



Genome-wide association mapping and genomic prediction of yield-related traits and starch pasting properties in cassava

Chalermopol Phumichai^{1,2,3} · Pornsak Aiemnaka⁴ · Piyaporn Nathaisong⁴ · Sirikan Hunsawattanukul^{1,2,3} · Phasakorn Fungfoo¹ · Chareinsuk Rojanaridpiched⁴ · Vichan Vichukit⁴ · Pasajee Kongsil¹ · Piya Kittipadukul¹ · Wannasiri Wannarat¹ · Julapark Chunwongse⁵ · Pumipat Tongyoo² · Chookiat Kijkhunasatian⁶ · Sunee Chotineeranat⁶ · Kuakoon Piyachomkwan⁶ · Marnin D. Wolfe⁸ · Jean-Luc Jannink⁷ · Mark E. Sorrells⁸

Received: 6 June 2021 / Accepted: 25 September 2021 / Published online: 18 October 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Key message GWAS identified eight yield-related, peak starch type of waxy and wild-type starch and 21 starch pasting property-related traits (QTLs). Prediction ability of eight GS models resulted in low to high predictability, depending on trait, heritability, and genetic architecture.

Abstract Cassava is both a food and an industrial crop in Africa, South America, and Asia, but knowledge of the genes that control yield and starch pasting properties remains limited. We carried out a genome-wide association study to clarify the molecular mechanisms underlying these traits and to explore marker-based breeding approaches. We estimated the predictive ability of genomic selection (GS) using parametric, semi-parametric, and nonparametric GS models with a panel of 276 cassava genotypes from Thai Tapioca Development Institute, International Center for Tropical Agriculture, International Institute of Tropical Agriculture, and other breeding programs. The cassava panel was genotyped via genotyping-by-sequencing, and 89,934 single-nucleotide polymorphism (SNP) markers were identified. A total of 31 SNPs associated with yield, starch type, and starch properties traits were detected by the fixed and random model circulating probability unification (FarmCPU), Bayesian-information and linkage-disequilibrium iteratively nested keyway and compressed mixed linear model, respectively. GS models were developed, and forward predictabilities using all the prediction methods resulted in values of -0.001 – 0.71 for the four yield-related traits and 0.33 – 0.82 for the seven starch pasting property traits. This study provides additional insight into the genetic architecture of these important traits for the development of markers that could be used in cassava breeding programs.

Introduction

Cassava (*Manihot esculenta* Crantz) is an economically important, staple tropical root crop. It is one of the most important dietary energy sources in most tropical countries (Ceballos et al. 2020). Cassava is robust, resistant to

Communicated by Damaris Odeny.

✉ Chalermopol Phumichai
chalermopol.ph@ku.th

¹ Department of Agronomy, Faculty of Agriculture, Kasetsart University, Bangkok 10900, Thailand

² Center for Agricultural Biotechnology, Kasetsart University, Kamphaeng Saen Campus, Nakhon Pathom 73140, Thailand

³ Center of Excellence On Agricultural Biotechnology: (AG-BIO/MHESI), Bangkok 10900, Thailand

⁴ Thai Tapioca Development Institute, Lumpini Tower, 1168/26 Rama IV Road, Bangkok 10120, Thailand

⁵ Department of Horticulture, Faculty of Agriculture Kamphaeng Saen, Kasetsart University, Nakhon Pathom 73140, Thailand

⁶ Cassava and Starch Technology Research Team, National Center for Genetic Engineering and Biotechnology, Pathumthani 12120, Thailand

⁷ United States Department of Agriculture - Agriculture Research Service, Ithaca, NY 14850, USA

⁸ Plant Breeding and Genetics Section, Cornell University, Ithaca, NY 14850, USA

droughts and floods, and grows well in degraded soils. Historically, cassava was grown for human consumption (e.g., gari, fufu, sago, table consumption, etc.). However, demand from worldwide markets for cassava (e.g., as a starch, animal feed, or bioethanol) has strengthened over the years. Cassava is now the second most important source of starch worldwide (Stapleton 2012). The global production of fresh root equivalents in 2018 amounted to 277.1 million tons (FAO 2018). Currently, there are important breeding efforts in Brazil, Colombia, China, Ghana, India, Nigeria, Kenya, Mozambique, Tanzania, Thailand, Uganda, and Vietnam (Ceballos et al. 2020).

