Check for updates

Validation of KASP-SNP markers in cassava germplasm for marker-assisted selection of increased carotenoid content and dry matter content

Adenike D. Ige[1,2], Bunmi Olasanmi[3], Guillaume J. Bauchet[4], Ismail S. Kayondo[1], Edwige Gaby Nkouaya Mbanjo[1], Ruth Uwugiaren[1,5], Sharon Motomura-Wages[6], Joanna Norton[6], Chiedozie Egesi[1,7], Elizabeth Y. Parkes[1], Peter Kulakow[1], Hernán Ceballos[8], Ibnou Dieng[1] and Ismail Y. Rabbi[1*]

[1]International Institute of Tropical Agriculture (IITA), Ibadan, Oyo State, Nigeria, [2]Pan African University Life and Earth Sciences Institute (including Health and Agriculture), University of Ibadan, Ibadan, Nigeria, [3]Department of Crop and Horticultural Sciences, University of Ibadan, Ibadan, Nigeria, [4]Boyce Thompson Institute, Ithaca, NY, United States, [5]Molecular Plant Sciences program, Washington State University, Pullman, WA, United States, [6]College of Tropical Agriculture and Human Resources, University of Hawaii at Manoa, Hilo, HI, United States, [7]Cornell University, Ithaca, NY, United States, [8]The Alliance of Bioversity International and the International Center for Tropical Agriculture (CIAT), Cali, Colombia

Provitamin A biofortification and increased dry matter content are important breeding targets in cassava improvement programs worldwide. Biofortified varieties contribute to the alleviation of provitamin A deficiency, a leading cause of preventable blindness common among pre-school children and pregnant women in developing countries particularly Africa. Dry matter content is a major component of dry yield and thus underlies overall variety performance and acceptability by growers, processors, and consumers. Single nucleotide polymorphism (SNP) markers linked to these traits have recently been discovered through several genome-wide association studies but have not been deployed for routine marker-assisted selection (MAS). This is due to the lack of useful information on markers' performances in diverse genetic backgrounds. To overcome this bottleneck, technical and biological validation of the loci associated with increased carotenoid content and dry matter content were carried out using populations independent of the marker discovery population. In the present study, seven previously identified markers for these traits were converted to a robust set of uniplex allele-specific polymerase chain reaction (PCR) assays and validated in two independent pre-breeding and breeding populations. These assays were efficient in discriminating marker genotypic classes and had an average call rate greater than 98%. A high correlation was observed between the predicted and observed carotenoid content as inferred by root yellowness intensity in the breeding (r = 0.92) and pre-breeding (r = 0.95) populations. On the other hand,

dry matter content-markers had moderately low predictive accuracy in both populations (r< 0.40) due to the more quantitative nature of the trait. This work confirmed the markers' effectiveness in multiple backgrounds, therefore, further strengthening their value in cassava biofortification to ensure nutritional security as well as dry matter content productivity. Our study provides a framework to guide future marker validation, thus leading to the more routine use of markers in MAS in cassava improvement programs.

1 Introduction

Cassava (*Manihot esculenta* Crantz) is a principal starchy root crop for both the rural and urban populations in the tropics, particularly in sub-Saharan Africa. The continent accounts for more than half of the total world's production of 303 million tonnes (FAOSTAT, 2020). Due to its ability to grow with few agricultural inputs in marginal environments characterized by poor soils and water stress, the crop takes on the crucial role of being a key food security crop in sub-Saharan Africa (Burns et al., 2010). In Africa, cassava roots are usually consumed fresh after short boiling and are also processed into various fermented products such as gari and fufu or unfermented products such as flour and starch. Besides its role as food, cassava is increasingly relied upon globally as an industrial raw material for the production of paper, textiles, plywood, glue, biofuel, animal feed and beverages (Balagopalan, 2002).

Among the major staple sources of carbohydrates, cassava has one of the longest breeding cycles, ranging from five to eight years (Ceballos et al., 2004; Ceballos et al., 2012). This is due to its long growth cycle of 12 - 18 months; clonal propagation, which results in low multiplication rates of planting propagules; its high levels of heterozygosity; and difficulty in making crosses due to poor and asynchronous flowering as well as low seed set per cross (Jennings and Iglesias, 2002; Ceballos et al., 2012). These challenges notwithstanding, breeding programs around the world have developed improved varieties that address

various production constraints, including biotic and abiotic stresses, improved yield and dry matter content (Kawano, 2003; Okechukwu and Dixon, 2008), as well as enhanced micronutrient content, particularly of provitamin A carotenoid (Ilona et al., 2017; Andersson et al., 2017). However, as the demand for cassava for food, feed, and industrial raw materials continues to grow due to an increase in the population (Anyanwu et al., 2015; Parmar et al., 2017), breeding programs need to adopt modern breeding technologies and tools such as marker-assisted selection or genomic selection to increase the rate of genetic gain to meet the demands in an ecologically sustainable manner (Ceballos et al., 2015).

Marker Assisted Selection (MAS) is one of the most important applications of molecular marker technology in plant breeding (Collard and Mackill, 2008). It facilitates the indirect selection of new plants based on the presence of a favorable allele at a marker that is closely linked to a trait of interest (Collard and Mackill, 2008). In cassava, MAS can be used at the early stages of the breeding scheme to select individuals with favorable alleles for storage-root traits that would otherwise only be phenotypically evaluated at maturity. This has several advantages, namely: 1) reduction in the time it takes to decide to advance a clone to the next stage of testing; 2) reduction in the number of clones to be advanced to larger plot trials, thereby saving scarce phenotyping resources, and 3) in some cases, the cost of marker assay is lower than those that are usually expended on the actual trait phenotyping. A good example is carotenoid quantification using the spectrophotometry method and High-Performance Liquid Chromatography (HPLC) which can be many-fold more expensive than a SNP assay (Semagn et al., 2014; Andersson et al., 2017). Therefore, the adoption of MAS can increase the efficiency of selection, leading to a more rapid rate of genetic gain, and fewer cycles of phenotypic evaluation, thus, reducing the time for varietal development (Collard and Mackill, 2008).

The prerequisite for the application of MAS is the identification of major genes or genomic regions associated with a trait of interest. Over the last 15 years, quantitative trait

---

Abbreviations: BLAST, Basic Local Alignment Search Tool; BLUP, Best Linear Unbiased Prediction; CET, Clonal Evaluation Trial; GWAS, Genome-Wide Association Study; HPLC, High-Performance Liquid Chromatography; IITA, International Institute of Tropical Agriculture; KASP, Kompetitive Allele-Specific PCR; MAE, Mean Absolute Error; MAS, Marker-Assisted Selection; PCR, Polymerase Chain Reaction; *PSY2*, Phytoene Synthase 2; QTL, Quantitative Trait Loci; REML, Restricted Maximum Likelihood; RMSE, Root Mean Square Error; SD, Standard Deviation; SN, Seedling Nursery; SNP, Single Nucleotide Polymorphism.

loci (QTL) mapping studies of different traits in cassava have been published (Fregene et al., 2001; Akano et al., 2002; Balyejusa et al., 2007; Okogbenin et al., 2012; Morillo et al., 2013; Rabbi et al., 2014). Most of these studies used segregating populations developed from either selfed or bi-parental crosses between parents with contrasting trait levels (Rabbi et al., 2014). More recently, association or linkage disequilibrium mapping using a genome-wide association study (GWAS) has become an approach for unraveling the molecular genetic basis underlying the natural phenotypic variation (Davey et al., 2011). The advantages of GWAS over QTL mapping are the higher mapping resolution and the identification of a broader set of alleles in large and diverse germplasm (Yu and Buckler, 2006). Several GWAS have been conducted on key cassava traits, including cassava mosaic disease resistance (Wolfe et al., 2016; Rabbi et al., 2020), carotenoids content (Esuma et al., 2016; Rabbi et al., 2017; Ikeogu et al., 2019; Rabbi et al., 2020), and dry matter content (Rabbi et al., 2020) in diverse cassava populations to discover significant loci. Despite this progress, the output from discovery research has not been translated into assays that breeders can easily use to support selection decisions (Chagné et al., 2019). To overcome this bottleneck and bridge the gap between discovery and routine usage, new trait-linked markers must be technically and biologically validated, preferably using independent populations (Platten et al., 2019; Ige et al., 2021). This process informs the breeder whether the expected allelic phenotypic effects are reproducible in different genetic backgrounds from the one in which the marker-trait association was originally identified (Li et al., 2013).

Cassava is very efficient in carbohydrate production, but its starchy roots lack essential micronutrients, including provitamin A carotenoid (Sayre et al., 2011; Ceballos et al., 2017). Vitamin A deficiency is a public health problem in more than half of all countries, especially in Africa and South-East Asia (WHO, 2022). This deficiency often leads to several severe health and economic consequences, including increased incidence of night blindness; suppressed immunity, leading to an increased mortality rate, especially among pregnant women and young children as well as reduced productivity (Sayre et al., 2011; WHO, 2022). Dry matter content is a crucial yield component and is a key determinant of variety acceptance by growers, processors, and consumers (Sánchez et al., 2014; Bechoff et al., 2018). Varieties with low dry matter content (less than 30%) are often less preferred than those with moderate to high dry matter content. Like carotenoid content, dry matter content can only be assessed on mature storage roots at the end of the growing season. Marker-assisted selection is expected to provide breeders with the ability, for example, to screen either for genotypes with high levels of these traits or eliminate those with undesirable levels at the early stages of testing, thereby allocating their limited field plots to high-value genotypes. The objective of the study was, therefore, to convert and validate SNP markers associated with increased provitamin A carotenoid

