Article

Genome-Wide Association Study of Root Quality Traits in Cassava (*Manihot esculenta* Crantz)

Kwame Obeng Dankwa [1,2,*], Jorge Luis Luna Melendez [2], Jhon Larry Moreno Alzate [2], Maria Alejandra Ospina [2], Luis Londoño [2], Sandra Milena Salazar [2], Nelson Morante [2], Bunmi Olasanmi [3], Winnie Gimode [2]

[1] Plant Breeding Program, Life and Earth Sciences Institute (Including Health and Agriculture), Pan African University, Ibadan 200132, Oyo State, Nigeria

[2] International Center for Tropical Agriculture (CIAT), Km 17, Recta Cali-Palmira, Palmira 763537, Colombia; J.L.Luna@cgiar.org (JLLM); j.l.moreno@cgiar.org (JLMA); Maria Alejandra Ospina (MAO); L.Londono@cgiar.org (LL); S.M.Salazar@cgiar.org (SMS); N.Morante@cgiar.org (NM); W.Gimode@cgiar.org (WG)

[3] Department of Crop and Horticultural Sciences, Faculty of Agriculture, University of Ibadan, Ibadan 200005, Nigeria; bunminadeco@yahoo.com (BO)

* Correspondence: Kwame Obeng Dankwa, Email: anionlives@gmail.com.

ABSTRACT

Cassava (*Manihot esculenta* Crantz) is a major staple crop globally, valued for its storage roots, which serve as a key source of food, feed, and industrial products such as starch and flour. Consequently, root quality traits are central to the cassava value chain. This study aimed to identify genomic regions associated with five important root quality traits: dry matter content (DMC), total carotenoid content (TCC), beta-carotene content (BCC), hydrogen cyanide content (HCN), and water absorption during boiling for 20 and 30 min (WAB_20min, WAB_30min). A panel of 300 cassava accessions was phenotyped for these traits and genotyped using genotyping-by-sequencing (GBS: 30,046 SNPs) and DArTag (2526 SNPs) platforms for genome-wide association analysis. Correlation analysis revealed generally low to moderate relationships among traits, except for a strong positive association between TCC and BCC, and between the WAB measurements, indicating relative independence among most root quality traits. Under the GBS platform, 10 significant SNPs were detected for each of HCN and BCC, 4 for TCC, and 2 for each of DMC and WAB_20min. Under the DArTag platform, 9 significant SNPs were identified for HCN, 5 for TCC, 5 for BCC, 1 for DMC, and 1 for WAB_20min. Across both marker platforms, no significant SNP was detected for WAB_30min. Overall, these findings provide insights into the genetic architecture of cassava root quality traits and highlight the potential of integrating genomic tools to enhance selection efficiency in cassava breeding programs.

Open Access

Received: 21 Aug 2026
Accepted: 28 Sep 2026
Published: 09 Nov 2026

Copyright © 2026 by the author. Licensee Hapres, London, United Kingdom. This is an open access article distributed under the terms and conditions of Creative Commons Attribution 4.0 International License.

KEYWORDS: cassava root quality traits; genome-wide association analysis; marker-assisted selection

INTRODUCTION

Cassava (*Manihot esculenta* Crantz) is a crucial staple crop in tropical and subtropical regions, serving as a primary caloric source for millions, particularly in sub-Saharan Africa [1,2]. Its adaptability to diverse environments underpins its significance for global food security [2]. Although cassava leaves are consumed by humans or used as animal feed in some parts of the world [3], the crop is mainly cultivated for its storage roots, which are the most economically important part of the plant [4]. The economic and nutritional value of cassava roots are determined by inter-related quality parameters, among which are starch content, dry matter content (DMC), total carotenoid content (TCC), beta-carotene content (BCC), hydrogen cyanide content (HCN), and water absorption during boiling (WAB) [5–8].

Dry matter content is a critical determinant of dry root yield which determines the product yield, with a strong correlation to starch content [6,9]. High DMC is generally preferred by the industrial sector [10]. Studies have shown that DMC can vary significantly, ranging from 14.1% to 51.0% across different root sections and genotypes [11]. Efforts to biofortify cassava with provitamin A carotenoids, such as beta-carotene, have sometimes resulted in a negative correlation with DMC, posing a challenge for breeding programs aiming to improve both traits simultaneously [1,2,7,12]. This negative correlation implies that increasing provitamin A content might inadvertently reduce the dry matter, affecting overall yield and product quality.

Total carotenoid content and beta-carotene content are essential nutritional attributes, as beta-carotene is a precursor to vitamin A, crucial for addressing vitamin A deficiency, a prevalent micronutrient deficiency in populations reliant on cassava [1,13,14]. Research has focused on enhancing these contents through conventional breeding and genetic engineering, with some genotypes accumulating up to 16.6 μg/g beta-carotene on fresh-weight basis [1,5,15]. Due to its nutritional importance, the inheritance of carotenoid content in cassava and its genetic basis have received significant research attention [12,14,16,17].

The presence of cyanogenic glycosides, which can yield hydrogen cyanide (HCN) when broken down by linamarase, is a significant safety concern in cassava [6,18]. Hydrogen cyanide is toxic and contributes to the bitter taste of some cassava cultivars [6]. Various processing techniques are employed to reduce HCN levels, including fermentation, drying, and heat processing (boiling or roasting) which can decrease cyanide content by up to 95% in roots and significantly in leaves [18–20]. The effectiveness of HCN reduction is critical for ensuring the safety of cassava-based foods [21].

Cassava root water absorption (WAB) is another important factor, particularly in the context of food product development and material science [8]. It is a key determinant of cooking quality in cassava roots and is directly linked to the mealiness and texture that consumers prioritize. Boiling is one of the simplest and most common modes of cassava consumption across Latin America, Africa, and Asia, where consumers prefer varieties that cook rapidly and soften during boiling, making mealiness and texture of boiled roots a high-priority trait for breeders [22]. Water absorption is the main parameter determining the quality of cassava cultivars during cooking, with mealy cooking types absorbing significantly more water than hard cooking types [23]. Mealiness in cassava is closely related to firmness, cell cohesiveness, and water uptake during boiling, and at the molecular level, pectins and cell wall characteristics have been linked to this trait [22]. Water absorption has been validated as a suitable instrumental method for evaluating mealiness and hardness of boiled cassava, and consumer preference scores have been translated into acceptability thresholds based on this measure, underscoring its practical utility in breeding programs [24]. Despite its importance as a sensory attribute, root mealiness has received comparatively little research attention in cassava relative to other crops, partly due to the predominant focus on yield and disease resistance in breeding programs [25].

Genome-wide association studies (GWAS) have elucidated the genetic architecture of several cassava root quality traits. For HCN, two major loci, S14_6050078 and S16_773999, on chromosomes 14 and 16 respectively were identified encoding an ATPase and a MATE transporter, which collectively contribute up to 37% of root HCN variation, with the MATE protein playing a dominant role in vacuolar cyanide transport [26].

Regarding DMC, a major QTL explaining ~37% of the phenotypic variance has been mapped to approximately 24.1 Mbp on chromosome 1 that co-localizes with major locus controlling TCC [12]. Linkage mapping also identified six additional DMC loci across four linkage groups, one of which co-localized with an HCN QTL on linkage group 10, suggesting pleiotropy or tight linkage [27].

For carotenoid traits, GWAS was used to identify a major QTL on chromosome 1 (23.4–24.7 Mbp) associated with TCC and beta-carotene, encompassing the candidate gene *Manes.01G124200* (phytoene synthase, *PSY*), which catalyzes beta-carotene accumulation [28,29]. Additional carotenoid-associated regions were found on chromosomes 2, 4, 13, 14, and 15 [29]. In a recent study, six significant SNPs consistently associated with both carotenoids and DMC were detected. They were mapped to candidate genes involved in ABA signaling, hydrolase activity, fatty-acid biosynthesis, cell-wall remodeling, and glycolytic metabolism, jointly explaining 75%–76% of the phenotypic variance [30].

While the genomic basis of water absorption during boiling (WAB) remains far less characterized than that of other root quality traits, SNPs associated with mealiness have been reported on chromosomes 6 and 13 in previous studies [25]. The identified SNPs on chromosome 13 are located in the vicinity of the candidate gene *Manes.13G026900* which is annotated as a plant invertase/pectin methylesterase inhibitor superfamily protein responsible for cell adhesion and the mealiness or crispness of various fruits and vegetables. Similar pectin-related gene activity has been observed to impact the texture of cooked potatoes [31–33]. Although WAB is reported to have limited genetic variation among cassava genotypes, most of the observed phenotypic variation was attributed to genotype-by-environment (G × E) interactions rather than additive genetic effects, suggesting a strong environmental influence on the trait [22]. Nevertheless, WAB is a critical post-harvest quality attribute that directly affects the cooking quality and consumer acceptance of boiled cassava and yam [23]. Therefore, improving WAB remains essential for increasing the adoption of improved cassava varieties in fresh consumption markets.

In this study, a genome wide association analysis was performed for five cassava root quality traits—DMC, TCC, BCC, HCN, and WAB using a multi-parental panel across three growing seasons with two genotyping platforms; genotyping by sequencing (GBS) and DArTag markers. Compared with biparental populations, multi-parental panels capture greater allelic diversity and more historical recombination, providing higher mapping resolution, stronger statistical power, and more broadly transferable marker–trait associations across diverse genetic backgrounds [34]. Identifying genetic markers associated with economically important traits provides valuable resources for marker-assisted selection (MAS) [35,36]. We hypothesized that integrating complementary genotyping platforms would improve the identification and validation of genomic regions controlling root quality traits while providing robust marker-trait associations for marker-assisted selection. The identified loci contribute to a better understanding of the genetic architecture of cassava root quality and provide valuable genomic resources to accelerate the development of cassava varieties with improved nutritional quality, processing characteristics, and consumer-preferred cooking attributes.

MATERIALS AND METHODS

Experimental Design and Field Establishment of Mapping Population

A field trial involving a multi-parental mapping population of 300 genotypes was conducted at Palmira, Colombia, using a partially replicated experimental design with two replications for the replicated entries. Each genotype was established in a single-row plot consisting of six plants, with 0.8 m spacing between plants within rows and 1.2 m spacing between adjacent plots. The population was evaluated over three

consecutive growing seasons: 2023/2024, 2024/2025, and 2025/2026. Roots were harvested 11 months after planting in each season. All evaluations were maintained under comparable management conditions. The average temperature, relative humidity, and mean daily precipitation recorded during the 2023/2024 season were 32.75 °C, 80.32%, 2.43 mm respectively. In the 2024/2025 season, the corresponding values were 27.07 °C, 80.88%, and 0.07 mm, respectively, while in 2025/2026 season, the average values of 24.34 °C, 78.15%, and 2.87 mm.

