

Genome-wide association study identifies novel genomic regions associated with yield-related traits in cassava (*Manihot esculenta* Crantz)

Bright Gyamfi Adu

`adubright1992@g.ecc.u-tokyo.ac.jp`

The University of Tokyo

Richard Akromah

Kwame Nkrumah University of Science and Technology

Stephen Amoah

Kwame Nkrumah University of Science and Technology

Andrzej Kilian

Diversity Array Technology

Richard Adu Amoah

Council for Scientific and Industrial Research-Plant Genetics Resources Research Institute

Joseph Adjebeng-Danquah

Council for Scientific and Industrial Research -Savanna Agricultural Research Institute

Alex Yeboah

Council for Scientific and Industrial Research -Savanna Agricultural Research Institute

Daniel Nyadanu

Cocoa Research Institute of Ghana

Fakhrul Islam Monshi

The University of Tokyo

Toru Fujiwara

The University of Tokyo

Research Article

Keywords: *Manihot esculenta* Crantz, single-nucleotide polymorphism (SNP), population structure, genome-wide association analysis (GWAS), haplotypes

Posted Date: December 1st, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-8007567/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Cassava (*Manihot esculenta* Crantz) is a vital calorie source in the tropics due to its adaptation to marginal agroecological conditions. Cassava's dual role as a food security and industrial crop has stimulated extensive research into increasing yield. Genetic gain can be accelerated through investment in marker discovery and identification of genetic loci controlling important traits for use in breeding programs. Using the Diversity Array Technology genotyping-by-sequencing (DArTseq) platform, a population of 94 cassava accessions, comprising local landraces from Ghana and exotic lines from the International Institute of Tropical Agriculture (IITA), was genotyped with more than 30,000 SNP markers. A genome-wide association study (GWAS) was carried out for five yield-related traits, namely number of storage roots (NSR), mean storage root weight (MRW), root weight per plant (RW), harvest index (HI), and dry matter content (DMC) to survey the genome for the putative loci associated with these traits. A total of 55 significant marker-trait associations were detected, and haplotype analysis showed that favorable alleles at each locus had stronger genetic effects on yield-related traits, leading to the prediction of candidate genes. The identification of *Manes.07G112000* (TPL-binding domain protein) and *Manes.02G156800* (F-box domain) as candidate genes associated with MRW and DMC, respectively, highlights their potential roles in jasmonate signaling pathways. This connection suggests the existence of a defense-growth trade-off influencing yield traits in cassava. These findings lay a valuable foundation for the development of molecular markers to assist in breeding programs aimed at enhancing yield potential and overall productivity in cassava cultivars.

Introduction

With the global population continuously increasing and climate changing rapidly, food insecurity poses a major challenge worldwide. Cassava (*Manihot esculenta* Crantz) is a starchy root crop that is mostly vegetatively propagated and serves as a staple food for over 800 million people across tropical regions, particularly in sub-Saharan Africa, Southeast Asia, and Latin America. It is also used as an industrial raw material for starch and alcohol production (Jarvis et al. 2012; Parmar et al. 2017). Global cassava production has nearly doubled over the past two decades, from 175.8 million tons in 2000 to 333.8 million tons in 2023 partly due to the introduction of new improved varieties with superior adaptive features to thrive in marginal environments, tolerance to drought and ability to thrive in poor soils which makes it an important food security crop (Jarvis et al. 2012; FAO, 2025). In addition, its ability to remain in the ground post-harvest makes cassava uniquely suited to smallholder farming systems, where it contributes significantly to both food and income security (Burns et al. 2010).

Over the past four decades, cassava breeding programs in Africa, Asia, and Latin America have developed improved varieties with resistance to biotic and abiotic stresses, as well as higher yield and starch content (Kawano, 2003; Okechukwu and Dixon, 2008). Although phenotype-based recurrent selection has achieved some success, the rate of genetic gain remains low due to several biological constraints, including asynchronous flowering, low seed set per cross, long growth cycles of 12–24 months, and low multiplication rates of planting materials (Ceballos et al. 2012). These challenges limit the breeding program's ability to respond rapidly to changing human needs under volatile climatic and environmental

conditions. The adoption of modern breeding tools such as marker-assisted recurrent selection and genomic selection has helped accelerate genetic gains by shortening selection cycles and increasing selection intensity, particularly in the early stages of cassava breeding (Ferguson et al. 2012; Ceballos et al. 2015; García-Ruiz et al. 2016; Bredeson et al. 2016). However, integration of molecular markers into cassava breeding pipelines requires investment in discovery research to identify major-effect loci as targets of selection.

Advances in next-generation sequencing technologies and the completion of the cassava genome draft (Bredeson et al. 2016) have made it cost-effective to generate genome-wide marker data through biparental quantitative trait loci (QTL) analysis or GWAS in large natural populations (Ferguson et al. 2012). Combined with phenotypic data, this enables the identification and mapping of agriculturally important genes and QTLs at the whole-genome level (Varshney et al. 2014). GWAS has been successfully applied in cassava to map loci associated with biotic stress resistance, nutritional traits, and morphological characteristics. Previous studies identified genomic regions linked to cassava mosaic disease (CMD) resistance (Rabbi et al. 2014; Wolfe et al. 2016), cassava green mite (CGM) resistance (Ezenwaka et al. 2018), and root quality traits such as dry matter content and carotenoid concentration (Esuma et al. 2016; Ikeogu et al. 2019). These efforts have laid the foundation for integrating genomic tools into cassava improvement pipelines. For example, Rabbi et al. (2020) mapped genomic regions conferring resistance to CMD using African cassava germplasm through biparental QTL mapping and GWAS. Further analysis of SNPs (S12_7926132 and S12_7926163 on chromosome 12 linked to the major CMD2 locus, and S14_4626854 on chromosome 14) using Kompetitive Allele-Specific Polymerase Chain Reaction (KASP) assays showed that genotypes carrying at least one resistant allele at the CMD2 locus had significantly higher yields (Ige et al. 2021). Beyond disease resistance, genomic regions associated with root quality traits such as mealiness, fiber content, adhesiveness, taste, aroma, color, and firmness after boiling have been reported, revealing candidate genes linked to carbohydrate metabolism, cell adhesion, secondary cell wall formation, and proteolytic activity. For nutritional improvement, SNPs in *Manes.01G124200* (phytoene synthase) have been used to select cassava lines with higher carotenoid content.

While such studies have improved understanding of disease resistance and root quality traits, there is a pressing need to prioritize root and yield-related traits to guide breeding strategies for enhancing cassava productivity in Africa, where the crop forms a major part of daily calorie intake with limited export value (Otekunrin, 2024). Developing high-yielding cultivars is a central goal of most breeding programs; thus, elucidating the genetic architecture of yield is critical to improving this complex trait. Key yield components such as storage root weight, root number, root diameter, root length, and dry matter content directly influence total yield and are essential for developing farmer-preferred varieties with improved market value (Adjebeng-Danquah et al. 2020; Adu et al. 2020). These traits often display significant genotype-by-environment interactions and moderate heritability, which makes improvement through conventional phenotypic selection challenging (Sinclair, 1998; Adjebeng-Danquah et al. 2020). The integration of GWAS with high-throughput genotyping technologies can help dissect the genetic basis of these traits by exploiting natural variation to identify candidate genes and favorable alleles for marker-assisted breeding (MAS) (Elshire et al. 2011; Hamblin and Rabbi, 2014).

SNPs associated with number of storage roots, storage root weight, dry matter content, and starch content have been reported using 158 cassava accessions by Zhang et al. (2018). Similarly, Mbe et al. (2024) identified 52 SNPs associated with nitrogen-use efficiency and yield-related traits, including stay-green ability, chlorophyll content, and fresh and dry root yield. Continued discovery of yield-related genes and loci will facilitate molecular breeding for cassava yield improvement. Landraces, in particular, serve as valuable reservoirs of favorable alleles for specific traits and remain an important resource for modern breeding. By integrating high-density SNP genotyping data with multi-year phenotypic evaluations, the present study aims to identify genomic regions and SNPs associated with natural variation in local and exotic cassava accessions for five root- and yield-related traits, and to suggest candidate genes for molecular breeding. These findings will promote more efficient use of molecular markers in cassava improvement.

Materials and methods

Plant material and field trials

The panel for GWAS consisted of 94 cassava accessions with 26 breeding lines from IITA, 22 improved or released varieties from Ghana and 46 landraces were obtained through the Council for Scientific and Industrial Research (CSIR-Ghana). These accessions showed high variability in levels of branching of the main stem and color of root pulp (storage root flesh color) ranging from white to yellow (Supplementary Table S1). The field work study was carried out at the research field of the CSIR-Plant Genetic Resources Research Institute at Bunso in the Eastern Region of Ghana in 2019 and 2020 years. Bunso (lat. 06° 46' N, long. 01 01' W, 149 m above sea level) lies in the semi-deciduous forest zone of Ghana with the soil type Nta series (FAO: Gleyic Arensol) (Jones et al. 2013). The experiment was laid in a randomized complete block design with three replications. The land was ploughed and harrowed before planting. Cassava cuttings measuring about 30 cm were planted horizontally in each hole. Each accession was planted in a single row consisting of five stands with an intra and inter row spacing 1 m × 1 m to give a plot size of 5m². No fertilizer was applied but weeding was done twice before harvesting (12 MAP).