biofortification and dry matter content; two important traits under active improvement in many breeding programs in the world.
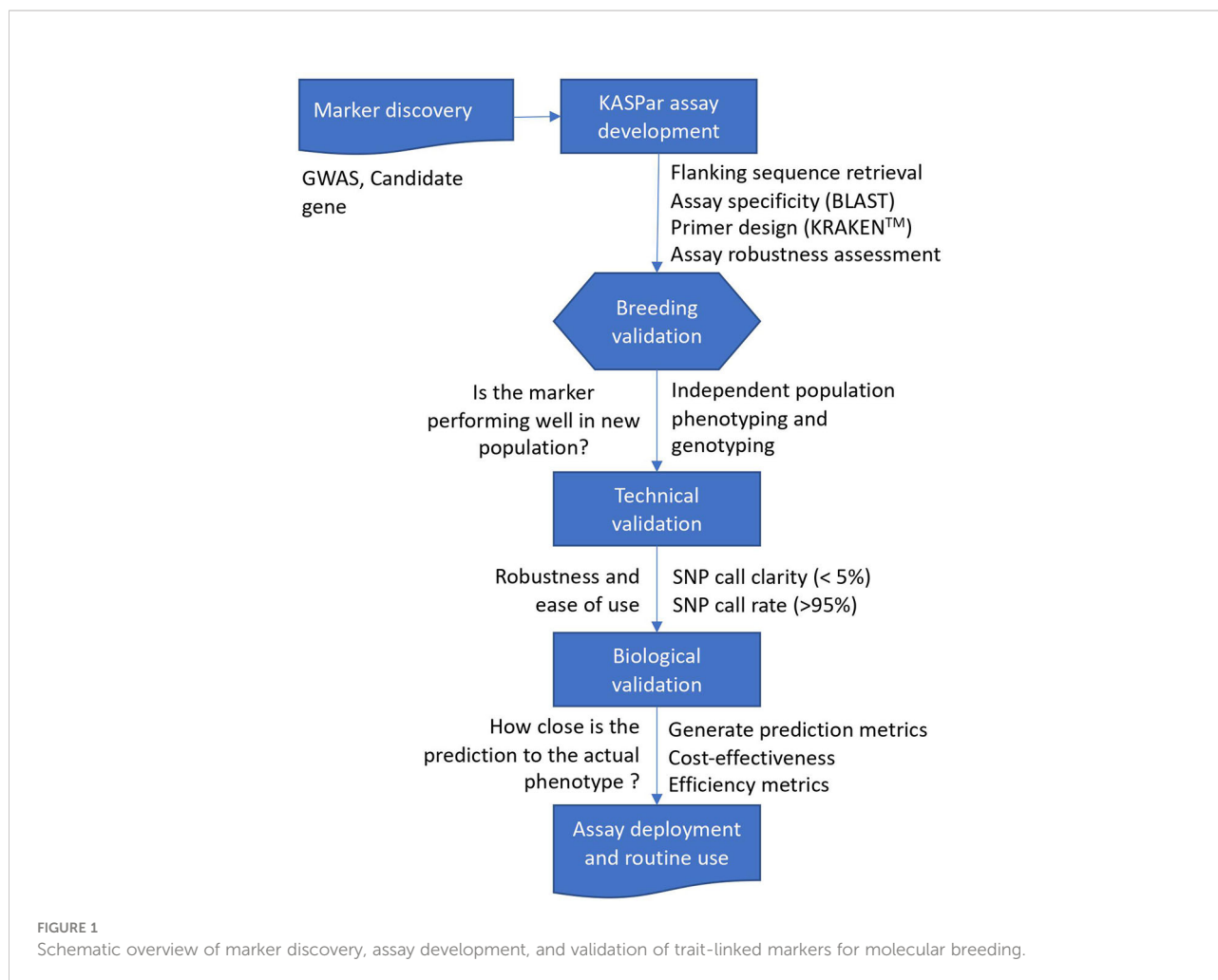
2 Materials and methods

2.1 Retrieving significant loci linked to increased carotenoid and dry matter contents

The marker discovery, development, and validation workflow used in the present study is presented in Figure 1. The SNP markers linked to increased carotenoid and dry matter contents validated in the present study (Table 1) were derived from (Udoh et al., 2017; Rabbi et al., 2020). Sequencing of four carotenoid pathway candidate genes in 167 cassava accessions from the International Institute of Tropical Agriculture (IITA), Nigeria, uncovered two important SNPs on phytoene synthase 2 (PSY2) (Udoh et al., 2017). The most significant SNP on PSY2 (position 572) is a causal mutation resulting in a non-synonymous amino acid substitution (Welsch et al., 2010). This marker was converted to a Kompetitive allele-specific PCR (KASP) assay and renamed as per its chromosomal position on the version 6.1 reference genome to S1_24155522. Additional markers associated with the study traits were obtained from a recent GWAS that used a large panel of 5130 diverse clones developed at IITA in Nigeria (Rabbi et al., 2020). The population was genotyped at more than 100K genome-wide SNP markers via genotyping-by-sequencing. For carotenoid content, a major locus on chromosome 1 tagged by three markers (S1_24159583, S1_24636113, and S1_30543962) as well as five new genomic regions associated with this trait on chromosomes 5, 8, 15, and 16 were identified. Of these, three (S1_30543962, S5_3387558, and S8_25598183) were selected for KASP conversion and validation in the present study. The markers associated with dry matter content were S1_24197219, S6_20589894, and S12_5524524.

2.2 Development of kompetitive allele-specific PCR assays

Fifty nucleotide bases flanking the target SNP on each side were obtained from the cassava (Manihot esculenta) reference genome (version 6.1) available at https://phytozome-next.jgi.doe.gov/info/Mesculenta_v6_1. Then, a nucleotide-nucleotide Basic Local Alignment Search Tool (BLAST) was used to check for locus-specificity of the assays to minimize the possibility of cross-amplification of the marker in non-target regions of the genome. Primers were designed using a proprietary Kraken™ software system from LGC Biosearch Technologies, UK, with the default parameters.

FIGURE 1
Schematic overview of marker discovery, assay development, and validation of trait-linked markers for molecular breeding.

Assay technical validation was carried out using a panel of 188 genetically diverse cassava genotypes that are known to segregate at the SNP assays. A no-template control was included in the SNP genotyping. The robustness of the assays was assessed under four DNA concentrations (Dilution 1 = 10X, Dilution 2 = 100X, Dilution 3 = 24X, Dilution 4 = 240X) using metrics such as ease of scoring the three expected genotype classes, tightness, and distinctiveness of the genotypic classes on cluster plots, percentage call rate, and percentage clarity.

2.3 Validation of kompetitive allele-specific PCR assays in independent populations

The KASP assays' performances were assessed in two independent populations from IITA, Nigeria. These populations, consisting of breeding and pre-breeding germplasm, were different from the panel used for GWAS marker discovery.

2.3.1 Description of the study populations

The breeding population is part of IITA's regular recurrent selection pipeline and was derived from controlled crosses among elite genotypes carried out in 2017. Yield, multiple stress tolerance, and dry matter content are the major traits for improvement in this population. The cohort was evaluated initially at the seedling nursery (SN) stage consisting of 22,420 progenies from 563 families (mean family size of 40, ranging from 1 to 220) in 2018 in Ibadan, Nigeria (7°24′ N, 3°54′ E; 200 m above sea level). The SN trial was planted at a spacing of 1 m × 0.25 m and harvested 12 months after planting; a selection of 1599 genotypes based on disease resistance, plant vigor, plant architecture, and root yield was advanced to clonal evaluation trial (CET) at Ikenne, Nigeria (6°52′ N 3°42′ E; 61 m above sea level).

The pre-breeding population was developed using a polycross hybridization between twenty-three (23) IITA and nineteen (19) CIAT (International Center for Tropical Agriculture) parental clones. To ensure safe germplasm exchange between Africa and Latin America, the hybridization

TABLE 1 Description of the seven SNP markers, their flanking sequences and KASP primers.

Traits	SNP name*	Chr	Position (bp)	Flanking sequences	Favorable allele	Unfavorable allele	Forward primer (Allele X and Allele Y)	Common Reverse Primer
Carotenoid content	S1_24155522	1	24155522	GACAGATGAGCTTGTTGATGGACCT AATGCTTCACACATAACGCCAACAG [A/C]TTTAGATAGGTGGGAAGCAAG GTTGGAAGATATGTTTCGAGGTCGT CCCT	A	C	Allele X: ATGCTTCACACATAACGCCAACAGA Allele Y: TGCTTCACACATAACGCCAACAGC	CCAACCTTGCTTCCCACCTATCTAA
	S1_30543962	1	30543962	GGAGGTTTTTTTATGTGGCATTCTCA GCAGCTGCAGGAATCTCATTGTTCTT TACAATTCCAAGGCTCTTTCTTGCAA TTAAAGGTGGGGAAGGTGCCCC[A/G] GACCTCTGGGGAACTGCTGGAAATGC TGCCATTAATATTGGTGGTAAATGCTT TAACCTTTCTCTGTCATATGAAGAAAA TGAGTTAATTGATGTATAAT	G	A	Allele X: TCCAGCAGTTCCCCAGAGGTCT Allele Y: CAGCAGTTCCCCAGAGGTCC	CTTTACAATTCCAAGGCTCTTTCTTGCAA
	S5_3387558	5	3387558	GTTACACTTAGACCCTTGTCATTAAAC ATTACTGAGGCTGCAGTTGAAGTGTAA ACAACTCTTTTCACTGTCTTTGATTCCA AGCATGTCCTTAATATCC[C/T]TAGCAA TCCATCCACGGCTATTTTGGT CACACTTTCTTCAGGTTCTTTTCCATAA TGATCCATTGGGTGAGCCACATGGAAG ACTCCAATACAACCTTCA	T	C	Allele X: TGATTCCAAGCATGTCCTTAATATCCC Allele Y: TTGATTCCAAGCATGTCCTTAATATCCT	CCAAAATAGCCGTGGATGGATTGCTA
	S8_25598183	8	25598183	TAAATTCTGACTGTCTTGGCATGACTGT CCAGGTAGTCCCCGAAAATGAGAATGC TGCTCTCTACTCCACTCATTCATTCAAG ATTTTGTTCAAGGAAGG[G/T]GGTTGTG GAACCTTCATTCCGCTCTTTT TCAACTTGCTCTCTTCAGTAAGGCAATA CAATCAGCAACAAACCTCTGGAATGGG GCCCCAGATGAACCCTT	T	G	Allele X: CGGAATGAAGGTTCCACAACCC Allele Y: GCGGAATGAAGGTTCCACAACCA	CTCTCTACTCCACTCATTCATTCAAGATT
Dry matter content	S1_24197219	1	24197219	GATGTAGGCATGTTACATATAAGGGCT ACATACACATTAGCAGCTAAAATGAGA CCCGGATACCGAGCAATGCCATCAATT GAGAGATGAACTCAGGGTG[C/T]CCTG GCCATGCAGCTCCAGTAACCAAA TTTTCATGAGTGTAGCAACGATGTATT GGATCAGGTTCTAGCCATGTTGCCCCA GCCAAGACCACGTTAATCT	C	T	Allele X: AATTGAGAGATGAACTCAGGGTGC Allele Y: TCAATTGAGAGATGAACTCAGGGTGT	TCATGAAAATTTGGTTACTGGAGCTGCAT
	S6_20589894	6	20589894	ATTGATGATTTTTTATTCATGATATGTA GCTATCAAAGTTACTCAGCAATGTCCTT GTTTTAGCCATGCTAGCAGCATGTTTTG TTGCGACAACAGTTGG[A/G]AGTTGTAT	G	A	Allele X: CATGTTTTGTTGCGACAACAGTTGGA Allele Y: ATGTTTTGTTGCGACAACAGTTGGG	CTGCCCAATGATATTCTGCATACAAGATA