Trait Phenotyping

The dry matter content (DMC), total carotenoid content (TCC), and beta-carotene content (BCC) were estimated using near infrared spectroscopy (NIRS) as described by Belalcazar et al. (2016) [37]. At least three representative roots from each genotype were peeled and thoroughly washed to remove adhering soil. The roots were gently wiped dry to eliminate excess moisture which could affect the accuracy of data collected by the NIRS equipment. Using the SKYMSEN food processor (model PAEI) fitted with the smallest grating disc, the roots were grated until a homogenous paste was obtained. Approximately 8 g of the paste was transferred to a FOSS NIRS ring cup, ensuring complete and uniform coverage of the quartz window without exceeding the rim of the cup. The lid was secured so that it sat flush with the cup, and the sample was scanned using a FOSS NIRS DS2500 instrument (FOSS, Hillerød, Denmark). Two independently prepared ring cups were analyzed for each cassava genotype as technical replicates, and the mean predicted value was used for statistical analysis. The DMC, TCC, and BCC were predicted on fresh-weight basis using the corresponding NIRS calibration models.

Hydrogen cyanide content (HCN) was determined using the colorimetric method described by Essers et al. (1993) [38]. Fresh cassava root samples (40 g, fresh weight) were homogenized in 50 mL of extraction solution containing 0.1 M orthophosphoric acid in 25% (*v/v*) ethanol and 75% (*v/v*) distilled water. The homogenate was centrifuged at 6000 rpm for 10 min at 25 °C, after which 0.1 mL of the supernatant was mixed with 0.4 mL of 0.1 M phosphate buffer (pH 7.0) and 0.1 mL of linamarase enzyme. The mixture was incubated at 30 °C for 15 min to hydrolyze cyanogenic glucosides, followed by the addition of 0.6 mL of 0.2 M sodium hydroxide. After standing for 5 min at room temperature, 2.8 mL of 0.1 M phosphate buffer (pH 6.0), 0.1 mL of chloramine-T, and 0.6 mL of isonicotinate-1,3-dimethylbarbituric acid reagent were added sequentially, with incubation for 5 and 10 min, respectively. Absorbance was measured at 605 nm using a spectrophotometer, and HCN concentration was determined from a standard calibration curve and expressed as mg/kg dry-weight.

To determine water absorption at 20 and 30 min (WAB_20min and WAB_30min), eight roots of each genotype were peeled, cut into proximal, central, and distal parts. These were then weighed and submerged in boiling water for the respective durations. The weight increase following immersion was recorded and expressed as a percentage of the initial weight [22].

Extracting BLUPs for Each Trait

For all traits, the three-year data were collapsed into a single best linear unbiased prediction (BLUPs) using a linear mixed model $Y_{ijk} = \mu + \beta_j + g_i + (g\beta)_{ij} + rk_{(j)} + \varepsilon_{ijk}$ within the lme4 package in RStudio [39,40]. Where Y_{ijk} = observed trait value for the i-th genotype, in the j-th year, and k-th rep. μ = overall intercept (grand mean). β_j = fixed effect of the j-th year (Year is fixed). g_i = random effect of the i-th genotype. $(g\beta)_{ij}$ = genotype × environment interaction. $rk_{(j)}$ = random effect of the k-th replication nested within the j-th year. ε_{ijk} = residual error.

Trait Association Analysis Using Pearson Correlation

To evaluate the relationships among measured traits, pairwise correlations were computed using the Pearson correlation coefficient, which measures the strength and direction of linear associations between continuous variables. Prior to analysis, missing values were handled using complete case analysis, whereby only observations with no missing values across the selected traits were retained. This approach ensured that correlation estimates were not biased by incomplete data. Statistical significance of the pairwise correlations were assessed using two-tailed tests, with ($p < 0.05$), ($p < 0.01$), and ($p < 0.001$) considered statistically significant. The resulting correlation matrix was visualized using the corrplot package in R [40].

Extraction of Variance Components and Heritability Estimation

Variance components were estimated for all traits using linear mixed-effects models fitted in R with the lme4 package. For the combined analysis across years, the observed phenotypic value was modeled as a function of the overall mean, random genotypic effect, random year effect, random genotype × year interaction effect, random replication effect, and residual error. The general model was expressed as:

$$Y_{ijk} = \mu + G_i + E_j + (G \times E)_{ij} + Rk_{(j)} + \varepsilon_{ijk}$$

where Y_{ijk} is the observed trait value of the ith genotype in the j-th year and k-th replication; μ is the overall mean; G_i is the random genotypic effect; E_j is the random environmental (year) effect; $(G \times E)_{ij}$ is the random genotype × environment interaction effect; $R_{k(j)}$ is the random replication effect nested within year; and ε_{ijk} is the residual error. The coefficient of variation (CV) for each year was calculated independently using a year-specific linear mixed-effects model with genotype and replication fitted as

random effects. The CV was calculated as the ratio of the residual standard deviation to the corresponding year-specific trait mean and expressed as a percentage:

$$CVy = \frac{\sigma_{\varepsilon,y}}{\bar{Y}_y} \times 100$$

where $\sigma_{\varepsilon,y}$ is the residual standard deviation estimated for year y, and $\bar{Y}_y$ is the corresponding year-specific trait mean.

The broad-sense heritability across environment was calculated as

$$H^2 = = \frac{\sigma_G^2}{\sigma_G^2 + (\sigma_{GE}^2/e) + (\sigma_E^2/re)}$$

where H^2 is the broad-sense heritability, σ_G^2 is genotypic variance, σ_{GE}^2 is genotype × environment interaction variance, σ_E^2 is Residual (error) variance, e is number of environments (years), r is number of replications per environment (year).

Genotyping, SNP Calling, and Quality Control

The SNP datasets used in this study were previously generated using DArTag and genotyping-by-sequencing (GBS) platforms in an earlier study involving the same cassava population [40]. Marker positions for the GBS dataset were aligned to the cassava reference genome assembly version 8.1 (v8.1), whereas the DArTag marker positions were based on the cassava mid-density panel, which was originally mapped to version 7.1 (v7.1). The SNP filtering was performed based on missing data and minor allele frequency (MAF) thresholds, resulting in 2526 high-quality SNPs for the DArTag dataset and 30,046 SNPs for the GBS dataset. Detailed procedures for SNP calling, sequence alignment, variant discovery, and marker filtering are described in Dankwa et al. (2026) [41].

Marker Distribution, Linkage Disequilibrium, and Population Structure

Genome-wide marker distribution, SNP density, linkage disequilibrium (LD) decay, principal component analysis (PCA), and kinship estimation for the DArTag and GBS datasets were previously assessed and reported for the same germplasm population [41]. The smoothed LD decay curves reached the background threshold ($r^2 \approx 0.10$) at approximately 685,090 bp for the DArTag markers and 625,262 bp for the GBS markers. Furthermore, PCA demonstrated a high degree of concordance between the two marker platforms; the first two principal components explained 21.3% and 22.2% of the total genetic variation for the DArTag and GBS datasets, respectively, reflecting comparable large-scale population structure. Both platforms consistently resolved the same major genetic clusters, with accessions grouping according to shared parental backgrounds. Consequently, these previously generated kinship matrices and population structure covariates were incorporated into the downstream GWAS model to account for genetic relatedness and minimize confounding population stratification.

Association Analyses

The phenotypic (BLUP values) and genotypic data were integrated to identify marker–trait associations using the Genomic Association and Prediction Integrated Tool (GAPIT) package version 3.4.0 in R, employing the BLINK model [42]. The model BLINK has been demonstrated to provide greater statistical power than several widely used GWAS models, while substantially reducing computational time for large-scale genomic analyses. In addition, it is highly robust for complex genetic architectures, effectively controlling for p value inflation arising from population structure and cryptic relatedness [43,44]. Notably, BLINK performs comparably to, or outperforms traditional mixed linear models in minimizing false-positive associations while improving the detection of true genetic signals [44]. Significant marker–trait associations were identified using a Bonferroni-corrected significance threshold ($p < 0.05/m$), where m represents the total number of SNPs tested.

Significant Marker Distribution Analysis

Significant marker distribution analysis was performed to evaluate the phenotypic effects of allelic variation at loci identified through genome-wide association analysis. Significant SNPs detected using the BLINK model in GAPIT were extracted and integrated with corresponding phenotypic and genotypic data in R within RStudio. For each significant SNP, individuals were grouped according to allelic classes, and the distribution of phenotypic values for DMC, HCN, TCC, BCC, and WAB_20min was visualized using violin plots generated with the ggplot2 package.

Local Linkage Disequilibrium Analysis

Local linkage disequilibrium (LD) was evaluated around the major genome-wide association signals for total carotenoid content (TCC), beta-carotene content (BCC), and hydrogen cyanide content (HCN) identified from the GBS and DArTag datasets. For each lead SNP, pairwise LD was calculated between the lead marker and neighboring SNPs within a ±500-kb physical window using PLINK v1.9 [45]. LD was measured as the squared correlation coefficient (r^2) between marker genotypes. An r^2 threshold of 0.20 was used to identify neighboring markers showing LD above the defined threshold with the lead SNP, consistent with the threshold used to characterize genome-wide LD decay. The extent of local LD support was assessed based on the physical distance between the lead SNP and neighboring markers with $r^2 \geq 0.20$. Local LD patterns were compared between GBS and DArTag to assess whether major associations occurred within shared LD-supported genomic regions across marker platforms.

Conversion of Significant SNP Coordinates between Cassava Genome Assemblies

Significant SNP markers identified from DArTag were converted from the *Manihot esculenta* v7.1 reference genome (https://phytozome-next.jgi.doe.gov/info/Mesculenta_v7_1) to *Manihot esculenta* v8.1 reference genome (https://phytozome-next.jgi.doe.gov/info/Mesculenta_v8_1) using a BLAST-based workflow. For each marker, a 2001 bp sequence centered on its v7.1 coordinate, comprising 1 kb upstream and 1 kb downstream, was extracted and aligned against the v8.1 reference genome using BLASTn. Alignments were required to have at least 95% sequence identity and an E-value ≤ 1×10^{-20}. When multiple alignments were recovered, the best supported position was selected based primarily on the highest bit score and, in the event of a tie, the longest alignment. After selecting the highest-ranking alignment, the v8.1 marker coordinate was assigned to the genomic position corresponding to the central nucleotide of the aligned flanking sequence, taking the alignment orientation into account.

RESULTS

Cassava Root Quality Traits Distribution

The distribution of best linear unbiased prediction (BLUP) values for the evaluated cassava root quality traits revealed substantial phenotypic variation among genotypes (Figure 1). The DMC exhibited an approximately normal distribution with moderate variation around a mean of 40.21%, ranging from 33.48% to 45.41%. HCN showed a positively skewed distribution with a mean of 62.90 μg/g and values spanning 1.40 to 209.18 μg/g. The carotenoids, TCC and BCC displayed similarly skewed distributions, with mean values of 1.23 μg/g and 0.88 μg/g, respectively. TCC ranged from 0.14 to 4.23 μg/g, whereas BCC varied from 0.06 to 2.47 μg/g, with most genotypes concentrated at lower values and only a few exhibiting markedly elevated carotenoid levels. The low carotenoid levels were expected as most of the genotypes in the population were white-fleshed. WAB_20min and WAB_30min also showed continuous variation among genotypes. WAB_20min displayed moderate positive skewness with a mean of 8.36% and values ranging from 3.39% to 17.04%, while WAB_30min approached a near-normal distribution with a mean of 20.19% and a range of 5.57% to 33.04%.