Since the cassava breeding cycle is long, with a low reproduction rate, it takes 7–8 years to develop a new cassava variety (Rojanaridpiched et al. 2010). The first cassava evaluations occur in seedlings, single row trials (SRT), followed by preliminary (PYT) and advanced (AYT) yield trials, and culminating in multi-location, multi-year uniform yield trials (UYT; Rojanaridpiched et al. 2010; Gracen et al. 2018; Ceballos et al. 2020). Genetic gains from cassava breeding efforts have been very limited in the last century compared to other crops (Ceballos et al. 2004). Factors limiting cassava's breeding efficiency include its heterozygous progenitors, long breeding cycles, clonal propagation, and non-recovery of recurrent genomes after single trait introgressions (Ceballos et al. 2016; Kuon et al. 2019). Although cassava is predominantly clonally propagated, it outcrosses with plants still capable of sexual reproduction. The intentional or unintentional inclusion of seedlings into clonally propagated stock continually generates new genotypes within a population, thus increasing haplotypic variation (McKey et al. 2010). Storage root yield is the key objective in cassava breeding programs. Farmers attach the highest importance to good storage root yields with high dry matter content (DMC) in the wild-type and waxy (amylose-free) cassava starch widely used in food and industrial applications (Sanchez et al. 2010; Aiemnaka et al. 2012). Cassava starch has many remarkable characteristics that are advantageous for industrial applications, including high paste viscosity and clarity and high freeze–thaw stability (Nwokocha et al. 2009; Sanchez et al. 2010; Toae et al. 2019).

Genome-wide association study (GWAS) is a powerful approach based on the linkage disequilibrium (LD) resulting from the association of target traits and haplotype loci. GWAS identifies genes, alleles, or haplotypes related to specific agronomic traits in complex environments and has been successfully used in many plant studies, including studies of wheat (Juliana et al. 2018), maize (Zhang et al. 2015), rice (Begum et al. 2015), and tomatoes (Sauvage et al. 2014). In cassava, GWAS has been successfully used to dissect the genetic basis of several traits (Rabbi et al. 2020), including dry matter content, total carotenoid content (Rabbi et al. 2017; Ikeogu et al. 2019), resistance to CMD (*CMD2*),

phytoene synthase (*PSY*), root number, shoot weight, harvest index (Rabbi et al. 2017; Okeke et al. 2017; Zhang et al. 2018; Somo et al. 2020; Yonis et al. 2020), cassava green mite resistance (Ezenwaka et al. 2018), cassava brown streak disease resistance (Kayondo et al. 2018), cassava mosaic disease resistance (Wolfe et al. 2016), and cyanide content (Ogbonna et al. 2021).

Genomic selection (GS) has been proposed as a promising tool to overcome breeding challenges (Meuwissen et al. 2001). GS is an alternative to traditional marker-assisted selection (MAS) for quantitative traits (Jannink et al. 2010; Hickey et al. 2017). GS aims to combine genome-wide molecular markers and phenotypes in a training population to predict the genetic values of future individuals for selection purposes. Since no significance test is required, eliminating biases in marker effect estimates, the breeding cycle is accelerated (Desta and Ortiz 2014). Contrary to MAS, GS is suitable for quantitative traits controlled by a large number of small-effect genes. GS has been introduced in many aspects of plant breeding, such as inbred performance predictions and hybrid predictions (Riedelsheimer et al. 2012; Crossa et al. 2014; Xu et al. 2014; Wang et al. 2017; Xu 2017). Accurate predictions are essential for the successful application of GS. The predictability (i.e., the prediction accuracy) obtained from cross-validation in training populations has been previously evaluated in maize, wheat, barley, and cassava (VanRaden 2008; Crossa et al. 2017; Wolfe et al. 2017; Andrade et al. 2019). These studies indicated that predictability is affected by various genetic factors, including heritability, relatedness, sample size, marker density, and genetic architecture. The predictabilities of low heritability traits, such as yield, were consistently lower than high heritability traits, such as kernel weight and plant height (Xu et al. 2018).