(Continued)

TABLE 1 Continued

Traits	SNP name*	Chr	Position (bp)	Flanking sequences	Favorable allele	Unfavorable allele	Forward primer (Allele X and Allele Y)	Common Reverse Primer
	S12_5524524	12	5524524	GAATATTGTTTTATCTTGTA TGCAGAATATCATTGGGCAGGAAGCAG GGAAAAGCGTGATTGAGGAATATTTAC GTCGTAGGGGTCACTCAG TGAATTATTTTAACTCTTTGATTGCTTC GCCAGTGCCTGGTCTCCAGAATGTGTG TGTTGCTTTGGTTTGTAGTTCCAAAGG TGAGCTGTGGCAATTTTA[T/C]TGCAGC CCCACTGGCATTAGACGCAGT AAATTATATCAGGACGAAGTAAGTTCA TCCTTCAAAGGAAATGATAATGGTCAA TTTGTGGGGAGCAAAGGTT	C	T	Allele X: TCTAATGCCAGTGGGGCTGCAA Allele Y: TCTAATGCCAGTGGGGCTGCAG	GTTCCAAAGGTGAGCTGTGGCAATT

*SNP marker position in base pairs (bp) is based on the cassava reference genome v6.1 (Bredeson et al., 2016).

was carried out in Hilo, Hawaii (19°38'24.57"N 155°4'57.76"W; 204m above sea level), which has a mild tropical climate that is suitable for cassava survival as well as prolific flowering. The objective of developing the population was to enhance provitamin A biofortification by introgressing a new source of novel alleles for Africa and to develop germplasm incorporating resistance to cassava mosaic disease, high content of provitamin A and starch, and tolerance to acid soils and drought for Latin America. Like the breeding population, a SN evaluation trial was established in Ibadan for 5,608 genotypes planted at a spacing of 1 m x 0.25 m. The mean family size was 16, ranging from 1 to 165 clones. The trial was harvested 10 months after planting and approximately 14% of the genotypes (790) were advanced to CET at Ikenne, Nigeria (6°52′ N 3°42′ E; 61 m above sea level) based on vigor alone.

2.3.2 Field experiment and phenotyping of cassava storage roots for carotenoid and dry matter contents

Genotypes at the first CET were used for the validation study. A CET was preferred because of the large size (typically several hundred) and diversity of most of the traits. The trials were laid out in an augmented design to accommodate a large number of entries. Each genotype was planted at a spacing of 1 m between rows and 0.5 m within rows. For the breeding population, the experiment comprised of 58 to 60 plots per 30 sub-blocks with five checks (IITA-TMS-IBA00070, IITA-TMS-IBA30572, TMEB419, IITA-TMS-IBA982101, IITA-TMS-IBA980581) randomly assigned to each sub-block. This trial was planted in June 2018 and harvested in June 2019. The pre-breeding population trial carried out between October 2018 and October 2019 consisted of 900 plots (50 plots per 18 sub-blocks) with four checks (TMEB419, IITA-TMS-IBA30572, IITA-TMS-IBA070593, and IITA-TMS-IBA000070) in each block. All field management practices were performed according to the technical recommendations and standard agricultural practices for cassava (Abass et al., 2014; Atser et al., 2017).

Direct estimation of total carotenoid content using laboratory extraction followed by spectrophotometry and HPLC is not only expensive but also has low throughput for routine germplasm screening, particularly at the early stages of breeding selection. Due to a large number of genotypes in this study, we used two color-based methods to assess the variation among the cassava genotypes for carotenoid content. Utilization of color intensity as a proxy for the carotenoids content in cassava is justified because of the well-established linear relationship between root yellowness and total carotenoids content (Pearson's coefficient, r, ranges from 0.81 to 0.84) (Iglesias et al., 1997; Chávez et al., 2005; Marín Colorado et al., 2009; Sánchez et al., 2014; Esuma et al., 2016) as well as with total beta-carotene (Udoh et al., 2017). Moreover, 80 to 90% of total carotenoid content in cassava is provitamin A compared to other crops, making color-based assessment a good proxy for

estimating not only total carotenoids content but also total β-carotene content (Wong et al., 2004; Ceballos et al., 2017; Jaramillo et al., 2018). In maize, kernel color is not correlated with the primary carotenoid of interest, that is, β-carotene, which has the highest pro-vitamin A activity due to the presence of other carotenoids such as β-cryptoxanthin, zeaxanthin, and lutein (Wong et al., 2004).

The first method is a standard visual assessment of the yellowness of root parenchyma using a color chart with a scale ranging from 1 (white root) to 7 (orange root) (Supplementary Figure 1). The second method is a surface color measurement using a CR-410 chromameter (Konica Minolta). The chromameter's three-dimensional color space defined by l*, a*, and b* coordinates provides a more objective and precise assessment of surface color and its intensity. The Commission Internationale de l'Éclairage (CIELAB) l* coordinate value represents sample lightness ranging from 0 (black) to 100 (diffuse white). The a* values represent either red (positive coordinate values) or green (negative coordinate values). Of importance in our study is the b* coordinate, whose positive values measure the degree of yellowness and therefore provide an indirect estimate of carotenoid content.

For the chromameter color measurements, eight roots per plot were peeled, washed, grated, and thoroughly mixed. A subsample was transferred into a transparent sampling bag (Whirl-Pak™) and scanned at four independent positions. The CR-410 chromameter was calibrated each day using a white ceramic and illuminant D65 was used as a source of light.

Root dry matter content was assessed using the oven-drying method. Eight fully developed roots were randomly selected from each plot, peeled, washed, grated, and thoroughly mixed. For each sample, 100 g was weighed and oven-dried for 72 h at 80°C. The dry samples were then weighed, and the dry matter content was expressed as the percentage of dry weight relative to fresh weight.

2.3.3 Genotyping

Young leaves were sampled three months after planting from the evaluation plots. Three 6mm diameter leaf discs were obtained from each genotype into 96-well plates on ice, and freeze-dried for at least 72 hours. The samples were shipped to a genotyping service provider (Intertek, Sweden) for automated DNA extraction and SNP genotyping using four markers linked to increased carotenoid content and three markers linked to increased dry matter content (Table 1) using the KASP assay. Two blank controls were included in each plate during genotyping.

The KASP assay protocol is provided in the KASP manual (LGC Genomics, 2013). In brief, genotyping was carried out using the high-throughput PCR SNPline workflow using 1 µL reaction volume in 1536-well PCR plates. The KASP genotyping reaction mix is comprised of three components: (i) sample DNA (10 ng); (ii) marker assay mix consisting of target-specific

primers; and (iii) KASP-TF^{TM} Master Mix containing two universal fluorescence resonant energy transfer cassettes (FAM and HEX), passive reference dye (ROX^{TM}), Taq polymerase, free nucleotides, and MgCl$_2$ in an optimized buffer solution. The SNP assay mix is specific to each marker and consists of two kompetitive allele-specific forward primers and one common reverse primer (Table 1). After PCR, the plates were fluorescently read, and allele calls were made using KRAKEN^{TM} software.

2.3.4 Data analysis

2.3.4.1 Phenotypic data analysis

A linear mixed model was implemented using restricted maximum likelihood (REML) to estimate the best linear unbiased predictions (BLUPs) for each genotype in the CETs of breeding and pre-breeding populations. The model was fitted using the *asreml* package (Butler, 2020) in R software version 4.0.3 (R Development Core Team, 2020). The mathematical model used for the incomplete block design analysis is represented as follows:

$$Y_{ijk} = \mu + G_i + R_k + B_{jk} + e_{ijk}$$

where Y_{ijk} is the vector of phenotype data of the ith genotype of the jth block nested into the kth replication, μ is the overall mean, G_i is the effect of the ith genotype, R_k is the effect of the kth replication, B_{jk} is the effect of the jth block nested into the kth replication, and e_{ijk} is the residual, modeled as a sum of measurement error and a spatially dependent random process. A first-order auto-regressive process in both row and column directions was used for the spatial trend (Gilmour et al). All effects except μ were assumed to be random.

Broad-sense heritability was calculated as:

$$H^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_e^2)$$

where H^2 is the broad-sense heritability; σ_g^2 and σ_e^2 are the variance components for the genotype effect and the residual error, respectively.

Pairwise correlation analysis of the traits using the BLUP estimates was determined using the *corr.test* function in the psych package (R Development Core Team, 2020).

2.3.4.2 Technical and biological validation of kompetitive allele-specific PCR markers

Technical performance metrics used to assess the robustness of markers include SNP call rate and call clarity. Call rate is the proportion of samples with non-missing genotype calls. Call clarity is defined by the ease of assigning samples to a genotype class based on their position on a fluorescence cluster Cartesian plot. The tighter and more distinct the cluster, the easier and more consistent it is to call the respective genotype class, namely homozygous for either allele 1 or 2 or heterozygous in the case of biallelic SNPs and a diploid genome.