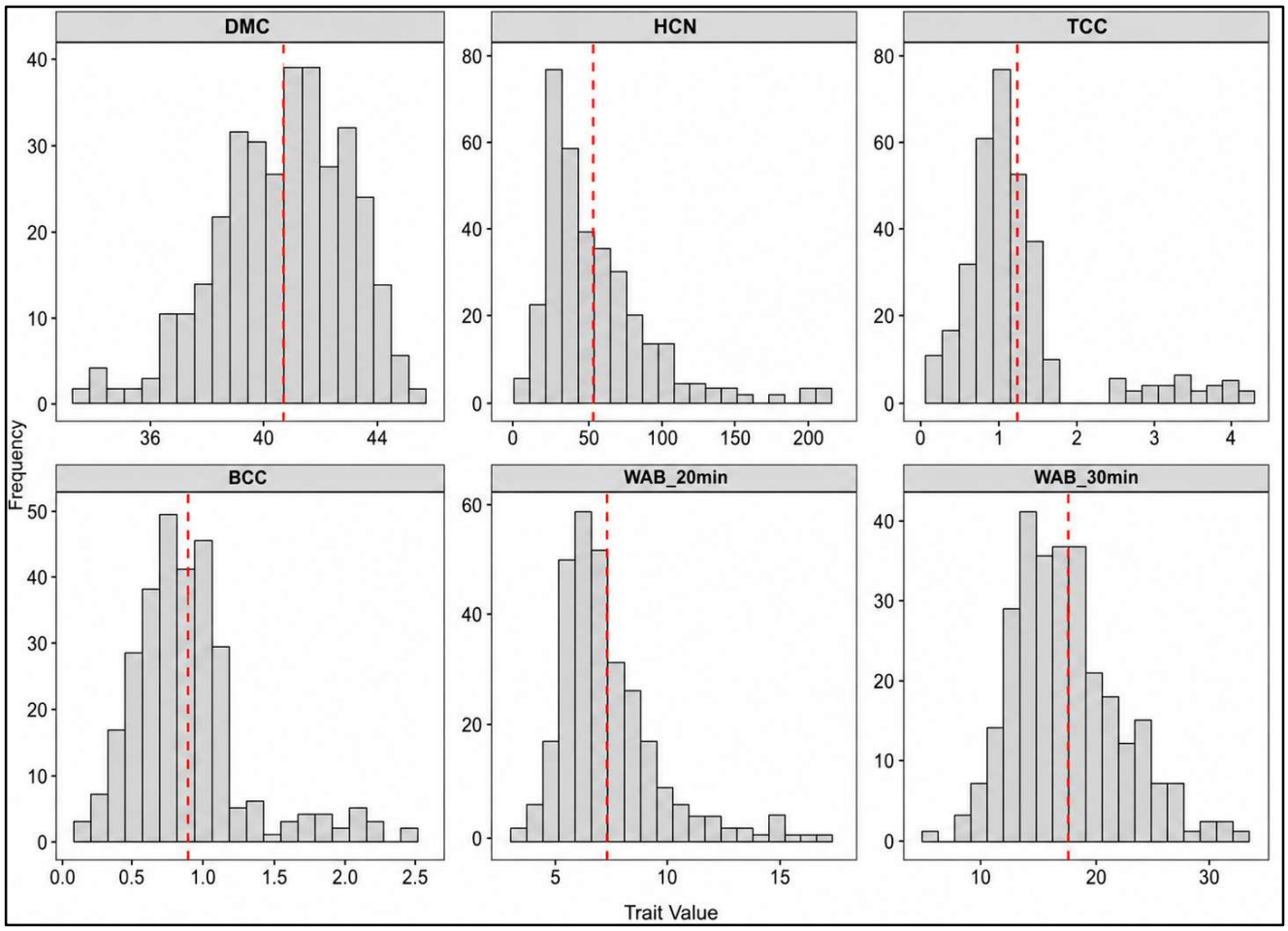


Figure 1. Phenotypic distribution of best linear unbiased predictors (BLUPs) for five cassava root quality traits. The dashed red vertical line in each panel indicates the population mean. DMC: dry matter content (%); HCN: hydrogen cyanide content (μg/g); TCC: total carotenoid content (μg/g); BCC: beta-carotene content (μg/g); WAB_20min and WAB_30min: water absorption at 20 and 30 min (%), respectively.

Correlation Analysis

Pairwise correlation analysis among the evaluated cassava root quality traits revealed distinct patterns of association (Figure 2). A strong and positive correlation was observed between water absorption measured at 20 min and 30 min (r = 0.80), indicating consistency between the two time points. Water absorption traits showed weak to moderate negative correlations with both dry matter content (DMC) (r = −0.18 for WAB_20min; r = −0.27 for WAB_30min) and hydrogen cyanide content (HCN) (r = −0.14 for WAB_20min; r = −0.16 for WAB_30min).

Total carotenoid content (TCC) exhibited a very strong positive correlation with beta-carotene content (BCC) (r = 0.95). In contrast, carotenoid traits (TCC and BCC) displayed negligible correlations with DMC (r = −0.05 and 0.04, respectively), HCN (r = 0.04 and 0.03, respectively), and water absorption metrics ($r \leq 0.03$).

Additionally, DMC showed a weak positive correlation with HCN (r = 0.27), pointing to a slight trend where genotypes with higher dry matter also possess increased cyanide content. Overall, most trait pairs exhibited

low-to-moderate correlations with the notable exceptions of the TCC–BCC and WAB_20–WAB_30min associations.

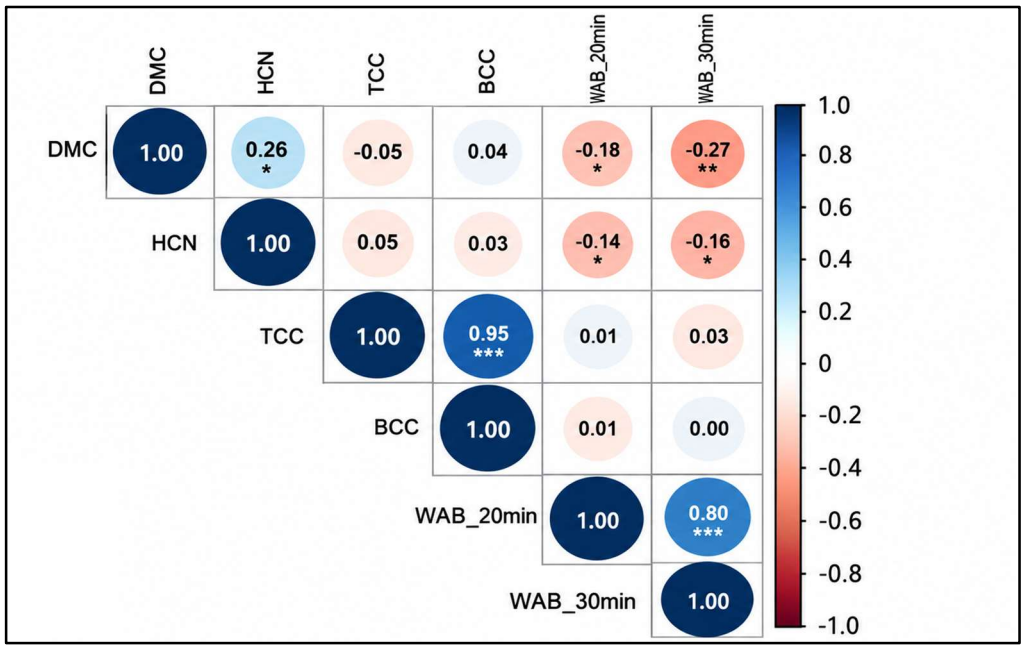


Figure 2. Correlation matrix of best linear unbiased predictors (BLUPs) among cassava root quality traits. Values within cells represent Pearson correlation coefficients (*r*), with blue indicating positive correlations and red indicating negative correlations. DMC: dry matter content; HCN: hydrogen cyanide content; TCC: total carotenoid content; BCC, beta-carotene content; WAB_20min and WAB_30min: water absorption at 20 and 30 min, respectively. Statistical significance is indicated as blank ($p \geq 0.05$), * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$).

Marker Density and Distribution, Linkage Disequilibrium, and Population Structure

The filtered DArTag and GBS marker datasets provided comprehensive genome-wide coverage suitable for association mapping. Detailed information on SNP density and chromosomal distribution across the cassava genome has been reported previously [43]. Likewise, the linkage disequilibrium (LD) decay analyses for both marker platforms demonstrated sufficient mapping resolution for genome-wide association studies (GWAS) and were used to define candidate genomic intervals surrounding significant marker–trait associations [43]. Population structure analysis of the germplasm panel also revealed moderate genetic stratification among the evaluated genotypes [43]. To account for this underlying structure and the relatedness among individuals, both population structure and kinship were incorporated into the GWAS model, thereby reducing the likelihood of spurious marker–trait associations.

Quantitative Genetic Parameters

The quantitative genetic parameters for all evaluated traits are summarized in Table 1. Genotypic variance (σ^2G) exceeded both genotype-

by-environment interaction variance (σ^2GE) and residual variance (σ^2E) for most traits, particularly for HCN, TCC, and BCC, indicating substantial genetic contribution to phenotypic variation. The WAB components, WAB_20min and WAB_30min exhibited σ^2GE values (11.04 and 27.41, respectively) that exceeded or approached their respective genotypic variances.

Broad-sense heritability (H^2) estimates were generally high, ranging from 0.58 to 0.96. TCC exhibited the highest heritability (0.96), closely followed by BCC (0.95), HCN (0.92), and DMC (0.90), indicating strong genetic control of these traits. Lower heritability estimates were observed for WAB_20min (0.58) and WAB_30min (0.70).

The coefficient of variation (CV) varied across traits and seasons. DMC consistently exhibited the lowest CV values across seasons (3.05%–5.46%). In contrast, HCN and water absorption traits displayed comparatively higher CV values, particularly HCN in Season 3 (38.70%) and WAB_20min in Season 2 (42.40%). TCC and BCC showed moderate CV values across seasons, ranging from 14.30% to 28.50%.

Table 1. Estimates of mean, variance components, broad-sense heritability, and coefficients of variation for root quality traits across three seasons in 300 cassava genotypes.

Trait	Mean	σ^2_G	σ^2_{GE}	σ^2_E	H^2	CV (%)		
						Season 1	Season 2	Season 3
DMC	40.21%	5.83	0.15	3.62	0.90	5.46	3.05	4.46
HCN	62.90 µg/g	1242.90	59.74	503.35	0.92	33.60	28.80	38.70
TCC	1.23 µg/g	0.61	0.03	0.09	0.96	26.20	14.30	20.30
BCC	0.88 µg/g	0.18	0.01	0.05	0.95	28.50	20.70	25.00
WAB_20min	8.36%	6.98	11.04	7.60	0.58	36.00	42.40	39.10
WAB_30min	20.19%	30.28	27.41	23.96	0.70	32.90	27.70	31.50

σ^2_G = genotypic variance. σ^2_{GE} = genotype × environment interaction variance. σ^2_E = residual error variance. H^2 = broad-sense heritability. CV = coefficient of variation.

