

# Genome-Wide Association Mapping of Correlated Traits in Cassava: Dry Matter and Total Carotenoid Content

Ismail Y. Rabbi,\* Lovina I. Udoh, Marnin Wolfe, Elizabeth Y. Parkes, Melaku A. Gedil, Alfred Dixon, Punna Ramu, Jean-Luc Jannink, and Peter Kulakow

## Abstract

Cassava is a starchy root crop cultivated in the tropics for fresh consumption and commercial processing. Primary selection objectives in cassava breeding include dry matter content and micronutrient density, particularly provitamin A carotenoids. These traits are negatively correlated in the African germplasm. This study aimed at identifying genetic markers associated with these traits and uncovering whether linkage and/or pleiotropy were responsible for observed negative correlation. A genome-wide association mapping using 672 clones genotyped at 72,279 single nucleotide polymorphism (SNP) loci was performed. Root yellowness was used indirectly to assess variation in carotenoid content. Two major loci for root yellowness were identified on chromosome 1 at positions 24.1 and 30.5 Mbp. A single locus for dry matter content that colocalized with the 24.1 Mbp peak for carotenoids was identified. Haplotypes at these loci explained 70 and 37% of the phenotypic variability for root yellowness and dry matter content, respectively. Evidence of megabase-scale linkage disequilibrium (LD) around the major loci of the two traits and detection of the major dry matter locus in independent analysis for the white- and yellow-root subpopulations suggests that physical linkage rather than pleiotropy is more likely to be the cause of the negative correlation between the target traits. Moreover, candidate genes for carotenoid (*phytoene synthase*) and starch biosynthesis (*UDP-glucose pyrophosphorylase* and *sucrose synthase*) occurred in the vicinity of the identified locus at 24.1 Mbp. These findings elucidate the genetic architecture of carotenoids and dry matter in cassava and provide an opportunity to accelerate breeding of these traits.

## Core Ideas

- Cassava, a starchy root crop, is a major source of dietary calories in the tropics.
- Most varieties consumed are poor in micronutrients, including pro-vitamin A.
- Dry matter and carotenoid content are governed by few major loci on chromosome 1.
- Genetic linkage, rather than pleiotropy, is the most likely cause of their negative correlation.

**C**ASSAVA (*Manihot esculenta* Crantz) is one of the most important food and feed crops in the tropics, and Africa accounts for more than half of the total worldwide production of 270.3 million tonnes (<http://faostat3.fao.org/>, accessed 26 Mar. 2016, verified 28 June 2017). Because of its remarkable tolerance to drought (El-Sharkawy, 1993), its ability to grow in poor soils (Cock, 1982), and its

I.Y. Rabbi, L.I. Udoh, E.Y. Parkes, M.A. Gedil, A. Dixon, and P. Kulakow, International Institute of Tropical Agriculture (IITA), PMB 5320 Ibadan, Oyo, Nigeria; M. Wolfe and J.-L. Jannink, Dep. of Plant Breeding and Genetics, Cornell Univ., Ithaca, NY 14853; P. Ramu, Institute of Genomic Diversity, Cornell Univ., Ithaca, NY 14853; J.-L. Jannink, USDA-ARS, R.W. Holley Center for Agriculture and Health, Ithaca, NY 14853. Received 15 Sept. 2016. Accepted 12 June 2017. \*Corresponding author (I.Rabbi@cgiar.org).

**Abbreviations:** BLUP, best linear unbiased predictor; GBS, genotyping-by-sequencing; GLM, general linear model; GWAS, genome-wide association studies; LD, linkage disequilibrium; MAS, marker-assisted selection; MLM, mixed-linear model; PC, principal component; QTL, quantitative trait locus; SNP, single nucleotide polymorphism; TMS, tropical *Manihot* selection; UDP, uridine diphosphate.

Published in Plant Genome  
Volume 10. doi: 10.3835/plantgenome2016.09.0094

© Crop Science Society of America  
5585 Guilford Rd., Madison, WI 53711 USA  
This is an open access article distributed under the CC BY license  
(<https://creativecommons.org/licenses/by/4.0/>).

perennial nature which allows it to be harvested as and when needed, this heterozygous and clonally propagated species plays a particularly important role in food security for millions of small-holder farmers in developing countries. Moreover, cassava is increasingly being cultivated for commercial processing to convert its storage roots into dehydrated chips, flour, and starch (Balagopalan, 2002). Dry matter content, of which a large proportion is starch, is therefore a primary factor that defines adoption of new cassava varieties by farmers and the market value of harvested roots (Okechukwu and Dixon, 2008). As a result, breeding of improved varieties with high dry matter content is one of the primary objectives of cassava genetic improvement programs in the world.

Another important target trait for cassava improvement in developing countries is biofortification of micronutrients (Pfeiffer and McClafferty, 2007; Saltzman et al., 2013). Most varieties grown and consumed throughout the world have white storage roots with negligible amounts of micronutrients in general, and provitamin A in particular (Welsch et al., 2010). Dietary diversification and breeding of farmer-preferred improved varieties with higher nutritional density are complementary approaches used in addressing potential micronutrient deficiency associated with consumption of cassava as the sole staple food (Sayre et al., 2011). The crop's gene pool exhibits considerable natural variation for storage root carotenoids that can be tapped for breeding of biofortified varieties, with some breeding populations reported to accumulate as much as 25.8  $\mu\text{g/g}$  fresh root weight (Ceballos et al., 2013; Sánchez et al., 2014).

Despite availability of natural genetic diversity in the global germplasm that is relevant to breeding for increased dry matter and total carotenoid contents, improving these traits through phenotype-based recurrent selection is a lengthy process due to the breeding complexities associated with the species, including an annual cropping cycle of 12 to 24 mo and low multiplication rate of planting materials. Understanding the genetic basis of variation in these traits is essential for increasing their selection efficiency and the rate of genetic gain. More importantly, several studies using diverse germplasm, particularly from Africa, have reported that dry matter and carotenoid content are negatively correlated, with  $r$  values ranging from  $-0.1$  to  $-0.5$  (Marín Colorado et al., 2009; Akinwale et al., 2010; Esuma et al., 2012; Njoku et al., 2015). On the contrary, the two traits are considered to be independent from each other in Latin American cassava (Ceballos et al., 2013; Sánchez et al., 2014). Despite its significant implication in breeding, it is currently not known whether the genetic basis of this correlation is due to genetic linkage or pleiotropy.

Several mapping studies using either bulk segregant analysis or quantitative trait loci (QTL) mapping of S1 or F1 populations have been reported separately for dry matter and carotenoid content (Balyejusa Kizito et al., 2007; Marín Colorado et al., 2009; Welsch et al., 2010; Morillo et al., 2013; Njoku et al., 2014). The mapping

resolution from single-cross experimental populations is expected to be limited due to the use of sparse genetic maps and the limited number of recombination events observed (Hamblin et al., 2011). Moreover, QTLs from such biparental populations may not provide insight into the tremendous genetic and phenotypic variation of the larger cassava gene pool (Zhao et al., 2011). The increased availability of genomic resources for cassava, including the chromosome-scale reference genome and integrated linkage map (Prochnik et al., 2012; International Cassava Genetic Map Consortium, 2015) and high-density genotyping using next-generation sequencing (Rabbi et al., 2014a, 2014b) makes it possible to use genome-wide association studies (GWAS) to dissect the phenotypic diversity of cassava germplasm with respect to dry matter and carotenoid content. GWAS, which takes advantage of natural LD generated by ancestral mutation, drift, and recombination events in diverse germplasm, offers the possibility to overcome the shortcomings of traditional biparental QTL mapping. These advantages allow GWAS to reveal a broader spectrum of trait-linked allelic variation and thus may provide the most useful markers for marker-assisted selection (MAS). Indeed, GWAS has already been applied in other crops such as maize to study the genetic architecture of carotenoid accumulation (Harjes et al., 2008; Owens et al., 2014; Suwarno et al., 2015). In cassava, Esuma et al. (2016) performed a GWAS study using a panel of partial S1 and S2 generation inbreds produced from eight clones. Using this limited number of parents, they reported a single genomic region on chromosome 1 that is associated with the variation in total carotenoid content. However, no joint association analysis examining carotenoids and dry matter content has hitherto been reported. Here, we present the results of a GWAS using a collection of 672 genotyped cassava clones representing diverse African germplasm genotyped at high-density using genotyping-by-sequencing (Elshire et al., 2011). The population was phenotyped in two locations for three consecutive field seasons. The results of this study will be used to develop efficient strategies to breed for varieties with high dry matter and increased provitamin A content.

## Methods

### Germplasm

The present work was performed using the tropical *Manihot* selection (TMS) cultivars developed at the International Institute of Tropical Agriculture (IITA) in Nigeria (Supplementary Table 1). This population, also known as the Genetic Gain collection, consists of advanced breeding lines and key landraces selected since 1970 (Okechukwu and Dixon, 2008; Ly et al., 2013). The pedigree of the collection is mainly composed of crosses between germplasm from West Africa and early introductions of cassava mosaic disease tolerant clones arising from interspecific hybridization between *Manihot glaziovii* and cultivated cassava at the Amani station in Tanzania (Hahn et al.,

1980). The collection also includes hybrid germplasm from Latin America (Wolfe et al., 2016).

## Locations and Experimental Design

The Genetic Gain population was planted using an incomplete block design with two checks per block and single row of either 5 or 10 plants. Spacing between plants within rows and between rows was 1 m. Data used for this study was collected in the 2012–2013, 2013–2014, and 2014–2015 field seasons in Ibadan (7°24' N, 3°54' E) and Ubiaja (6°39' N, 6°22' E). The trials are usually planted in June, in the middle of the raining season in Southwest Nigeria, and harvested in June of the following year.

## Assessment of Dry Matter Content and Yellow Color Intensity of Storage Roots

Dry matter content was assessed using the oven-drying method. Eight fully developed roots were randomly selected from each plot, peeled, chipped, and thoroughly mixed. For each sample, 100 g was weighed and oven-dried for 48 h at 104°C. The samples were then reweighed, and the dry matter content was expressed as the percentage of dry weight relative to fresh weight.