Trait measurement

The accessions were phenotyped for five yield-related traits including number of storage roots per plant (NSR), mean storage root weight (MRW), root weight per plant (RW), harvest index (HI) and root dry matter content (DMC) as indicated in the standard cassava descriptor (Fukuda et al. 2010). The three middle plants were harvested and measured for the selected traits after maturity (Supplementary Table S1). For DMC measurement, triplicate of 100 grams of root pulp (mid-section) were dried at 80 °C for 48 hours after which the dry weight taken and expressed as a percentage of the original fresh weight.

Genomic DNA extraction and DArTseq genotyping

Genomic DNA extraction and DArTseq genotyping

DNA samples were extracted from the youngest fully expanded leaves of each of the 94 cassava genotypes two weeks after planting using the DArT DNA extraction protocol (Kilian et al. 2012). The concentration of extracted DNA was checked using the Nanodrop 2000c spectrophotometer (NanoDrop Lite, LT2878, Thermo Scientific, USA). DNA samples were diluted between 50–100 ng/μl, packaged and shipped to Diversity Array Technology corporation (Canberra, Australia) for DArTSeq genotyping. Genotyping-by-Sequencing, and SNP calling were performed for each sample using the DArTseq genotyping platform (<https://www.diversityarrays.com/technology-and-resources/dartreseq/>). The sequences of the genomic representations were aligned to cassava reference genome v6.1, resulting in the selection of 31865 raw SNP markers, however, SNPs with over 5% missing data and/or minor allele frequency below 5% were removed. After filtering, a total of 24790 SNP markers were obtained and used for downstream analyses.

Population structure analysis

Population structure was estimated using the Bayesian model of the Markov chain Monte Carlo (MCMC) implemented in STRUCTURE v.2.3.4 (Pritchard et al. 2010) based on 24790 SNP points. For each run, the initial burn-in period was set to 20,000 followed by 30,000 MCMC (Markov chain Monte Carlo) replications, with no prior information on the origin of individuals. Five iterations were performed for each number of hypothetical populations (k) tested from 1 to 10. The STRUCTURE results for the assumed population (1–10) were subsequently analysed online using the STRUCTURESELECTOR (Li et al. 2018) to identify a distinct peak in the change of likelihood (ΔK) at the true value of K.

Marker–Trait Association Mapping

GWAS analysis was performed using a unified mixed-model approach implemented in the “*rrBLUP*” package in R version 4.5 (Endelman 2011). The model includes a random effect to account for population structure and relatedness, which is critical for reducing false positives, and a fixed effect for batches to ensure accurate detection of genetic associations by integrating data from two experimental datasets to mitigate batch effects (Kang et al. 2008). We established population parameters previously determined (P3D) as FALSE to allow the model to reestimate variance components for each marker, providing a more accurate estimation of marker-trait associations. Significant SNPs linked to traits were identified, with peaks surpassing a threshold of $-\log_{10}(\text{p-value}) \geq 4$. Manhattan plots for association mapping were visualized using the “*qqman*” package in R (Turner 2018).

Haplotype analysis and candidate gene identification

Haplotype and candidate gene analysis were performed using significant SNPs selected from the GWAS results. SNPs with the strongest association with the target signal were used to perform the haplotype analysis and the phenotypes of the accessions with different haplotypes were compared. For the candidate gene analysis, mapping of these selected SNP markers onto genes was done using the SNP location and gene description from the *M.esculenta*_305_v6.1.gene.gff3 file of the cassava reference

genome available in Phytozome_v14 (Goodstein et al. 2012), and the intersect function from bedtools (Quinlan and Hall 2010). Gene ontology annotation was carried out on the plant Ensembl website (<https://plants.ensembl.org/index.html>) and UniProtKB tool (<https://www.uniprot.org/>).

Statistical analysis

All phenotypic data were subjected to various statistical analyses using R Statistical Software v4.5. The differences in the different years for each trait were estimated from the significance of the mean square for years from the analysis of variance (ANOVA) while the effect of alleles at significant SNPs was assessed by comparing phenotypic data for the haplotype groups ($p < 0.05$). Using the “lme4” package (De Boeck et al. 2011), we considered the genotype as a random effect to obtain the variance components of all the traits while considering years as environment (Piepho 1998).

A linear mixed model was used to obtain the best linear unbiased predictions (BLUPs) for each genotype. The genetic source of phenotypic variance was indicated as broad sense heritability (H^2) according to Falconer and Mackay (1996).

$$H^2 = \frac{\sigma^2_g}{\sigma^2_g + \sigma^2_y + \sigma^2_{gy} + \sigma^2_e}$$

Where σ^2_g is genotype variance, σ^2_y is variance across years, σ^2_{gy} is the genotype by year interaction variance and σ^2_e is error variance.

Pearson’s correlation analysis between different traits was performed using the “corrplot” package (Wei et al. 2017) while data were visualized in R using the “ggplot2” package (Wickham 2016).

Results

Phenotypic variations and relationships among measured traits

Understanding the variation that exists in traits and the underlying structure of the population is necessary for GWAS and other trait-marker association studies. Here we characterize five yield-related traits in cassava; number of storage root per plant (NSR), mean storage root weight (MRW), storage root weight per plant (RW), harvest index (HI) and storage root dry matter content (DMC). Summary statistics showed significant variation in these traits in the association panel based on the average performance of each genotype, which is useful for deciphering their genetic architectures (Table 1 and Fig. 1A). Except for DMC, all measured exhibited large coefficient of variation (> 30) an indication of broad phenotypic variability within the association panel. The data showed significant genetic variance and environmental variance for all traits except NSR, while genotype by environment interaction was significant for RW (Table 1). Analysis of the phenotypic classes of the panel showed that all measured traits followed a normal distribution though DMC were slightly skewed towards the tail (Fig. 1A)

Broad sense (H^2) estimates ranged from 59.2% (RW) to 76.1% (DMC) suggesting considerable potential for improvement of the phenotypes across all traits through targeted selection (Table 1). Pearson's correlation (r^2) between traits estimates showed that all traits had positive correlations with RW, with that of NSR and MRW exceeding 0.60 ($p < 0.001$) (Fig. 1B). MRW was also significantly positively correlated with HI and DMC making it an important trait to improve field and economic yield of cassava.

Table 1

Summary statistics and heritability for yield-related traits assessed among the cassava genotypes

Traits	min	max	mean	std	CV	MSg	MSgy	σ^2g	σ^2p	H^2
NSR	1	13	5.52	1.99	35.94	7.30***	2.69ns	1.140	1.812	0.628
MRW (kg)	0.133	1.53	0.54	0.21	39.22	0.09***	0.03ns	0.015	0.023	0.656
RW (kg)	0.21	8.11	3.05	1.50	49.19	4.56***	1.86**	0.674	1.140	0.592
HI (%)	7.22	80.9	46.10	13.44	31.09	0.06***	0.02ns	0.010	0.014	0.714
DMC (%)	19.91	44.35	37.02	4.34	15.06	58.36***	12.94ns	11.102	14.589	0.761

NSR = number of storage roots, MRW = mean storage root weight, RW = root weight per plant, HI = harvest index, and DMC = storage root dry matter content min = minimum, max = maximum, Std = standard deviation, CV = coefficient of variation, MSg = mean square of genotypes, MSgy = mean square of genotype x year interaction, σ^2g = genotypic variance, σ^2p = phenotypic variance, H^2 = broad sense heritability.

Distribution of SNP markers and population structure analysis

A total of 31865 SNPs distributed across the 18 chromosomes were generated through DArT sequencing platform ranging from 1251 on chromosome12 to 3697 on chromosome 1 (Fig. 2A). Considering the exclusion of SNPs with less than 95% call rate and/or less than 5% minor allele frequency (MAF), 24790 SNP markers that passed the quality test were used to estimate the genetic structure of the cassava population using the Bayesian clustering model implemented in the computer software STRUCTURE and GWAS. Population structure analysis is required in genome-wide association to avoid false-positive associations. The population stratification inferred by assuming admixture model-based clustering method indicated the presence of four ($K = 4$) subgroups (Fig. 2B). Though 65% of the accessions showed admixture (with $\geq 1\%$ of ancestry from any of the subgroups) composition, assignment of accessions into subgroups ranged from 10 in subgroup2 (green) comprising of IITA lines to 36 in subgroup4 (orange) which were mainly landraces from Ghana. The expected heterozygosity was used to compute the diversity between individuals in each subgroup. Subgroup3 (blue-black) and subgroup4 (orange) showed a similar average genetic diversity (0.35) between individuals within each group defined by the expected

heterozygosity indicating the diverse nature of the groups (Supplementary Table S1). Again, the highest difference was observed between subgroups occurred between subgroup1 and subgroup2 (Supplementary Table S1) indicating large genetic base of the study population and present opportunity to detect both favorable and unfavorable alleles which enhances the power of GWAS to detect true associations.

Marker-trait association mapping and haplotype analysis

To identify genomic regions and potential SNP markers associated with variation in the five yield related traits in cassava, GWAS analysis for the traits was performed. To mitigate the batch effect that could arise from the two experimental sets, a unified mixed-model approach was employed for GWAS. The resulting Manhattan plot displayed multiple peaks across various chromosomes for all the traits except for RW (Fig. 3). A total of 55 significant SNPs associated with yield-related traits were detected at a threshold of $-\log(P) = 4$ (Supplementary Table S1) comprising of 20, 5, 1, 12, and 17 for NSR, MRW, RW, HI and DMC respectively.

