candidate gene presents a prospect for the validation of genes involved in plant architecture and flowering in cassava breeding. Yield improvement is a critical component of all crop improvement programs. Plant breeders are continually developing new strategies and technologies to enhance crop productivity and ensure global food security. Fresh root yield and root weight are important metrics for evaluating cassava yield. The candidate gene Manes.08G058000 identified in FYLD and RTWT encodes an MYB-like (myeloblastosis) DNA-binding protein (Goodstein et al., 2014). MYB is a transcription factor that plays a crucial role in regulating various cellular processes in plants, including the cell cycle and cell morphogenesis (Ambawat et al., 2013), biotic and abiotic stress responses (Roy, 2016), secondary metabolism, such as anthocyanin biosynthesis (Roy, 2016), and plant development (Katiyar et al., 2012; Roy, 2016). Similarly, Manes.02G069600, a candidate gene identified in DYLD, encodes a C2H2-type domain-containing protein (Aleksander et al., 2023). C2H2-type domain-containing proteins affect the metabolic pathways involved in photosynthetic processes (Lakshmanan et al., 2014; Habibpournmehraban et al., 2022).

The discovered novel candidate genes Manes.12G007100 (ANGBR9), Manes.08G075650 (PLTHT9), Manes.08G058000 (FYLD

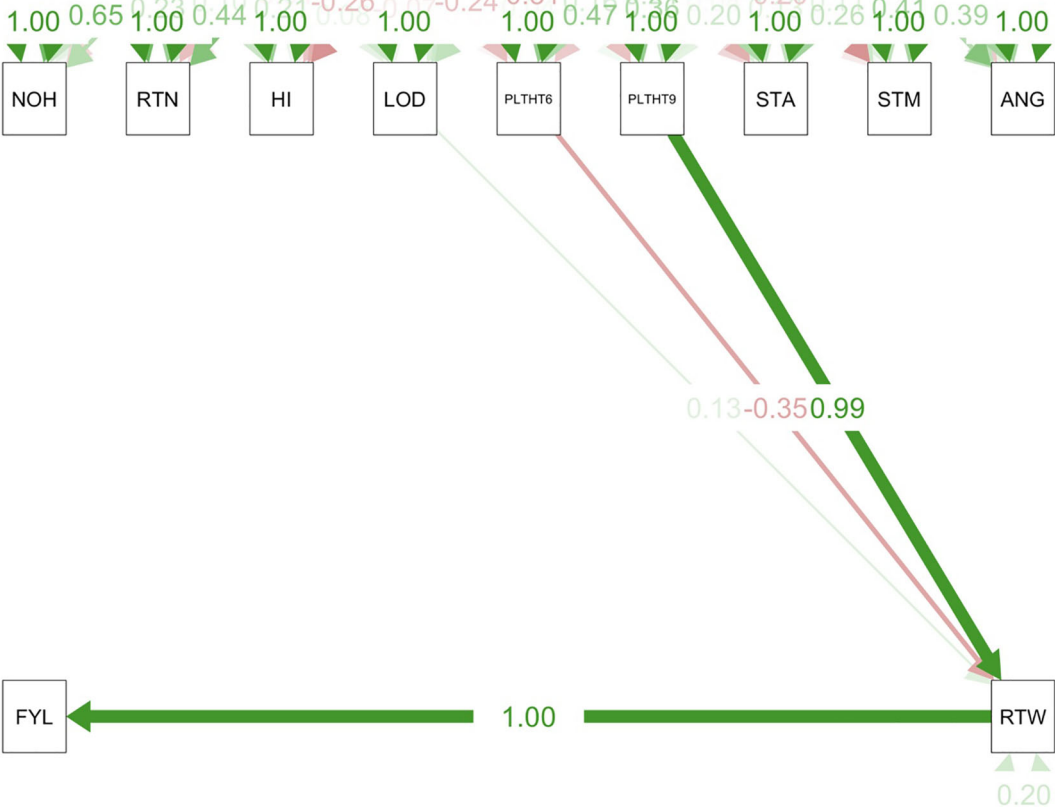


FIGURE 11
Path coefficient analysis plot of plant architecture and yield traits in Onne trial. NOH, number of harvested plants per plot; RTN, number of harvested roots per plot; RTW, root weight; SC, starch content; FYL, fresh root yield; HI, harvest index; LOD, number of lodged plants per plot; PLTH6, plant height at 6 months after planting; PLTH9, plant height at 9 months after planting; ANG, angle of branching; STM, stem diameter.

TABLE 2 Summary of top significant SNPs for plant architecture and yield traits of cassava.