Because of the well-established linear relationship between intensity of yellow color and carotenoid content in cassava storage roots (Pearson's coefficient,  $r$ , ranges from 0.81 to 0.89), we used root yellowness as an indirect measure of carotenoid content (Iglesias et al., 1997; Chávez et al., 2005; Marín Colorado et al., 2009; Akinwale et al., 2010; Sánchez et al., 2014). The relative difference among clones in the Genetic Gain population was assessed using two complementary methods. The first was a visual gradation of yellow color using a standard color chart (Supplementary Fig. 1) starting from 1 (white) to 7 (deep yellow). Due to the potential subjectivity inherent in visual color scores, we complemented the color chart method through the use of a Minolta CR-410 chromameter (Konica Minolta Sensing Americas, Ramsey, NJ). Approximately 100 g of grated samples from freshly peeled roots were placed in transparent Nasco (Fort Atkinson, WI) Whirl-Pak sampling bags and four chromameter measurements were taken in different sections of the bag. We chose the Commission Internationale de l'Éclairage (CIELAB) method that records color values in a three-dimensional color space, where the  $L^*$  coordinate corresponds to a lightness coordinate, and the  $a^*$  coordinate corresponds either to red (positive values) or to green (negative values). Of importance to this study was the  $b^*$  coordinate, whose positive values represent yellow while the negative values represent blue. The illuminant used was D65 and calibration was done each day with a white ceramic.

## SNP Genotyping

DNA was extracted as described in Rabbi et al. (2014a) and genotyping-by-sequencing was performed as described by (Elshire et al., 2011). DNA from each individual in the Genetic Gain collection was digested with *ApeKI*, a

methylation sensitive restriction enzyme that recognizes a five base-pair sequence (GCWGC, where W is either A or T). Unique nucleotide adapters (barcodes) were ligated to the digested DNA and samples were pooled to create 95-plex genotyping-by-sequencing (GBS) libraries which were amplified using polymerase chain reaction. Sequencing was performed using Illumina HiSeq2500. The sequenced reads were assigned to genotypes using the unique barcode sequences and aligned to v. 6.0 of the cassava reference genome ([www.phytozome.org/cassava](http://www.phytozome.org/cassava), verified 29 June 2017) with the Bowtie 2 (Langmead and Salzberg, 2012). The SNPs were discovered using the GBS pipeline v. 2 implemented in TASSEL software (Glauibitz et al., 2014) and converted to dosage format (0 = homozygous reference, 1 = heterozygous, 2 = homozygous nonreference alleles). Missing data were filtered, as described in Wolfe et al. (2016) and imputed with the glmnet algorithm (Wong et al., 2014) in R (R Core Team, 2017).

## Phenotypic Data Analysis

The phenotypic data across two locations and 3 yr was collapsed to single best linear unbiased predictor (BLUP) values for each clone by fitting the following mixed linear model with the *lme4* package in R:

$$y_{ij} = \mu + c_i + \beta_i + c_l \times \beta_i + \varepsilon_{ij}$$

Here,  $y_{ij}$  represents raw phenotypic observations,  $\mu$  is the grand mean,  $c_i$  is a random effects term for genotype with  $c_i \sim N(0, \sigma_1^2)$ ,  $\beta_i$  is a fixed effect for the combination of location and year harvested,  $c_l \times \beta_i$  is a random effect for genotype-by-environment variance, and  $\varepsilon_{ij}$  is the residual variance, assumed to be random and distributed  $N(0, \sigma_e^2)$ . Broad-sense heritability for dry matter content and yellow color intensity was calculated according to Ly et al. (2013). Genetic correlation among traits was also calculated from BLUP values.

## Population Structure and Genome-Wide Association Analyses

Inherent population structure and cryptic relatedness can lead to spurious associations in GWAS (Astle and Balding, 2009). To control for these confounding factors, three standard GWAS models were compared: a simple one-way ANOVA model with no correction (naïve model); a general linear model (GLM) with the first five principal components (PCs) of the SNP matrix as covariates (GLM + 5 PCs); and a mixed-linear model (MLM) using the five PCs and marker-estimated kinship matrix (Yu et al., 2006). The association analyses were implemented in TASSEL (Bradbury et al., 2007; Zhang et al., 2010). Association test P-values were considered significant when more extreme than the Bonferroni threshold with experiment-wise type I error rate of 0.05.

The patterns and extent of LD in a population not only determines the obtainable resolution in association mapping studies (Hamblin et al., 2011), but also has strong implication in the interpretation of association peaks. Therefore, the level of LD decay and the local

**Table 1. Summary of phenotype variation, variance components ( $\pm$ SE) and broad-sense heritability ( $H^2$ ) for dry matter (DM), color chart, and chromameter Commission Internationale de l'Éclairage (CIELAB) readings.†**

Trait‡	DM, %	Color chart	L*	a*	b*
Minimum	8.4	1.0	69.5	-3.0	11.1
Average	28.6	1.7	84.4	-0.3	20.8
Maximum	45.4	7.0	90.2	4.6	40.8
N	3232	4237	1360	1360	1360
$V_G$	16.52 (4.06)	1.11 (1.05)	5.22 (2.29)	0.86 (0.92)	16.52 (4.06)
$V_{G \times E}$	3.53 (1.88)	0.06 (0.24)	0.77 (0.88)	0.16 (0.41)	2.87 (1.69)
$V_E$	1.91 (1.38)	0.01 (0.11)	0.53 (0.73)	0.08 (0.28)	1.11 (1.05)
Residual	10.27 (3.20)	0.17 (0.41)	3.26 (1.80)	0.30 (0.55)	2.71 (1.65)
$H^2$	0.51	0.82	0.53	0.61	0.87

† CIELAB readings: L\*, lightness; a\*, red/green coordinate; b\*, yellow/blue coordinate.

‡ N, number of plots from which data was obtained for the various traits;  $V_G$ , genotype variance;  $V_E$ , environment variance;  $V_{G \times E}$ , genotype-by-environment variance.

patterns of LD along each chromosome were determined by calculating intrachromosomal pairwise squared correlation ( $r^2$ ) using PLINK (Purcell et al., 2007).

## Results

### SNP Genotyping

A total of 72,279 genome-wide SNP markers were called for the 672 Genetic Gain individuals after filtering for minor allele frequency threshold of 0.05. The high-density coverage of SNPs resulted in an average of 4015 markers per chromosome, ranging from 3101 on chromosome 16 to 5880 on chromosome 1.

### Phenotypic Variability

We investigated the phenotypic variation in dry matter content, as well as carotenoid-based intensity of yellow root color using a visual color chart and chromameter. The dry matter content varied widely in the Genetic Gain population, ranging from 8.4 to 45% (average 28.6%, Table 1). About two-thirds of the evaluated clones have white storage roots while the remaining showed a range of yellow color, suggesting varying levels of carotenoid content. On average, the visual score was 1.7 and ranged from 1 (white) to 7 (yellow). The average chromameter measure of yellow color intensity ( $b^*$  value) was 20.8 and ranged from 11.1 (white) to 40.8 (yellow). Dry matter was approximately normally distributed while chromameter  $b^*$  values showed a bimodal distribution (Fig. 1) in which the first peak ( $b^*$  values from 10 to 20) is associated with the white clones, while the second peak ( $b^*$  values from 20 to 40) is associated with the variations among the yellow clones.

Broad-sense heritability was high for root yellow color ( $H^2 = 0.87$  and  $0.82$  for color chart and chromameter  $b^*$  values, respectively) but moderate for dry matter content ( $H^2 = 0.51$ ). These values are within the range of heritabilities reported previously for these traits (Balyejusa Kizito et al., 2007; Ceballos et al., 2013). The relative importance of genotype-by-environment variance ( $V_{G \times E}$ ) compared with genotype ( $V_G$ ) variance was measured by the ratio  $V_{G \times E}/V_G$ . For all traits, the  $V_G$  component was larger than the  $V_{G \times E}$ . The interaction is minimal for the yellow color

measurements (0.054 and 0.173 for color chart and chromameter  $b^*$  values, respectively). For dry matter content, we observed a slightly higher interaction ratio of 0.214.

The BLUPs for dry matter content and gradation of yellow color were negatively correlated in our germplasm collection (Pearson's correlation coefficient,  $r = -0.59$ ;  $P$ -value  $< 0.0001$ ), indicating that clones with higher carotenoid content are more likely to have low dry matter content (Fig. 2), which confirms previous findings in cassava (Marín Colorado et al., 2009; Akinwale et al., 2010; Esuma et al., 2012; Njoku et al., 2015). On the other hand, we found a positive association between dry matter content and color lightness (chromameter  $L^*$  value,  $r = 0.60$ ,  $P$ -value  $< 0.0001$ ). The two measures of yellow color (i.e., color chart and chromameter  $b^*$  axis) were strongly correlated ( $r = 0.96$ ,  $P$ -value  $< 0.0001$ ).

### Population Structure

Analysis of population structure in 672 accessions genotyped across 72,279 SNPs using PC analysis detected subtle genetic differentiation in the Genetic Gain collection, with the first 10 PCs explaining about 23% of the genetic variation. The first two PCs, which accounted for 8% of the genetic variation, revealed genetic differentiation between white and yellow-root clones along the PC2 axis (Fig. 3).

### Linkage Disequilibrium

Several regions of extensive megabase-scale LD were discovered in chromosomes 1, 4, and 10, as well as smaller regions in other chromosomes (Fig. 4). Excluding results from chromosomes with large LD blocks (i.e., chromosomes 1, 4, and 10), we found that, on average, LD drops almost to background levels ( $r^2 < 0.1$ ) at around 2 Mb in the Genetic Gain population (Fig. 5).