Genome-wide association study for variation in NSR revealed only genomic regions on chromosome 5 (1.66–5.45 Mb) to be significantly associated with the trait (Fig. 3A, Table 2). A total of 20 markers were significantly associated with NSR on chromosome 5. First (SNP_2747056, $p = 8.73E-08$) and second (SNP_2749412, $2.31E-07$) peak SNPs were only 2.3 Kb apart and averagely explained about 21% (R^2) of the trait variation. We observed in both planting seasons that, accessions homozygote for the top SNP allele (SNP_2747056, TT) had higher number of storage roots per plant (> 9) followed by the heterozygote (GT) and the least in the lines homozygote for reference allele (GG) (Fig. 3F).

For MRW, five significant SNPs localized between 24.02 to 24.58 Mb on Chromosome 7 were identified to be significantly associated with the variation in the traits (Fig. 3B, Table 2). The topmost SNP (SNP_24033698, $p = 1.91E-05$) occurred in the same gene (*Manes.07G112000*) as the third one (SNP_24026637, $5.42E-05$). The mean storage root weight of the genotype's homozygote for the top SNP allele (SNP_24033698, CC) or the heterozygotes (TC) was $> 33\%$ higher than those harboring the two of the reference alleles (TT) (Fig. 3G).

A single SNP tagged as SNP_5559383 ($p = 1.91E-07$) around 5.55 Mb of chromosome 14 was found to be associated with RW (Fig. 3C, Table 2). The SNP explained 19% of the total phenotypic variance in the root weight among the lines. Contrast to NSR and MRW, individual's homozygote for the reference allele (SNP_5559383, CC) possessed higher root weight per plant (> 3.3 kg) relative to the alternative alleles (< 2.6 kg for CT or TT) (Fig. 3H).

Association analysis for harvest index identified significant SNPs from different genomic regions (Fig. 3D, Table 2). A total of 12 significant marker traits associations were detected for HI. The top three SNPs on chromosomes 12, 14 and 16 explained on average 16.7% of the total variation in the trait (Table 2). The HI of the accession's homozygote for the top SNP (SNP_21013081, CC) was 18% higher than those without (Fig. 3I).

GWAS for variation in dry matter content of the storage roots uncovered several regions scattered across 6 chromosomes. A total of 17 SNPs were significantly associated with the trait; however, the first two significant SNPs (SNP_11795901 and SNP_493861) occurred on chromosome 2 (Fig. 3E, Table 2). Cassava lines homozygote for the SNP allele (SNP_11795901, TT) was on average 13% higher in dry matter content than those with homozygote reference allele (GG) (Fig. 3J).

Table 2							
Summary statistics of top significant SNPs associated with yield-related traits							
Traits	SNP_ID	Chr	Position	Allele (Ref > SNP)	p-value	MAF	R ²
NSR	SNP_2747056	5	2747056	G > T	8.73E-08	0.08	0.21
NSR	SNP_2749412	5	2749412	C > T	2.31E-07	0.08	0.20
NSR	SNP_2373946	5	2373946	G > T	2.31E-07	0.08	0.18
NSR	SNP_2814227	5	2814227	G > A	3.07E-07	0.06	0.15
MRW	SNP_24033698	7	24033698	T > C	1.91E-05	0.53	0.19
MRW	SNP_24218174	7	24218174	A > T	3.21E-05	0.33	0.18
MRW	SNP_24026637	7	24026637	A > G	5.43E-05	0.28	0.17
MRW	SNP_24587896	7	24587896	A > G	5.61E-05	0.47	0.17
MRW	SNP_24069335	7	24069335	T > C	7.95E-05	0.28	0.14
RW	SNP_5559383	14	5559383	C > T	1.91E-07	0.23	0.19
HI	SNP_21013081	12	21013081	T > C	5.75E-06	0.69	0.18
HI	SNP_22383767	14	22383767	T > G	7.22E-06	0.06	0.16
HI	SNP_5919578	16	5919578	G > C	9.29E-06	0.05	0.14
DMC	SNP_11795901	2	11795901	G > T	2.23E-07	0.47	0.17
DMC	SNP_493861	2	493861	T > C	4.87E-07	0.11	0.14
DMC	SNP_3087896	12	3087896	A > G	9.39E-07	0.59	0.14
DMC	SNP_26380158	11	26380158	A > G	1.78E-06	0.17	0.13

Chr = chromosome number, MAF = minor allele frequency, R² = proportion of phenotypic variation explained by SNPs. The alleles in bold letters correspond to the favorable alleles.

Candidate gene identification

Genomic regions localized by the top significant SNPs were explored to identify putative candidate genes (Table 3) using the *M.esculenta*_305_v6.1.gene.gff3 file of the cassava reference genome available on phytozome.

For NSR, the top two SNPs (SNP_2747056 and SNP_2749412) occurred in the exons of the same gene (*Manes.05G037900*) annotated as malate dehydrogenase making it the potential candidate for NSR. Similarly, the first (SNP_24033698) and third (SNP_24026637) top SNPs associated with MRW co-localized in the 5' untranslated region (UTR) and second exon respectively of *Manes.07G112000* encoding the TPL-binding domain in jasmonate signaling (Table 3, Fig. 4A). Haplotypes generated from the combination of the alleles from two SNPs (SNP_24033698 and SNP_24026637) localized within *Manes.07G112000* revealed the haplotypes had significantly different MRW across the two separate sets of experiments. Cultivars carrying haplotype E and F (with one or two favorable alleles at each SNP point) showed significantly high MRW relative to others (Fig. 4). The candidate genes *Manes.14G068300* and *Manes.14G068200* encoding ubiquitin-conjugating enzyme E2 protein and protein kinase domain-containing protein (Pkinase), were found closest to the only significant SNP (SNP_5559383) identified for RW. The SNPs; SNP_21013081 localized in the exon of *Manes.12G101200* (expressed protein with a Myb domain) while SNP_5919578 occurred in the intron of *Manes.16G042900* (SF7 complex intermediate associated protein), hence, were suggested as candidate genes for HI (Table 3). For DMC, candidate genes for the top three SNPs are reported. SNP_11795901 occurred 9.3Kb upstream of *Manes.02G156800* encoding F-box domain protein, SNP_493861 localized in the exon of *Manes.02G004000* encoding SF2-N6-adenosine-methyltransferase while SNP_26380158 was the closest to *Manes.11G153600* annotated as PLATZ transcription factor family protein (Table 3).

(A) Exon structure of *Manes.07G112000* and DNA polymorphisms in the gene resulting in six different haplotypes. The two lines indicate the position of the two SNPs (SNP_24026637 and SNP_24033698) within the gene. The haplotypes indicate nucleotide or SNP combinations. Boxplots for mean storage root weight of the haplotypes for the (B) 2019 and (C) 2020 planting seasons are shown. In the boxplots, the central lines denote the average. Significant levels were determined using a least significant difference test and the different lowercase letters above the boxplots represent significant differences ($P \leq 0.05$).

Table 3
List of potential candidate genes identified in the vicinity of the GWAS hits for the traits

Traits	Chr	Position	Candidate gene	Localization of SNP	Annotation
NSR	5	2747056	<i>Manes.05G037900</i>	exon	Malate dehydrogenase
NSR	5	2749412	<i>Manes.05G037900</i>	exon	Malate dehydrogenase
NSR	5	2373946	<i>Manes.05G032800</i>	exon	SNF2 family N-terminal domain protein
NSR	5	2814227	<i>Manes.05G039500</i>	downstream (2.9Kb)	COP9 signalosome complex subunit
MRW	7	24033698	<i>Manes.07G112000</i>	exon	TPL-binding domain in jasmonate signalling
MRW	7	24218174	<i>Manes.07G113800</i>	upstream (6Kb)	SEL-1-like protein
MRW	7	24026637	<i>Manes.07G112000</i>	5'UTR	TPL-binding domain in jasmonate signalling
MRW	7	24069335	<i>Manes.07G112400</i>	exon	GTP-binding protein SEC4, Ras family GTP-binding proteins
RW	14	5559383	<i>Manes.14G068300</i>	upstream (0.1Kb)	Ubiquitin-conjugating enzyme e2
RW	14	5559383	<i>Manes.14G068200</i>	upstream (2.9Kb)	Protein kinase domain (Pkinase) // Leucine rich repeat (LRR_8)
RW	14	5559383	<i>Manes.14G067800</i>	downstream (29.8Kb)	Alpha, alpha-trehalose-phosphate synthase (UDP-forming)
HI	12	21013081	<i>Manes.12G101200</i>	exon	Uncharacterized Myb domain
HI	14	22383767	<i>Manes.14G166100</i>	upstream (18Kb)	ATP binding / protein kinase-related
HI	16	5919578	<i>Manes.16G042900</i>	intron	SF7 complex intermediate associated protein
DMC	2	11795901	<i>Manes.02G156800</i>	upstream (9.3Kb)	F-box domain protein
DMC	2	493861	<i>Manes.02G004000</i>	exon	SF2 - N6-adenosine-methyltransferase
DMC	12	3087896	<i>Manes.12G037300</i>	downstream (0.8Kb)	Thioredoxin
DMC	11	26380158	<i>Manes.11G153600</i>	downstream (10.7Kb)	PLATZ transcription factor family protein

Chr = chromosome number, UTR = untranslated region.