In addition to genetic factors, statistical factors influence predictability. Parametric methods widely used in GS to predict genetic values include genomic best linear unbiased prediction (GBLUP; VanRaden 2008), least absolute shrinkage, selection operator (LASSO; Tibshirani 1996), and partial least squares (PLS; Geland and Kowalski 1986), and nonparametric methods, including random forest (Svetnik et al. 2003), neural networks (NN; Ehret et al. 2015), support vector machines (SVM; Maenhout et al. 2007), and reproducing kernel Hilbert spaces (RKHS) regressions (de Los Campos et al. 2010). Several investigators have compared the predictive performance of these methods using simulation and empirical data (Heslot et al. 2012; VanRaden 2008; Riedelsheimer et al. 2012; Howard et al. 2014; Wang et al. 2015). The key limitation that GS has in the case of cassava is the use of heterozygous progenitors that distinguish cassava from most of the crops (Ceballos et al. 2016 and 2020) where GS has been proven to be useful as highlighted and the strong nonadditive genetic effect for fresh root yield

as reported by articles published by Wolfe and co-workers (2017, 2019).

The objectives of this study were to (1) understand the genetic architecture of yield-related traits and starch pasting property traits in a cassava breeding population and identify novel QTLs loci associated with the above traits using GWAS; (2) explore the potential utility of using GS for future cassava breeding programs.

Materials and methods

Plant materials and field trials

We obtained 276 cassava genotypes, including both 247 wild-type and 29 waxy cassava starch types, from the germplasm collection developed at the Department of Agriculture, Kasetsart University, Bangkok, Thailand, Thai Tapioca Development Institute (TTDI). The cassava genotypes were from International Center for Tropical Agriculture (CIAT), International Institute of Tropical Agriculture (IITA), and other breeding programs (Supplementary Table S1).

Each clone was represented by 10 plants in a single row spaced with 1 m × 1 m apart. KU50 and HB80, two commercial varieties, were planted after every ten rows in incomplete blocks for 2 consecutive years at Tapioca Development Institute (TDI) station (2015 and 2016; experimental stations: TDI (15.1577245 101.5028433; N15°9.46347, E101°30.170598) and Nakhon Ratchasima Province, Thailand). The plots were treated with the fertilizer N–P–K: 15–7–18 at 312.5 kg ha⁻¹. Weed control was primarily manual 1 month after planting. The plants were planted in March or April before the rainy season, then harvested 10 to 12 months later in the following year.

Phenotypic evaluation and statistical analysis

The total fresh root weight (FRW) of the storage roots harvested per row was measured in kilograms. The starch content (SC; %) was measured by Riemann balance, 5 kg of bulked roots for each genotype from single row plot was weighted in air and immersed in water (Kawano et al. 1987). The harvest index (HI) was computed as the total fresh root weight to total biomass at harvest ratio. The average number of root per plant (RP) was computed as total number of root in plot divided by number of harvested plant. The mixed model augmented design with un-replicated entries used in this study.