Trait	SNP_ID	Chromosome number	Position	Value of P	Minor allele frequency	Phenotypic variance explained
ANGBR9	CassV7_chr12_436985	12	436985	2.37E ⁻⁰⁶	0.0782	10.01
	CassV7_chr12_674448	12	674448	1.43E ⁻⁰⁶	0.0687	4.62
	CassV7_chr14_18810275	14	18810275	1.65E ⁻⁰⁵	0.3949	7.53
PLTH6	CassV7_chr08_17260363	8	17260363	9.89E ⁻⁰⁶	0.4866	21.01
PLTH9	CassV7_chr08_17260363	8	17260363	8.56E ⁻⁰⁶	0.4866	19.57
FYLD	CassV7_chr08_6710616	8	6710616	1.70E ⁻⁰⁵	0.3165	0
	CassV7_chr08_7101298	8	7101298	6.80E ⁻⁰⁶	0.2918	12.7
BRNLEV3	CassV7_chr15_15530801	15	15530801	1.40E ⁻⁰⁵	0.491	16.18
BRNLEV9	CassV7_chr01_27643909	1	27643909	1.06E ⁻⁰⁵	0.2737	0
	CassV7_chr01_28103607	1	28103607	5.42E ⁻⁰⁶	0.2737	0
	CassV7_chr01_28262168	1	28262168	6.50E ⁻⁰⁶	0.3431	5.06
	CassV7_chr01_29155618	1	29155618	9.29E ⁻⁰⁶	0.4982	9.22

(Continued)

TABLE 2 Continued

Trait	SNP_ID	Chromosome number	Position	Value of P	Minor allele frequency	Phenotypic variance explained
	CassV7_chr01_29944386	1	29944386	9.97E ⁻⁰⁷	0.3303	0
	CassV7_chr01_30312867	1	30312867	9.97E ⁻⁰⁷	0.3303	10.54
DYLD	CassV7_chr02_6125643	2	6125643	1.76E ⁻⁰⁵	0.3859	12.57
RTWT	CassV7_chr08_6710616	8	6710616	1.70E ⁻⁰⁵	0.3165	0
	CassV7_chr08_7101298	8	7101298	6.80E ⁻⁰⁶	0.2918	12.7
BRNHB9	CassV7_chr12_674448	12	674448	1.14E ⁻⁰⁵	0.0691	21.72
SHTWT	CassV7_chr06_25850871	6	25850871	7.15E ⁻⁰⁶	0.1906	23.81
TYLD	CassV7_chr06_25850871	6	25850871	7.15E ⁻⁰⁶	0.1906	23.81

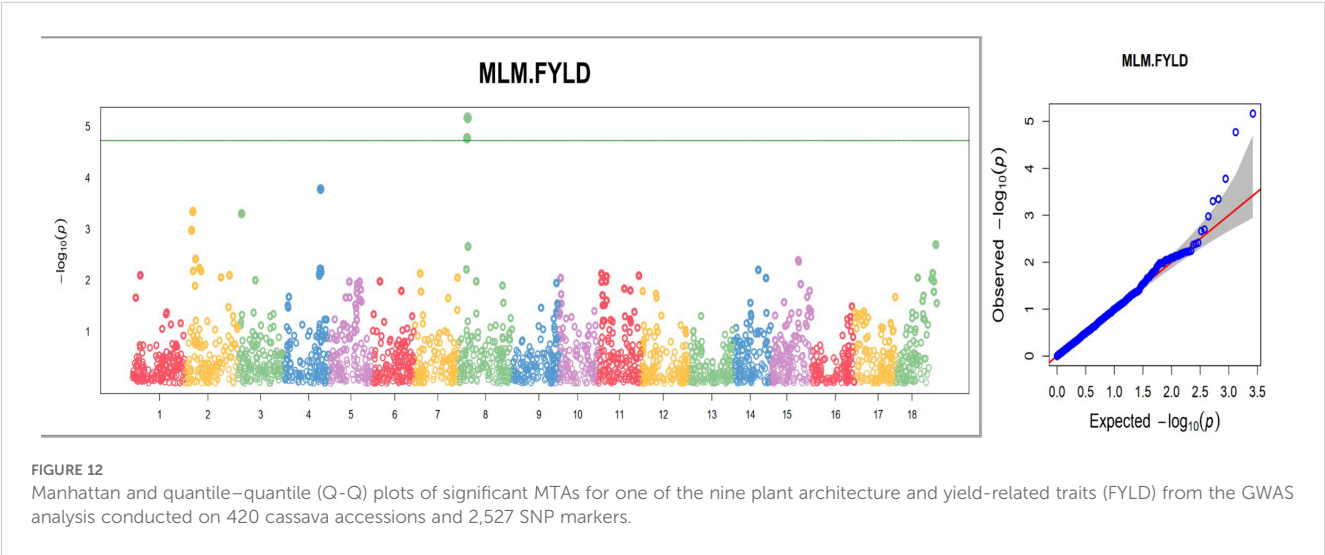


TABLE 3 Matrix of traits by environments of significant SNPs for plant architecture and yield traits of cassava across the test environments.

Traits/environments	Ikenne	Mokwa	Onne	MET
Level of Branching at 3 Months after Planting		◊		
Level of Branching at 9 Months after Planting	◊		◊	
Fresh Root Yield			◊	
Fresh Root Weight			◊	
Shoot Weight			◊	
Top Yield			◊	
Dry Yield		◊	◊	
Plant Height at 6 Months after Planting	◊			◊
Plant Height at 9 Months after Planting	◊			◊
Angle of Branching		◊		
Branching Habit				◊

MET, Multi-Environment Trial.

TABLE 4 Gene annotation of the significant SNPs for plant architecture and yield traits.