Biological validation of the converted markers was assessed using three complementary approaches. First, the allele substitution effect was visualized using boxplots, and the difference in carotenoid and dry matter content BLUP values among the genotypic classes at each marker locus was assessed using a pairwise t-test. Second, the predictive ability of the SNP markers was estimated using a multiple linear regression model. As shown in the linear model below, marker alleles and the observed phenotypes were considered as the independent and response variables, respectively.

$$Y = \mu + \mu_1 + \mu_2 + \dots + \mu_n + e$$

where: Y = phenotypic observations of traits, μ = overall mean of the population, μ_1, μ_2, μ_n = marker effects, e = residual value.

Bootstrap resampling was carried out to obtain robust estimates of model parameters, specifically the magnitude and confidence intervals of the allele-substitution effects for the markers associated with the two traits (Davison and Hinkley, 1997). The *reg_intervals* function in the tidymodels R package (Kuhn and Wickham, 2020) was used to generate 1000 bootstrap resamples and fit the multiple linear regression model on each one.

Finally, a 5-fold cross-validation analysis repeated 10 times was carried out to obtain marker performance metrics including predictive accuracy (R^2), root mean square error (RMSE, the square root of the mean squared difference between observed and predicted trait values), and mean absolute error (MAE, the average absolute difference between the predictions made by the model and the actual observations). To achieve this, the breeding and pre-breeding population data were partitioned into training and testing sets in a 3:1 ratio with a stratification based on the target traits (chromameter b* value or dry matter content). The regression model developed in the training set was used to predict the trait values in the hold-out testing set. All model training and cross-validation analyses were implemented in the R caret package (Kuhn, 2008).

3 Results

3.1 Phenotypic variation for root yellowness and dry matter content

Out of the evaluated clones, 81% of the breeding population and 52% of the pre-breeding population had white storage roots, while the remaining showed a range of yellow color (visual score of between 2 and 5), suggesting varying levels of carotenoid content (Figure 2). The average visual score of root yellowness was 1.30 (standard deviation (sd) = 0.72) in the breeding population and 1.74 (sd = 0.95) in the pre-breeding population. The chromameter b* values showed a bi-modal distribution in the two populations (Figure 2). The first peak (b*values from 11 to 22) is associated with clones that produced
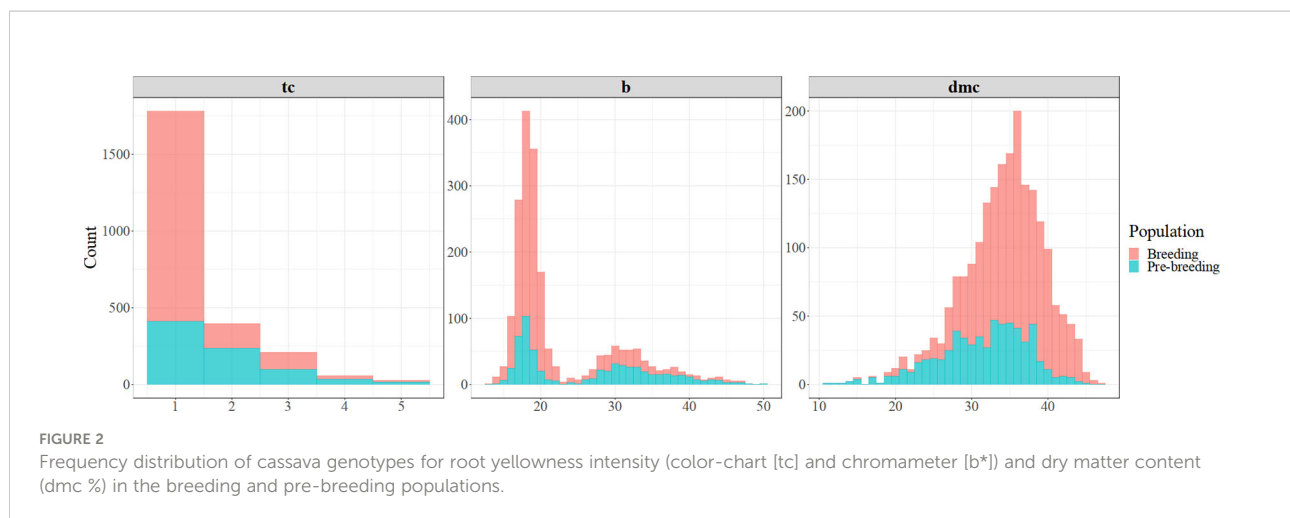
FIGURE 2
Frequency distribution of cassava genotypes for root yellowness intensity (color-chart [tc] and chromameter [b*]) and dry matter content (dmc %) in the breeding and pre-breeding populations.

white roots, while the second peak (b* values from 22 to 50) is associated with the variations among clones with yellow roots. The average chromameter measures of yellow color intensity were 21.0 (sd = 6.12) and 26.2 (sd = 8.82) for breeding and pre-breeding populations, respectively. The dry matter content of the clones evaluated in the two populations was normally distributed (Figure 2), ranging from 11.2 to 47.4, with averages of 31.5 (sd = 5.92) in the pre-breeding population and 35.1 (sd = 4.80) in the breeding population.

The broad-sense heritability of the visual assessment from the color chart and chromameter values were 0.87 and 0.88, respectively, for the breeding population, and 0.81 and 0.93, respectively for the pre-breeding population (Table 2). The heritability estimate for dry matter content in the pre-breeding population (0.61) was lower than that of the breeding population (0.70) (Table 2).

The two measures of root yellowness intensity; visual assessment and chromameter b* value were significantly and positively correlated (~0.90) in the two populations suggesting that visual scoring is a good proxy for yellow-color intensity. Significant negative correlations ranging from -0.27 to -0.20 were observed between root yellowness and dry matter content in the two populations. However, a lower magnitude of correlation coefficient was observed between visual assessment and dry matter content (-0.20) as well as between chromameter b* value and dry matter content (-0.23) in the pre-breeding population.

3.2 Technical validation of kompetitive allele-specific PCR assays

3.2.1 SNP call rate, call clarity, and genotypic frequencies

All markers were successfully converted to allele-specific KASP assays. The call rate and clarity were high for a wide range of DNA dilution levels tested during marker development, indicating that the assays are robust and suitable for routine use (Supplementary Figure 2). The overall call rate was above 98% for all the markers in the two populations genotyped (mean = 99%, sd = 0.53) (Supplementary Table 1). As expected, three distinct clusters were observed for all the SNPs except for marker S5_3387558 where the frequency of cluster TT was very low (Supplementary Figure 3).

Allelic and genotypic frequencies of the markers are presented in Supplementary Figures 4, 5, respectively. The favorable alleles across all the carotenoid-linked markers were more common in the pre-breeding population (ranging from 11 to 34%) compared to the breeding population (ranging from 3 to 11%) (Supplementary Figure 4). The favorable allele A at marker S1_24155522 had a frequency of 34% and 11% in the pre-breeding and breeding populations, respectively (Supplementary Figure 4). More than 15% of the individuals were homozygous for allele A at this marker in the pre-breeding population (Supplementary Figure 5).

TABLE 2 Broad-sense heritability calculated on a mean plot basis for root visual assessment, chromameter value, and dry matter content in the two cassava populations.

Traits	Breeding population			Pre-breeding population		
	σ_g^2	σ_e^2	H^2	σ_g^2	σ_e^2	H^2
Visual assessment	0.500	0.072	0.87	0.637	0.145	0.81
Chromameter b* value	34.641	4.572	0.88	65.676	4.879	0.93
Dry matter content	16.231	6.831	0.70	21.806	14.039	0.61

σ_g^2 is the clonal genotypic variance, σ_e^2 is the residual variance, and H^2 is the broad-sense heritability.

The percentage was much lower in the breeding population with only 2.3% of the individuals fixed for the same allele. In the two populations, between 0.4 to 7.3% of the individuals were fixed for the favorable alleles at the three remaining markers suggesting an opportunity to use these markers to increase their frequencies in the population (Supplementary Figure 5). For dry matter content, the favorable alleles at the linked SNPs occurred at intermediate to high frequencies ranging from 28 to 76% in both populations (Supplementary Figure 4). The percentage of individuals that were fixed for the favorable alleles was higher in the breeding than in the pre-breeding population for this trait (Supplementary Figure 5). About 27 to 53% of the individuals in the pre-breeding population were fixed for the unfavorable alleles (Supplementary Figure 5).

3.2.2 Biological validation

3.2.2.1 Allelic substitution effects on carotenoid and dry matter contents

Significant pairwise differences between genotypic classes at all the markers associated with carotenoid content were observed (Figure 3). Most of the markers displayed an additive mode of action with individuals carrying two copies of the favorable alleles having a higher intensity of root yellowness (b*) than those with only one copy while those that are fixed for non-favorable alleles had white roots. For instance, the mean b*

values for genotype classes AA, CA, and CC for marker S1_24155522 were 38.53 ± 2.85, 31.64 ± 3.89, and 18.37 ± 2.48, respectively in the pre-breeding population (Figure 3B).

The genotype classes at the dry matter content-linked markers were not as differentiated as those for carotenoid content (Figure 4). Nonetheless, significant differences were observed among the genotypes at marker S6_20589894 in the two populations. In the pre-breeding population, there was no significant difference among CC, CT, and TT at marker S12_5524524 (Figure 4).