### Population-Wide GWAS

#### Variation in Carotenoid Content Estimated by Root Yellowness

Of the three standard GWAS models tested, the model correcting for kinship and 5 PCs had the lowest inflation factors as determined from quartile-quartile (QQ) plots,

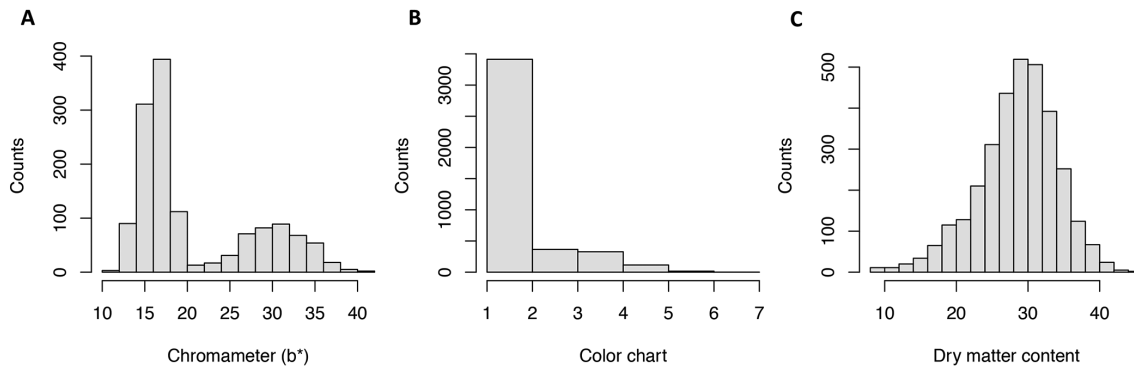


Fig. 1. Distribution of phenotype for chromameter ( $b^*$ ), color chart, and dry matter content.

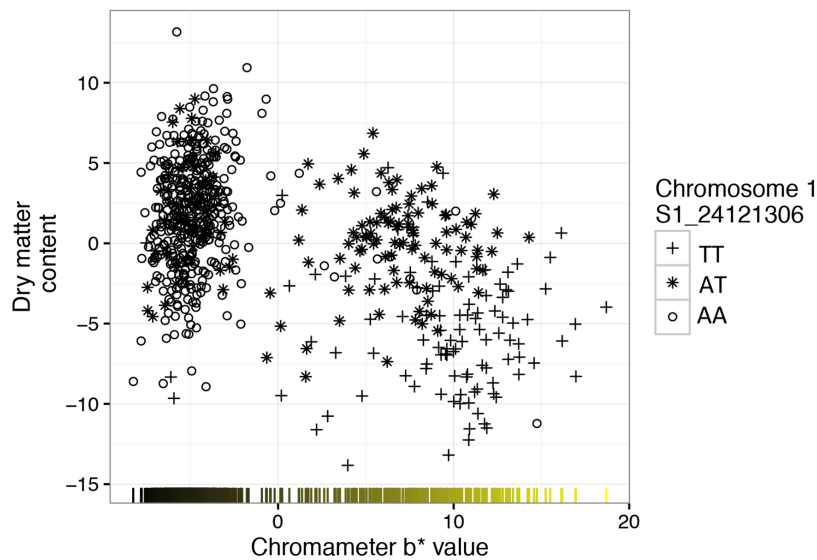


Fig. 2. Relationship between dry matter content and root yellowness best linear unbiased predictors (BLUPs, expressed as  $b^*$  value of chromameter measurement). Different symbols denote the genotype at marker S1\_24121306 that is associated with both dry matter content and root color intensity.

and therefore had the lowest false-discovery rate (Supplementary Fig. 2). The MLM-based GWAS analysis for yellow color in the storage root parenchyma using both the color chart and the chromameter-based methods uncovered the same major association regions occurring at 24.1 and 30.5 Mbp of chromosome 1 (Fig. 6). This is not surprising, given the high correlation between the two color assessment methods. The first major peak is tagged by marker S1\_24121306 for visual gradation of color ( $P$ -value =  $1.74 \times 10^{-22}$ ) and marker S1\_24159585 ( $P$ -value =  $1.79 \times 10^{-19}$ ) for chromameter  $b^*$  value (Table 2). The second peak was tagged by the same marker, S1\_30543382, for both measures of color peak intensity and occurred 6.5 Mbp away from the first peak ( $P$ -value of  $2.884 \times 10^{-11}$  and  $1.659 \times 10^{-11}$ , for color chart and chromameter, respectively). All other SNPs between the two major regions were not significant at the Bonferroni significance threshold ( $P = 6.92 \times 10^{-7}$ ; Fig. 7). The LD between SNPs S1\_24121306 and S1\_30543382 was 0.3, suggesting moderate nonrandom segregation of alleles of the two markers in these regions.

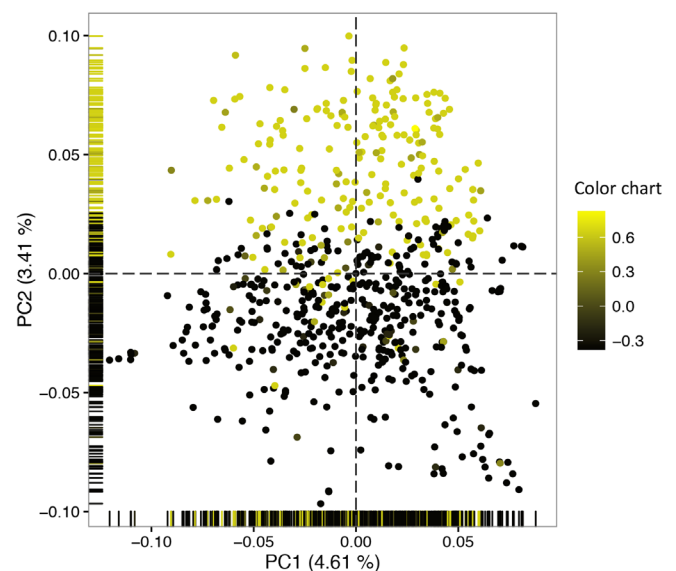


Fig. 3. Population structure of the Genetic Gain collection as revealed by the first two axes of principal component analysis (PCA) biplot. Yellow color highlights accessions with yellow roots.

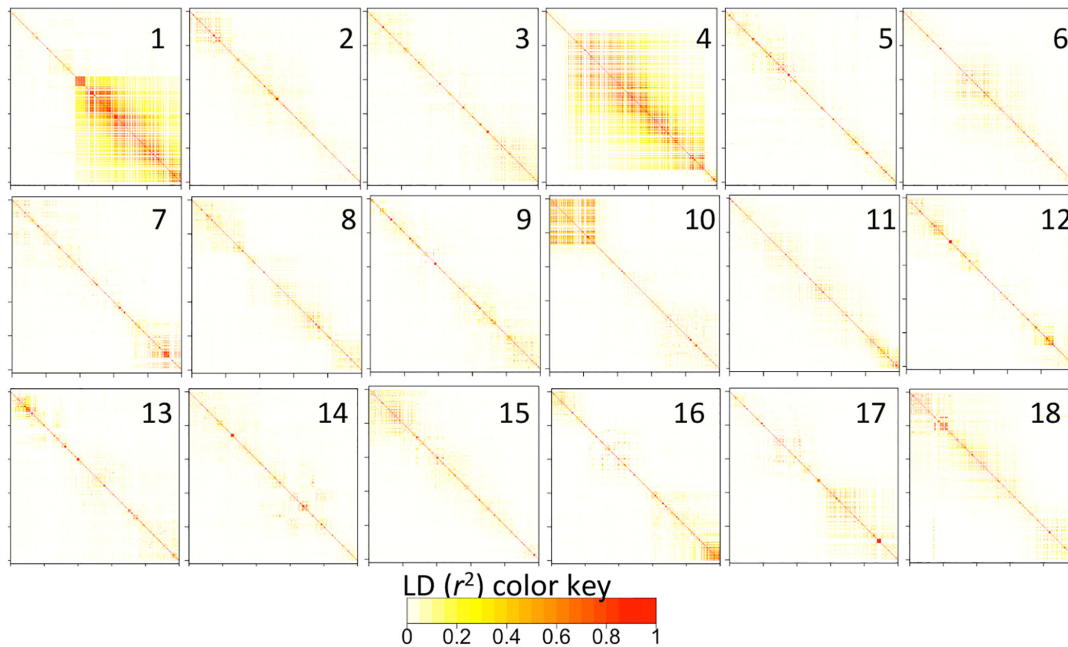
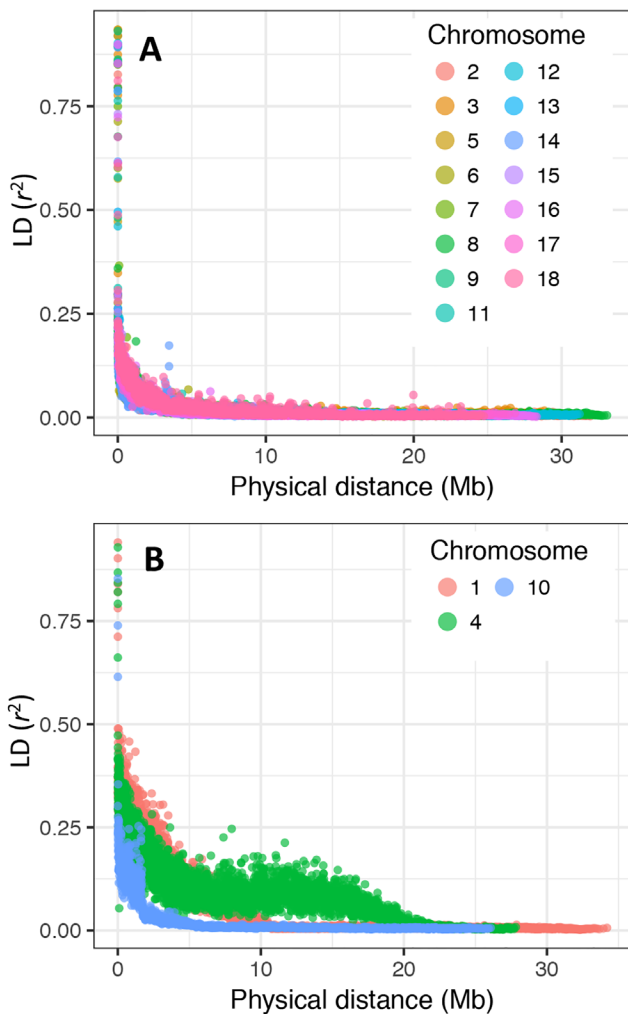


Fig. 4. Local pattern of linkage disequilibrium ( $r^2$ ) along each of the 18 cassava chromosomes. Note the large linkage disequilibrium blocks in chromosomes 1, 4, and 10. Single nucleotide polymorphisms are arranged according to their order and not their physical position.