Trait	SNP_ID	Chromosome number	Gene ID	Gene annotation
PLTHT9	CassV7_chr08_17260363	8	Manes.08G075800	Oligopeptide transporter 1-like
	CassV7_chr08_17260363	8	Manes.08G075650	Uncharacterized protein
ANGBR9	CassV7_chr12_436985	12	Manes.12G003700	Thaumatococcus-like protein
	CassV7_chr12_436985	12	Manes.12G003900	Legume lectin domain-containing protein
	CassV7_chr12_436985	12	Manes.12G003600	Protein TIC 20
	CassV7_chr12_436985	12	Manes.12G003800	Thaumatococcus-like protein
	CassV7_chr12_674448	12	Manes.12G007000	Glutaredoxin domain-containing protein
	CassV7_chr12_674448	12	Manes.12G007100	Uncharacterized protein
	CassV7_chr12_674448	12	Manes.12G006900	Exocyst subunit Exo70 family protein
	CassV7_chr14_18810275	14	NA	
BRNLEV3	CassV7_chr15_15530801	15	Manes.15G171966	Uncharacterized protein
BRNLEV9	CassV7_chr01_27643909	1	Manes.01G122000	Zinc finger ZPR1-type domain-containing protein
	CassV7_chr01_28103607	1	Manes.01G128600	BED-type domain-containing protein
	CassV7_chr01_28262168	1	Manes.01G130800	PH domain-containing protein
	CassV7_chr01_29155618	1	Manes.01G142400	Uncharacterized protein
	CassV7_chr01_29944386	1	Manes.01G154200	K04523 - ubiquitin (UBQLN, DSK2)
	CassV7_chr01_30312867	1	Manes.01G159600	SMP-LTD domain-containing protein
FYLD	CassV7_chr08_7101298	8	Manes.08G058000	Uncharacterized protein
	CassV7_chr08_6710616	8	NA	
RTWT	CassV7_chr08_7101298	8	Manes.08G058000	Uncharacterized protein
	CassV7_chr08_6710616	8	NA	
SHTWT	CassV7_chr06_25850871	6	Manes.06G122900	Membrane-anchored ubiquitin-fold protein
DYLD	CassV7_chr02_6125643	2	Manes.02G069600	DUF4005 domain-containing protein
	CassV7_chr02_6125643	2	Manes.02G069700	Uncharacterized protein MANES_02G069700
	CassV7_chr02_6125643	2	Manes.02G069800	C2H2-type domain-containing protein
TYLD	CassV7_chr06_25850871	6	Manes.06G122900	Membrane-anchored ubiquitin-fold protein

and RTWT), Manes.02G069700 (DYLD), Manes.01G142400, and Manes.15G171966 (BRNLEV3 and BRNLEV9) provide insights into the genetic pathways through which these traits could be improved.

Conclusion

The genetic underpinnings of plant architecture and yield-associated traits in cassava were examined in this study through a genome-wide association study (GWAS). This investigation involved

420 cassava accessions and employed two genotyping platforms, DArTseq and DArTag, across three distinct environments. Sixteen significant, validated marker-trait associations were discovered using the DArTag markers, compared to only one reported using the DArTseq markers. This highlights the importance of maintaining a balance between the density and distribution of markers for more reliable GWAS results.

The significant marker-trait associations were linked to important genomic regions that could enhance marker-assisted selection for suitable plant architecture and increased yield in