3.2.2.2 Marker-trait regression, confidence intervals, and models' predictive performances

The estimates of marker-trait regression parameters from bootstrap resampling analysis for the two traits are presented in Figures 5 and 6. The regression model with all the four markers for carotenoid variation produced R^2 values of 0.85 in the breeding population and 0.91 in the pre-breeding population. However, in a subset of the breeding and pre-breeding populations consisting of only genotypes with yellow roots, the R^2 values decreased to 0.46 and 0.53, respectively. SNP S1_24155522 had the strongest effect on variation in root yellowness. The effect size of having a single copy of a favorable allele (A) on the increase in root yellowness intensity (chromameter b* value) was 10.8 and 12.1 in the breeding and pre-breeding populations, respectively. Having two copies of the same
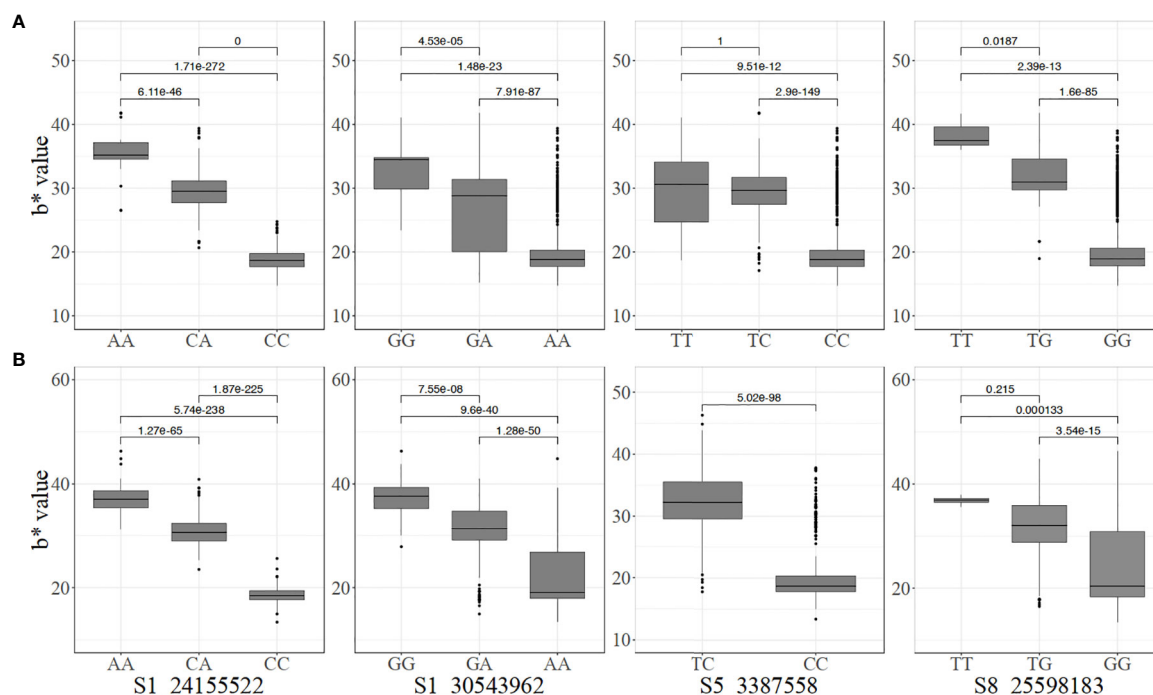

FIGURE 3
Allelic substitution effects of the markers associated with increased carotenoid content in the (A) breeding, and (B) pre-breeding populations (For marker S5_3387558, the mean and standard deviation cannot be estimated because one genotype had TT).
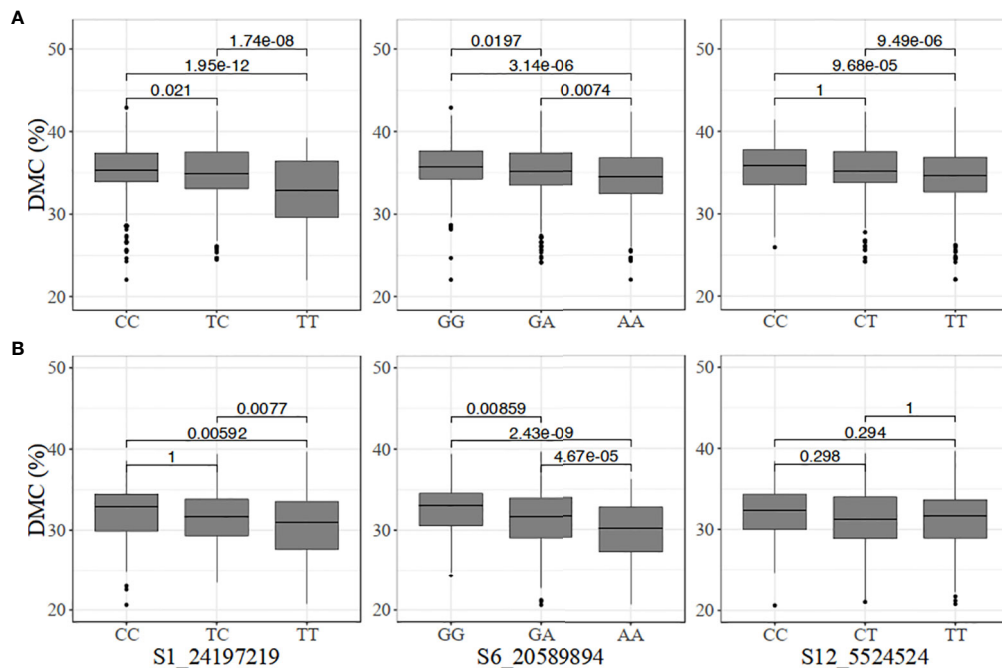
Allelic substitution effects of the markers associated with increased dry matter content (DMC) in the **(A)** breeding, and **(B)** pre-breeding populations.

allele resulted in an even larger effect size of 15.5 and 17.8, respectively, in the two populations. The confidence intervals of these marker genotypes were narrow, indicating higher precision of the marker prediction. After controlling for the major locus (S1_24155522) effect in the two populations, the other three markers had a low to moderate effect on the trait (Supplementary Figure 6). The effect sizes of the minor SNPs were more significant in the breeding compared to the pre-breeding population, particularly for markers S5_3387758 and S8_25598183.

The regression model with all three markers for dry matter content produced low R^2 values of 0.06 in the breeding and 0.09 in the pre-breeding population. Having two copies of favorable alleles across all SNPs was associated with an increase in dry matter content percentage from between 1.01 and 2.50 percentage units in the breeding population. A similar direction of effects was observed in the pre-breeding population except for marker S12_5524524 which did not contribute to the multiple regression model. A notable observation is a reversal in the effects of markers S1_24197219 and S6_20589894 across the two populations, suggesting a QTL by genetic background interaction.

The predictive accuracy of the carotenoid markers from the cross-validation regression analysis ranged from 0.84 to 0.91 with a mean of 0.87. In the pre-breeding population, the value was higher and approximately 0.90 in the training and testing sets (Table 3, Supplementary Figure 7). However, low predictive accuracy values were obtained for dry matter content-linked markers in the breeding population (0.07 for the training set and 0.05 for the testing set) and pre-breeding population (0.08 for the training set and 0.07 for the testing set) (Table 3, Supplementary Figure 7). In the breeding population, RMSE and MAE values for carotenoid markers were 1.88 and 1.43, respectively, in the training set, and 2.03 and 1.52, respectively, in the testing set (Table 3). The values of RMSE and MAE were 2.31 and 1.71, respectively, in the training set, and 2.35 and 1.68 in the testing set of the pre-breeding population. These values were higher for dry matter content-markers in both populations compared to those of carotenoid content-markers. The use of RMSE and MAE is very common in model evaluation, and they are good measures of prediction accuracy.

4 Discussion

The present study focused on the development and validation of markers for carotenoids and dry matter contents, two traits that are of primary importance to cassava breeding programs worldwide (Sánchez et al., 2006; Okechukwu and Dixon, 2008; Bouis et al., 2011; Saltzman et al., 2013; Talsma et al., 2013). Similar to our observations, several studies that used diverse cassava germplasm, particularly from Africa have reported that dry matter content and carotenoid content parameters such as total
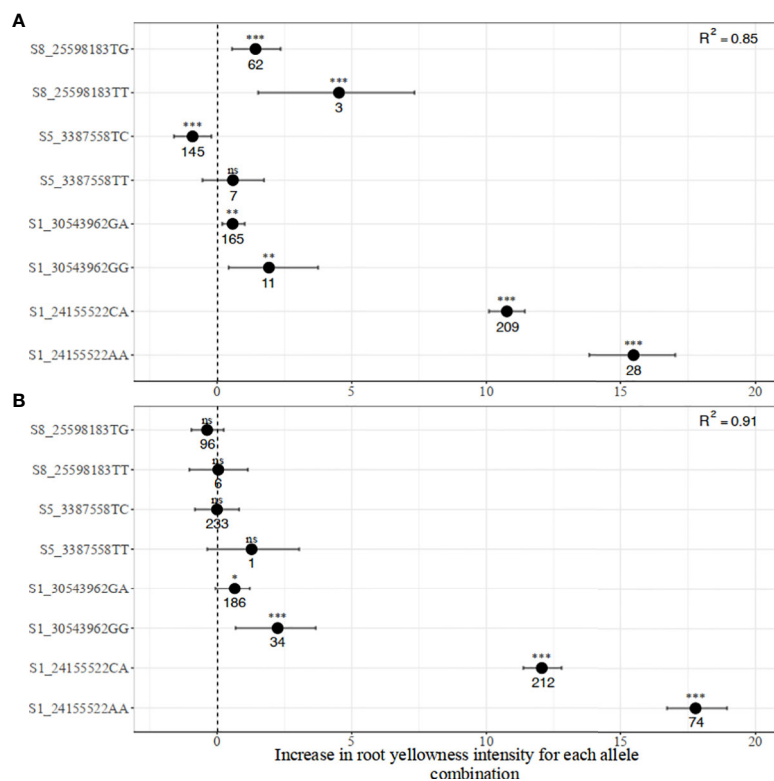
FIGURE 5
Distribution of marker allelic effects associated with increased carotenoid content in **(A)** breeding, and **(B)** pre-breeding populations.

carotenoid content, root yellowness intensity, and visual assessment of storage roots are negatively correlated with r-values ranging from 0.1 to 0.6 (Marín Colorado et al., 2009; Akinwale et al., 2010; Njoku et al., 2015; Esuma et al., 2016; Rabbi et al., 2017). On the contrary, these traits are independent in Latin American cassava populations (Ceballos et al., 2013; Sánchez et al., 2014). Although the selection of genotypes based on high intensities of root yellowness at the early stage of the breeding scheme saves time and costs associated with carotenoid quantification, it would indirectly select for lower dry matter content (Sánchez et al., 2014).