The model was

$$Y_{ij} = \mu + \beta i + \tau j + \varepsilon_{ij}$$

where β and τ are the effects of blocks and entries, respectively. Broad-sense heritability (H^2) was estimated using BLUPs and the formula:

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

where σ_g^2 is the genotype variance, σ_e^2 is the residual variance, and r is the number of environments. The variance components were calculated using the *lmer* function in the R package (*lme4*, v1.1–7; Bates et al. 2015). The variance component of combined analysis for year and locations was estimated using the REML method explained above for each genotype which were used in GWAS and GS analysis. The genetic correlation between traits was estimated using best linear unbiased predictions (BLUPs) of measured traits. The Pearson correlation was performed using the *ggcorrplot* package in R.

Starch pasting properties

Starch extraction

Starch extraction was prepared following Chaengsee et al. (2020). Washed fresh cassava roots were peeled and the upper and lower edges were cut off and then were chopped into small pieces before being crushed in water at a ratio of 1:2 of cassava to water. The crushed paste was separated from the water by passing through a cloth and 90 μ m screen, respectively. The paste was washed in water before being oven-dried at 50 °C for 24 h. The dried starch was milled and screened through a 90 μ m screen before further analysis.

RVA profile parameters

Viscosity profiles of wild-type and waxy cassava starch dispersions were analyzed with a Rapid Visco Analyzer (model RVA-4 Series, Newport Scientific, Warriewood, Australia) according to the method of (1995). Starch samples (2.58 g, dry basis) were dispersed in distilled water to prepare 9.21% starch suspension with a total weight of 28 g. Viscosity was recorded under the temperature profile: The samples were started at 50 °C 1 min then heated from 50 to 95 °C with a heating rate increase of 12 °C/min and then maintained at 95 °C for 2.5 min. The hot paste was subsequently cooled to 50 °C with a cooling rate of 12 °C/min and then held for 2 min (Standard program No. 1). Suspension was constantly stirred at 160 rpm, and measurements for Peak viscosity (PV), trough viscosity (TV), breakdown viscosity (BD), final viscosity (FV), setback from trough (SB), pasting temperature (PT), and peak temperature (PeT). The wild-type and waxy cassava's phenotypic distribution and starch pasting properties were visualized with the R package "ggplot2" (Wickham, 2016).

Genotyping-by-sequencing

Sequence alignment, SNP calling, and SNP imputation

DNA extraction and genotyping-by-sequencing Genomic DNA was extracted using the CTAB method from Doyle & Doyle (1990) with some minor modifications. Briefly, we applied the chloroform:isoamyl alcohol (24:1) extraction step twice to improve the removal of the phenolic compounds. The DNA samples were then quantified using the NanoDrop D-1000 per the manufacturer's instructions. The DNA samples were diluted to 20 ng/ μ l, then subsequently used for GBS library preparation. The GBS library construction protocol for the germplasms followed Elshire et al. (2011). The GBS library was performed at the Institute of Genomic Diversity (Cornell University, Ithaca, NY, USA) for 101-cycle single-end sequencing on one lane of a 16-lane flow cell Illumina HiSeq 2000 (Illumina).

The sequencing reads were processed with the GBS Discovery Pipeline for species with a reference genome implemented in TASSEL version 3.0 (Bradbury et al. 2007) following pipeline documentation (Glaubitz et al. 2014). In conjunction with the barcode adapter key file, the pipeline identified high-quality, unique sequence reads (termed tags), which contained a barcode, a cut site, and an inserted genomic sequence. The pipeline then merged these tags, indexed them, and aligned them to the reference genome. The sequence tags for our GBS library were aligned to the version 1.0 release of the cassava V6 reference genome (Bredeson et al. 2016) by implementing the Burrows–Wheeler alignment (Li and Durbin 2009) command of the pipeline. The TASSEL 3.0 Discovery SNP Caller (Glaubitz et al. 2014) was used to align the multiple sequence tags from the same physical locations across the genome, to call SNPs at these locations across the individual samples, and to output this data into one HapMap format file (.hmp.txt) per chromosome for downstream analysis. Missing SNP data were imputed with the FIL-LIN algorithm (Swarts et al. 2014) in TASSEL version 5.0 (Bradbury et al. 2007).