Markers Linked to Root Quality Traits

Genome-wide association analysis utilizing both GBS and DArTag markers identified significant loci associated with cassava root quality traits under the BLINK model (Tables 2 and 3). The two marker platforms demonstrated good agreement especially in chromosomes associated with the studied traits. Both datasets identified significant associations for DMC, HCN, TCC, BCC, and WAB_20min, whereas no significant loci were detected for WAB_30min.

For the carotenoid-related traits, both GBS and DArTag detected a major genomic region on chromosome 1, demonstrating strong concordance between the two marker systems. In the GBS dataset, SNPs S01_25897096 and S01_31747016 were consistently associated with both TCC and BCC. Notably, S01_31747016 exhibited highly significant associations, explaining 48.11% and 49.56% of the phenotypic variance for TCC and BCC, respectively. Similarly, DArTag analysis detected regions on chromosome 1 for TCC and BCC via the marker CassV7_chr01_27813884,

which was also consistently associated with both traits and accounted for over 80% of the phenotypic variance.

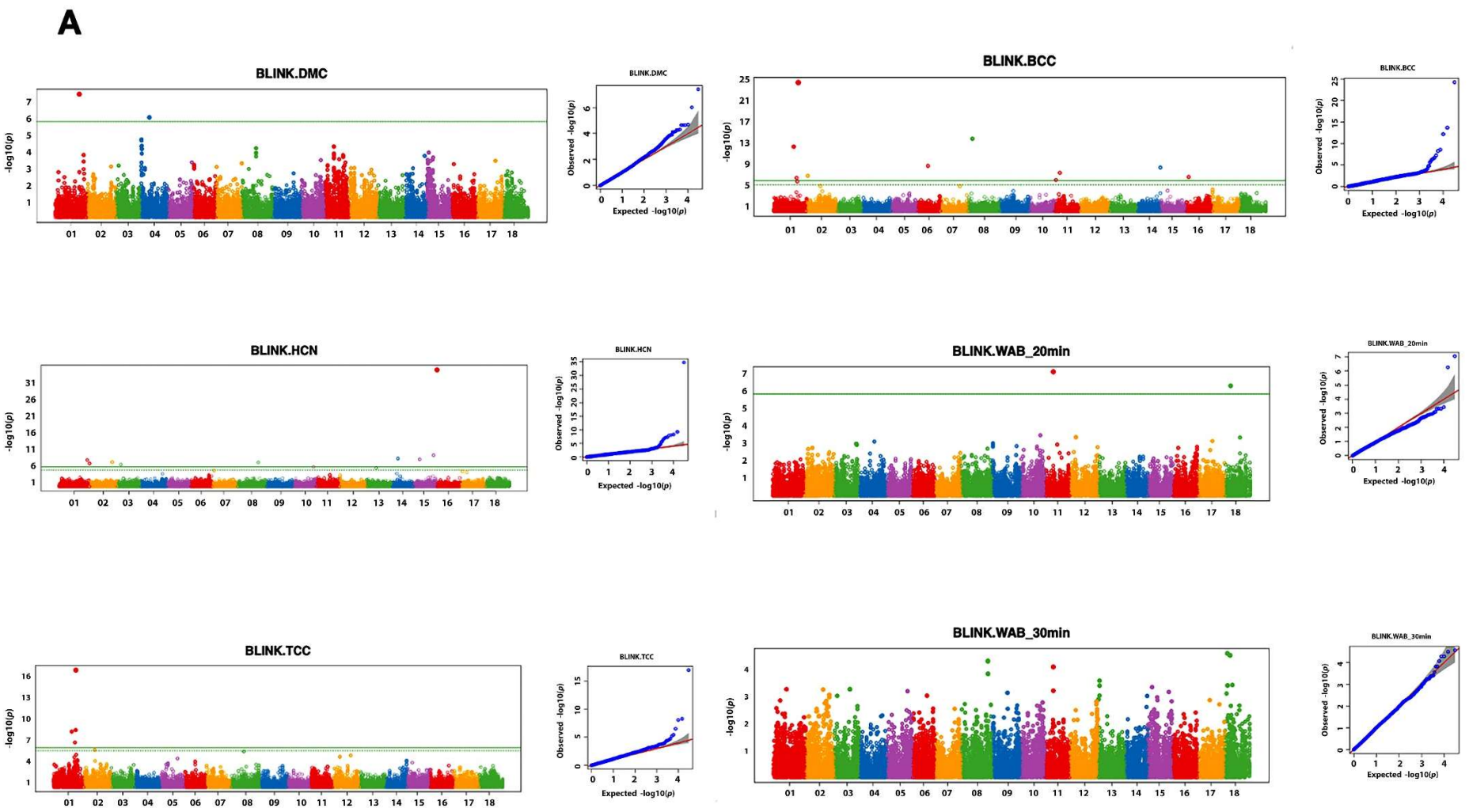
Similarly, both marker platforms identified chromosome 16 as the primary genomic region associated with hydrogen cyanide (HCN) content. Specifically, GBS detected marker S16_810004, whereas DArTag identified CassV7_chr16_693845. The DArTag marker explained up to 36.31% of the phenotypic variance.

For dry matter content (DMC), GBS detected significant loci on chromosomes 1 and 4, whereas DArTag detected significant associations on chromosome 14. For WAB_20min, GBS markers showed association on chromosomes 11 and 18 (S11_11002934, S18_7289602) whileDArTag identified a single peak on chromosome 7 (CassV7_chr07_29527635).

The identified positions associated with TCC and BCC on chromosome 1 and HCN on chromosome 16 showed concordance between the two genotyping platforms, whereas the major associations detected for DMC and WAB_20min differed between GBS and DArTag. For TCC and BCC, the significant positions identified (v8.1) using GBS and DArTag were located at 31.75 Mb and 31.96 Mb, respectively, corresponding to a physical distance of approximately 0.21 Mb (210 kb). Similarly, the HCN-associated positions on chromosome 16 were located at 0.810 Mb and 0.699 Mb in GBS and DArTag, respectively, with a distance of approximately 0.111 Mb (111 kb).

The local LD analysis showed that the major TCC, BCC, and HCN associations were generally supported by neighboring markers, although the extent of LD support differed between platforms. The TCC/BCC-associated SNP S01_31747016 on chromosome 1 showed moderate LD with nearby markers, with the strongest neighboring associations reaching r^2 = 0.31 within 91–365 kb. A second chromosome 1 signal, S01_31747046, showed stronger local LD, with several neighboring markers exhibiting r^2 > 0.20 and values up to 0.61 within approximately 475 kb. In the DArTag dataset, the corresponding chromosome 1 signal (CassV7_chr01_31974032) showed strong LD with a neighboring marker 290 kb away (r^2 = 0.71). For HCN, the chromosome 16 GBS-associated SNP (S16_810004) exhibited extensive local LD, with several neighboring markers showing r^2 > 0.20 and maximum values of approximately 0.80 within 500 kb. Similarly, the DArTag-associated SNP (CassV7_chr16_693845) showed strong LD with neighboring markers, including a marker 89 kb away with r^2 = 1.00 and another approximately 496 kb away with r^2 = 0.34. Overall, these results indicate that the major TCC, BCC, and HCN associations occurred within LD-supported genomic regions rather than representing completely isolated markers, with stronger and more extensively characterized local LD observed in the higher-density GBS dataset.

The Manhattan plots displayed distinct association peaks corresponding to these significant loci across traits, and Q–Q plots indicated adequate control of population structure with minimal false-positive inflation (Figure 3).


A
BLINK.DMC
BLINK.BCC
BLINK.HCN
BLINK.WAB_20min
BLINK.TCC
BLINK.WAB_30min
-log10(p)
Observed -log10(p)
Expected -log10(p)

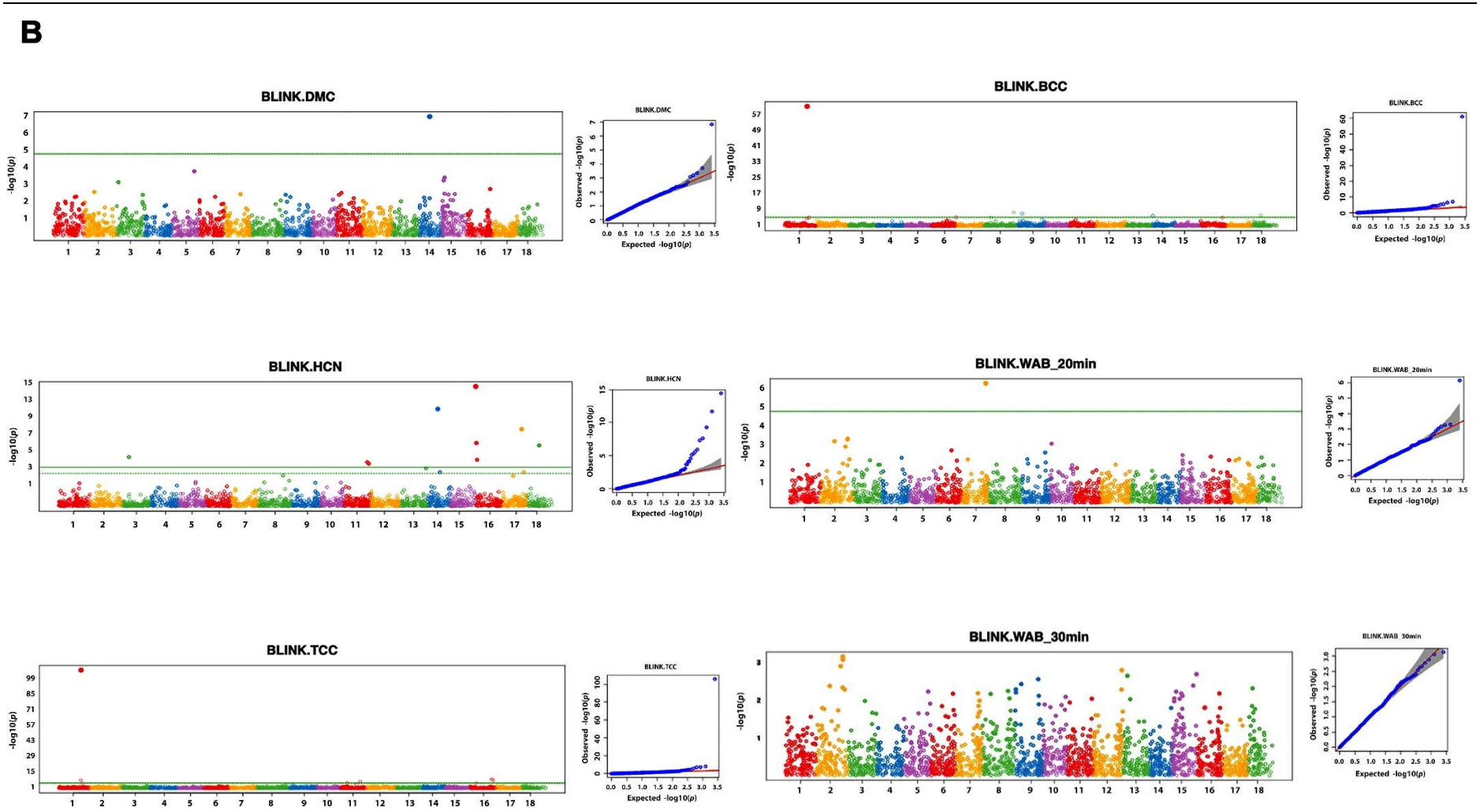


Figure 3. Manhattan and quantile–quantile (Q–Q) plots from genome-wide association studies (GWAS) for cassava root quality traits using the BLINK model in 300 cassava accessions: (**A**) GBS markers and (**B**) DArTag markers. DMC: dry matter content; HCN: hydrogen cyanide content; TCC: total carotenoid content; BCC, beta-carotene content; WAB_20min and WAB_30min: water absorption at 20 and 30 min, respectively. Horizontal green lines indicate the significance threshold.