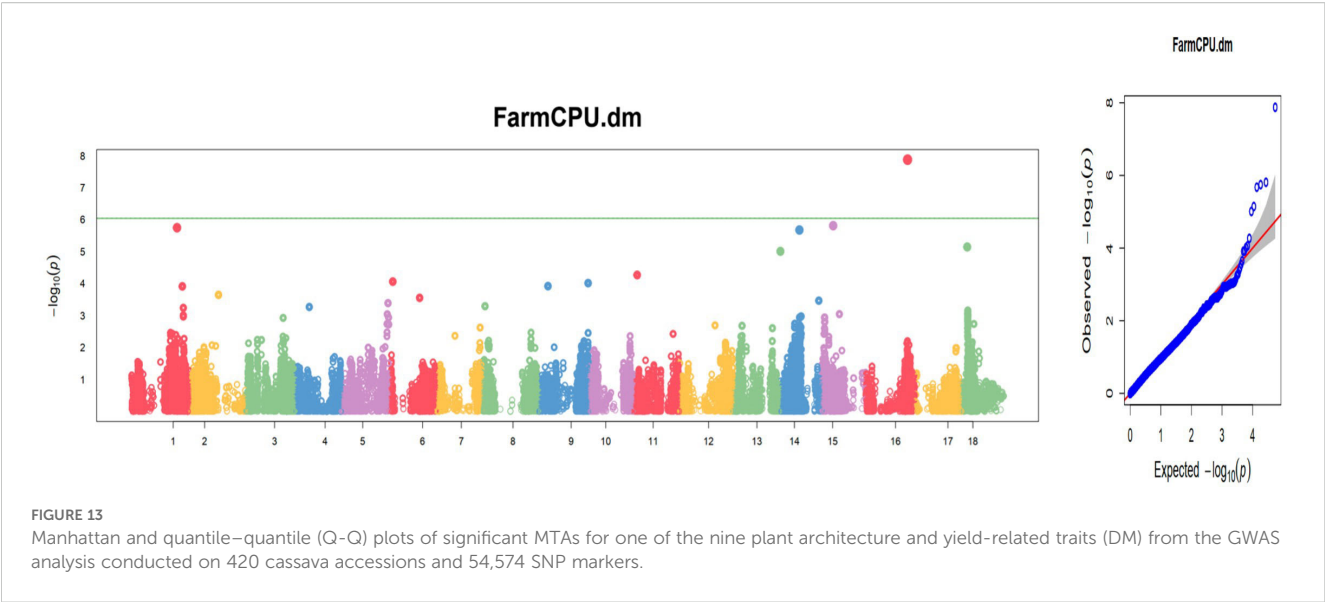


TABLE 5 Summary of top significant SNPs for plant architecture and yield traits of cassava.

Trait	SNP_ID	Chromosome number	Position	Value of P	Minor allele frequency	Phenotypic variance explained
ANGBR9	S1_19494690	1	19494690	6.77E ⁻⁰⁷	0.15944	0.85688
	S1_23883230	1	23883230	4.58E ⁻⁰⁷	0.19515	18.5628
PLTTP	S3_1388806	3	1388806	1.17E ⁻⁰⁷	0.46495	0
	S5_10702830	5	10702830	1.80E ⁻⁰⁸	0.44393	3.88491
	S15_10073295	15	10073295	2.78E ⁻⁰⁷	0.1729	1.60299
VR	S4_5944464	4	5944464	2.94E ⁻⁰⁸	0.42523	1.12237
	S16_2295914	16	2295914	2.73E ⁻⁰⁸	0.30958	1.87443
	S17_17302312	17	17302312	5.68E ⁻⁰⁸	0.43107	5.10115
BRNLEV6	S8_27519983	8	27519983	5.53E ⁻⁰⁷	0.15278	18.6092
	S11_16097795	11	16097795	5.67E ⁻⁰⁸	0.10301	15.8714
BRNLEV9	S5_22148410	5	22148410	1.20E ⁻⁰⁹	0.45718	0.61819
	S11_16097795	11	16097795	3.30E ⁻⁰⁸	0.10301	22.6755
DM	S16_24715889	16	24715889	1.33E ⁻⁰⁸	0.14419	24.3844
HI	S9_3493565	9	3493565	1.32E ⁻¹³	0.31481	5.37937
	S12_29499402	12	29499402	1.81E ⁻⁰⁷	0.4456	5.60214
	S16_22907596	16	22907596	1.43E ⁻⁰⁸	0.06134	7.05957
NOHAV	S6_21336567	6	21336567	9.00E ⁻⁰⁷	0.08912	32.7769

cassava breeding. These include putative candidate genes for angle of branching (7), plant height (2), level of branching (7), fresh root yield and weight (1), fresh shoot weight and top yield (1), and dry yield (3). These candidate genes exhibit various functions related to plant architecture, adaptation, yield (root development), plant growth, and stress response. Out of the 21 putative candidate genes identified in this study, six novel genes (Manes.08G075650, Manes.12G007100, Manes.15G171966, Manes.01G142400,

TABLE 6 Gene annotation of the significant SNPs for plant architecture and yield traits.

Trait	Marker	Position	Chromosome	Gene ID	Gene annotation
ANGBR9	S1_19494690	19494690	1	Manes.01G056000	glucan endo-1,3-beta-D-glucosidase
				Manes.01G056100	Ubiquitin carboxyl-terminal hydrolase
	S1_23883230	23883230	1	Manes.01G077800	Nitrate-transporting ATPase
				Manes.01G077900	Uncharacterized Manes.01G077900
PLTTP	S3_1388806	1388806	3	Manes.03G016100	DYW domain-containing protein
				Manes.03G016200	AAA+ ATPase domain-containing protein
				Manes.03G016300	Adenine Nucleotide Transporter
	S15_10073295	10073295	15	Manes.15G122400	9-cis-epoxycarotenoid dioxygenase
VR	S4_5944464	5944464	4	Manes.04G041500	RBR-type E3 ubiquitin transferase
	S16_2295914	2295914	16	Manes.16G020700	F-box domain-containing protein
BRNLEV6	S11_16097795	16097795	11	Manes.11G089900	Triosephosphate isomerase
BRNLEV9	S5_22148410	22148410	5	Manes.05G148900	Cytokinin riboside 5'-monophosphate phosphoribohydrolase
				Manes.05G148800	Endoglucanase
				Manes.05G148700	Myb/SANT-like domain-containing protein
	S11_16097795	16097795	11	Manes.11G089900	Triosephosphate isomerase
DM	S16_24715889	24715889	16	NA	NA
HI	S9_3493565	3493565	9	Manes.09G018900	Glycosyltransferase
	S12_29499402	29499402	12	Manes.12G109600	RING-type E3 ubiquitin transferase
	S16_22907596	22907596	16	Manes.16G062600	JmjC domain-containing protein
NOHAV	S6_21336567	21336567	6	Manes.06G073700	TauD/TfdA-like domain-containing protein