Discussion

Root yield and dry matter content of cassava are important traits in breeding for subsistence and commercial use; therefore, understanding their genetic architecture and underlying genomic regions influencing variations in these traits could accelerate genetic improvement. The population showed large phenotypic variation within all traits, and the magnitude of broad-sense heritability estimates for the traits were comparable to previously reported estimates value (Adjebeng-Danquah et al. 2020; Adu et al. 2020) indicating the presence of substantial genetic component of these traits to allow for selection. Root weight (RW), which defines actual fresh yield per plant is a component of number of storage roots (NSR) and the individual root weight (MRW). The significant positive correlation between MRW and both RW and DMC makes MRW a significant trait to simultaneously improve both RW and DMC hence selection for high MRW could potentially improve both traits (Tumuhimbise et al. 2015; Adjebeng-Danquah et al. 2016, Adu et al. 2020). Knowledge of the structure underlying the population for association analysis could eliminate the effect of false positives associations (Yu et al. 2006). Estimation of population structure revealed that most of the accessions were in admixture state with their genetic composition coming from the different subgroups (K = 4). More importantly, each subgroup consists of accessions from different sources/origins indicating broad genetic base of the study population and its suitability for association analysis. GWAS provides insights into the genetic basis for complex traits where it highlights signals of associations between SNPs and phenotypic traits in diverse population (Bangarwa et al. 2020). NSR and MRW which are directly related to root yield, were highly associated with SNPs on chr5 and chr7 respectively, similar to other reports (Zhang et al. 2018). The variation in the haplotype's performance and more importantly the localization of the top significant SNPs within the exonic regions of candidate genes could serve as important steps towards the development of functional markers (Uchendu et al. 2021). Though none of the SNP markers showed pleiotropic effect (a single SNP significantly influencing more than one trait), loci with favorable alleles influencing individual traits mark the beginning of marker discovery for the trait.

The most notable candidate genes identified in this study were found in the vicinity occupied by SNP_2747056 and SNP_2749412 on chromosome 5 for NSR as well as SNP_24033698 and SNP_24026637 on chromosome 7 for MRW.

SNP_2747056 and SNP_2749412 for NSR co-located with *Manes.05G037900* a malate dehydrogenase (MDH) protein involved in the interconversion of oxaloacetate and malate a critical step in tricarboxylic acid (TCA) cycle and photosynthesis directly implicating it in energy production and carbon flux distribution in plants (Scheibe, 2004; Tomaz et al. 2010; Baird et al. 2024; Martinez-Vaz et al. 2024). MDH enzymes are localized in chloroplasts, mitochondria, peroxisomes, and the cytosol of plants (Gietl 1992). In mitochondria, MDH enzymes are reported to participate in the tricarboxylic acid (Krebs) cycle (Gietl 1992), provision of NAD⁺ for glycine oxidation (Journet et al. 1981) and provision of CO₂ for carbon fixation in the bundle sheath cells of some C₄ plants (Hatch and Osmond 1976). Mutations in mitochondrial MDH affected photorespiration efficiency, leading to growth retardation and reduced ATP production in

Arabidopsis. Knocked out lines (*mmdh1* and *mmdh2*) in *Arabidopsis thaliana* were sensitive to MDH activity with impaired root growth, delayed development, and reduced respiration rates (Tomaz et al. 2010). In tomato, the expression of the antisense fragment of mMDH revealed low root dry weight and low respiration in the roots (van der Merwe et al. 2009). This suggests a distinct impact of MDH disruption on roots development (Tesfaye et al. 2001; Tomaz et al. 2010) which serves as storage organ of carbon in cassava. In addition, the MDH enzyme was recently reported to be associated with the LIKE SEX FOUR 1-malate dehydrogenase complex involved in starch degradation which included glucan phosphatase and β -amylase (Liu et al. 2024) suggesting sideline roles for MDH in starch metabolism that demand further analysis. These putative gene has not been well studied in cassava and further investigation will be needed to explore their potential roles in cassava in relation to root development and carbon partition.

For MRW, SNP_24026637 and SNP_24033698 on chromosome7 were localized in the 5'UTR and exon region respectively of *Manes.07G112000* gene encoding a TOPLESS (TPL) binding domain involved in jasmonate signaling. *Manes.07G112000* (TPL-binding domain protein) has a C-terminal zinc binding domain from the NINJA (Novel Interactor of JAZ) protein which interacts with the TIFY domain of JAZ1 (Jasmonate Zim-domain protein 1). TOPLESS (TPL) is a transcriptional co-repressor that plays a key role in jasmonate (JA) signaling by suppressing the expression of JA-responsive genes under non-stress conditions to promote growth where NINJA (Novel Interactor of JAZ) acts as a bridge, connecting JAZ to TPL to form a JAZ–NINJA–TPL Repressor Complex (Long et al. 2006; Pauwels et al. 2010; Acosta et al. 2013). Several evidence suggests JA-signaling regulates defense-growth trade-off (Noir et al. 2013, Attaran et al. 2014, Major et al. 2017; Guo et al. 2018) suggesting JA-signaling as a regulatory hub mediating metabolic reprogramming in response to changing environmental conditions. The role of JA-signaling in the regulation of carbohydrates were evident in poplar tree leaves (Babst et al. 2005), tobacco (Hanik et al. 2010), cabbage leaves (Tytgat et al. 2013) and tobacco plants (Wang et al. 2014) with low starch concentration as result of impaired JA signaling. In Arabidopsis, mutants with non-functional *TPL* showed stunted growth, short roots, and enhanced defense gene expression, due to uncontrolled JA signaling (Pauwels et al. 2010; Acosta et al. 2013). These independent findings suggest the role of the JAZ-NINJA-TPL complex and JA signaling mediated modulation of carbohydrate metabolism in the regulation of plant growth and induction of defense responses. Herein, cassava accession harboring the two SNPs (SNP_24033698 and SNP_24026637; haplotype E) localized in *Manes.07G112000* (TPL-binding domain) had bigger tubers (high MRW) compared to those with reference alleles or without the SNPs (Fig. 4) making it ideal for marker development. The potential role of *Manes.07G112000* (TPL-binding domain) in jasmonate signaling coupled with the reported effect of the hormone in the regulation of carbon in other crops through resource allocation (growth-defense) makes it a potential gene for further studies in cassava in relation to yield.

DMC and HI on the other hand were associated with SNPs on different chromosomes revealing the complexity of the traits (Rabbi et al. 2022). Chromosomes14 was associated with both RW and HI suggesting that these traits could be coinherited. Previous studies linked stem diameter and dry mass content to Chromosome14 (Zhang et al 2018). Candidates have been suggested for DMC, including *Manes.02G156800*, annotated as F-box domain (F-box) protein 9.3Kb upstream of SNP_11795901 and

Manes.02G004000 (N6-adenosine-methyltransferase) containing SNP_493861, both on chromosome 2. The gene *Manes.02G004000* which encodes N6-adenosine-methyltransferase harbored the SNP_493861 in the seventh exon was reported to be involved in mRNA modifications via methylation. Methylation of adenosine residues in mRNA is reported to influence embryonic development in *Arabidopsis* (Zhong et al. 2008; Bodi et al. 2012). Recently, *Arabidopsis thaliana* plants lacking mRNA adenosine methylase were reported to exhibit heightened sensitivity to drought implicating it in stress response (Ganguly et al. 2024). *Manes.02G156800* has F-box domain found in *Arabidopsis thaliana* protein as ATTENUATED FAR-RED RESPONSE (AT2G24540.1, AtAFR) also known as SKP1-interacting partner29 (AtSKIP29) a component of SCF (SKP1/ASK-cullin-F-box protein) E3 ubiquitin ligase complexes involved in the ubiquitination and subsequent proteasomal degradation of target proteins thereby regulating a wide range of physiological processes including hormone signaling, development, and stress responses (Guo et al. 2003; Thines et al. 2007). More importantly, the SCF complex through its interaction with *COI1* (CORONATINE INSENSITIVE1) targets JAZ repressors following jasmonate perception to regulate growth and defense trade-offs (Xu et al. 2002; Thines et al. 2007). However, these putative genes have not been well studied in cassava and would need further investigation in relation to dry matter accumulation in the roots of cassava.

These findings suggest some genetic basis of yield-related traits in cassava. Through GWAS, we identified significant SNPs for the studied traits and promising genes for the MRW and DMC to be associated with jasmonate signaling which could modulate carbohydrate profiles in plant growth-defense conflicts. These SNPs and genes could be potential targets for breeders in marker-assisted breeding to improve yield in cassava after further validation.

Conclusion

The SNPs and candidate genes identified for yield-related traits provide valuable insights into the genetic basis of cassava yield improvement and offer potential markers for MAS breeding of high-yielding cultivars. The studies identified significant SNPs associated with natural variation in traits in the cassava population. The haplotype analysis highlighted *Manes.05G037900* on chromosome 5, *Manes.07G112000* on chromosome 7 and *Manes.02G156800* on chromosome 2 as potential target for enhancing number of storage roots, mean storage root weight and dry matter content in cassava respectively through marker-assisted and genomic selection. Cassava is a relatively hardy and drought-tolerant crop, whether the disruption of *Manes.07G112000* reported to be involved in plant defense through jasmonate signaling, will enhance yield via a growth-defense tradeoff or through novel biological pathways remains to be exploited. Further validation of these associations and candidate genes will be essential to accelerate cassava genetic improvement.