### Dry Matter Content

Genetic variation in dry matter content was found to be associated with a major locus occurring at 24.1 Mbp region of chromosome 1 and tagged by marker S1\_24121306 ( $P$ -value of  $1.862 \times 10^{-12}$ ). Importantly, this locus for dry matter content colocalizes with one of the two peaks found to be associated with carotenoid content (Fig. 7). The genomic colocalization of the major loci for dry matter content and root yellowness suggests either a strong physical linkage between the genes underlying these important traits, or a pleiotropic effect. Distinguishing between these two possible causes is important in cassava improvement efforts that target both traits. We therefore attempted to unravel the genetic cause of the observed association by (i) exploring the underlying LD patterns in the QTL regions on chromosome 1; (ii) searching for plausible biological explanation by identifying candidate genes for both traits in the target region; and (iii) carrying out independent association analysis for dry matter content within the white root and yellow root subpopulations. We would expect that the GWAS peak for dry matter content to disappear in the white root subpopulation if the main cause of the negative correlation between the two traits is pleiotropy.

Exploration of the LD landscape along chromosome 1 uncovered a mega-base-scale region of low recombination extending from 22 to 33 Mb surrounding the association peaks for dry matter content and yellow color (Fig. 7). This region was recently shown to coincide with a large

Fig. 5 (left). A Moving-average based linkage disequilibrium (LD) decay profile in the Genetic Gain population. (A) All chromosomes except 1, 4, and 10; (B) chromosomes 1, 4, and 10.

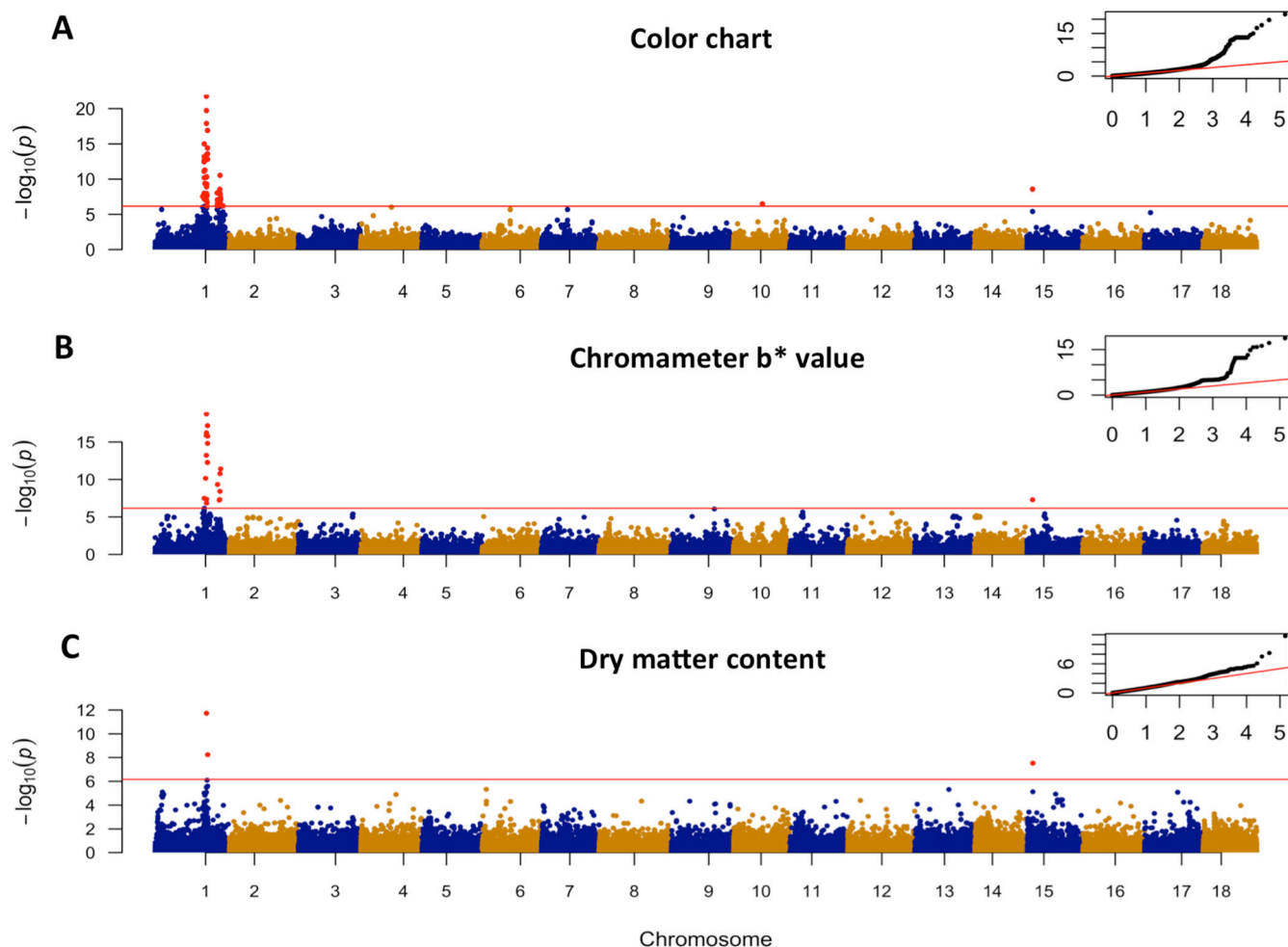


Fig. 6. Genome-wide association results. Manhattan and quantile-quantile plot of the mixed-linear model (MLM) for root yellowness estimated using (A) chromameter b\* value, (B) color chart method, and (C) dry matter content. The red horizontal line indicates the genome-wide significance threshold.

**Table 2. Summary of significant associations between selected traits and SNP markers from the mixed-linear model (MLM) analysis. Only results in the major loci from chromosome 1 are shown.**

Trait	SNP	Allele	Position, bp	P-value	Candidate genes and mid-position
Color chart	S1_24121306	A/T†	24,121,306	$1.74 \times 10^{-22}$	Phytoene synthase (Manes.01G124200; 24,155,070 bp)
Color chart	S1_30543382	A†/G	30,543,382	$2.91 \times 10^{-11}$	NA‡
Chromameter b*	S1_24159585	C/T†	24,159,585	$1.79 \times 10^{-19}$	Phytoene synthase (Manes.01G124200; 24,155,070 bp)
Chromameter b*	S1_30543382	A†/G	30,543,382	$1.66 \times 10^{-11}$	NA
Dry matter	S1_24121306	A†/T	24,121,306	$1.86 \times 10^{-12}$	UDP-glucose pyrophosphorylase (Manes.01G123000; 24,061,652 bp); sucrose synthase (Manes.01G123800; 24,142,314 bp)

† Nucleotide linked to the favorable allele at each SNP.

‡ NA, not available; no known candidate gene was found around the locus.

*Manihot glaziovii* introgression segment (Bredeson et al., 2016) that traces back to early breeding for resistance to cassava mosaic and cassava brown streak viruses in the 1930s (Hahn et al., 1980). Clustering of the Genetic Gain population based on identity-by-descent relationship (i.e., a measure of how many alleles at any marker in each of the two samples came from the same ancestral chromosomes) calculated using only markers from this extensive LD region (2150 SNPs from markers S1\_21567540 to

S1\_34950326) revealed at least two major groups of accessions (Supplementary Fig. 3), indicating presence of few major haplotypes associated with the LD blocks.

### GWAS for Dry Matter Content in White Root and Yellow Root Subpopulation

If the phenotypic association between dry matter and carotenoid contents and the collocation of their association signals (~24.1 Mbp region) is largely caused by

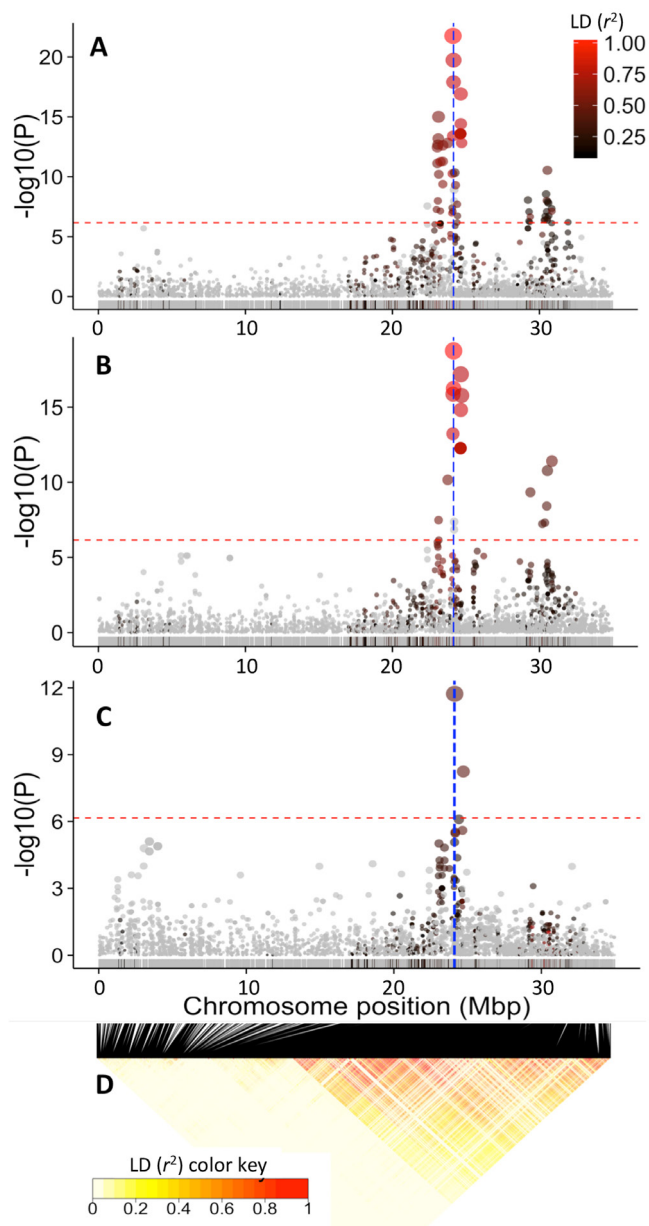


Fig. 7. Genome-wide association studies (GWAS) results for chromosome 1. Manhattan plot of the mixed-linear model for: (A) root yellowness as measured by chromameter  $b^*$  value, (B) color chart, (C) dry matter content, and (D) linkage disequilibrium (LD) heatmap showing pairwise squared correlation of alleles between markers along chromosome 1. Note the common peak at ~24.1 Mbp region for the three traits. Red horizontal line indicates the genome-wide significance threshold. The SNPs are colored according to their degree of linkage disequilibrium ( $r^2$ ) with the leading variant (i.e., top SNP for the first peak at 24.1 Mbp). The vertical blue line in (A) and (B) denote the position of the carotenoid biosynthesis gene, phytoene synthase (24,155,070 bp), and the two adjacent blue lines in (C) denotes the positions of the UDP-glucose pyrophosphorylase (24,061,652 bp) and sucrose synthase (24,142,314 bp).