Table 2. Summary of identified significant SNPs for cassava root quality traits using GBS markers with their genomic inflation factor (λ).

Trait	SNP_ID	Chromosome	Position (v8.1)	*p* value	Minor allele frequency	Phenotypic variance explained (%)
DMC (λ = 1.01)	S01_32444435	1	32444435	3.84×10^{-8}	0.49	14.85
	S04_11557168	4	11557168	9.50×10^{-7}	0.08	46.63
HCN (λ = 1.06)	S01_39539583	1	39539583	1.31×10^{-8}	0.12	4.99
	S01_42779500	1	42779500	1.40×10^{-7}	0.09	4.77
	S02_31580928	2	31580928	5.81×10^{-8}	0.18	3.81
	S03_5308326	3	5308326	3.16×10^{-7}	0.23	4.22
	S08_30022353	8	30022353	7.22×10^{-8}	0.36	5.75
	S10_28393883	10	28393883	1.48×10^{-6}	0.31	3.99
	S14_7913036	14	7913036	4.83×10^{-9}	0.46	6.27
	S15_9189016	15	9189016	8.50×10^{-9}	0.15	3.10
	S15_28733703	15	28733703	5.12×10^{-10}	0.42	5.66
	S16_810004	16	810004	1.73×10^{-35}	0.23	42.78
TCC (λ = 1.22)	S01_25897096	1	25897096	8.42×10^{-9}	0.20	38.67
	S01_30421344	1	30421344	3.08×10^{-7}	0.07	6.88
	S01_31502986	1	31502986	4.99×10^{-9}	0.18	0.40
	S01_31747016	1	31747016	1.10×10^{-17}	0.06	48.11
BCC (λ = 1.12)	S01_25897096	1	25897096	6.71×10^{-13}	0.20	3.36
	S01_29461308	1	29461308	5.40×10^{-7}	0.48	0.54
	S01_31747016	1	31747016	5.47×10^{-25}	0.06	49.56
	S02_1246166	2	1246166	2.15×10^{-7}	0.49	0.54
	S06_14143048	6	14143048	2.88×10^{-9}	0.26	0.87
	S08_5797008	8	5797008	2.16×10^{-14}	0.09	41.19
	S11_2245363	11	2245363	1.30×10^{-6}	0.16	0.39
	S11_7248761	11	7248761	5.81×10^{-8}	0.44	0.78
	S14_28977948	14	28977948	5.63×10^{-9}	0.15	0.33
	S16_3315007	16	3315007	3.60×10^{-7}	0.21	0.29
WAB_20min (λ = 0.96)	S11_11002934	11	11002934	9.14×10^{-8}	0.21	9.51
	S18_7289602	18	7289602	5.62×10^{-7}	0.11	64.91
WAB_30min	---	---	---	---	---	---

SNP positions are based on the *Manihot esculenta* reference genome assembly v8.1.

Table 3. summary of identified significant SNPs for cassava root quality traits using DArTag markers with their genomic inflation factor (λ).

Trait	SNP_ID	Chromosome	Position (v 7.1)	Position (v8.1)	*p* value	Minor allele frequency	Phenotypic variance explained (%)
DMC (λ = 1.08)	CassV7_chr14_12795714	14	12795714	9475089	1.45×10^{-7}	0.20	71.20
HCN (λ = 1.18)	CassV7_chr03_8150269	3	8150269	7778125	1.24×10^{-6}	0.43	1.74
	CassV7_chr11_28673541	11	28673541	28543652	5.07×10^{-6}	0.36	3.54
	CassV7_chr11_30512451	11	30512451	30369172	7.66×10^{-6}	0.21	3.25
	CassV7_chr14_12449260	14	12449260	10509960	2.00×10^{-12}	0.29	6.30
	CassV7_chr16_693845	16	693845	698512	3.73×10^{-15}	0.30	36.31
	CassV7_chr16_1585893	16	1585893	1715759	2.52×10^{-8}	0.28	10.92
	CassV7_chr16_2243532	16	2243532	2241622	2.59×10^{-6}	0.38	5.73
	CassV7_chr17_24332833	17	24332833	27463679	5.55×10^{-10}	0.47	4.43
	CassV7_chr18_14814434	18	14814434	12711859	5.01×10^{-8}	0.23	7.94
TCC (λ = 1.20)	CassV7_chr01_27491928	1	27491928	31638400	5.30×10^{-8}	0.31	0.79
	CassV7_chr01_27813884	1	27813884	31960251	1.79×10^{-106}	0.05	89.59
	CassV7_chr11_24974203	11	24974203	24830647	1.44×10^{-6}	0.31	0.88
	CassV7_chr16_27306321	16	27306321	28827261	8.34×10^{-9}	0.20	1.30
	CassV7_chr16_29232527	16	29232527	30627538	7.30×10^{-8}	0.28	5.04
BCC (λ = 1.11)	CassV7_chr01_27813884	1	27813884	31960251	1.21×10^{-61}	0.05	82.67
	CassV7_chr08_37070625	8	37070625	37890606	1.06×10^{-7}	0.35	0.86
	CassV7_chr09_6567149	9	6567149	7082110	3.75×10^{-7}	0.30	9.37
	CassV7_chr14_321954	14	321954	1345410	5.82×10^{-6}	0.29	0.50
	CassV7_chr18_10920912	18	10920912	8764628	3.80×10^{-6}	0.44	1.39
WAB_20min (λ = 1.09)	CassV7_chr07_29527635	7	29527635	31166514	7.07×10^{-7}	0.42	25.06
WAB_30min	---	---	---		---	---	---

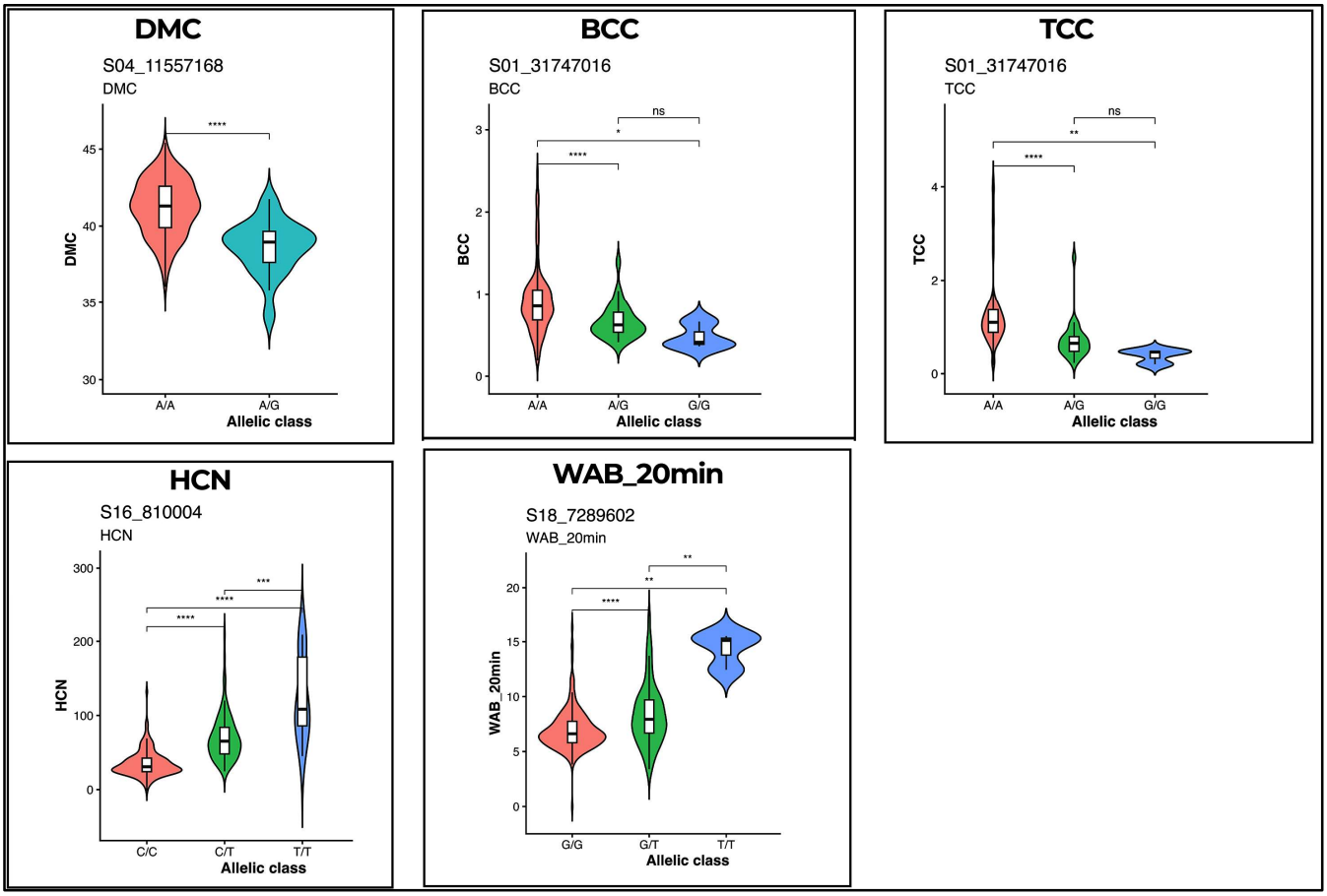
DArTag assay design and original coordinates correspond to *Manihot esculenta* reference genome v7.1, with updated positions provided for v8.1 to allow direct comparison with GBS markers.