Declarations

Author contributions

AGB, RA, SA conceived and designed the project, AGB, RA, SA, RAA, AY, and JAD collected the cassava landraces and contributed to the field evaluation of the accessions. AGB, RA, AK and DA contributed to the

data analysis, AGB wrote the draft manuscript and DA, RA, FIM and TF revised the paper. All authors read and approved of the final manuscript.

Funding

The authors gratefully acknowledge the Alliance for a Green Revolution in Africa, through the Improved Masters in Cultivar Development Programme (IMCDA; 2014 PASS-012), Faculty of Agriculture, Kwame Nkrumah University of Science and Technology, Ghana for providing funds for this study. AGB is supported by JSPS as a foreign post-doctoral fellow.

Data availability

The datasets supporting the conclusions of this article are provided within the article and its additional files.

Declarations Conflict of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Acosta IF, Gasperini D, Chételat A, Stolz S, Santuari L, Farmer EE (2013) Role of NINJA in root jasmonate signaling. *Proc Natl Acad Sci USA* 110(38). <https://doi.org/10.1073/pnas.1307910110>
2. Adjebeng-Danquah J, Gracen VE, Offei SK, Asante IK, Manu-Aduening J (2016) Agronomic performance and genotypic diversity for morphological traits among cassava genotypes in the Guinea Savannah Ecology of Ghana. *J Crop Sci Biotechnol* 19(1). <https://doi.org/10.1007/s12892-015-0095-8>
3. Adjebeng-Danquah J, Asante IK, Manu-Aduening J, Agyare RY, Gracen VE, Offei SK (2020) Genotypic variability in some morpho-physiological traits in different environments and their relationship with cassava (*Manihot esculenta* Crantz) root yield. *Int. J. Agron.* (2020), pp. 1–19, 10. <https://doi.org/1155/2020/5871351>
4. Adu BG, Yeboah A, Akromah R, Bobobee E, Amoah S, Kena AW, Amoah RA (2020) Whole genome SNPs and phenotypic characterization of cassava (*Manihot esculenta* Crantz) germplasm in the semi-deciduous forest ecology of Ghana. *Ecological Genetics and Genomics*, 17. <https://doi.org/10.1016/j.egg.2020.100068>
5. Adu SV, Asamoah RD (1992) Soils of ayensu/densu basin, nta series (fao: gleyic arenosol), Soil Research Institute Memoir. 9
6. Attaran E, Major IT, Cruz JA, Rosa BA, Koo AJK, Chen J, Kramer DM, He SY, Howe GA (2014) Temporal dynamics of growth and photosynthesis suppression in response to jasmonate signaling. *Plant Physiol* 165(3). <https://doi.org/10.1104/pp.114.239004>
7. Babst BA, Ferrieri RA, Gray DW, Lerdau M, Schlyer DJ, Schueller M, Thorpe MR, Orians CM (2005) Jasmonic acid induces rapid changes in carbon transport and partitioning in *Populus*. *New Phytol* 167(1). <https://doi.org/10.1111/j.1469-8137.2005.01388.x>

8. Baird LM, Berndsen CE, Monroe JD (2024) Malate dehydrogenase in plants: evolution, structure, and a myriad of functions. *Essays Biochem* 68(2):221–233. <https://doi.org/10.1042/EBC20230089>
9. Bangarwa SK, Kumawat R, Barupal Hlal, Kumar A, Yadav MK (2020) Association Mapping in Crops Plants. *Int J Curr Microbiol Appl Sci* 9(9). <https://doi.org/10.20546/ijcmas.2020.909.161>
10. Bodi Z, Zhong S, Mehra S, Song J, Graham N, Li H, May S, Fray RG (2012) Adenosine methylation in *Arabidopsis* mRNA is associated with the 3' end and reduced levels cause developmental defects. *Front Plant Sci* 3(MAR). <https://doi.org/10.3389/fpls.2012.00048>
11. Bredeson Jv, Lyons JB, Prochnik SE, Wu GA, Ha CM, Edsinger-Gonzales E, Grimwood J, Schmutz J, Rabbi IY, Egesi C, Nauluvula P, Lebot V, Ndunguru J, Mkamilo G, Bart RS, Setter TL, Gleadow RM, Kulakow P, Ferguson ME, Rokhsar DS (2016) Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nat Biotechnol* 34(5). <https://doi.org/10.1038/nbt.3535>
12. Burns A, Gleadow R, Cliff J, Zacarias A, Cavagnaro T (2010) Cassava: The drought, war and famine crop in a changing world. *Sustainability* 2(11). <https://doi.org/10.3390/su2113572>
13. Ceballos H, Ramirez J, Jarvis A, Alvarez E (2012) Adaptation of cassava to changing climates. *Food Secur* 3(1):61–71. <https://doi.org/10.1002/9780470960929.ch28>
14. Ceballos H, Kawuki RS, Gracen VE et al (2015) Conventional breeding, marker-assisted selection, genomic selection and inbreeding in clonally propagated crops: a case study for cassava. *Theor Appl Genet* 128:1647–1667. <https://doi.org/10.1007/s00122-015-2555-4>
15. de Boeck P, Bakker M, Zwitser R, Nivard M, Hofman A, Tuerlinckx F, Partchev I (2011) The estimation of item response models with the lmer function from the lme4 package in R. *J Stat Softw* 39(12). <https://doi.org/10.18637/jss.v039.i12>
16. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6(5). <https://doi.org/10.1371/journal.pone.0019379>
17. Endelman JB (2011) Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *Plant Genome* 4(3). <https://doi.org/10.3835/plantgenome2011.08.0024>
18. Esuma W, Kawuki RS, Herselman L, Labuschagne MT (2016) Diallel analysis of provitamin A carotenoid and dry matter content in cassava (*Manihot esculenta* Crantz). *Breed Sci* 66(4). <https://doi.org/10.1270/jsbbs.15159>
19. Ezenwaka L, Carpio DP, del, Jannink JL, Rabbi I, Danquah E, Asante I, Danquah A, Blay E, Egesi C (2018) Genome-wide association study of resistance to cassava green mite pest and related traits in cassava. *Crop Sci* 58(5). <https://doi.org/10.2135/cropsci2018.01.0024>
20. Falconer D, Mackay T (1996) Introduction to quantitative genetics, Longman. Essex, England :254–256
21. Ferguson M, Rabbi I, Kim DJ, Gedil M, Lopez-Lavalle LAB, Okogbenin E (2012) Molecular Markers and Their Application to Cassava Breeding: Past, Present and Future. In *Tropical Plant Biology* (Vol. 5, Issue 1). <https://doi.org/10.1007/s12042-011-9087-0>

22. Food and Agriculture Organization of the United Nations (2025) Our World in Data. Cassava production. Production: Crops and livestock products. Accessed on July 23, 2025 from <https://archive.ourworldindata.org/20250718-111234/grapher/cassava-production.html>
23. Fukuda WMG, Guevara CL, Kawuki R, Ferguson ME (2010) Selected Morphological and Agronomic Descriptors for the Characterization of Cassava. International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria, p 19
24. Ganguly DR, Li Y, Bhat SS, Tiwari S, Ng PJ, Gregory BD, Sunkar R (2025) mRNA ADENOSINE METHYLASE promotes drought tolerance through N6-methyladenosine-dependent and independent impacts on mRNA regulation in Arabidopsis. *New Phytol* 245:183–199. <https://doi.org/10.1111/nph.20227>
25. Gietl C (1992) Partitioning of malate dehydrogenase isoenzymes into glyoxysomes, mitochondria, and chloroplasts. *Plant Physiol* 100(2). <https://doi.org/10.1104/pp.100.2.557>
26. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res* 40(D1). <https://doi.org/10.1093/nar/gkr944>
27. Guo H, Ecker JR (2003) Plant responses to ethylene gas are mediated by SCFEBF1/EBF2- dependent proteolysis of EIN3 transcription factor. *Cell* 115(6). [https://doi.org/10.1016/S0092-8674\(03\)00969-3](https://doi.org/10.1016/S0092-8674(03)00969-3)
28. Guo Q, Major IT, Howe GA (2018) Resolution of growth–defense conflict: mechanistic insights from jasmonate signaling. *Current Opinion in Plant Biology*, 44. <https://doi.org/10.1016/j.pbi.2018.02.009>
29. Hatch MD, Osmond CB (1976) Compartmentation and Transport in C4 Photosynthesis. In *Transport in Plants III*. https://doi.org/10.1007/978-3-642-66417-5_5
30. Ige AD, Olasanmi B, Mbanjo EGN, Kayondo IS, Parkes EY, Kulakow P, Egesi C, Bauchet GJ, Ng E, Lopez-Lavalle LAB, Ceballos H, Rabbi IY (2021) Conversion and validation of uniplex snp markers for selection of resistance to cassava mosaic disease in cassava breeding programs. *Agronomy* 11(3). <https://doi.org/10.3390/agronomy11030420>
31. Ikeogu UN, Akdemir D, Wolfe MD, Okeke UG, Chinedozi A, Jannink JL, Egesi CN (2019) Genetic Correlation, Genome-Wide Association and Genomic Prediction of Portable NIRS Predicted Carotenoids in Cassava Roots. *Front Plant Sci* 10. <https://doi.org/10.3389/fpls.2019.01570>
32. Jarvis A, Ramirez-Villegas J, Campo H, B.V. et al (2012) Is Cassava the Answer to African Climate Change Adaptation? *Trop Plant Biol* 5:9–29. <https://doi.org/10.1007/s12042-012-9096-7>
33. Journet E-P, Neuburger M, Douce R (1981) Role of Glutamate-oxaloacetate Transaminase and Malate Dehydrogenase in the Regeneration of NAD + for Glycine Oxidation by Spinach leaf Mitochondria. *Plant Physiol* 67(3). <https://doi.org/10.1104/pp.67.3.467>
34. Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E (2008) Efficient control of population structure in model organism association mapping. *Genetics* 178(3). <https://doi.org/10.1534/genetics.107.080101>
35. Kawano K (2003) Thirty years of cassava breeding for productivity - Biological and social factors for success. *Crop Sci* 43(4). <https://doi.org/10.2135/cropsci2003.1325>

36. Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H, Caig V, Heller-Uszynska K, Jaccoud D, Hopper C, Aschenbrenner-Kilian M, Evers M, Peng K, Cayla C, Hok P, Uszynski G (2012) Diversity arrays technology: A generic genome profiling technology on open platforms. *Methods in Molecular Biology*, 888. https://doi.org/10.1007/978-1-61779-870-2_5
37. Li YL, Liu JX (2018) StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. *Mol Ecol Resour* 18(1). <https://doi.org/10.1111/1755-0998.12719>
38. Liu J, Wang X, Guan Z, Wu M, Wang X, Fan R, Zhang F, Yan J, Liu Y, Zhang D, Yin P, Yan J (2024) The LIKE SEX FOUR 1-malate dehydrogenase complex functions as a scaffold to recruit β -amylase to promote starch degradation. *Plant Cell* 36(1). <https://doi.org/10.1093/plcell/koad259>
39. Long JA, Ohno C, Smith ZR, Meyerowitz EM (2006) TOPLESS regulates apical embryonic fate in *Arabidopsis*. *Science* 312(5779). <https://doi.org/10.1126/science.1123841>
40. Major IT, Yoshida Y, Campos ML, Kapali G, Xin XF, Sugimoto K, de Oliveira Ferreira D, He SY, Howe GA (2017) Regulation of growth–defense balance by the JASMONATE ZIM-DOMAIN (JAZ)-MYC transcriptional module. *New Phytol* 215(4). <https://doi.org/10.1111/nph.14638>
41. Martinez-Vaz BM, Howard AL, Jamburuthugoda VK, Callahan KP (2024) Insights into the regulation of malate dehydrogenase: inhibitors, activators, and allosteric modulation by small molecules. *Essays Biochem* 68(2):173–181. <https://doi.org/10.1042/EBC20230087>
42. Mbe JO, Dzidzienyo DK, Abah SP, Njoku DN, Onyeka J, Tongoona P, Egesi C (2024) Novel SNP markers and other stress-related genomic regions associated with nitrogen use efficiency in cassava. *Front Plant Sci* 15:1376520. <https://doi.org/10.3389/fpls.2024.1376520>
43. Noir S, Bömer M, Takahashi N, Ishida T, Tsui TL, Balbi V, Shanahan H, Sugimoto K, Devoto A (2013) Jasmonate controls leaf growth by repressing cell proliferation and the onset of endoreduplication while maintaining a potential stand-by mode. *Plant Physiol* 161(4). <https://doi.org/10.1104/pp.113.214908>
44. Okechukwu RU, Dixon AGO (2008) Genetic gains from 30 years of cassava breeding in Nigeria for storage root yield and disease resistance in elite cassava genotypes. *J Crop Improv* 22(2). <https://doi.org/10.1080/15427520802212506>
45. Otekunrin OA (2024) Cassava (*Manihot esculenta* Crantz): a global scientific footprint-production, trade, and bibliometric insights. *Discov Agric* 2:94. <https://doi.org/10.1007/s44279-024-00121-3>
46. Parmar A, Sturm B, Hensel O (2017) Crops that feed the world: Production and improvement of cassava for food, feed, and industrial uses. In *Food Security* (Vol. 9, Issue 5). <https://doi.org/10.1007/s12571-017-0717-8>
47. Pauwels L, Barbero GF, Geerinck J, Tilleman S, Grunewald W, Pérez AC, Chico JM, Bossche Rvanden, Sewell J, Gil E, García-Casado G, Witters E, Inzé D, Long JA, de Jaeger G, Solano R, Goossens A (2010) NINJA connects the co-repressor TOPLESS to jasmonate signalling. *Nature* 464(7289). <https://doi.org/10.1038/nature08854>
48. Piepho HP (1998) Methods for comparing the yield stability of cropping systems - A review. *J Agron Crop Sci* (Vol 180. <https://doi.org/10.1111/j.1439-037X.1998.tb00526.x>

49. Pritchard JK (2007) Documentation for structure software: Version 2. 2. *Statistics*
50. Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6). <https://doi.org/10.1093/bioinformatics/btq033>
51. Rabbi IY, Hamblin MT, Kumar PL, Gedil MA, Ikpan AS, Jannink JL, Kulakow PA (2014) High-resolution mapping of resistance to cassava mosaic geminiviruses in cassava using genotyping-by-sequencing and its implications for breeding. *Virus Research*, 186. <https://doi.org/10.1016/j.virusres.2013.12.028>
52. Rabbi IY, Kayondo SI, Bauchet G, Yusuf M, Aghogho CI, Ogunpaimo K, Uwugiaren R, Smith IA, Peteti P, Agbona A, Parkes E, Lydia E, Wolfe M, Jannink JL, Egesi C, Kulakow P (2022) Genome-wide association analysis reveals new insights into the genetic architecture of defensive, agro-morphological and quality-related traits in cassava. *Plant Mol Biol* 109(3). <https://doi.org/10.1007/s11103-020-01038-3>
53. Scheibe R (2004) Malate valves to balance cellular energy supply. *Physiol Plant* 120(1). <https://doi.org/10.1111/j.0031-9317.2004.0222.x>
54. Sinclair TR (1998) Historical changes in harvest index and crop growth: a case study for cassava. *Field Crops Res* 57(3):213–221
55. Tesfaye M, Temple SJ, Allan DL, Vance CP, Samac DA (2001) Overexpression of malate dehydrogenase in transgenic alfalfa enhances organic acid synthesis and confers tolerance to aluminum. *Plant Physiol* 127(4). <https://doi.org/10.1104/pp.010376>
56. Thines B, Katsir L, Melotto M, Niu Y, Mandaokar A, Liu G, Nomura K, He SY, Howe GA, Browse J (2007) JAZ repressor proteins are targets of the SCFCOI1 complex during jasmonate signalling. *Nature* 448(7154). <https://doi.org/10.1038/nature05960>
57. Tomaz T, Bagard M, Pracharoenwattana I, Lindén P, Lee CP, Carroll AJ, Ströher E, Smith SM, Gardeström P, Millar AH (2010) Mitochondrial malate dehydrogenase lowers leaf respiration and alters photorespiration and plant growth in *Arabidopsis*. *Plant Physiol* 154(3). <https://doi.org/10.1104/pp.110.161612>
58. Tumuhimbise R, Shanahan P, Melis R, Kawuki R (2015) Genetic variation and association among factors influencing storage root bulking in cassava. *J Agric Sci* 153(7). <https://doi.org/10.1017/S0021859614000999>
59. Turner SD (2018) qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *J Open Source Softw* 3(25). <https://doi.org/10.21105/joss.00731>
60. Tytgat TOG, Verhoeven KJF, Jansen JJ, Raaijmakers CE, Bakx-Schotman T, McIntyre LM, van der Putten WH, Biere A, van Dam NM (2013) Plants Know Where It Hurts: Root and Shoot Jasmonic Acid Induction Elicit Differential Responses in *Brassica oleracea*. *PLoS ONE* 8(6). <https://doi.org/10.1371/journal.pone.0065502>
61. Uchendu K, Njoku DN, Paterne A, Rabbi IY, Dzidzienyo D, Tongoona P, Offei S, Egesi C (2021) Genome-Wide Association Study of Root Mealiness and Other Texture-Associated Traits in Cassava. *Frontiers in Plant Science*, 12. <https://doi.org/10.3389/fpls.2021.770434>
62. van der Merwe MJ, Osorio S, Moritz T, Nunes-Nesi A, Fernie AR (2009) Decreased mitochondrial activities of malate dehydrogenase and fumarase in tomato lead to altered root growth and