As part of the breeders' toolbox for MAS, markers validated can be used to select for the study traits simultaneously and are expected to address the challenges associated with vitamin A deficiency and higher demand for varieties with higher dry matter content. Vitamin A deficiency is a widespread nutritional public health problem in sub-Saharan Africa, with women and children being the most affected (Gegios et al., 2010; Stephenson et al., 2010). Breeding of clones with enhanced carotenoid levels is one of the most cost-effective and sustainable approaches to helping the communities burdened by vitamin A deficiency (Pfeiffer and McClafferty, 2007; Bouis

et al., 2011; Talsma et al., 2013). While we have explored the performance of the markers in the IITA pre-breeding and breeding populations, these assays should have wide application in other breeding programs where the QTLs are present and are linked to the same SNP alleles. More importantly, these markers can be used for rapid mobilization of the favorable alleles in new populations developed using parents that are known to carry the associated trait alleles.

Trait discovery in cassava has been an active area of research with the advent of genome-wide SNP markers from genotyping-by-sequencing (Wolfe et al., 2016; Esuma, 2016; Rabbi et al., 2017; Udoh et al., 2017; Ikeogu et al., 2019; Rabbi et al., 2020). However, these trait discoveries have not been translated into deployable assays, obscuring their utility in MAS. Here, we have provided a framework for translating the outputs from genetic mapping to a set of easy-to-use, robust, and predictive allele-specific uniplex assays. The framework includes both technical and biological validation of the assays in a range of diverse germplasm to ascertain the relevance of the markers for predicting the trait values in independent populations. The KASP SNP platform was chosen due to its amenability for genotyping of any combination of individual samples and marker assays, and ease of automation to achieve high-
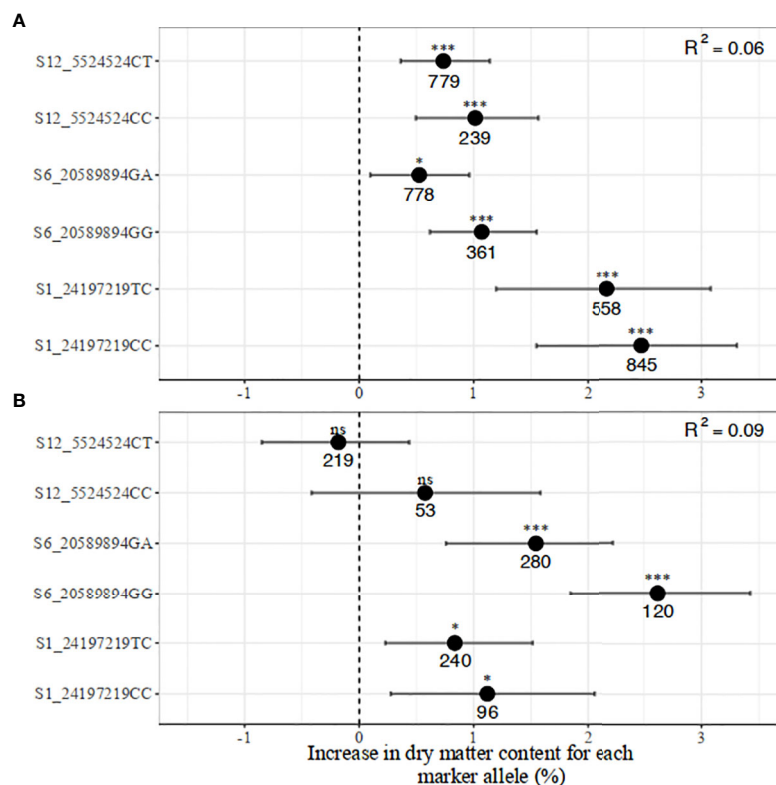
FIGURE 6
Distribution of the marker allelic effects associated with increased dry matter content in **(A)** breeding, and **(B)** pre-breeding populations.

throughput population screening (Semagn et al., 2014; Ogbonna et al., 2020; Ige et al., 2021). The designed SNP assays were found to work under a wide range of DNA concentrations. Even though the tightness of the cluster plots differed between the standard and low DNA concentrations, they were sufficiently distinct to allow for a high genotype call rate and call clarity. This suggests that the assays are expected to work under diverse DNA concentrations and most likely from different sample preparation methods, including fresh, frozen, lyophilized, or oven-dried (Semagn et al., 2014).

The best way to measure the predictive ability of a model is to test it on a dataset that is independent of the data used to train the model (Wani et al., 2018). The k-fold cross-validation, where the original dataset is randomly partitioned into equally sized k-subsets (a single subset is retained as the validation data for testing the model, and the remaining k - 1 subsets are used as training data), is one of the most commonly used cross-validation methods (Refaeilzadeh et al., 2009; Mathew et al., 2015). It is routinely used to assess genomic prediction accuracies (Okeke et al., 2017; de Andrade et al., 2019;

TABLE 3 Prediction performance metrics of the markers associated with increased carotenoid and dry matter contents in the training and testing sets of the breeding and pre-breeding populations.

Traits	Populations		N	R^2	RMSE	MAE
Chromameter b* value (Carotenoid content)	Breeding	Training set	1030	0.84	1.88	1.43
		Testing set	345	0.84	2.03	1.52
	Pre-breeding	Training set	396	0.91	2.31	1.71
		Testing set	133	0.90	2.35	1.68
Dry matter content (%)	Breeding	Training set	1102	0.07	3.13	2.48
		Testing set	368	0.05	3.20	2.48
	Pre-breeding	Training set	402	0.08	3.70	3.00
		Testing set	136	0.07	3.07	2.53

N, Number of observations; R^2, Prediction accuracy; RMSE, Root mean square error; MAE, Mean absolute error.

Phumichai et al., 2022). To our knowledge, this is the first study to use this metric for marker validation in cassava. In the present study, the performance of the regression model in an independent data set, that is, the testing set in terms of predictive accuracy for chromameter b* values were 0.84 in the breeding population and 0.90 in the pre-breeding population. These values are quite similar to those obtained in the training sets, suggesting that the models developed are stable and reliable. The low values of RMSE and MAE recorded in the breeding population compared to that of the pre-breeding population indicated that the markers are more accurate in predicting the carotenoid content in the breeding population. Both measures of cross-validation accuracy for this trait suggest that the designed assays can be deployed for routine use in breeding pipelines with carotenoid biofortification as a breeding goal. On the other hand, the predictive accuracy of the dry matter content markers (mean = 0.07) across populations was lower than the values obtained for carotenoid content markers. This could be due to the quantitative nature of dry matter content (Kawano et al., 1987). In the discovery population (Rabbi et al., 2020), also reported low predictive ability ($R^2<$ 0.11) of these markers.

Moreover, for both traits, we used a bootstrapping regression approach to provide robust estimates of allele substitution effects and their confidence intervals (Fox and Weisberg, 2018). The multiple regression analysis of carotenoid content markers revealed that marker S1_24155522 was the main driver in carotenoid accumulation while the other markers played additional but minor roles. This result is consistent with earlier observations that the *PSY2* gene, which hosts marker S1_24155522 is a key rate-limiting step in the carotenoid pathway in cassava (Welsch et al., 2010; Rabbi et al., 2020). In a candidate gene-based association study, Udoh et al. (2017) reported that total carotenoid content and β-carotene were significantly associated with this marker, which occurs at position 572 of the *PSY2* gene. Indeed, the previously identified SNPs from other candidate genes such as *lcyE, lcyB,* and *crtRB* were hardly significantly associated with the trait (Udoh et al., 2017). On the other hand, markers S1_24197219 and S6_20589894 had small but significant effects on dry matter content in both populations, while marker S12_5524524 showed an effect in the pre-breeding population. Marker S6_20589894 was reported to occur close to the gene Manes.06G103600 (Bidirectional sugar transporter Sweet4-Related) which mediates fructose transport across the tonoplast of roots (Rabbi et al., 2020).

While we have assessed the performance of selected markers across the two diverse populations, we acknowledge that these markers may be tagging only a subset of major loci underlying the studied traits, particularly dry matter content. Ongoing and future GWAS and biparental QTL mapping studies will likely uncover additional QTLs. Such markers can be validated using the framework provided in this study and incorporated into the breeders' toolset, thus increasing the accuracy of predicting these traits. Moreover, other traits that are of importance for which major associations have recently been reported but not converted to marker assays include cassava green mite (Rabbi et al., 2020), cassava brown streak disease (Kayondo et al., 2018) and root mealiness (Uchendu et al., 2021). A major caveat of our study is the use of single-marker assays to tag each major locus for the two traits. The top SNPs at these loci are expected to be tightly linked to the causal allele based on the large GWAS population used in the discovery, with more than 5000 individuals genotyped at more than 100K genome-wide positions. However, factors such as independent emergence or evolution of favorable alleles at specific genes and nearby SNP can result in non-perfect association, hence resulting in false-positive and false-negative. This and other limitations of single marker analysis can be addressed by a haplotype-based approach through, for example, amplicon sequencing (AmpSeq) of targeted genomic regions (Yang et al., 2016). Further work is required to establish the viability of Amplicon Sequencing as a platform for haplotype-based MAS in cassava.

Data availability statement

The data presented in this study can be found in an online repository. The names of the repository and accession number (s) can be found at: https://cassavabase.org/ftp/manuscripts/Ige_et_al_2022/Ige_et_al_2022.csv.

Author contributions

ADI, BO, IYR, and PK designed the study. SM-W and JN developed the pre-breeding population. ADI and RU performed the experiment. ADI, GB, ID, and IYR analyzed the data. ADI drafted the manuscript. BO, GB, ISK, EGNM, CE, EP, PK, HC, and IYR revised the manuscript. All authors have read, edited, and approved the current version of the manuscript.

Funding

Acknowledgments

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1016170/full#supplementary-material

References

Abass, A. B., Towo, E., Mukuka, I., Okechukwu, R. U., Ranaivoson, R., Tarawali, G., et al. (2014). *Growing cassava: a training manual from production to postharvest* (Ibadan, Nigeria: International Institute of Tropical Agriculture).