Allelic Effects and Genotypic Distribution of Significant Loci

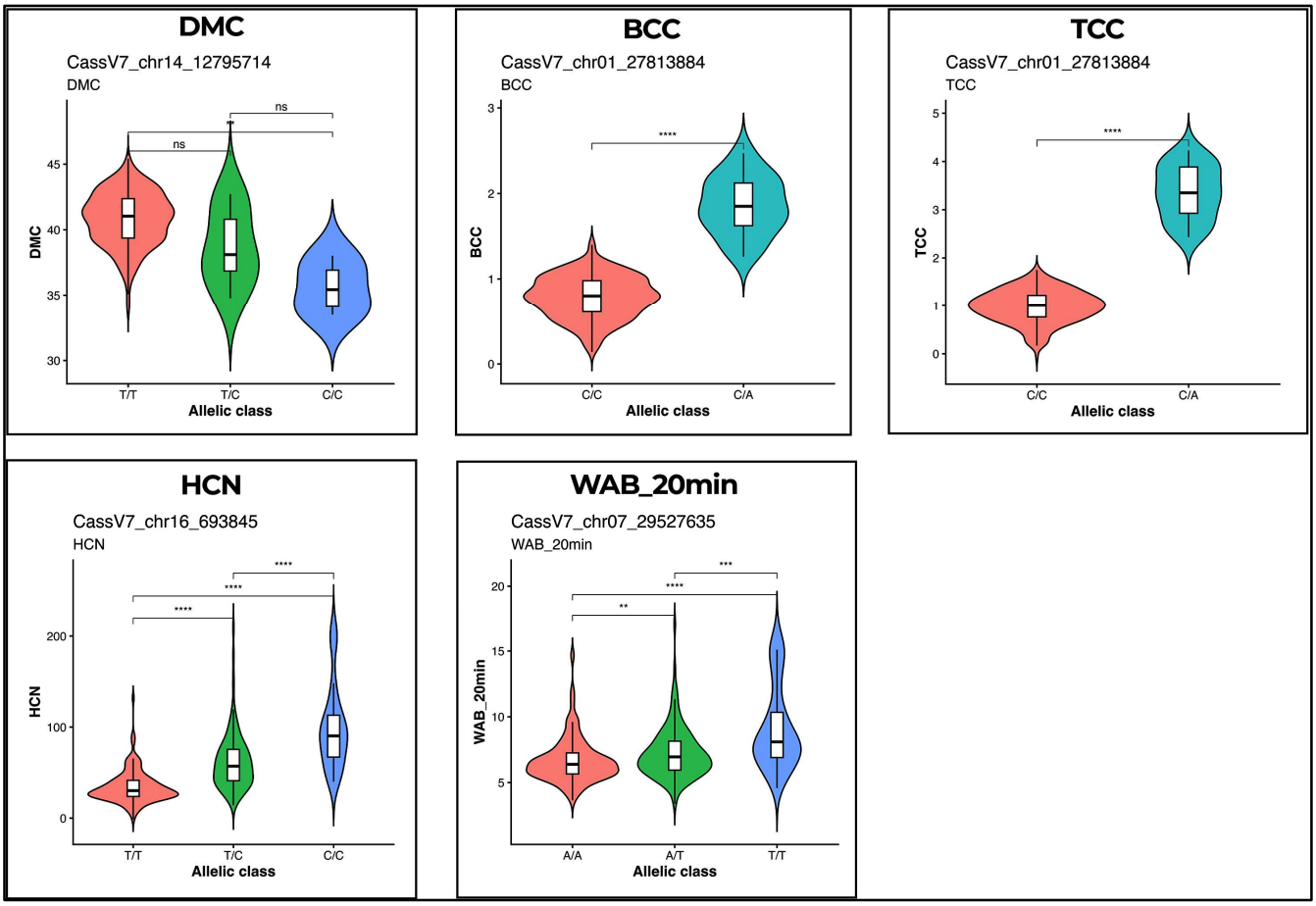
Marker-trait association analyses of the significant SNPs revealed distinct allelic effects on root quality traits, though the magnitude, direction and distribution varied among loci and traits (Figure 4; Figure S1). Figure 4 illustrates markers explaining the highest phenotypic variance (PVE) for each trait (Tables 2 and 3). For GBS markers (Figure 4A), the DMC-associated marker S04_11557168 (PVE = 46.63%) displayed two allelic classes (A/A and A/G), with the homozygous A/A class dominating the population ($p < 0.0001$).

TCC and BCC marker S01_31747016 (PVE = 48.11% and 49.56% respectively) showed three allelic classes (A/A, A/G, and G/G) with the G/G class sparsely populated, as evidenced by the narrow and compressed violin shape. This could drive the high PVE at this locus.

The HCN marker S16_810004 (PVE = 42.78%) showed a progressive increase in cyanogenic potential from C/C to T/T ($p < 0.001$). However, the T/T class was underrepresented and exhibited a wide phenotypic spread (up to >300 mg/kg), contributing to within-class heterogeneity. Finally, for WAB_20min, marker S18_7289602 (PVE = 64.91%) showed that the T/T homozygous class exhibited the highest median water absorption.



(A)



(**B**)

Figure 4. Representative violin plots for the significant SNP markers with the highest phenotypic variance explained (PVE) identified for each trait for (**A**) GBS markers and (**B**) DArTag markers. Each panel represents a significant marker locus, with violin shapes illustrating phenotype distribution, embedded boxplots indicating the median and interquartile range. DMC: dry matter content; HCN: hydrogen cyanide content; TCC: total carotenoid content; BCC, beta-carotene content; WAB_20min and WAB_30min water absorption at 20 min. Statistical significance is indicated as ns ($p \geq 0.05$), * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$), and **** ($p < 0.0001$).

For the DArTag markers (Figure 4B), the DMC marker CassV7_chr14_12795714 (PVE = 71.20%) displayed three allelic classes (T/T, T/C, and C/C). The non-significant difference (ns) between T/T and T/C, alongside a significant reduction in DMC for C/C ($p < 0.01$), suggests a dominant gene action for the T allele in driving high dry matter content. The HCN marker CassV7_chr16_693845 (PVE = 36.31%) displayed a highly significant stepwise increase in HCN from T/T to C/C ($p < 0.0001$), though the low-HCN T/T class was notably low in frequency. TCC and BCC—both mapping to the same locus, CassV7_chr01_27813884 (PVE = 89.59% and 82.67%, respectively), showed only two allelic classes. This fixed state concentrates the variance into a single pairwise contrast, yielding exceptionally high PVE estimates (>80%). For the WAB_20min, marker (CassV7_chr07_29527635) (PVE = 25.06%) showed a clear dosage-

dependent increase in water absorption capacity from A/A to T/T (p < 0.0001).

DISCUSSION

Cassava root is widely considered the most economically important part of the plant, accounting for roughly 50% of the mature plant's biomass and serving as the primary storage organ for starch, which is the basis of its economic value [46]. In this study, the identified loci for all evaluated root quality traits were distributed across different chromosomes, underscoring the polygenic nature and broad genomic control underlying cassava root quality.

High dry matter content (DMC), which is directly linked to starch yield in cassava, is a critical quality characteristic that benefits both industrial processors and subsistence farmers [47,48]. Higher DMC increases economic value per unit mass for industrial applications (such as starch and ethanol extraction) and provides higher caloric content with lower water weight for subsistence farm households [49,50]. Previous studies [12] identified a major locus for DMC on chromosome 1 that co-localized with a peak for provitamin A carotenoids, demonstrating a genetic association between high dry matter and root carotenoid accumulation. Consistent with these findings, the present study identified a significant GBS marker associated with DMC at position 32,444,435 bp on chromosome 1, while another marker at 31,747,016 bp on the same chromosome was significantly associated with both TCC and BCC.

The proportion of phenotypic variance explained by the GBS-associated SNPs varied widely across traits, with DMC and carotenoid-related traits consistently showing the largest individual locus effects—suggesting that these quality traits may be influenced by a few major-effect genes. Identifying multiple significant SNPs associated with the same root quality trait across the genotyping platforms (GBS and DArTag) confirms the polygenic nature of these traits [11,51]. The varying proportions of phenotypic variance explained by the identified significant SNPs suggests the presence of both major- and minor-effect loci influencing root quality. Marker effects observed across different chromosomes, demonstrates broad genome-wide control of these traits. The high phenotypic variance explained (PVE) by several markers in this study may be inflated due to the Beavis effect, whereby effect sizes of significant loci are often overestimated in studies with limited population sizes [52,53]. In our empirical observations of allelic distributions (Figure 4), this bias was exacerbated by rare or underrepresented genotypic classes. For instance, at marker CassV7_chr01_27813884, the complete absence of the alternative homozygous class (A/A) restricted variance partitioning to a single pairwise contrast, slightly elevating the PVE. Similarly, lower minor allele frequencies may have disproportionately inflated between-group variance components [54]. Previous studies have shown that PVE inflation decreases with increasing sample size, and larger populations of

approximately 1000 individuals may be required to obtain precise estimates for polygenic traits [55,56]. Some of the PVE estimates reported here, may represent upper-bound values that require validation in larger, independently sampled populations with more balanced allele frequencies. However, the relatively elevated genomic inflation factor (λ) observed for some traits does not necessarily indicate poor model fit or the presence of bias and may arise from genuine polygenic signals rather than confounding effects. Complex traits are often influenced by numerous loci distributed across the genome, and the cumulative effects of these loci can legitimately inflate test statistics, resulting in higher λ values even when population structure and relatedness are adequately controlled [57].

Although a negative phenotypic correlation between DMC and carotenoid content has been widely reported in cassava breeding populations [17,58,59] the phenotypic correlation observed in the present panel was negligible ($r = -0.05$ for TCC, $r = 0.04$ for BCC; Figure 2). This apparent discrepancy may reflect differences in population composition. Most genotypes in this panel were white- or cream-fleshed with low carotenoid content, restricting the phenotypic range over which a trade-off between DMC and carotenoids could be detected. Alternatively, it may indicate that the previously reported trade-off is population-specific rather than a general genetic constraint. Some studies suggest that this relationship may result from physical linkage, particularly megabase-scale linkage disequilibrium on chromosome 1 [12,14], whereas others have proposed pleiotropy as the underlying cause [17]. The identification of distinct but close genomic positions on chromosome 1 for DMC and TCC in this present study may support the hypothesis that the association between these traits is driven by physical linkage rather than pleiotropy. However, the genetic basis of this relationship remains unresolved and warrants further investigation.

The shared detection of loci for TCC and BCC (Tables 2 and 3) is also biologically plausible because beta-carotene constitutes a major component of total carotenoids in cassava storage roots [60,61]. Consequently, genomic regions controlling beta-carotene biosynthesis are expected to substantially influence total carotenoid levels. The close proximity of the TCC and BCC associations detected by GBS (31.75 Mb) and DArTag (31.96 Mb) on v8.1 reference assembly suggests a common underlying genomic region, providing evidence of cross-platform replication and supporting the robustness of this genomic region for TCC and BCC.

Consumer acceptability is a key breeding objective. While dry matter and water absorption are very important traits for acceptability by consumers who boil or process the roots [62], high levels of carotenoids and beta-carotene is essential for improving the nutritional value of diets in regions where cassava is a staple food [62,63].

Several positions were identified for HCN on chromosome 16 across genotyping platforms and models. This confirms several reports that primarily associate cassava HCN to chromosomes 16 and 14 [26,64]. The genetic architecture underlying HCN levels in cassava differs markedly between chromosomes 14 and 16. Loci on chromosome 16 exhibit an additive pattern of inheritance, contributing substantially to the natural variation in HCN content. In contrast, chromosome 14 displays a dominance pattern of inheritance, with a much smaller contribution to the observed natural variation in HCN levels [64]. Besides these, this study identified novel positions on chromosomes 2, 6, 7, 8, and 13 for HCN.