- architecture via diverse mechanisms. *Plant Physiol* 149(2). <https://doi.org/10.1104/pp.108.130518>
63. van Zonneveld M, Dawson I, Thomas E, Scheldeman X, van Etten J, Loo J, Hormaza JI (2014) Application of molecular markers in spatial analysis to optimize in situ conservation of plant genetic resources. In *Genomics of Plant Genetic Resources: Volume 1. Managing, Sequencing and Mining Genetic Resources*. https://doi.org/10.1007/978-94-007-7572-5_4
 64. Varshney RK, Terauchi R, McCouch SR (2014) Harvesting the Promising Fruits of Genomics: Applying Genome Sequencing Technologies to Crop Breeding. *PLoS Biol* 12(6). <https://doi.org/10.1371/journal.pbio.1001883>
 65. Wang W, Liu G, Niu H, Timko MP, Zhang H (2014) The F-box protein COI1 functions upstream of MYB305 to regulate primary carbohydrate metabolism in tobacco (*Nicotiana tabacum* L. cv. TN90). *J Exp Bot* 65(8). <https://doi.org/10.1093/jxb/eru084>
 66. Wei T, Simko V, Levy M, Xie Y, Jin YJ, Zemla J (2017) Visualization of a Correlation Matrix. R package corrplot. *Statistician*, 56
 67. Welsch R, Arango J, Bär C, Salazar B, Al-Babili S, Beltrán J, Chavarriaga P, Ceballos H, Tohme J, Beyera P (2010) Provitamin a accumulation in cassava (*Manihot esculenta*) roots driven by a single nucleotide polymorphism in a phytoene synthase gene. *Plant Cell* 22(10). <https://doi.org/10.1105/tpc.110.077560>
 68. Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. In *Media* (Vol. 35, Issue July)
 69. Wolfe MD, Rabbi IY, Egesi C, Hamblin M, Kawuki R, Kulakow P, Lozano R, Carpio DP, del, Ramu P, Jannink J (2016) Genome-Wide Association and Prediction Reveals Genetic Architecture of Cassava Mosaic Disease Resistance and Prospects for Rapid Genetic Improvement. *Plant Genome* 9(2). <https://doi.org/10.3835/plantgenome2015.11.0118>
 70. Xu L, Liu F, Lechner E, Genschik P, Crosby WL, Ma H, Peng W, Huang D, Xie D (2002) The SCF(COI1) ubiquitin-ligase complexes are required for jasmonate response in Arabidopsis. *Plant Cell* 14(8):1919–1935. <https://doi.org/10.1105/tpc.003368>
 71. Zhang S, Chen X, Lu C, Ye J, Zou M, Lu K, Feng S, Pei J, Liu C, Zhou X, Ma P, Li Z, Liu C, Liao Q, Xia Z, Wang W (2018) Genome-wide association studies of 11 agronomic traits in cassava (*Manihot esculenta* crantz). *Frontiers in Plant Science*, 9. <https://doi.org/10.3389/fpls.2018.00503>
 72. Zhong S, Li H, Bodi Z, Button J, Vespa L, Herzog M, Fray RG (2008) MTA is an Arabidopsis messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. *Plant Cell* 20(5). <https://doi.org/10.1105/tpc.108.058883>

Figures

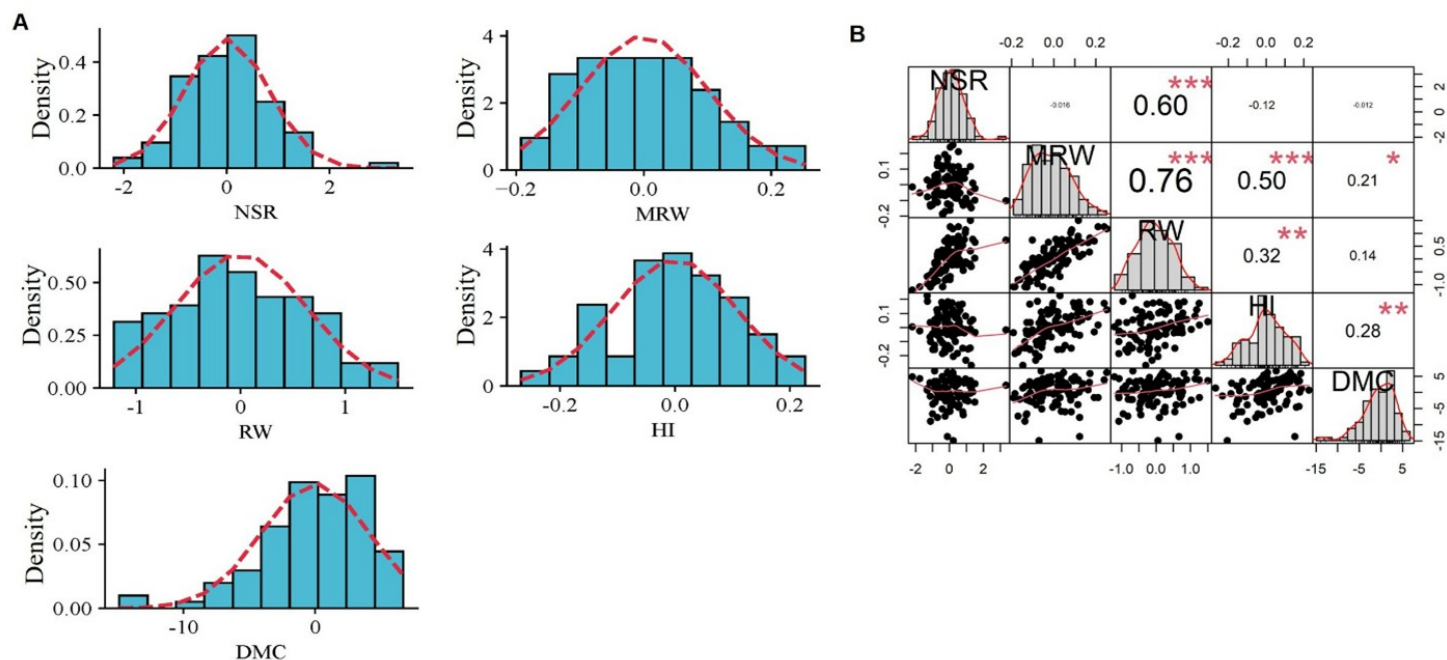


Figure 1

Distribution of traits and relationship among the traits based on best linear unbiased predictions (BLUPs). (A) Density distribution of variation in five traits in cassava germplasm, (B) correlation coefficients among traits. *, **, *** indicate significance at 0.05, 0.01 and 0.001. NSR= number of storage roots, MRW= mean storage root weight, RW= root weight per plant, HI= harvest index, and DMC= storage root dry matter content.

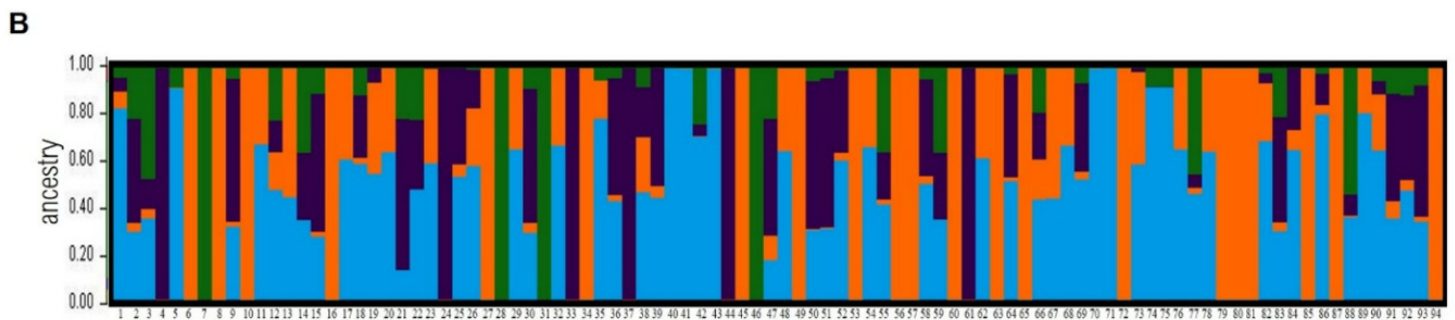
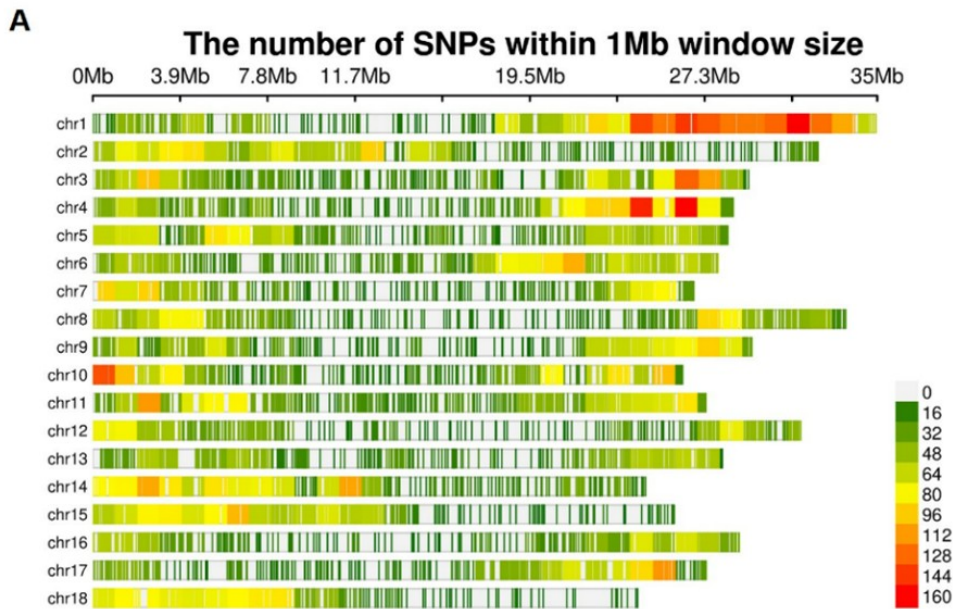


Figure 2

Overview of SNP genotyping data and population structure based on SNPs. (A) Genome-wide single nucleotide polymorphism (SNP) distributions on 18 chromosomes of cassava. The horizontal axis displays the chromosome length in base pairs; the vertical axis on the left displays the number of SNPs (inserted bar indicates the number of SNP within 1 Mb window). (B) Individual ancestry estimated by ADMIXTURE analysis ($K = 4$). The colored sections within each bar indicate membership of the genotypes in the different subpopulations (subgroup1=blue, subgroup2=green, subgroup3=blue-black, subgroup4=orange).