Akano, O., Dixon, O., Mba, C., Barrera, E., and Fregene, M. (2002). Genetic mapping of a dominant gene conferring resistance to cassava mosaic disease. *Theor. Appl. Genet.* 105, 521–525. doi: 10.1007/s00122-002-0891-7

Akinwale, M. G., Aladesanwa, R. D., Akinyele, B. O., Dixon, A. G. O., and Odiyi, A. C. (2010). Inheritance of carotene in cassava (Manihot esculenta crantz). *Int. J. Genet. Mol. Biol.* 2, 198–201. doi: 10.5897/IJGMB.9000034

Andersson, M. S., Saltzman, A., Virk, P. S., and Pfeiffer, W. H. (2017). Progress update: crop development of biofortified staple food crops under HarvestPlus. *Afr J. Food Agric. Nutr. Dev.* 17, 11905–11935. doi: 10.18697/ajfand.78.HarvestPlus05

Anyanwu, C. N., Ibeto, C. N., Ezeoha, S. L., and Ogbuagu, N. J. (2015). Sustainability of cassava (Manihot esculenta crantz) as industrial feedstock, energy and food crop in Nigeria. *Renew Energy* 81, 745–752. doi: 10.1016/j.renene.2015.03.075

Atser, G., Dixon, A., Ekeleme, F., Chikoye, D., Dashiell, K. E., Ayankanmi, T. G., et al. (2017). *The ABC of weed management in cassava production in Nigeria: a training manual* (Ibadan, Nigeria: International Institute of Tropical Agriculture).

Balagopalan, C. (2002). *Cassava utilization in food, feed and industry*. CABI

Balyejusa, K. E., Rönnberg-Wästljung, A.-C., Egwang, T., Gullberg, U., Fregene, M., and Westerbergh, A. (2007). Quantitative trait loci controlling cyanogenic glucoside and dry matter content in cassava (Manihot esculenta crantz) roots. *Hereditas* 144, 129–136. doi: 10.1111/j.2007.0018-0661.01975.x

Bechoff, A., Tomlins, K., Fliedel, G., Lopez-lavalle, L. A. B., Westby, A., Hershey, C., et al. (2018). Cassava traits and end-user preference: Relating traits to consumer liking, sensory perception, and genetics. *Crit. Rev. Food Sci. Nutr.* 58, 547–567. doi: 10.1080/10408398.2016.1202888

Bouis, H. E., Hotz, C., McClafferty, B., Meenakshi, J. V., and Pfeiffer, W. H. (2011). Biofortification: A new tool to reduce micronutrient malnutrition. *Food Nutr. Bull.* 32, S31–S40. doi: 10.1177/15648265110321S105

Bredeson, J. V., Lyons, J. B., Prochnik, S. E., Wu, G. A., Ha, C. M., Edsinger-Gonzales, E., et al. (2016). Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nat. Biotechnol.* 34, 562–570. doi: 10.1038/nbt.3535

Burns, A., Gleadow, R., Cliff, J., Zacarias, A., and Cavagnaro, T. (2010). Cassava: The drought, war and famine crop in a changing world. *Sustainability* 2, 3572–3607. doi: 10.3390/su2113572

Butler, D. (2020). "Asreml: fits the linear mixed model," in *R package version 4.1*, (UK: VSNi) vol. 0.143.

Ceballos, H., Davrieux, F., Talsma, E. F., Belalcazar, J., Chavarriaga, P., and Andersson, M. S. (2017). *Carotenoids in cassava roots* (UK: IntechOpen). doi: 10.5772/intechopen.68279

Ceballos, H., Iglesias, C. A., Perez, J. C., and Dixon, A. G. (2004). Cassava breeding: opportunities and challenges. *Plant Mol. Biol.* 56, 503–516. doi: 10.1007/s11103-004-5010-5

Ceballos, H., Kawuki, R. S., Gracen, V. E., Yencho, G. C., and Hershey, C. H. (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. doi: 10.1007/s00122-015-2555-4

Ceballos, H., Kulakow, P., and Hershey, C. (2012). Cassava breeding: Current status, bottlenecks and the potential of biotechnology tools. *Trop. Plant Biol.* 5, 73–87. doi: 10.1007/s12042-012-9094-9

Ceballos, H., Morante, N., Sánchez, T., Ortiz, D., Aragón, I., Chávez, A. L., et al. (2013). Rapid cycling recurrent selection for increased carotenoids content in cassava roots. *Crop Sci.* 53, 2342–2351. doi: 10.2135/cropsci2013.02.0123

Chagné, D., Vanderzande, S., Kirk, C., Profitt, N., Weskett, R., Gardiner, S. E., et al. (2019). Validation of SNP markers for fruit quality and disease resistance loci in apple (Malus $\times$ domestica borkh.) using the OpenArray® platform. *Hortic. Res.* 6, 30. doi: 10.1038/s41438-018-0114-2

Chávez, A. L., Sánchez, T., Jaramillo, G., Jm, B., Echeverry, J., Bolaños, E. A., et al. (2005). Variation of quality traits in cassava roots evaluated in landraces and improved clones. *Euphytica* 143, 125–133. doi: 10.1007/s10681-005-3057-2

Collard, B. C. Y., and Mackill, D. J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos. Trans. R Soc. Lond B Biol. Sci.* 363, 557–572. doi: 10.1098/rstb.2007.2170

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., and Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* 12, 499–510. doi: 10.1038/nrg3012

Davison, A. C., and Hinkley, D. V. (1997). *Bootstrap methods and their application* (UK: Cambridge university press).

de Andrade, L. R. B., Sousa, M. B., Oliveira, E. J., de Resende, M. D. V., and Azevedo, C. F. (2019). Cassava yield traits predicted by genomic selection methods. *PloS One* 14, e0224920. doi: 10.1371/journal.pone.0224920

Esuma, W. (2016). *Genetic analysis and genome-wide association mapping of carotenoid and dry matter content in cassava* (South Africa: PhD Thesis. University of the Free State). 66, 627–635

Esuma, W., Herselman, L., Labuschagne, M. T., Ramu, P., Lu, F., Baguma, Y., et al. (2016). Genome-wide association mapping of provitamin a carotenoid content in cassava. *Euphytica* 212, 97–110. doi: 10.1007/s10681-016-1772-5

Esuma, W., Kawuki, R. S., Herselman, L., and Labuschagne, M. T. (2016). Diallel analysis of provitamin a carotenoid and dry matter content in cassava (Manihot esculenta crantz). *Breed Sci.* 15159 (66), 627–635. doi: 10.1270/jsbbs.15159

FAOSTAT (2020). Available at: http://www.fao.org/faostat/en/#data/QC (Accessed January 11, 2021).

Fox, J., and Weisberg, S. (2018). *An r companion to applied regression* (USA: Sage publications).

Fregene, M., Okogbenin, E., Mba, C., Angel, F., Suarez, M. C., Janneth, G., et al. (2001). Genome mapping in cassava improvement: Challenges, achievements and opportunities. *Euphytica* 120, 159–165. doi: 10.1023/A:1017565317940

Gegios, A., Amthor, R., Maziya-Dixon, B., Egesi, C., Mallowa, S., Nungo, R., et al. (2010). Children consuming cassava as a staple food are at risk for inadequate zinc, iron, and vitamin a intake. *Plant Foods Hum. Nutr.* 65, 64–70. doi: 10.1007/s11130-010-0157-5

Gilmour, A. R., Cullis, B. R., and Verbyla, A. P. (1997) Accounting for natural and extraneous variation in the analysis of field experiments. *J. Agric. Biol. Environ. Stat.* 1997, 269–293. doi: 10.2307/1400446

Ige, A. D., Olasanmi, B., Mbanjo, E. G. N., Kayondo, I. S., Parkes, E. Y., Kulakow, P., et al. (2021). Conversion and validation of uniplex SNP markers for selection of resistance to cassava mosaic disease in cassava breeding programs. *Agronomy* 11, 420. doi: 10.3390/agronomy11030420

Iglesias, C., Mayer, J., Chavez, L., and Calle, F. (1997). Genetic potential and stability of carotene content in cassava roots. *Euphytica* 94, 367–373. doi: 10.1023/A:1002962108315

Ikeogu, U. N., Akdemir, D., Wolfe, M. D., Okeke, U. G., Chinedozi, A., Jannink, J.-L., et al. (2019). Genetic correlation, genome-wide association and genomic prediction of portable NIRS predicted carotenoids in cassava roots. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01570

Ilona, P., Bouis, H. E., Palenberg, M., Moursi, M., and Oparinde, A. (2017). Vitamin a cassava in Nigeria: crop development and delivery. *Afr J. Food Agric. Nutr. Dev.* 17, 12000–12025. doi: 10.4314/ajfand.v17i2

Jaramillo, A. M., Londoño, L. F., Orozco, J. C., Patiño, G., Belalcazar, J., Davrieux, F., et al. (2018). A comparison study of five different methods to measure carotenoids in biofortified yellow cassava (Manihot esculenta). *PloS One* 13, e0209702. doi: 10.1371/journal.pone.0209702

Jennings, D. L., and Iglesias, C. (2002). *Cassava: Biology, production and utilization* (Wallingford UK: CABI) 18, 149–166. doi: 10.1079/9780851995243.0149

Kawano, K. (2003). Thirty years of cassava breeding for productivity–biological and social factors for success. *Crop Sci.* 43, 1325–1335. doi: 10.2135/cropsci2003.1325

Kawano, K., Fukuda, W., and Cenpukdee, U. (1987). Genetic and environmental effects on dry matter content of cassava Root1. *Crop Sci.* 27, 69–74. doi: 10.2135/cropsci1987.0011183X002700010018x

Kayondo, S. I., Del Carpio, D. P., Lozano, R., Ozimati, A., Wolfe, M., Baguma, Y., et al. (2018). Genome-wide association mapping and genomic prediction for CBSD resistance in manihot esculenta. *Sci. Rep.* 8, 1–11. doi: 10.1038/s41598-018-19696-1

Kuhn, M. (2008). Building predictive models in r using the caret package. *J. Stat. Softw* 28, 1–26. doi: 10.18637/jss.v028.i05

Kuhn, M., and Wickham, H. (2020). "Tidymodels: Easily install and load the'Tidymodels' packages," in *R package version 0.1. 0.* Available at: https://tidymodels.tidymodels.org/.