For the GBS markers, the SNP S01_32444435 on chromosome 1, which was associated with DMC, is located within the putative gene *Manes.01G131400*. This gene has been reported to be involved in shaping tubular endoplasmic reticulum (ER) membranes and regulating localized membrane tension. Given that the tubular ER is an important site for the synthesis of proteins and lipids involved in cellular growth and storage, *Manes.01G131400* represents a positional candidate gene that may be relevant to DMC variation. However, the observed marker–trait association does not establish a causal role for this gene, and its potential contribution to dry matter accumulation requires further functional validation [65,66]. The TCC-associated SNP S01_25897096 on chromosome 1 is located near the putative gene *Manes.01G061900*, annotated as a Bromo-Adjacent Homology (BAH) domain-containing protein. BAH domain-containing proteins are associated with chromatin recognition and epigenetic regulation of gene expression, making *Manes.01G061900* a positional candidate for TCC variation. However, a direct role of this gene in regulating carotenoid biosynthesis or degradation has not been established, and its potential contribution to carotenoid accumulation requires experimental validation [67]. The DArTag marker CassV7_chr01_27813884, which was associated with both TCC and BCC, was located within the putative MePSY1 gene (*Manes.01G124200*) on chromosome 1 at approximately 27.81 Mb on the v7.1 reference assembly. The MePSY1 gene encodes phytoene synthase (PSY), a key enzyme in the carotenoid biosynthetic pathway. PSY catalyzes the first committed step of carotenoid biosynthesis by converting geranylgeranyl diphosphate to phytoene and a naturally occurring functional variant in PSY has been shown to increase enzymatic activity and carotenoid accumulation in yellow-fleshed cassava roots [68,69]. The detection of a marker associated with both TCC and BCC within PSY1 therefore identifies this gene as a biologically plausible candidate contributing to variation in these carotenoid-related traits.

Previous GWAS based on the Cassava v6.1 reference assembly identified a major region on chromosome 1 around 24.1 Mb that was associated with variation in both carotenoid and DMC [12]. In the present study, the DMC-associated SNP S01_32444435 was located at 32.44 Mb on the v8.1 assembly, while the TCC-associated DArTag marker CassV7_chr01_27491928 corresponded to 31.64 Mb after coordinate conversion to v8.1. These two associations are separated by approximately 0.81 Mb (806 kb), placing them within a relatively close genomic interval on chromosome 1. Their physical proximity raises the possibility that the DMC and TCC associations may reflect linkage disequilibrium among neighboring loci rather than necessarily resulting from pleiotropic effects of a single gene [12]. Among the DArTag markers, CassV7_chr07_29527635, associated with WAB_20, was located within the putative gene *Manes.07G105500*, annotated as a GTPase-activating protein GYP2. GYP2 proteins have been reported to participate in vesicle trafficking and intracellular transport processes, which may influence the distribution of starches, proteins, and other storage compounds within plant cells. Given the potential role of these processes in cellular turgor, tissue texture, and mealiness, *Manes.07G105500* represents a positional candidate gene for variation in water absorption during boiling. However, the marker–trait association alone does not establish a causal relationship, and the potential contribution of this gene to cooking quality traits requires further functional validation [70]. The comparatively weak genetic signal observed for water absorption traits in this study—reflected in fewer significant loci and the complete absence of significant associations for WAB_30min aligns with previous reports describing water absorption and cassava cooking quality as traits strongly influenced by environmental and genotype × location effects [22]. Similar patterns have been reported in other root and tuber crops, where cooking-related textural traits such as mealiness are governed by complex, multigene pathways involving cell wall and pectin metabolism, rather than single major-effect loci [31–33].

Although some significant SNPs appeared isolated in the Manhattan plots, local LD analysis showed that the major TCC, BCC, and HCN associations were supported by neighboring markers. This was particularly evident for the higher density GBS loci, whereas the lower marker density of the DArTag platform provided fewer markers for characterizing local LD. These findings suggest that the apparent isolation of some associations in the Manhattan plots does not necessarily indicate unreliable signals [35]. Cross-platform concordance varied among traits: while the TCC and BCC associations on chromosome 1 and the HCN association on chromosome 16 were consistent across both platforms, major associations for DMC and WAB_20min mapped to different chromosomes. These discrepancies are likely driven by differences in marker density and genomic distribution between the platforms, as well as distinct local LD patterns and the underlying genetic architecture of traits governed by multiple loci.

Collectively, while these loci represent genuinely significant genomic regions controlling cassava root quality traits, high PVE values in markers with rare or underrepresented allelic classes should be interpreted with caution. Validation in independent, larger biparental or diversity populations will be essential to refine these variance components.

CONCLUSIONS

This study provides key insights into the genetic architecture of five cassava root quality traits. In addition to confirming previously reported chromosomal regions associated with these traits, the study identified novel association signals, thereby expanding current knowledge and offering new perspectives on the underlying genetic basis of cassava root quality. Collectively, these findings contribute to a more comprehensive understanding of cassava trait inheritance and provide a valuable foundation for future genetic improvement efforts in cassava breeding.

SUPPLEMENTARY MATERIALS

The phenotypic data (PPD BLUPs) and genotypic datasets (DArTag and GBS) used to run the GWAS have been uploaded as supplementary files. The following supplementary materials are available online, Sheet S1: Geno_GBS_markers, Sheet S2: Geno_DArTag_markers, Sheet S3: Pheno_BLUPs_GBS_all traits, Sheet S4: Pheno_BLUPs_DArTAG_all traits, Figure S1: Violin plots showing the distribution of root quality traits across allelic classes for the remaining significant SNP markers identified using (A) GBS and (B) DArTag markers. Each panel represents a significant marker locus, with violin plots illustrating the distribution of phenotypic values and embedded boxplots indicating the median and interquartile range. DMC, dry matter content; HCN, hydrogen cyanide content; TCC, total carotenoid content; BCC, beta-carotene content; and WAB_20min, water absorption after 20 min. Statistical significance is indicated as ns ($p \geq 0.05$), * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$), and **** ($p < 0.0001$).

DATA AVAILABILITY

Field trial details are uploaded at www.cassavabase.org under trial names 202379PDF1C_ciat, 202440PDEPR_ciat, and 202517PDEPR_ciat for seasons 1, 2, and 3 respectively.

AUTHOR CONTRIBUTIONS

KOD, BO, and WG conceptualized the study. JLLM, JLMA, MAO, and LL collected and organized all data. SMS and NM designed and established experimental field. NM developed plant populations. KOD analysed the data and wrote the paper with inputs from all authors.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

FUNDING

The authors declare financial support was received for the research and/or publication of this article. This research was fully funded by the Gates Foundation at CIAT, under the Project Number: INV-054653. The PhD studies of the lead author in Plant breeding at the Pan African University, Life and Earth Sciences Institute (Including Health and Agriculture), Ibadan, Nigeria was sponsored by the African Union Commission (AUC) scholarship.

ACKNOWLEDGMENTS

The authors are grateful to the staff of CIAT cassava program for their support and effort.

REFERENCES

1. Beyene G, Solomon FR, Chauhan RD, Gaitán-Solis E, Narayanan N, Gehan J, et al. Provitamin A biofortification of cassava enhances shelf life but reduces dry matter content of storage roots due to altered carbon partitioning into starch. Plant Biotechnol J. 2018;16(6):1186-200.
2. Olayide P, Large A, Stridh L, Rabbi I, Baldermann S, Stavolone L, et al. Gene expression and metabolite profiling of thirteen Nigerian cassava landraces to elucidate starch and carotenoid composition. Agronomy. 2020;10(3):424.
3. Latif S, Müller J. Potential of cassava leaves in human nutrition: a review. Trends Food Sci Technol. 2015;44(2):147-58.
4. Otálora A, Garces-Villegas V, Chamorro A, Palencia M, Combatt EM, Mendoza JS. 'Cassava, manioc or yuca' (Manihot esculenta): an overview about its crop, economic aspects and nutritional relevance. J Sci Technol Appl. 2022;16:1-0.
5. Sanchez T, Ceballos H, Dufour D, Ortiz D, Morante N, Calle F, et al. Prediction of carotenoids, cyanide and dry matter contents in fresh cassava root using NIRS and Hunter color techniques. Food Chem. 2014;151:444-51.
6. Alamu EO, Dixon GA, Adesokan M, Maziya-Dixon B. Correlate the cyanogenic potential and dry matter content of cassava roots and leaves grown in different environments. Sci Rep. 2023;13(1):15382.
7. Udemba IO, Olasanmi B, Iluebbey P. Tandem effect of bio-fortification on dry matter content and gari yield of some yellow root cassava (*Manihot esculenta* Crantz) varieties. Asian J Adv Agric Res. 2023;23(2):11-9.
8. Abdullah MI, Wan Mohamed WZ. The effect of lotus root starch content on the water barrier properties and biodegradability of cassava bioplastic films. Mater Sci Forum. 2024;1114:89-95.

9. Beyene G, Chauhan RD, Gehan J, Siritunga D, Taylor N. Cassava shrunken-2 homolog MeAPL3 determines storage root starch and dry matter content and modulates storage root postharvest physiological deterioration. Plant Mol Biol. 2022;109(3):283-99.
10. Sánchez T, Chavez AL, Ceballos H, Rodriguez-Amaya DB, Nestel P, Ishitani M. Reduction or delay of post-harvest physiological deterioration in cassava roots with higher carotenoid content. J Sci Food Agric. 2006;86(4):634-9.
11. Ceballos H, Luna J, Escobar AF, Ortiz D, Pérez JC, Sanchez T, et al. Spatial distribution of dry matter in yellow fleshed cassava roots and its influence on carotenoid retention upon boiling. Food Res Int. 2012;45(1):52-9.
12. Rabbi IY, Udoh LI, Wolfe M, Parkes EY, Gedil MA, Dixon A, et al. Genome-wide association mapping of correlated traits in cassava: dry matter and total carotenoid content. Plant Genome. 2017;10(3): plantgenome2016.09.0094.
13. Parkes E, Aina O, Kingsley A, Iluebbey P, Bakare M, Agbona A, et al. Combining ability and genetic components of yield characteristics, dry matter content, and total carotenoids in provitamin A cassava F1 cross-progeny. Agronomy. 2020;10(12):1850.
14. Ceballos H, Morante N, Sánchez T, Ortiz D, Aragon I, Chavez AL, et al. Rapid cycling recurrent selection for increased carotenoids content in cassava roots. Crop Sci. 2013;53(6):2342-51.
15. Edoh NL, Ukpabi UJ, Igoli JO, Evans BS. Biochemical characteristics of fresh roots of transgenic high beta-carotene cassava. Niger Agric J. 2018;49(2):131-5.
16. Meléndez-Martínez AJ. An overview of carotenoids, apocarotenoids, and vitamin A in agro-food, nutrition, health, and disease. Mol Nutr Food Res. 2019;63(15):1801045.
17. Villwock SS, Parkes EY, Mbanjo EG, Rabbi IY, Jannink JL. Mediated pleiotropy drives the negative correlation of total carotenoid and dry matter contents in cassava (Manihot esculenta). Genetics. 2025;231(2):iyaf162.
18. Patra A, Prasath VA. Isolation of detoxified cassava (Manihot esculenta L.) leaf protein by alkaline extraction-isoelectric precipitation: optimization and its characterization. Food Chem. 2024;437:137845.
19. Umuhozariho MG, Shayo NB, Msuya JM, Sallah PY. Cyanide and selected nutrients content of different preparations of leaves from three cassava species. Afr J Food Sci. 2014;8(3):122-9.
20. Thaweewong P, Anuntagool J. Change in free cyanide content of bitter cassava during incubation and drying and physical properties of dry-milled cassava flour. Food Bioprod Process. 2023;138:139-49.
21. Lambri M, Fumi MD, Roda A, De Faveri DM. Improved processing methods to reduce the total cyanide content of cassava roots from Burundi. Afr J Biotechnol. 2013;12(19):2685.
22. Tran T, Zhang X, Ceballos H, Moreno JL, Luna J, Escobar A, et al. Correlation of cooking time with water absorption and changes in relative density during boiling of cassava roots. Int J Food Sci Technol. 2021;56(3):1193-205.