Genome-wide association study GWAS was performed using trait BLUPs for 276 genotypes. To discover any associations between the genome-wide GBS SNPs and trait phenotypes in our germplasm, we used the compressed mixed linear model (CMLM; Zhang et al. 2010), the multi-locus mixed model (MLMM) developed by Segura et al. (2012), and an enhanced version of the fixed and random model circulating probability unification (FarmCPU) method (Liu et al. 2016), Bayesian-information and linkage-disequilibrium iteratively nested keyway (BLINK) method was employed to evaluate the dichotomous traits

of waxy and wild-type starch phenotype, implemented in the Genome Association and Prediction Integrated Tool (GAPIT) R package (Lipka et al. 2012) in R version 3.4.2 (R Core Team 2017). The percentage of phenotypic variance explained (PVE) by significant SNPs was calculated using *lm* function in R to obtain an adjusted R^2 .

We determined significant associations for each trait by adjusting the raw *p* value using the conservative Bonferroni error rate control method. Manhattan and Q–Q plots were visualized using R “CMplot” (Yin et al. 2021). The pairwise LD between the genome-wide markers on each chromosome was calculated using TASSEL v5.2.20 (Bradbury et al. 2007). Marker pairs with statistically significant LD ($p < 0.05$) were considered in the LD analysis. The r^2 values were plotted against distance (Mb). A LOESS smoothing line was fitted using R software package ggplot2 (Wickham, 2016). The Bonferroni-corrected threshold ($\alpha = 1, -\log_{10}(P) \geq 6.25$) was established at a probability of $0.05/89,943 = 5.55 \times 10^{-7}$ (equivalent to a $-\log_{10}P$ score of 6.25), which was used as the cut-off for the 89,943 GBS markers as being identification significant SNP-trait associations (Yang et al. 2014). Narrow-sense heritability estimates were computed from the variance components estimated by the genomic mixed models fit by TASSEL v5.2.20 (Bradbury et al. 2007), which compute an additive genetic relationship (i.e., kinship) matrix, as explained in Endelman and Jannink (2012).

Identification of candidate genes We identified candidate genes within the associated genomic region using the significant GWAS results and the phytozome 12 portal link to biomart (<https://phytozome.jgi.doe.gov/biomart/> accessed 15 Oct. 2019). We searched for genes located in windows of ± 50 kb around the most significant SNPs.

Genomic prediction methods

Evaluation of genomic prediction models using cross-validation

The genomic prediction was evaluated using eight statistical models. The parametric models included ridge regression BLUP (rrBLUP) from the rrBLUP package (Endelman 2011), BayesA (Meuwissen et al. 2001), BayesB (Meuwissen et al. 2001), and BayesC (Habier et al. 2011). Bayesian Lasso (BL; Park and Casella 2008), and Bayesian ridge regression (BRR; de los Campos et al. 2013) were executed using the BGLR R package developed by Pérez and de los Campos, (2014). Semi-parametric models based on RKHS methods were run using the BGLR R package (de los Campos et al. 2009, 2010). A nonparametric, random forest classification (RFC; Breiman, 2001), a machine learning method used for regression and classification, has been successfully

used for predictions (Wolfe et al. 2017; González-Camacho et al. 2018). We implemented RFC using the “randomForest” package in R (Liaw 2013).

Predictability or accuracy

The predictability for cassava performance was evaluated using a tenfold cross-validation. The sample was randomly partitioned into 10 parts: nine parts used to estimate parameters, and 1 part was held out of the model, and subsequently predicted and used to estimate accuracy. Random sample training and validation sets were repeated 30 times, and the means of the Pearson’s correlation coefficients were defined as the genomic prediction ability. The Studentized range statistic, Tukey’s honest significance test (R Core Team 2017) was used to test for significant differences in the predictive abilities of the eight models.