Manes.08G058000, Manes.02G069700) were discovered. This represents a significant contribution to the existing knowledge. These findings provide a gateway for exploring the genetic control of cassava plant architecture and yield. This research output will provide cassava breeders with the genetic and molecular leverage required to fast-track cassava improvement in terms of yield, productivity, and adaptation for mechanized cultivation and industrial use.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

AO: Formal analysis, Writing – review & editing, Validation, Writing – original draft, Methodology, Investigation, Data curation, Conceptualization. DD: Conceptualization, Supervision, Methodology, Writing – review & editing. EM: Conceptualization, Formal analysis, Methodology, Writing – review & editing. SO: Writing – review & editing, Supervision. PT: Supervision, Writing –

review & editing. ED: Writing – review & editing, Supervision. CE: Supervision, Funding acquisition, Writing – review & editing, Conceptualization. IR: Writing – review & editing, Funding acquisition, Formal analysis, Conceptualization, Supervision.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. Abiodun Fatai Olayinka carried out this research as part of his PhD studies in Plant Breeding at WACCI, University of Ghana, Legon, Accra, Ghana. This study was funded by the NextGen Cassava Breeding Project (Grant INV-007637), which received financial support from the Bill and Melinda Gates Foundation (BMGF) and the UK Foreign, Commonwealth & Development Office (FCDO).

Acknowledgments

We appreciate the efforts and support received from the efficient and industrious members of the staff of the cassava breeding program at IITA, Ibadan during field evaluation and laboratory analysis.