23. Kouadio OK, Nindjin C, Bonfoh B, N'dri D, Amani GN. Water absorption as an evaluation method of cooking quality for yam (Dioscorea alata) and cassava (*Manihot esculenta* Crantz). Procedia Food Sci. 2011;1:153-9.
24. Iragaba P, Adinsi L, Delgado LF, Nanyonjo AR, Nuwamanya E, Wembabazi E, et al. Definition of sensory and instrumental thresholds of acceptability for selection of cassava genotypes with improved boiling properties. J Sci Food Agric. 2024;104(8):4561-72.
25. Uchendu K, Njoku DN, Paterne A, Rabbi IY, Dzidzienyo D, Tongoona P, et al. Genome-wide association study of root mealiness and other texture-associated traits in cassava. Front Plant Sci. 2021;12:770434.
26. Ogbonna AC, Braatz de Andrade LR, Rabbi IY, Mueller LA, Jorge de Oliveira E, Bauchet GJ. Large-scale genome-wide association study, using historical data, identifies conserved genetic architecture of cyanogenic glucoside content in cassava (*Manihot esculenta* Crantz) root. Plant J. 2021;105(3):754-70.
27. Balyejusa Kizito E, Rönnberg-Wästljung AC, Egwang T, Gullberg U, Fregene M, Westerbergh A. Quantitative trait loci controlling cyanogenic glucoside and dry matter content in cassava (*Manihot esculenta* Crantz) roots. Hereditas. 2007;144(4):129-36.
28. Esuma W, Ozimati A, Kulakow P, Gore MA, Wolfe MD, Nuwamanya E, et al. Effectiveness of genomic selection for improving provitamin A carotenoid content and associated traits in cassava. G3. 2021;11(9):jkab160.
29. Ikeogu UN, Akdemir D, Wolfe MD, Okeke UG, Chinedozi A, Jannink JL, et al. Genetic correlation, genome-wide association and genomic prediction of portable NIRS predicted carotenoids in cassava roots. Front Plant Sci. 2019;10:1570.
30. Solarte Certuche DC, Mamedio de Freitas G, Jannink JL, Garcia Morales CF, Sousa Cerqueira T, Santos de Santana B, et al. Uncovering genomic regions controlling root quality traits in cassava (*Manihot esculenta* Crantz) using different GWAS models. bioRxiv. [Preprint]. 2026 June 15. doi: 10.64898/2026.06.11.731598
31. Parker CC, Parker ML, Smith AC, Waldron KW. Pectin distribution at the surface of potato parenchyma cells in relation to cell-cell adhesion. J Agric Food Chem. 2001;49(9):4364-71.
32. Jarvis MC, Briggs SP, Knox JP. Intercellular adhesion and cell separation in plants. Plant Cell Environ. 2003;26(7):977-89.
33. Ross HA, Morris WL, Ducreux LJ, Hancock RD, Verrall SR, Morris JA, et al. Pectin engineering to modify product quality in potato. Plant Biotechnol J. 2011;9(8):848-56.
34. Korte A, Farlow A. The advantages and limitations of trait analysis with GWAS: a review. Plant Methods. 2013;9(1):29.
35. Ige AD, Olasanmi B, Bauchet GJ, Kayondo IS, Mbanjo EG, Uwugiaren R, et al. Validation of KASP-SNP markers in cassava germplasm for marker-assisted selection of increased carotenoid content and dry matter content. Front Plant Sci. 2022;13:1016170.

36. Udoh LI, Gedil M, Parkes EY, Kulakow P, Adesoye A, Nwuba C, et al. Candidate gene sequencing and validation of SNP markers linked to carotenoid content in cassava (*Manihot esculenta* Crantz). Mol Breed. 2017;37(10):123.
37. Belalcazar J, Dufour D, Andersson MS, Pizarro M, Luna J, Londoño L, et al. High-throughput phenotyping and improvements in breeding cassava for increased carotenoids in the roots. Crop Sci. 2016;56(6):2916-25.
38. Essers SA, Bosveld M, Van Grift RM, Voragen AG. Studies on the quantification of specific cyanogens in cassava products and introduction of a new chromogen. J Sci Food Agric. 1993;63(3):287-96.
39. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015;67:1-48.
40. R Core Team. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing; 2020.
41. Dankwa KO, Melendez JL, Giraldo JC, Sánchez-Sarria CE, Selvaraj M, Olasanmi B, et al. Genome-wide association study of post-harvest physiological deterioration in cassava (*Manihot esculenta* Crantz) using visual and AI-powered phenotyping. Front Plant Sci. 2026;17:1807180.
42. Wang J, Zhang Z. GAPIT version 3: boosting power and accuracy for genomic association and prediction. Genomics Proteomics Bioinformatics. 2021;19(4):629-40.
43. Liu X, Huang M, Fan B, Buckler ES, Zhang Z. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. PLoS Genet. 2016;12(2):e1005767.
44. Huang M, Liu X, Zhou Y, Summers RM, Zhang Z. BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. Gigascience. 2019;8(2):giy154.
45. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559-75.
46. Fei S, Mahama AA, Singh AK, Singh A. Cassava breeding. In: Suza WP, Lamkey KR, editors. Crop improvement. Ames (IA, US): Iowa State University Digital Press; 2023.
47. FAO (Food and Agriculture Organization of the United Nations). Cassava varieties with high dry matter and starch content. STI Portal 2025. Available from: https://sti-portal.fao.org/innovations/cassava-varieties-high-dry-matter-and-starch-content. Accessed on 2025 Jul 10.
48. Dankwa KO, Peprah BB. Industrialization of cassava sector in Ghana: progress and the role of developing high starch cassava varieties. Ghana J Agric Sci. 2019;54(2):79-85.
49. Alliance of Bioversity International and CIAT. Tools for measuring the dry matter of fresh cassava roots. 2026. Available from: https://alliancebioversityciat.org/tools-innovations/tools-measuring-dry-matter-fresh-cassava-roots. Accessed on 2026 Jul 10.

50. Enesi RO, Hauser S, Pypers P, Kreye C, Tariku M, Six J. Understanding changes in cassava root dry matter yield by different planting dates, crop ages at harvest, fertilizer application and varieties. Eur J Agron. 2022;133:126448.
51. Ssemakula G, Dixon A. Genotype X environment interaction, stability and agronomic performance of carotenoid-rich cassava clones. Sci Res Essays. 2007;2(9):390-9.
52. Xu S. Theoretical basis of the Beavis effect. Genetics. 2003;165(4):2259-68.
53. Garner C. Upward bias in odds ratio estimates from genome-wide association studies. Genet Epidemiol. 2007;31(3):288-95.
54. Xie W, Xu S. Estimation of genetic variance contributed by a quantitative trait locus: correcting the bias associated with significance tests. G3. 2021;11(11):jkab290.
55. King EG, Long AD. The Beavis effect in next-generation mapping panels in Drosophila melanogaster. G3. 2017;7(6):1975-84.
56. Lenz PRN, Beaulieu J, Mansfield SD, Clément S, Desponts M, Bousquet J. Factors affecting the accuracy of genomic selection for growth and wood quality traits in an advanced-breeding population of black spruce (Picea mariana). Tree Genet Genomes. 2017;13:80.
57. Yang J, Weedon MN, Purcell S, Lettre G, Estrada K, Willer CJ, et al. Genomic inflation factors under polygenic inheritance. Eur J Hum Genet. 2011;19(7):807-12.
58. Esuma W, Rubaihayo P, Pariyo A, Kawuki R, Wanjala B, Nzuki I, et al. Genetic diversity of provitamin A cassava in Uganda. J Plant Stud. 2012;1(1):60.
59. Njoku DN, Gracen VE, Offei SK, Asante IK, Egesi CN, Kulakow P, et al. Parent-offspring regression analysis for total carotenoids and some agronomic traits in cassava. Euphytica. 2015;206(3):657-66.
60. Ceballos H, Davrieux F, Talsma EF, Belalcazar J, Chavarriaga P, Andersson MS. Carotenoids in cassava roots. Carotenoids. 2017;14:189-221.
61. Drapal M, Ovalle Rivera TM, Becerra Lopez-Lavalle LA, Fraser PD. Carotenoid composition and sequestration in cassava (Manihot esculentum Crantz) roots. PLoS ONE. 2024;19(11):e0312517.
62. Adetoro NA, Ogunbayo SA, Akinwale MO. Evaluation of agronomic performance of beta-carotene rich (yellow fleshed) cassava varieties in Nigeria. J Plant Breed Crop Sci. 2018;10(10):273-80.
63. Peprah BB. Genetic improvement of beta carotene in cassava (*Manihot esculenta* Crantz) landraces [dissertation]. Bloemfontein (South Africa): University of the Free State; 2020.
64. Kanaabi M, Mukasa SB, Nuwamanya E, Iragaba P, Baguma JK, Nanyonjo AR, et al. Validation of KASP markers associated with hydrogen cyanide in fresh cassava roots in Uganda cassava germplasm. Agronomy. 2024;14(12):2765.
65. Voeltz GK, Prinz WA, Shibata Y, Rist JM, Rapoport TA. A class of membrane proteins shaping the tubular endoplasmic reticulum. Cell. 2006;124(3):573-86.
66. Jacquemyn J, Cascalho A, Goodchild RE. The ins and outs of endoplasmic reticulum-controlled lipid biosynthesis. EMBO Rep. 2017;18(11):1905-21.

67. Wang JY, Lin PY, Al-Babili S. On the biosynthesis and evolution of apocarotenoid plant growth regulators. Semin Cell Dev Biol. 2021;109:3-11.
68. Welsch R, Arango J, Bär C, Salazar B, Al-Babili S, Beltran J, et al. Provitamin A accumulation in cassava (Manihot esculenta) roots driven by a single nucleotide polymorphism in a phytoene synthase gene. Plant Cell. 2010;22(10):3348-56.
69. Esuma W, Herselman L, Labuschagne MT, Ramu P, Lu F, Baguma Y, et al. Genome-wide association mapping of provitamin A carotenoid content in cassava. Euphytica. 2016;212(1):97-110.
70. Smith GR, Givan SA, Cullen P, Sprague Jr GF. GTPase-activating proteins for Cdc42. Eukaryot Cell. 2002;1(3):469-80.

How to cite this article:
Dankwa KO, Luna Melendez JL, Moreno Alzate JL, Ospina MA, Londono L, Salazar SM, et al. Genome-Wide Association Study of Root Quality Traits in Cassava (*Manihot esculenta* Crantz). Crop Breed Genet Genom. 2026;8(4):e260025. https://doi.org/10.20900/cbgg20260025.