physical linkage rather than by pleiotropy, the major dry matter locus should be detectable in both white root and yellow root germplasm when analyzed independently. We therefore split the Genetic Gain dataset into white

root ( $n = 427$ ) and yellow root ( $n = 210$ ) subpopulations and repeated the GWAS analysis. Clones that were at the borderline between yellow-root and white-root were excluded from these analyses. To mitigate the loss of power as a result of double-fitting markers in the MLM both as a fixed effect tested for association and as a random effect as part of the kinship (Lippert et al., 2011; Listgarten et al., 2012), the MLM analysis was performed using a kinship matrix calculated excluding markers from chromosome 1.

We recovered the major dry matter content association signal in both the white root and yellow-root subpopulation (Fig. 8). Though coinciding with the locus identified in the population-wide GWAS, the association signal in the white subpopulation was much broader, extending from 24 to 33 Mbp and generally overlaps with the broad LD region of the chromosome 1 (Fig. 8). Conversely, association signal for the yellow subpopulation was relatively narrow. A survey of the underlying LD pattern in the same chromosome region for the yellow subpopulation showed a recombination spot. This indicates normal recombination occurs in the region in this subpopulation and suggests that it is possible to break the linkage between PSY2 and the starch biosynthesis gene.

### Selection Sweep Associated with Breeding for Yellow-Root Varieties

To determine whether the breeding for carotenoid content trait in the Genetic Gain germplasm resulted in a selection sweep around the major QTL region, we quantified genome-wide nucleotide variation in the yellow root subpopulation ( $n = 210$ ) and the non-yellow subpopulation ( $n = 427$ ). A sliding-window scan of expected heterozygosity ( $\pi$ ) and Tajima's D detected a ~6 Mb region with decrease in nucleotide diversity in the yellow compared to white-root subpopulation around the first major carotenoid locus site (~24.1 Mbp) relative to its chromosomal neighborhood (Fig. 9).

### Proportion of Variance Explained by Markers QTL Haplotypes

To determine predictive ability of the discovered loci for yellow color intensity and dry matter content, we performed a multiple linear regression analysis using the *lm* function in R and considered the top markers for these traits as independent variables and the traits measurements as the response variables. A model considering the two major peaks associated with gradation of yellow color as assessed using a color chart (S1\_24121306 and S1\_30543382) returned an adjusted squared correlation ( $R^2$ ) of 0.81. For the measure of continuous variation in intensity of yellow color using chromameter ( $b^*$  value), the adjusted  $R^2$  from same genomic regions (S1\_24159585 and S1\_30543382) was 0.70 while that for dry matter content was moderate ( $R^2 = 0.37$ ). This finding suggests that the major loci on chromosome 1 would be useful in MAS breeding in cassava. Single or joint allelic substitution effects at the associated loci with



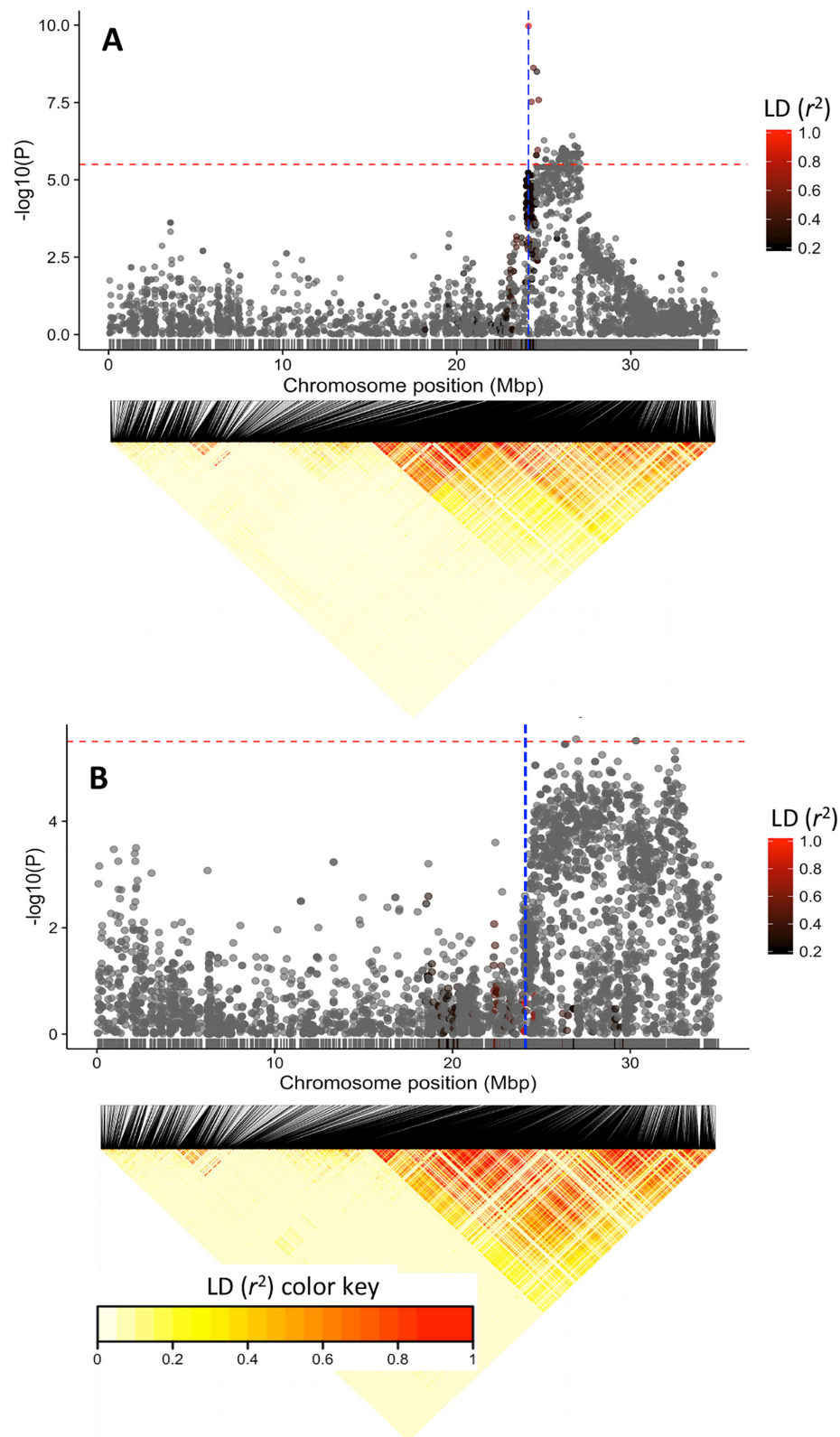


Fig. 8. Manhattan plots of the mixed-linear model (MLM) analysis of the yellow root (top) and white root subpopulations (bottom). Below each is an linkage disequilibrium (LD) heatmap showing pairwise squared correlation of alleles between markers along chromosome 1 for each subpopulation. Note the large number of single nucleotide polymorphisms (SNPs) showing significant association with dry matter in the white subpopulation compared to that of the yellow subpopulation. Red horizontal line indicates the genome-wide significance threshold. The vertical blue line in (A) denotes the position of the carotenoid biosynthesis gene, phytoene synthase (24,155,070 bp), and the two adjacent blue lines in (B) denote the positions of the UDP-glucose pyrophosphorylase (24,061,652 bp) and sucrose synthase (24,142,314 bp).

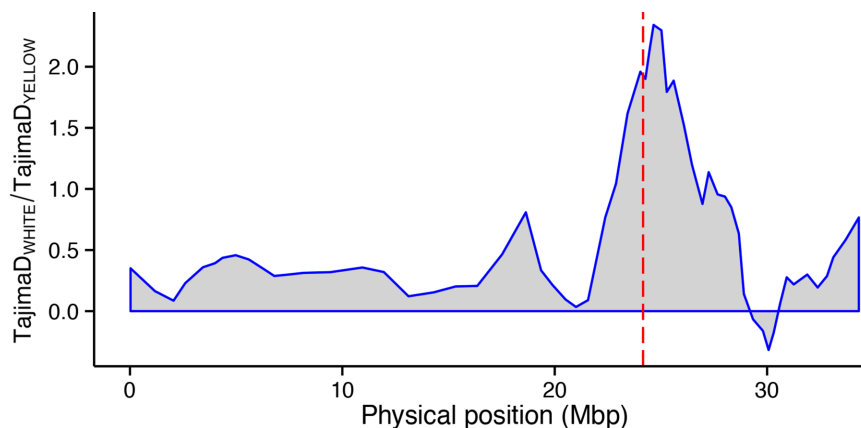


Fig. 9. Selection sweep associated with positive selection for provitamin A trait varieties. The red dashed line indicates the position of SNP S1\_24121306.

respect to chromameter  $b^*$  value, color chart, and dry matter content is shown in Fig. 10.