Conflict of interest

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

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

References

- Aalborg, T., Sverrisdóttir, E., Kristensen, H. T., and Nielsen, K. L. (2024). The effect of marker types and density on genomic prediction and GWAS of key performance traits in tetraploid potato. *Front. Plant Sci.* 15, 1340189. doi: 10.3389/fpls.2024.1340189
- Abah, S. P., Mbe, J. O., Dzidzienyo, D. K., Njoku, D., Onyeka, J., Danquah, E. Y., et al. (2024). Determination of genomic regions associated with early storage root formation and bulking in cassava. *Front. Plant Sci.* 15, 1391452. doi: 10.3389/fpls.2024.1391452
- Aghogho, C. I., Kayondo, S. I., Eleblu, S. J., Ige, A., Asante, I., Offei, S. K., et al. (2024). Genome-wide association study for yield and quality of granulated cassava processed product. *Plant Genome* 17, e20469. doi: 10.1002/tpg2.20469
- Aleksander, S. A., Balhoff, J., Carbon, S., Cherry, J. M., Drabkin, H. J., Ebert, D., et al. (2023). The gene ontology knowledgebase in 2023. *Genetics* 224, iyad031. doi: 10.1093/genetics/iyad031
- Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109
- Alvarado, G., Rodríguez, F. M., Pacheco, A., Burgueño, J., Crossa, J., Vargas, M., et al. (2020). META-R: A software to analyze data from multi-environment plant breeding trials. *Crop J.* 8, 745–756. doi: 10.1016/j.cj.2020.03.010
- Alves, A. A. C. (2002). "Cassava botany and physiology," in *Cassava: biology, production and utilization*. Eds. R. J. Hillocks, J. M. Tresh and A. C. Bellotti (CABI Publishing, London), 67–89. doi: 10.1079/9780851995243.0067
- Ambawat, S., Sharma, P., Yadav, N. R., and Yadav, R. C. (2013). MYB transcription factor genes as regulators for plant responses: an overview. *Physiol. Mol. Biol. Plants* 19, 307–321. doi: 10.1007/s12298-013-0179-1
- Baguma, J. K., Mukasa, S. B., Nuwamanya, E., Alicai, T., Omongo, C. A., Ochwo-Ssemakula, M., et al. (2024). Identification of genomic regions for traits associated with flowering in cassava (*Manihot esculenta* Crantz). *Plants* 13, 796. doi: 10.3390/plants13060796
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Software* 67, 1–48. doi: 10.18637/jss.v067.i01
- Bentley, A. R., Chen, C., and D'Agostino, N. (2022). Genome wide association studies and genomic selection for crop improvement in the era of big data. *Front. Genet.* 13, 873060. doi: 10.3389/fgene.2022.873060
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, s13742–s13015. doi: 10.1186/s13742-015-0047-8
- Dash, G. K., Sabarinathan, S., Donde, R., Gouda, G., Gupta, M. K., Behera, L., et al. (2021). "Status and perspectives of genome-wide association studies in plants," in *Bioinformatics in rice research: theories and techniques* (Springer Singapore, Singapore), 413–457.
- Edet, M. A., Tijani-Eniola, H., Lagoke, S. T. O., and Tarawali, G. (2015). Relationship of cassava growth parameters with yield, yield related components and harvest time in Ibadan, Southwestern Nigeria. *J. Natural Sci. Res.* 5, 87–92.
- Epskamp, S., Stuber, S., Nak, J., Veenman, M., and Jorgensen, T. D. (2019). semPlot: Path diagrams and visual analysis of various SEM packages' output. *R Package version* 1. doi: 10.32614/CRAN.package.semPlot
- FAO (2013). Cassava, a 21st century crop. Save and Grow: Cassava, a Guide to Sustainable Production Intensification. Available online at: <https://www.fao.org/3/i3278e/i3278e.pdf> (Accessed July 1, 2023).
- FAO (2020). *Food and agriculture organization of the united nations: crops* (Rome, Italy: FAOSTAT). Available online at: <https://www.fao.org/faostat/en/data/QCL> (Accessed July 1, 2023).
- FAO (2021). *Food and agriculture organization of the United Nations. World's Area Harvested and Production Quantity* (Rome, Italy: FAOSTAT). Available online at: <http://www.fao.org/faostat/en/> (Accessed July 1, 2023).
- Fei, J., Lu, J., Jiang, Q., Liu, Z., Yao, D., Qu, J., et al. (2022). Maize plant architecture trait QTL mapping and candidate gene identification based on multiple environments and double populations. *BMC Plant Biol.* 22, 110. doi: 10.1186/s12870-022-03470-7
- Fukuda, W. M. G., Guevara, C. L., Kawuki, R., and Ferguson, M. E. (2010). *Selected morphological and agronomic descriptors for the characterization of cassava* (Oyo State, Nigeria: IITA).
- Galili, T. (2015). dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* 31, 3718–3720. doi: 10.1093/bioinformatics/btv428
- Goodstein, D., Batra, S., Carlson, J., Hayes, R., Phillips, J., and Shu, S. (2014). *Phytozome comparative plant genomics portal* (Berkeley, California, United States: Lawrence Berkeley National Laboratory).
- Habibpourmehraban, F., Atwell, B. J., and Haynes, P. A. (2022). Unique and shared proteome responses of rice plants (*Oryza sativa*) to individual abiotic stresses. *Int. J. Mol. Sci.* 23, 15552. doi: 10.3390/ijms232415552
- Hardenbol, P., Banér, J., Jain, M., Nilsson, M., Namsaraev, E. A., Karlin-Neumann, G. A., et al. (2003). Multiplexed genotyping with sequence-tagged molecular inversion probes. *Nat. Biotechnol.* 21, 673–678. doi: 10.1038/nbt821
- Hardigan, M. A., Feldmann, M. J., Carling, J., Zhu, A., Kilian, A., Famula, R. A., et al. (2023). A medium-density genotyping platform for cultivated strawberry using DArTag technology. *Plant Genome* 16, e20399. doi: 10.1002/tpg2.20399
- Hohenfeld, C. S., Passos, A. R., de Carvalho, H. W. L., de Oliveira, S. A. S., and de Oliveira, E. J. (2022). Genome-wide association study and selection for field resistance to cassava root rot disease and productive traits. *PLoS One* 17, e0270020. doi: 10.1371/journal.pone.0270020
- Katiyar, A., Smita, S., Lenka, S. K., Rajwanshi, R., Chinnusamy, V., and Bansal, K. C. (2012). Genome-wide classification and expression analysis of MYB transcription factor families in rice and Arabidopsis. *BMC Genomics* 13, 544. doi: 10.1186/1471-2164-13-544
- Katoch, R., and Tripathi, A. (2021). Research advances and prospects of legume lectins. *J. Biosci.* 46, 104. doi: 10.1007/s12038-021-00225-8
- Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., et al. (2012). "Diversity arrays technology: a generic genome profiling technology on open platforms," in *Data production and analysis in population genomics: Methods and protocols* (Humana Press, Totowa, NJ), 67–89.
- Lakshmanan, M., Mohanty, B., Lim, S. H., Ha, S. H., and Lee, D. Y. (2014). Metabolic and transcriptional regulatory mechanisms underlying the anoxic adaptation of rice coleoptile. *AoB Plants* 6, plu026. doi: 10.1093/aobpla/plu026
- Laurans, M., Munoz, F., Charles-Dominique, T., Heuret, P., Fortunel, C., Isnard, S., et al. (2024). Why incorporate plant architecture into trait-based ecology? *Trends Ecol. Evol.* 39, 524–536. doi: 10.1016/j.tree.2023.11.011

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2025.1660789/full#supplementary-material>