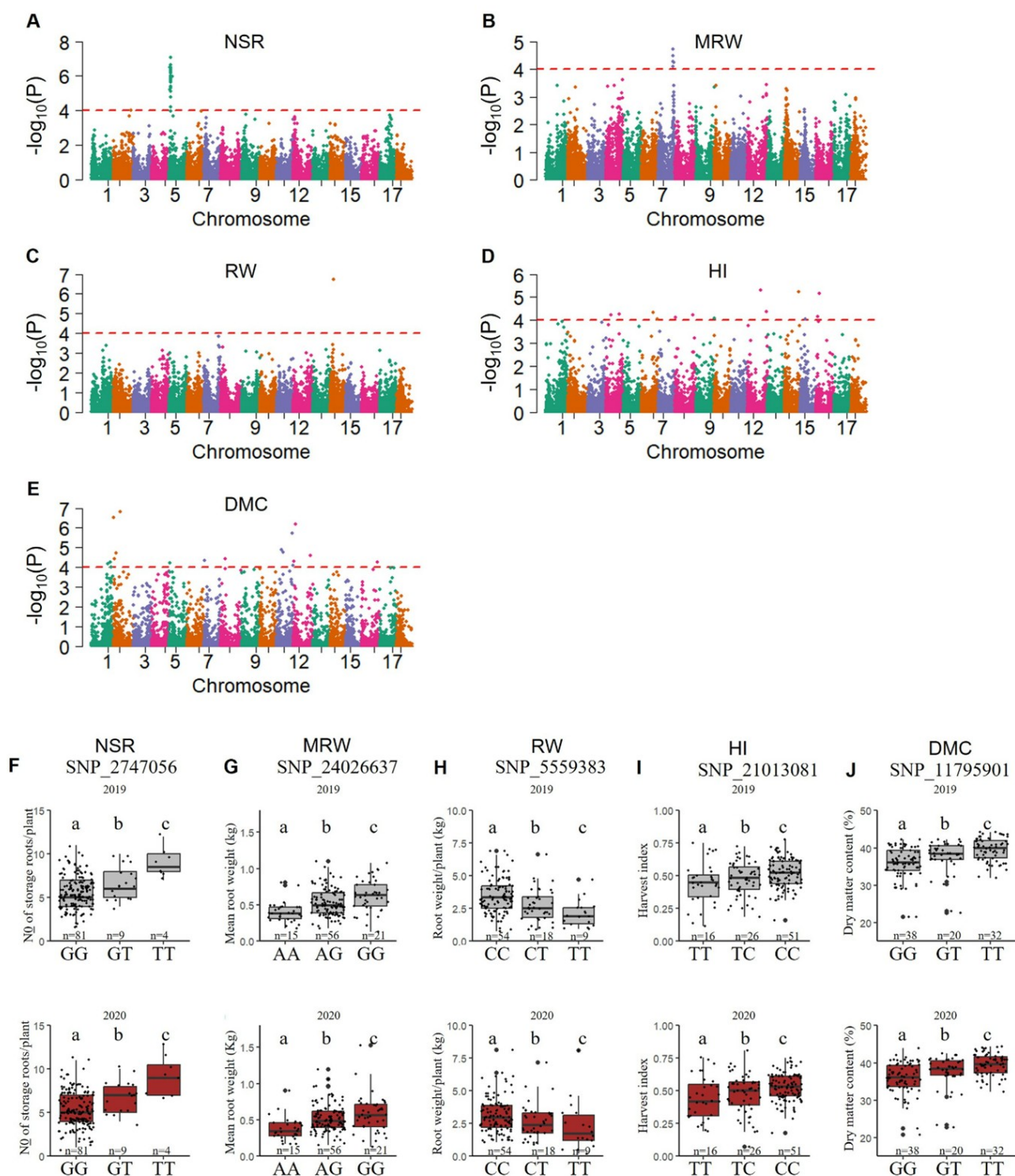


Figure 3

Genome-wide association and haplotype analysis for the five traits based on the topmost SNP. Manhattan plots of $-\log_{10}(P)$ -value) showing chromosomal positions of SNP markers for (A) number of storage root, NSR, (B) mean storage root weight, MRW, (C) root weight, RW, (D) harvest index, HI, and (E) dry matter content, DMC. The dashed horizontal red line indicates the genome-wide significance threshold. Allelic effect on (F) NSR, (G) MRW, (H) RW, (I) HI and (J) DMC based on the highest significant SNP associated

with the trait across the two years. For the haplotypes, from left to right is homozygote reference, heterozygote and homozygote SNP. In the boxplots, the central lines denote the average. Significant levels were determined using a least significant difference test and the different lowercase letters above the boxplots represent significant differences ($P \leq 0.05$).

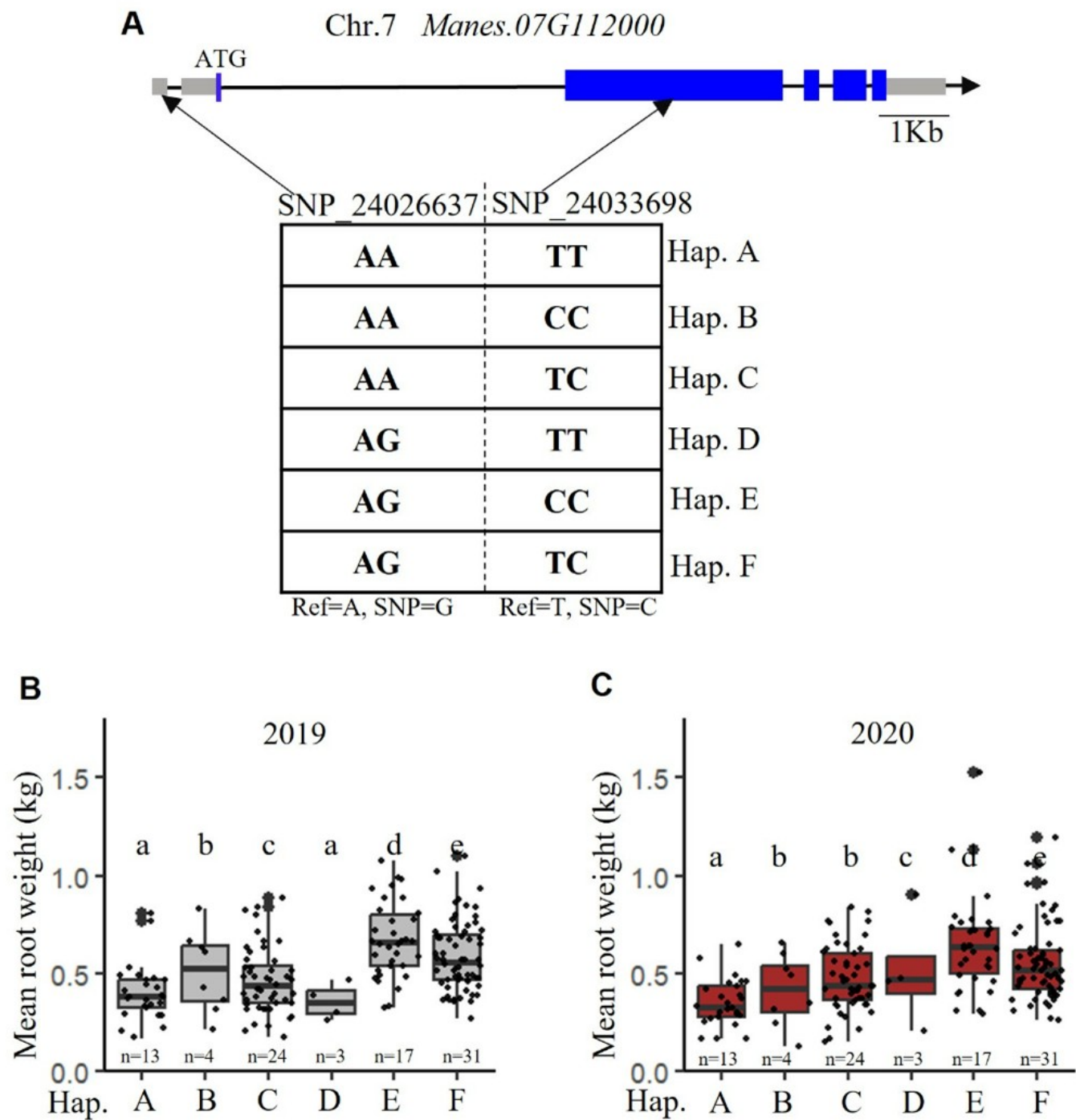


Figure 4

Gene structure and haplotype analysis for the mean storage root weight (MRW) based on the two topmost SNPs.

(A) Exon structure of *Manes.07G112000* and DNA polymorphisms in the gene resulting in six different haplotypes. The two lines indicate the position of the two SNPs (SNP_24026637 and SNP_24033698) within the gene. The haplotypes indicate nucleotide or SNP combinations. Boxplots for mean storage root weight of the haplotypes for the (B) 2019 and (C) 2020 planting seasons are shown. In the boxplots, the central lines denote the average. Significant levels were determined using a least significant difference test and the different lowercase letters above the boxplots represent significant differences ($P \leq 0.05$).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarydata.xlsx](#)