### Candidate Genes

The first of the two genomic regions associated with color intensity (tagged by SNP S1\_24159585) was found ~4.5 Kbp away from phytoene synthase 2 (*PSY2*) in the cassava v. 6 reference genome. The *PSY2* enzyme, named Manes.01G124200 and located at 24,155,070 bp, is involved in the first dedicated step of the carotenoid biosynthesis pathway in cassava roots, which converts geranylgeranyl diphosphate to phytoene (Welsch et al., 2010). Presence of the null versus the functional *PSY2* allele is responsible for the qualitative color difference between the white and the yellow roots, respectively (Welsch et al., 2010; Rabbi et al., 2014a). Our study suggests that allelic variation associated with increases in enzyme activity could contribute to deeper yellow by increasing the flux into the pathway. No known candidate genes were found in the vicinity of the second significant association signal on chromosome 1 occurring at 30.5 Mbp.

For dry matter content, we found two particular genes that are pivotal in central carbon metabolism in the vicinity of the top SNP linked to the trait. The first is uridine diphosphate (UDP)-glucose pyrophosphorylase (named Manes.01G123000 in the cassava reference genome). This gene, which occurs in the 24.06 Mbp region, plays a key role in carbohydrate metabolism and is strongly associated with the yield production both in grains and root crops (Smith, 2008; Zeeman et al., 2010). UDP-glucose pyrophosphorylase was recently found to be up-regulated during cassava storage root bulking (Yang et al., 2011; Wang et al., 2016). The second key carbohydrate metabolism gene was sucrose synthase (named Manes.01G123800), which occurs in the 24.14 Mbp region. Identifying these independent, potential candidate genes for carotenoid and carbohydrate biosynthesis strongly favors the possibility that the association between these two traits is caused by physical linkage rather than pleiotropy. This hypothesis warrants further investigation.

### Discussion and Conclusion

The present study revealed that the genetic architecture for dry matter content and intensity of yellow color resulting from carotenoid accumulation in cassava roots is governed by few major loci on chromosome 1 and explains the large repeatability estimates, particularly for yellow color. These findings expand on those from previous genetic mapping efforts for dry matter and carotenoid content. Using a candidate gene mapping approach, Welsch et al. (2010) reported that a SNP mutation in the *PSY2* gene, leading to amino-acid substitution, differentiates white and yellow cassava storage roots. Similarly, a biparental QTL mapping study that used two clones from the Genetic Gain collection (TMS-I961089A and TMEB117) also uncovered a single QTL peak whose confidence interval encompassed the same *PSY2* gene (Rabbi et al., 2014a). The F1 progeny from the TMS-I961089A × TMEB117 population, also genotyped using the GBS method, segregated at an approximately 1:1 ratio for white versus light-yellow roots, suggesting that the yellow-root parent was heterozygous for the functional allele at the *PSY2* locus. Kizito et al. (2007) reported a QTL for dry matter content in a biparental population genotyped using SSR markers that also corresponds to this region on chromosome 1. More recently, Esuma et al. (2016) reported a single genomic region on chromosome 1 underlies the variation in total carotenoid content in eight S1 and S2 partially inbred families. This peak, around 24.66 Mbp, is close to our first locus tagged by SNP S1\_24121306. However, that study did not look at genetic architecture for dry matter content.

While the amount of total carotenoids in the Genetic Gain collection was not directly estimated, previous studies of diverse cassava germplasm have consistently reported a strong linear relationship between yellow color and carotenoid content (Pearson's coefficient,  $r$ , ranging from 0.81 to 0.89; Iglesias et al., 1997; Chávez et al., 2005; Marín Colorado et al., 2009; Akinwale et al., 2010; Sánchez et al., 2014). Hence, the results obtained here should be useful for breeding efforts targeting breeding for improved carotenoid content. Nevertheless,

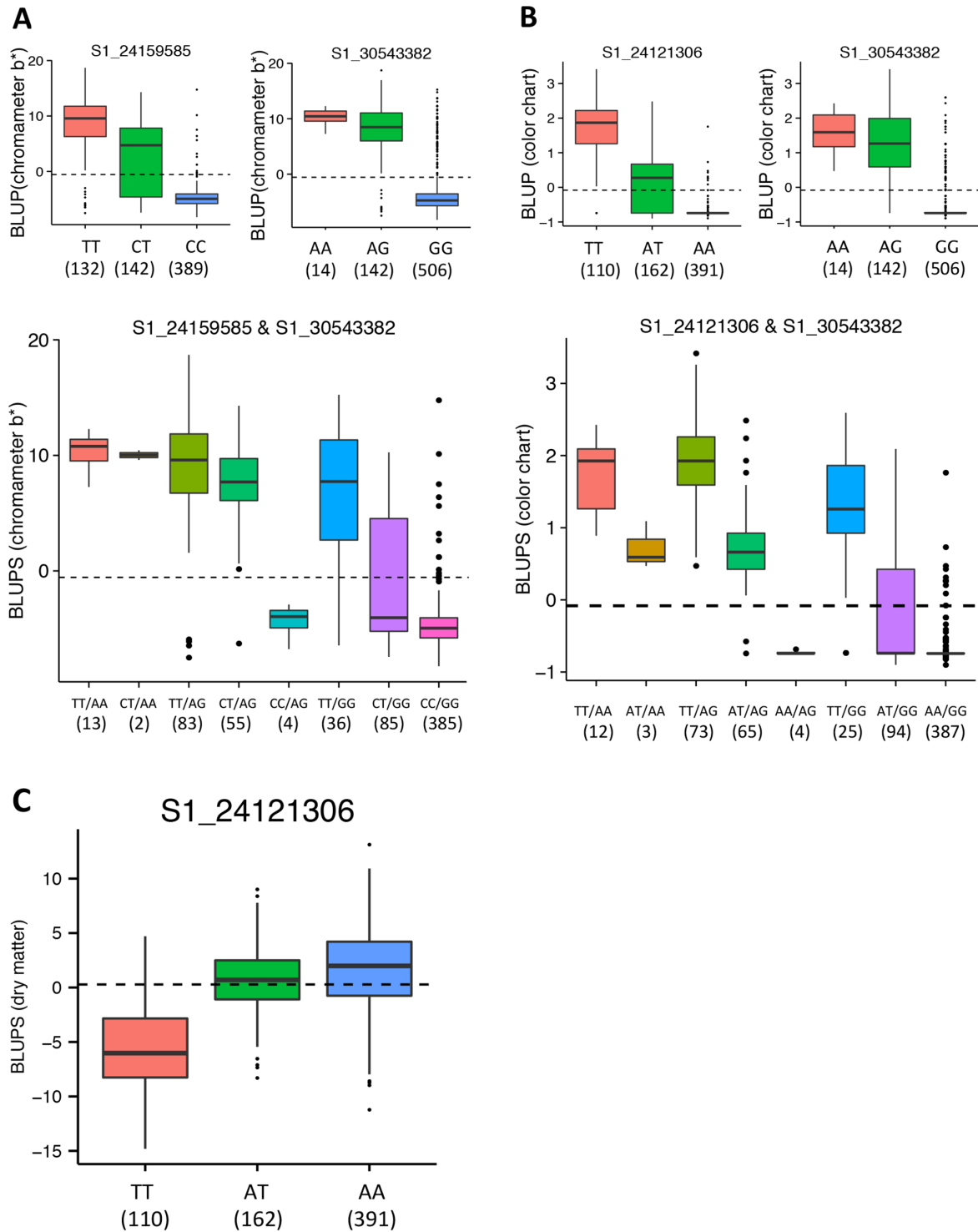


Fig. 10. Effect of the most significantly associated markers on the best linear unbiased predictors (BLUPs) for yellow color measured by (A) chromameter ( $b^*$  value) and (B) color chart in the Genetic Gain population. The boxplots show the effects of the most significantly associated SNPs at first and second peaks (above) and the two-locus haplotypes (below) on chromosome 1. (C) Effect of the most significantly associated markers on the BLUPs for dry matter content. The dashed line represents the population mean of the BLUPs. The numbers in parentheses below genotypic categories refer to the number of accessions for each genotype.

we propose to quantify total carotenoids and constituents as a future study to corroborate the current findings.

Given the importance of dry matter content in cassava and the fact that we found a single genomic region

associated with this trait, further studies are warranted to fine-map and validate the identity of the causal locus. To do this effectively would require different populations that are lacking the wild introgression segments in

chromosome 1. This will lead to reduced LD and allow higher mapping resolution. Additionally, specialized crossing scheme using strategically selected set of parents, such as the nested-association mapping population design (Yu et al., 2008) will reduce the confounding effect of population structure. Given our marker density and sample size, this study is sufficiently powered to find large effect alleles that are common in the studied germplasm. To detect additional QTLs with small effects will require a larger association panel genotyped at higher density.

The use of a broad cassava diversity panel in GWAS not only provides the foundation to map genomic regions associated with natural variation in dry matter and carotenoid content but also allows us to determine whether the negative correlation between these traits is caused by gene pleiotropy or genetic linkage. In the context of breeding to simultaneously increase carotenoid and dry matter content, the observed negative association between these traits in our germplasm is undesirable. Several lines of evidence suggested genetic linkage rather than pleiotropy to be the cause of the observed association. First, the genomic region harboring the QTLs for yellow color and dry matter content was found to occur in chromosomal segments that exhibit low overall recombination compared to the genome-wide patterns. Recent work by Bredeson et al. (2016) has shown that this chromosome 1 region harbors a large *M. glaziovii* introgression that commonly occurs in the Genetic Gain collection. Second, independent association analysis for dry matter content in the white and the yellow subpopulations detected the same association signal. However, the QTL in the white subpopulation was broader suggesting that the favorable alleles were located in nonrecombining haplotype. Third, strong candidate genes for dry matter (UDP-glucose pyrophosphorylase and sucrose synthase) and carotenoid content (phytoene synthase) were found in the major association region (24.1 Mbp) of chromosome 1. Presence of these distinct genes suggests that the two traits may be controlled through independent biological processes. These hypotheses need to be tested through functional genetics studies with these candidate genes. Taken together, these findings suggest that the phenotypic correlation between dry matter and carotenoid content is primarily caused by the physical linkage of loci underlying these traits. Moreover, Ortiz et al. (2011) found a fairly large positive correlation ( $r = 0.62$ ) between these traits. It is therefore possible that the nature of association (whether positive or negative) is dependent on the allelic status at the linked dry matter and carotenoid biosynthesis genes. We also detected a reduction of expected heterozygosity ( $\pi$ ) around the major gene region in the yellow versus white subpopulation. This suggests that the genetic base for sources of favorable alleles with respect to carotenoid biosynthesis at this locus is narrow, possibly arising from a single haplotype, which could be linked in *cis* to low-dry matter alleles in the dry matter locus. Alternatively, balancing selection of the *M. glaziovii* introgression in the white cassava subpopulation might be the cause of the higher levels of heterozygosity relative to the yellow subpopulation.