LGC Genomics (2013). *KASP genotyping chemistry user guide and manual* (Teddingt UK: LGC Ltd).

Li, L., Tacke, E., Hofferbert, H.-R., Lübeck, J., Strahwald, J., Draffehn, A. M., et al. (2013). Validation of candidate gene markers for marker-assisted selection of potato cultivars with improved tuber quality. *Theor. Appl. Genet.* 126, 1039–1052. doi: 10.1007/s00122-012-2035-z

Marín Colorado, J. A., Ramírez, H., and Fregene, M. (2009). Genetic mapping and QTL analysis for carotenes in a s1 population of cassava. *Acta Agronómica* 58, 15–21.

Mathew, B., Léon, J., and Sillanpää, M. J. (2015). Integrated nested Laplace approximation inference and cross-validation to tune variance components in estimation of breeding value. *Mol. Breed* 35, 99. doi: 10.1007/s11032-015-0248-y

Morillo, A. C., Morillo, C. Y., and Ceballos, L. H. (2013). Identification of QTLs for carotene content in the genome of cassava (Manihot esculenta crantz) and S1 population validation. *Acta Agronómica* 62, 196–206.

Njoku, D., Gracen, V., Offei, S., Asante, I., Egesi, C., Kulakow, P., et al. (2015). Parent-offspring regression analysis for total carotenoids and some agronomic traits in cassava. *Euphytica* 206, 657–666. doi: 10.1007/s10681-015-1482-4

Ogbonna, A. C., De Andrade, L. R. B., Rabbi, I. Y., Mueller, L. A., De Oliveira, E. J., and Bauchet, G. J. (2020). Genetic architecture and gene mapping of cyanide in cassava (Manihot esculenta crantz.). *BioRxiv.* doi: 10.1101/2020.06.19.159160

Okechukwu, R. U., and Dixon, A. G. O. (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, 181–208. doi: 10.1080/15427520802212506

Okeke, U. G., Akdemir, D., Rabbi, I., Kulakow, P., and Jannink, J.-L. (2017). Accuracies of univariate and multivariate genomic prediction models in African cassava. *Genet. Sel Evol.* 49, 88. doi: 10.1186/s12711-017-0361-y

Okogbenin, E., Egesi, C. N., Olasanmi, B., Ogundapo, O., Kahya, S., Hurtado, P., et al. (2012). Molecular marker analysis and validation of resistance to cassava

mosaic disease in elite cassava genotypes in Nigeria. *Crop Sci.* 52, 2576–2586. doi: 10.2135/cropsci2011.11.0586

Parmar, A., Sturm, B., and Hensel, O. (2017). Crops that feed the world: Production and improvement of cassava for food, feed, and industrial uses. *Food Secur* 9, 907–927. doi: 10.1007/s12571-017-0717-8

Pfeiffer, W. H., and McClafferty, B. (2007). HarvestPlus: breeding crops for better nutrition. *Crop Sci.* 47, S–88. doi: 10.2135/cropsci2007.09.0020IPBS

Phumichai, C., Aiemnaka, P., Nathaisong, P., Hunsawattanakul, S., Fungfoo, P., Rojanaridpiched, C., et al. (2022). Genome-wide association mapping and genomic prediction of yield-related traits and starch pasting properties in cassava. *Theor. Appl. Genet.* 135, 145–171. doi: 10.1007/s00122-021-03956-2

Platten, J. D., Cobb, J. N., and Zantua, R. E. (2019). Criteria for evaluating molecular markers: Comprehensive quality metrics to improve marker-assisted selection. *PloS One* 14, e0210529. doi: 10.1371/journal.pone.0210529

Rabbi, I., Hamblin, M., Gedil, M., Kulakow, P., Ferguson, M., Ikpan, A. S., et al. (2014). Genetic mapping using genotyping-by-sequencing in the clonally propagated cassava. *Crop Sci.* 54, 1384–1396. doi: 10.2135/cropsci2013.07.0482

Rabbi, I. Y., Kayondo, S. I., Bauchet, G., Yusuf, M., Aghogho, C. I., Ogunpaimo, K., et al. (2020). 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, 195–213. doi: 10.1101/2020.04.25.061440

Rabbi, I. Y., Udoh, L. I., Wolfe, M., Parkes, E. Y., Gedil, M. A., Dixon, A., et al. (2017). Genome-wide association mapping of correlated traits in cassava: Dry matter and total carotenoid content. *The Plant Genome* 10. doi: 10.3835/plantgenome2016.09.0094

R Development Core Team (2020). *R: A language and environment for statistical computing* (Vienna, Austria: R Foundation for Statistical Computing).

Refaeilzadeh, P., Tang, L., and Liu, H. (2009). Cross-validation. *Encycl Database Syst.* 5, 532–538. doi: 10.1007/978-0-387-39940-9_565

Saltzman, A., Birol, E., Bouis, H. E., Boy, E., Moura, F. F. D., Islam, Y., et al. (2013). Biofortification: Progress toward a more nourishing future. *Glob Food Secur* 2, 9–17. doi: 10.1016/j.gfs.2012.12.003

Sánchez, T., Ceballos, H., Dufour, D., Ortiz, D., Morante, N., Calle, F., et al. (2014). Prediction of carotenoids, cyanide and dry matter contents in fresh cassava root using NIRS and hunter color techniques. *Food Chem.* 151, 444–451. doi: 10.1016/j.foodchem.2013.11.081

Sánchez, T., Chávez, A., Ceballos, H., Rodriguez-Amaya, D., Nestel, P., and Ishitani, M. (2006). Reduction or delay of post-harvest physiological deterioration in cassava roots with higher carotenoid content. *J. Sci. Food Agric.* 86, 634–639. doi: 10.1002/jsfa.2371

Sayre, R., Beeching, J. R., Cahoon, E. B., Egesi, C., Fauquet, C., Fellman, J., et al. (2011). The BioCassava plus program: biofortification of cassava for sub-Saharan Africa. *Annu. Rev. Plant Biol.* 62, 251–272. doi: 10.1146/annurev-arplant-042110-103751

Semagn, K., Babu, R., Hearne, S., and Olsen, M. (2014). Single nucleotide polymorphism genotyping using kompetitive allele specific PCR (KASP): overview of the technology and its application in crop improvement. *Mol. Breed* 33, 1–14. doi: 10.1007/s11032-013-9917-x

Stephenson, K., Amthor, R., Mallowa, S., Nungo, R., Maziya-Dixon, B., Gichuki, S., et al. (2010). Consuming cassava as a staple food places children 2-5 years old at risk for inadequate protein intake, an observational study in Kenya and Nigeria. *Nutr. J.* 9, 9. doi: 10.1186/1475-2891-9-9

Talsma, E. F., Melse-Boonstra, A., de Kok, B. P. H., Mbera, G. N. K., Mwangi, A. M., and Brouwer, I. D. (2013). Biofortified cassava with pro-vitamin a is sensory and culturally acceptable for consumption by primary school children in Kenya. *PloS One* 8, 1–8. doi: 10.1371/journal.pone.0073433

Uchendu, K., Njoku, D. N., Paterne, A., Rabbi, I. Y., Dzidzienyo, D., Tongoona, P., et al. (2021). Genome-wide association study of root mealiness and other texture-associated traits in cassava. *Front. Plant Sci.* 12, 770434. doi: 10.3389/fpls.2021.770434

Udoh, L. I., Gedil, M., Parkes, E. Y., Kulakow, P., Adesoye, A., Nwuba, C., et al. (2017). Candidate gene sequencing and validation of SNP markers linked to carotenoid content in cassava (Manihot esculenta crantz). *Mol. Breed* 37, 123. doi: 10.1007/s11032-017-0718-5

Wani, F. J., Rizvi, S. E. H., Sharma, M. K., and Bhat, M. I. J. (2018). A study on cross validation for model selection and estimation. *Int. J. Agric. Sci.* 14, 165–172. doi: 10.15740/HAS/IJAS/14.1/165-172

Welsch, R., Arango, J., Bär, C., Salazar, B., Al-Babili, S., Beltrán, J., et al. (2010). Provitamin a accumulation in cassava (manihot esculenta) roots driven by a single nucleotide polymorphism in a phytoene synthase gene. *Plant Cell* 22, 3348–3356. doi: 10.1105/tpc.110.077560

WHO (2022) *Nutr landsc inf syst*. Available at: https://www.who.int/data/nutrition/nlis/info/vitamin-a-deficiency#:~:text=Deficiency%20of%20vitamin%

20A%20is,outcomes%20of%20pregnancy%20and%20lactation (Accessed January 11, 2021).

Wolfe, M. D., Rabbi, I. Y., Egesi, C., and Hamblin, M. (2016). Genome-wide association and prediction reveals genetic architecture of cassava mosaic disease resistance and prospects for rapid genetic improvement. *The Plant Genome.* 9. doi: 10.3835/plantgenome2015.11.0118

Wong, J. C., Lambert, R. J., Wurtzel, E. T., and Rocheford, T. R. (2004). QTL and candidate genes phytoene synthase and ζ-carotene desaturase associated with the accumulation of carotenoids in maize. *Theor. Appl. Genet.* 108, 349–359. doi: 10.1007/s00122-003-1436-4

Yang, S., Fresnedo-Ramírez, J., Wang, M., Cote, L., Schweitzer, P., Barba, P., et al. (2016). A next-generation marker genotyping platform (AmpSeq) in heterozygous crops: A case study for marker-assisted selection in grapevine. *Hortic. Res.* 3, 16002. doi: 10.1038/hortres.2016.2

Yu, J., and Buckler, E. S. (2006). Genetic association mapping and genome organization of maize. *Curr. Opin. Biotechnol.* 17, 155–160. doi: 10.1016/j.copbio.2006.02.003