- Lauri, P.É., and Lespinasse, J. M. (2001). Genotype of apple trees affects growth and fruiting responses to shoot bending at various times of year. *J. Am. Soc. Hortic. Sci.* 126, 169–174. doi: 10.21273/JASHS.126.2.169
- Lemmon, M. A. (2010). “Pleckstrin homology (PH) domains,” in *Handbook of cell signaling* (London, United Kingdom: Academic Press), 1093–1101.
- Li, S., Cui, Y., Zhou, Y., Luo, Z., Liu, J., and Zhao, M. (2017). The industrial applications of cassava: current status, opportunities and prospects. *J. Sci. Food Agric.* 97, 2282–2290. doi: 10.1002/jsfa.8287
- Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., et al. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28, 2397–2399. doi: 10.1093/bioinformatics/bts444
- Liu, T., Zeng, J., Xia, K., Fan, T., Li, Y., Wang, Y., et al. (2012). Evolutionary expansion and functional diversification of oligopeptide transporter gene family in rice. *Rice* 5, 12. doi: 10.1186/1939-8433-5-12
- Liu, X., Huang, M., Fan, B., Buckler, E. S., and Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet.* 12, e1005767. doi: 10.1371/journal.pgen.1005767
- Lu, X., Liu, P., Tu, L., Guo, X., Wang, A., Zhu, Y., et al. (2024). Joint-GWAS, linkage mapping, and transcriptome analysis to reveal the genetic basis of plant architecture-related traits in maize. *Int. J. Mol. Sci.* 25, 2694. doi: 10.3390/ijms25052694
- Mansaray, A., Babatunde, K. A., Yormah, T. B. H., Marie, Y., and Rahman, C. A. (2020). Effect of spatial arrangement, plant architecture and cropping system on the growth, yield and yield-related components of cassava (*Manihot esculenta* L.). *J. Global Agric. Ecol.* 10, 1–7.
- Mbe, J. O., Dzidzienyo, D. K., Abah, S. P., Njoku, D. N., Onyeka, J., Tongoon, P., et al. (2024). Novel SNP markers and other stress-related genomic regions associated with nitrogen use efficiency in cassava. *Front. Plant Sci.* 15, 1376520. doi: 10.3389/fpls.2024.1376520
- Mora Moreno, R. E., Soto, J. C., and López, C. (2016). Identification of QTLs associated to plant architecture in cassava (*Manihot esculenta* crantz). *Acta Biológica Colombiana* 21, 99–109. doi: 10.15446/abc.v21n1.49251
- Morgil, H., Gercek, Y. C., Tulum, I., Morgil, H., Gercek, Y. C., and Tulum, I. (2020). “Single nucleotide polymorphisms (SNPs) in plant genetics and breeding,” in *The recent topics in genetic polymorphisms* (London, England, United Kingdom: IntechOpen).
- Namakula, F. B., and Nuwamanya, E. (2022). Genetic variability and Heritability of starch content among white fleshed and Provitamin A cassava in Uganda. *Afr. J. Rural Dev.* 7, 201–208.
- Nweke, F. I. (2004). “The cassava transformation in Africa,” in *New challenges in the cassava transformation in Nigeria and Ghana*. Ed. F. I. Nweke (International Institute of Tropical Agriculture (IITA, Ibadan, Nigeria), 1–15.
- OgokBenin, E., and Fregene, M. (2003). Genetic mapping of QTLs affecting productivity and plant architecture in a full-sib cross from non-inbred parents in Cassava (*Manihot esculenta* Crantz). *Theor. Appl. Genet.* 107, 1452–1462. doi: 10.1007/s00122-003-1383-0
- Panahi, B., Jalali, H. M., and Hamid, R. (2024). Using next-generation sequencing approach for discovery and characterization of plant molecular markers. *Curr. Plant Biol.* 40, 100412. doi: 10.1016/j.cpb.2024.100412
- Patiño, B. O., Cadavid López, L. F., García, M., and Alcalde, C. (2002). *Mechanized systems for planting and harvesting cassava* (*Manihot esculenta* crantz) Vol. 3 (Cali, Colombia: Cassava in the Third Millennium), 74–94.
- Pineda, M., Yu, B., Tian, Y., Morante, N., Salazar, S., Hyde, P. T., et al. (2020). Effect of pruning young branches on fruit and seed set in cassava. *Front. Plant Sci.* 11, 1107. doi: 10.3389/fpls.2020.1107
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- Rabbi, I. Y., Udoh, L. I., Wolfe, M., Parkes, E. Y., Gedil, M. A., Dixon, A., et al. (2017). Genome-wide association mapping of correlated traits in cassava: dry matter and total carotenoid content. *Plant Genome* 10, plantgenome2016–09. doi: 10.3835/plantgenome2016.09.0094
- R Core Team (2020). *R: A language and environment for statistical computing* (Vienna, Austria: R Foundation for Statistical Computing).
- Rodriguez-Alvarez, M. X., Boer, M. P., van Eeuwijk, F. A., and Eilers, P. H. (2018). Correcting for spatial heterogeneity in plant breeding experiments with P-splines. *Spatial Stat* 23, 52–71. doi: 10.1016/j.spa.2017.10.003
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *J. Stat. Software* 48, 1–36. doi: 10.18637/jss.v048.i02
- Rouhier, N., Couturier, J., and Jacquot, J. P. (2006). Genome-wide analysis of plant glutaredoxin systems. *J. Exp. Bot.* 57, 1685–1696. doi: 10.1093/jxb/erl001
- Roy, S. (2016). Function of MYB domain transcription factors in abiotic stress and epigenetic control of stress response in plant genome. *Plant Signaling Behav.* 11, e1117723. doi: 10.1080/15592324.2015.1117723