Although cassava is a predominantly outcrossing species, its clonally propagated nature means that modern varieties have undergone relatively few recombination cycles compared with seed crops. Pedigree records show that most accessions in the Genetic Gain collection are not far removed from founder clones. Accordingly, the extent of LD in this study (~2 Mbp) is much greater than the LD in maize (<10 Kb; Yan et al., 2009) as well as in grape (<10 Kb; Myles et al., 2011), another clonal species. Moreover, the overall recombination pattern is far from homogeneous, owing to the persistent introgressions of *M. glaziovii* chromosomal segments that are legacies of the historical breeding program in East Africa (Hahn et al., 1980; Jennings, 1994). From these results, it is expected that the mapping resolution will vary widely across the cassava genome depending mainly on whether a locus-of-interest occurs in or outside the large LD blocks.

This study presents significant progress toward dissecting the genetic architecture of two key traits in cassava breeding. The major loci associated with carotenoid content variation and the single locus associated with dry matter content represent markers that will be useful for MAS of these traits. Although the results of the present study suggest genetic linkage is more likely to be responsible for the negative correlation between the studied traits, there is need for further investigations to confirm or reject this hypothesis. Possible follow-up experiments include investigating whether knocking out of PSY2 using gene-silencing methods (Lu et al., 2003; Burch-Smith et al., 2004; Fofana et al., 2004) in yellow root genotypes will result in increased dry-matter content. An alternate approach is to study the effect of activating PSY2 on dry matter content in white root genotypes using gene-editing technologies like CRISPR-CAS9 (Hsu et al., 2014; Sander and Joung, 2014).

### Data Availability

The SNP and phenotype datasets can be accessed through CassavaBase at [ftp://ftp.cassavabase.org/manuscripts/TPG\\_Rabbi\\_et\\_al\\_2017.zip](ftp://ftp.cassavabase.org/manuscripts/TPG_Rabbi_et_al_2017.zip)

### Supplemental Information Available

Supplementary Fig. S1. Standard color chart used for visual gradation of yellow color in cassava root parenchyma starting from 1 (white) to 7 (deep yellow).

Supplementary Fig. S2. Quantile–quantile plots for P-values obtained from simple GLM, GLM+5PCs and MLM model for (A) color chart, (B) chromameter  $b^*$  value, and (C) dry matter content.

Supplementary Fig. S3. Heatmap of identity-by-descent relationship using SNPs from large LD block in chromosome 1 around the major QTL region.

### Conflict of Interest Disclosure

The authors declare that there is no conflict of interest.

## Acknowledgments

We would like to acknowledge Oluwafemi Alaba and Ruth Uwugiaren for DNA extractions; the staff of the Cassava Breeding Unit of IITA for conducting field trials; Sharon Mitchell and Charlotte Acharya of Cornell's Genomic Diversity Facility for sequencing; Lukas Mueller and Guillaume Bauchet for SNP processing and imputation. We sincerely thank the editor and the four anonymous reviewers for their useful suggestions that helped us to improve the manuscript. This work was supported through HarvestPlus project, The CGIAR Research Programme on Roots, Tubers, and Bananas (CRP-RTB) and The Next Generation Cassava Breeding grant OPP1048542 from Bill and Melinda Gates Foundation and the United Kingdom Department for International Development.

## References

- Akinwale, M.G.B., R.D. Aladesanwa, B.O. Akinyele, A.G.O. Dixon, and A.C. Odiyi. 2010. Inheritance of  $\beta$ -carotene in cassava (*Manihot esculenta* Crantz). *Int. J. Genet. Mol. Biol.* 2(10):198–201.
- Astle, W., and D. Balding. 2009. Population structure and cryptic relatedness in genetic association studies. *Stat. Sci.* 24(4):451–471. doi:10.1214/09-STS307
- Balagopalan, C. 2002. Cassava utilization in food, feed and industry. Chapter 15. In: R.J. Hillocks and J.M. Thresh, editors, *Cassava: Biology, production and utilization*. CABI, Wallingford.
- Balyejusa Kizito, E., A.-C. Rönnberg-Wästljung, T. Egwang, U. Gullberg, M. Fregene, and A. Westerbergh. 2007. Quantitative trait loci controlling cyanogenic glucoside and dry matter content in cassava (*Manihot esculenta* Crantz) roots. *Hereditas* 144(4):129–136. doi:10.1111/j.2007.0018-0661.01975.x
- Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, and E.S. Buckler. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23(19):2633–2635. doi:10.1093/bioinformatics/btm308
- Bredeson, J.V., J.B. Lyons, S.E. Prochnik, G.A. Wu, C.M. Ha, E. Edinger-Gonzales, J. Grimwood, J. Schmutz, I.Y. Rabbi, C. Egesi, P. Nauluvula, V. Lebot, J. Ndunguru, G. Mkamilo, R.S. Bart, T.L. Setter, R.M. Gleadow, P. Kulakow, M.E. Ferguson, S. Rounsley, and D.S. Rokhsar. 2016. Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nat. Biotechnol.* 34(5):562–570. doi:10.1038/nbt.3535
- Burch-Smith, T.M., J.C. Anderson, G.B. Martin, and S.P. Dinesh-Kumar. 2004. Applications and advantages of virus-induced gene silencing for gene function studies in plants. *Plant J.* 39(5):734–746. doi:10.1111/j.1365-313X.2004.02158.x
- Ceballos, H., N. Morante, T. Sánchez, D. Ortiz, I. Aragón, A.L. Chávez, M. Pizarro, F. Calle, and D. Dufour. 2013. Rapid cycling recurrent selection for increased carotenoids content in cassava roots. *Crop Sci.* 53(6):2342. doi:10.2135/cropsci2013.02.0123
- Chávez, A.L., T. Sánchez, G. Jaramillo, J.M. Bedoya, J. Echeverry, E.A. Bolaños, H. Ceballos, and C.A. Iglesias. 2005. Variation of quality traits in cassava roots evaluated in landraces and improved clones. *Euphytica* 143(1–2):125–133. doi:10.1007/s10681-005-3057-2
- Cock, J.H. 1982. Cassava: A basic energy source in the tropics. *Science* 218(4574):755–762. doi:10.1126/science.7134971
- El-Sharkawy, M.A. 1993. Drought-tolerant cassava for Africa, Asia, and Latin America. *Bioscience* 43(7):441–451. doi:10.2307/1311903
- Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6(5):E19379. doi:10.1371/journal.pone.0019379
- Esuma, W., L. Herselman, M.T. Labuschagne, P. Ramu, F. Lu, Y. Baguma, E.S. Buckler, and R.S. Kawuki. 2016. Genome-wide association mapping of provitamin A carotenoid content in cassava. *Euphytica* 212(1):97–110. doi:10.1007/s10681-016-1772-5
- Esuma, W., P. Rubaihayo, A. Pariyo, R. Kawuki, B. Wanjala, I. Nzuki, J.J. Harvey, and Y. Baguma. 2012. Genetic diversity of provitamin A cassava in Uganda. *J. Plant Stud.* 1(1):60–71. doi:10.5539/jps.v1n1p60
- Fofana, I.B.F., A. Sangaré, R. Collier, C. Taylor, and C.M. Fauquet. 2004. A geminivirus-induced gene silencing system for gene function validation in cassava. *Plant Mol. Biol.* 56(4):613–624. doi:10.1007/s11103-004-0161-y
- Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, and E.S. Buckler. 2014. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS One* 9(2). doi:10.1371/journal.pone.0090346
- Hahn, S.K., E.R. Terry, and K. Leuschner. 1980. Breeding cassava for resistance to cassava mosaic disease. *Euphytica* 29(3):673–683. doi:10.1007/BF00023215
- Hamblin, M.T., E.S. Buckler, and J.L. Jannink. 2011. Population genetics of genomics-based crop improvement methods. *Trends Genet.* 27(3):98–106. doi:10.1016/j.tig.2010.12.003
- Harjes, C.E., T.R. Rocheford, L. Bai, T.P. Brutnell, C.B. Kandianis, S.G. Sowinski, A.E. Stapleton, R. Vallabhaneni, M. Williams, E.T. Wurtzel, J. Yan, and E.S. Buckler. 2008. Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science* 319:330–333. doi:10.1126/science.1150255
- Hsu, P.D., E.S. Lander, and F. Zhang. 2014. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157(6):1262–1278. doi:10.1016/j.cell.2014.05.010
- Iglesias, C., J. Mayer, L. Chavez, and F. Calle. 1997. Genetic potential and stability of carotene content in cassava roots. *Euphytica* 94(3):367–373. doi:10.1023/A:1002962108315
- International Cassava Genetic Map Consortium. 2015. High-resolution linkage map and chromosome-scale genome assembly for cassava (*Manihot esculenta* Crantz) from 10 Populations. *G3 (Bethesda)* 5(1):133–144. doi:10.1534/g3.114.015008
- Jennings, D.L. 1994. Breeding for resistance to African cassava mosaic geminivirus in East Africa. *Trop. Sci.* 34(1):110–122.
- Kizito, B.E., A.-C. Rönnberg-Wästljung, T. Egwang, U. Gullberg, M. Fregene, and A. Westerbergh. 2007. Quantitative trait loci controlling cyanogenic glucoside and dry matter content in cassava (*Manihot esculenta* Crantz) roots. *Hereditas* 144(4):129–36. doi:10.1111/j.2007.0018-0661.01975.x
- Langmead, B., and S.L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9(4):357–359. doi:10.1038/nmeth.1923
- Lippert, C., J. Listgarten, Y. Liu, C.M. Kadie, R.I. Davidson, and D. Heckerman. 2011. FaST linear mixed models for genome-wide association studies. *Nat. Methods* 8(10):833–837. doi:10.1038/nmeth.1681
- Listgarten, J., C. Lippert, C.M. Kadie, R.I. Davidson, E. Eskin, and D. Heckerman. 2012. Improved linear mixed models for genome-wide association studies. *Nat. Methods* 9(6):525–526. doi:10.1038/nmeth.2037
- Lu, R., A.M. Martin-Hernandez, J.R. Peart, I. Malcuit, and D.C. Baulcombe. 2003. Virus-induced gene silencing in plants. *Methods* 30(4):296–303. doi:10.1016/S1046-2023(03)00037-9
- Ly, D., M. Hamblin, I. Rabbi, G. Melaku, M. Bakare, H.G. Gauch, R. Okechukwu, A.G.O. Dixon, P. Kulakow, and J.L. Jannink. 2013. Relatedness and genotype  $\times$  environment interaction affect prediction accuracies in genomic selection: A study in cassava. *Crop Sci.* 53(4):1312–1325. doi:10.2135/cropsci2012.11.0653
- Marín Colorado, J.A., H. Ramírez, and M. Fregene. 2009. Genetic mapping and QTL analysis for carotenoids in a S1 population of cassava. *Acta Agron. Univ. Nac. Colomb.* 58(1):15–21.
- Morillo, A.C., Y. Morillo, and H. Ceballos. 2013. Identification of QTLs for carotene content in the genome of cassava (*Manihot esculenta* Crantz) and S1 population validation. *Acta Agronómica. Univ. Nac. Colomb.* 62(3):196–206.
- Myles, S., A.R. Boyko, C.L. Owens, P.J. Brown, F. Grassi, M.K. Aradhya, B. Prins, A. Reynolds, J.-M. Chia, D. Ware, C.D. Bustamante, and E.S. Buckler. 2011. Genetic structure and domestication history of the grape. *Proc. Natl. Acad. Sci. USA* 108(9):3530–3535. doi:10.1073/pnas.1009363108
- Njoku, D.N., V.E. Gracen, S.K. Offei, I.K. Asante, E.Y. Danquah, C.N. Egesi, and E. Okogbenin. 2014. Molecular marker analysis of F1 progenies and their parents for carotenoids inheritance in African cassava (*Manihot esculenta* Crantz). *Afr. J. Biotechnol.* 13(40):3999–4007. doi:10.5897/AJB2013.12244
- Njoku, D.N., V.E. Gracen, S.K. Offei, I.K. Asante, C.N. Egesi, P. Kulakow, and H. Ceballos. 2015. Parent-offspring regression analysis for total carotenoids and some agronomic traits in cassava. *Euphytica* 206(3):657–666. doi:10.1007/s10681-015-1482-4