Results

Phenotypic analysis

Figure 1 shows the phenotypic distributions of the yield-related traits. Table 1 provides descriptive statistics of fresh root weight, percentage of starch content, number

of average roots, and harvest index. The fresh root weight varied widely in the population, ranging from 0.1 to 11.3 kg/plant. The percentage of starch content ranged from 3.8 to 30.1%, with an average of 16.4%. The harvest index ranged from 0.1 to 0.9, with an average of 0.54. The number of average roots ranged from 2 to 27, with an average of 11 roots per plant. The broad-sense heritability was moderate to low for harvest index, fresh root weight, root number per plant, and percentage of starch content ($H^2 = 0.57, 0.43, 0.40,$ and $0.33,$ respectively). phenotypic variations of yield-related traits were affected more environmental factors than genetic factors. The pasting properties varied significantly among the different lines. The narrow-sense heritability ranged from 0.58 for SB to 0.85 for PV (Table 1). These results indicate that the phenotypic variations in the starch pasting properties were mainly affected by genetic factors and that this panel can be used for further genetic analyses.

Trait correlations of BLUPs for yield-related traits FRW with SC, HI, and RP were low ranging from 0.09 to 0.2. SC was positively correlated with RP ($r = 0.56, p < 0.001$) (Fig. 2). FRW was not significantly correlated with starch pasting properties traits except for starch pasting temperature ($r = -2.4, p < 0.001$). The correlation between HI and starch pasting properties traits was positive ranging from 0.48 to 0.57 ($p < 0.001$). There were moderately to highly significant