- Rymaszewski, W., Vile, D., Bedie, A., Dauzat, M., Luchaire, N., Kamrowska, D., et al. (2017). Stress-related gene expression reflects morphophysiological responses to water deficit. *Plant Physiol.* 174, 1913–1930. doi: 10.1104/pp.17.00318
- Sánchez-Sevilla, J. F., Horvath, A., Botella, M. A., Gaston, A., Foltá, K., Kilian, A., et al. (2015). Diversity arrays technology (DART) marker platforms for diversity analysis and linkage mapping in a complex crop, the octoploid cultivated strawberry (*Fragaria × ananassa*). *PLoS One* 10, e0144960.
- Santos, V. D. S., Fukuda, W. M. G., Oliveira, L. A. D., Pereira, M. E. C., Nuti, M. R., Carvalho, J. L. V. D., et al. (2023). Genetic parameters considering traits of importance for cassava biofortification. *Crop Breed. Appl. Biotechnol.* 23, e447923211. doi: 10.1590/1984-70332023v23n2a23
- Shao, P., Peng, Y., Wu, Y., Wang, J., Pan, Z., Yang, Y., et al. (2022). Genome-wide association study and transcriptome analysis reveal key genes controlling fruit branch angle in cotton. *Front. Plant Sci.* 13, 988647. doi: 10.3389/fpls.2022.988647
- Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redona, E., et al. (2015). Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLoS Genet.* 11, e1004982. doi: 10.1371/journal.pgen.1004982
- Srisawad, N., Sraphet, S., Suksee, N., Boontung, R., Smith, D. R., and Triwitayakorn, K. (2023). Use of Diversity Arrays Technology (DART) for detection of QTL underlying plant architecture and yield-related traits in cassava. *J. Crop Improvement* 37, 99–118. doi: 10.1080/15427528.2022.2058668
- Susmitha, P., Kumar, P., Yadav, P., Sahoo, S., Kaur, G., Pandey, M. K., et al. (2023). Genome-wide association study as a powerful tool for dissecting competitive traits in legumes. *Front. Plant Sci.* 14, 1123631. doi: 10.3389/fpls.2023.1123631
- Tang, D., Chen, M., Huang, X., Zhang, G., Zeng, L., Zhang, G., et al. (2023). SRplot: A free online platform for data visualization and graphing. *PLoS One* 18, e0294236. doi: 10.1371/journal.pone.0294236
- Uffelmann, E., Huang, Q. Q., Munung, N. S., De Vries, J., Okada, Y., Martin, A. R., et al. (2021). Genome-wide association studies. *Nat. Rev. Methods Primers* 1, 59. doi: 10.1038/s43586-021-00056-9
- Villwock, S. S., Parkes, E. Y., Nkouaya Mbanjo, E. G., Rabbi, I. Y., and Jannink, J. L. (2025). Bivariate genome-wide association study reveals polygenic contributions to covariance between total carotenoid and dry matter contents in yellow-fleshed cassava (*Manihot esculenta*). *bioRxiv*, 2025–2003.
- Wei, T., and Simko, V. (2021). *An introduction to corrplot package*. R Packag. Vienna, Austria: R Foundation for Statistical Computing (via CRAN).
- Wenzl, P., Carling, J., Kudrna, D., Jaccoud, D., Huttner, E., Kleinhofs, A., et al. (2004). Diversity Arrays Technology (DART) for whole-genome profiling of barley. *Proc. Natl. Acad. Sci.* 101, 9915–9920. doi: 10.1073/pnas.0401076101
- Wickham, H. (2016). “Programming with ggplot2,” in *Ggplot2: elegant graphics for data analysis* (Springer International Publishing, Cham), 241–253.
- Yan, W., Du, M., Zhao, W., Li, F., Wang, X., Enejí, A. E., et al. (2019). Relationships between plant architecture traits and cotton yield within the plant height range of 80–120 cm desired for mechanical harvesting in the Yellow River Valley of China. *Agronomy* 9, 587. doi: 10.3390/agronomy9100587
- Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76–82. doi: 10.1016/j.ajhg.2010.11.011
- Yang, G., Pan, Y., Pan, W., Song, Q., Zhang, R., Tong, W., et al. (2024). Combined GWAS and eGWAS reveals the genetic basis underlying drought tolerance in emmer wheat (*Triticum turgidum* L.). *New Phytol.* 242, 2115–2131. doi: 10.1111/nph.19589
- Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M., and Price, A. L. (2014). Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* 46, 100–106. doi: 10.1038/ng.2876
- Yu, M., Liu, Z. H., Yang, B., Chen, H., Zhang, H., and Hou, D. B. (2020). The contribution of photosynthesis traits and plant height components to plant height in wheat at the individual quantitative trait locus level. *Sci. Rep.* 10, 1–10. doi: 10.1038/s41598-020-69138-0
- Zhai, Z., Gayomba, S. R., Jung, H. I., Vimalakumari, N. K., Piñeros, M., Craft, E., et al. (2014). OPT3 is a phloem-specific iron transporter that is essential for systemic iron signaling and redistribution of iron and cadmium in Arabidopsis. *Plant Cell* 26, 2249–2264. doi: 10.1105/tpc.114.123737
- Zhang, Z., Ersoz, E., Lai, C. Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., et al. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* 42, 355–360. doi: 10.1038/ng.546
- Zhang, Y., Yu, C., Lin, J., Liu, J., Liu, B., Wang, J., et al. (2017). OsMPH1 regulates plant height and improves grain yield in rice. *PLoS One* 12, e0180825. doi: 10.1371/journal.pone.0180825
- Zhu, C., Gore, M., Buckler, E. S., and Yu, J. (2008). Status and prospects of association mapping in plants. *Plant Genome* 1. doi: 10.3835/plantgenome2008.02.0089