- Okechukwu, R.U., and A.G.O. Dixon. 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):181–208. doi:10.1080/15427520802212506
- Ortiz, D., T. Sánchez, N. Morante, H. Ceballos, H. Pachón, M.C. Duque, A.L. Chávez, and A.F. Escobar. 2011. Sampling strategies for proper quantification of carotenoid content in cassava breeding. *J. Plant Breed. Crop Sci.* 3(1):14–23.
- Owens, B.F., A.E. Lipka, M. Magallanes-Lundback, T. Tiede, C.H. Diepenbrock, C.B. Kandianis, E. Kim, J. Cepela, M. Mateos-Hernandez, C.R. Buell, E.S. Buckler, D. DellaPenna, M.A. Gore, and T. Rocheford. 2014. A foundation for provitamin A biofortification of maize: Genome-wide association and genomic prediction models of carotenoid levels. *Genetics* 198(4):1699–1716. doi:10.1534/genetics.114.169979
- Pfeiffer, W.H., and B. McClafferty. 2007. HarvestPlus: Breeding crops for better nutrition. *Crop Sci.* 47(Supplement\_3): S-88.
- Prochnik, S., P.R. Marri, B. Desany, P.D. Rabinowicz, C. Kodira, M. Mohiuddin, F. Rodriguez, C. Fauquet, J. Tohme, T. Harkins, D.S. Rokhsar, and S. Rounsley. 2012. The cassava genome: Current progress, future directions. *Trop. Plant Biol.* 5(1):88–94. doi:10.1007/s12042-011-9088-z
- Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A.R. Ferreira, D. Bender, J. Maller, P. Sklar, P.I.W. de Bakker, M.J. Daly, and P.C. Sham. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81(3):559–575. doi:10.1086/519795
- R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at <https://www.R-project.org/> (verified 6 July 2017).
- Rabbi, I., M. Hamblin, M. Gedil, P. Kulakow, M. Ferguson, A.S. Ikpan, D. Ly, and J.L. Jannink. 2014a. Genetic mapping using genotyping-by-sequencing in the clonally propagated cassava. *Crop Sci.* 54(4):1384–1396. doi:10.2135/cropsci2013.07.0482
- Rabbi, I.Y., M.T. Hamblin, P.L. Kumar, M.A. Gedil, A.S. Ikpan, J.L. Jannink, and P.A. Kulakow. 2014b. High-resolution mapping of resistance to cassava mosaic geminiviruses in cassava using genotyping-by-sequencing and its implications for breeding. *Virus Res.* 186:87–96. doi:10.1016/j.virusres.2013.12.028
- Saltzman, A., E. Birol, H.E. Bouis, E. Boy, F.F. De Moura, Y. Islam, and W.H. Pfeiffer. 2013. Biofortification : Progress toward a more nourishing future. *Glob. Food Secur.* 2(1):9–17. doi:10.1016/j.gfs.2012.12.003
- Sánchez, T., H. Ceballos, D. Dufour, D. Ortiz, N. Morante, F. Calle, T. Zum Felde, M. Domínguez, and F. Davrieux. 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
- Sander, J.D., and J.K. Joung. 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* 32(4):347–355. doi:10.1038/nbt.2842
- Sayre, R., J.R. Beeching, E.B. Cahoon, C. Egesi, C. Fauquet, J. Fellman, M. Fregene, W. Gruissem, S. Mallova, M. Manary, B. Maziya-Dixon, A. Mbanaso, D.P. Schachtman, D. Siritunga, N. Taylor, H. Vanderschuren, and P. Zhang. 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
- Smith, A.M. 2008. Prospects for increasing starch and sucrose yields for bioethanol production. *Plant J.* 54(4):546–558. doi:10.1111/j.1365-3113.2008.03468.x
- Suwarno, W.B., K.V. Pixley, N. Palacios-Rojas, S.M. Kaepler, and R. Babu. 2015. Genome-wide association analysis reveals new targets for carotenoid biofortification in maize. *Theor. Appl. Genet.* 128(5):851–864. doi:10.1007/s00122-015-2475-3
- Wang, X., L. Chang, Z. Tong, D. Wang, Q. Yin, D. Wang, X. Jin, Q. Yang, L. Wang, Y. Sun, Q. Huang, A. Guo, and M. Peng. 2016. Proteomics profiling reveals carbohydrate metabolic enzymes and 14-3-3 proteins play important roles for starch accumulation during cassava root tuberization. *Sci. Rep.* 6(January):19643. doi:10.1038/srep19643
- Welsch, R., J. Arango, C. Bär, B. Salazar, S. Al-Babili, J. Beltrán, P. Charvarriaga, H. Ceballos, J. Tohme, and P. Beyer. 2010. Provitamin A accumulation in cassava (*Manihot esculenta*) roots driven by a single nucleotide polymorphism in a phytoene synthase gene. *Plant Cell* 22(10):3348–3356. doi:10.1105/tpc.110.077560
- Wolfe, M.D., I.Y. Rabbi, C. Egesi, M. Hamblin, R. Kawuki, P. Kulakow, R. Lozano, D.P. Del Carpio, P. Ramu, and J.L. Jannink. 2016. Genome-wide association and prediction reveals genetic architecture of cassava mosaic disease resistance and prospects for rapid genetic improvement. *Plant Genome* 9(2):1–248. doi:10.3835/plantgenome2015.11.0118
- Wong, W.W.L., J. Griesman, and Z.Z. Feng. 2014. Imputing genotypes using regularized generalized linear regression models. *Stat. Appl. Genet. Mol. Biol.* 13(5). doi:10.1515/sagmb-2012-0044
- Yan, J., T. Shah, M.L. Warburton, E.S. Buckler, M.D. McMullen, and J. Crouch. 2009. Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PLoS One* 4(12):E8451. doi:10.1371/journal.pone.0008451
- Yang, J., D. An, and P. Zhang. 2011. Expression profiling of cassava storage roots reveals an active process of glycolysis/gluconeogenesis. *J. Integr. Plant Biol.* 53(3):193–211. doi:10.1111/j.1744-7909.2010.01018.x
- Yu, J., J.B. Holland, M.D. McMullen, and E.S. Buckler. 2008. Genetic design and statistical power of nested association mapping in maize. 551(January): 539–551.
- Yu, J., G. Pressoir, W.H. Briggs, I. Vroh Bi, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland, S. Kresovich, and E.S. Buckler. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38(2):203–208. doi:10.1038/ng1702
- Zeeman, S.C., J. Kossmann, and A.M. Smith. 2010. Starch: Its metabolism, evolution, and biotechnological modification in plants. *Annu. Rev. Plant Biol.* 61:209–234. doi:10.1146/annurev-arplant-042809-112301
- Zhang, Z., E. Ersoz, C.-Q. Lai, R.J. Todhunter, H.K. Tiwari, M.A. Gore, P.J. Bradbury, J. Yu, D.K. Arnett, J.M. Ordovas, and E.S. Buckler. 2010. Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* 42(4):355–360. doi:10.1038/ng.546
- Zhao, K., C.-W. Tung, G.C. Eizenga, M.H. Wright, M.L. Ali, A.H. Price, G.J. Norton, M.R. Islam, A. Reynolds, J. Mezey, A.M. McClung, C.D. Bustamante, and S.R. McCouch. 2011. Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat. Commun.* 2:467. doi:10.1038/ncomms1467