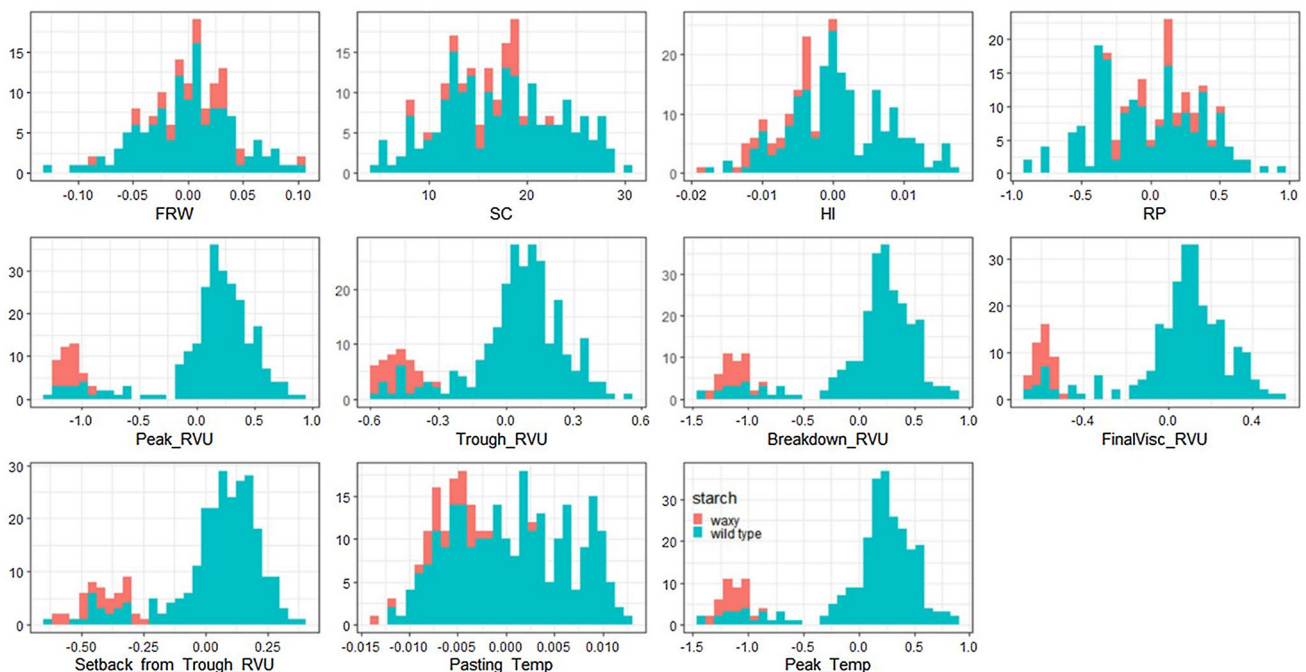


Fig. 1 Frequency distribution of yield-related traits: number of root per plant (RP), fresh root weight (FRW), harvest index (HI), starch content (SC), and starch pasting property traits: peak RVU, trough RVU, breakdown RVU, final viscosity RVU, setback from trough

RVU, pasting temp and peak temp. Each panel shows the distribution of the best linear unbiased estimates of the 276 cassava panel used for this study

Table 1 Summary information for phenotypic data of the four yield-related traits and starch pasting properties

Trait	Mean	Range	Heritability (H^2)	Heritability (h^2)
FRW (kg/plant)	3.3	0.1–11.3	0.43	0.10
SC (%)	16.4	3.8–30.1	0.33	0.25
HI	0.54	0.1–0.9	0.57	0.51
RP	11	2.0–27	0.40	0.30
Peak viscosity (RVU)	404.4	327–527	–	0.85
Trough viscosity (RVU)	139.8	112–175	–	0.77
Breakdown viscosity (RVU)	264.6	191–379	–	0.84
Final viscosity (RVU)	214	179–256	–	0.79
Setback viscosity (RVU)	74.3	56–99	–	0.58
Pasting temperature (°C)	71.7	68.45–75.25	–	0.62
Peak temperature (°C)	85.4	79.85–91.15	–	0.67

Heritability, means, and ranges of 276 genotypes across all trials

FRW Fresh root weight, SC Starch content, HI Harvest index, RP Number of roots per plant

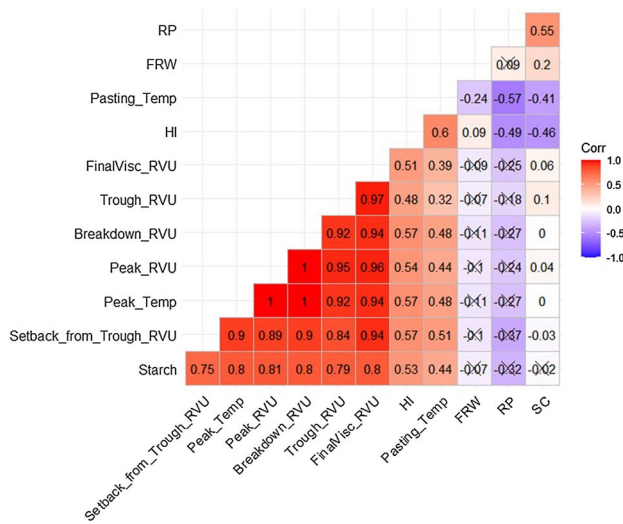


Fig. 2 Heatmap of genetic correlation for yield and starch properties-related traits using BLUP values

positive correlations among starch properties traits with values ranging from 0.39 to 1.0 ($p < 0.001$).

Starch pasting properties

The pasting characteristics and paste viscosity profiles of starch in 9.21% w/w analysis concentration are given in Table 2 and Fig. 3a. The pasting temperatures of the wild-type cassava starches were similar to the waxy cassava starches (68.45–75.25 °C and 67.75–73.36 °C, respectively). The peak viscosity of the wild-type cassava starches (327–527 RVU) was significantly higher than those of the waxy cassava starches (200–243 RVU). The breakdown viscosity of wild-type cassava starches (191–379 RVU) was higher than those of waxy cassava starches (102–140 RVU). After cooling, the final viscosities of the wild-type cassava starches were significantly higher than those of the waxy cassava starches (179–256 and 133–151 RVU, respectively). The final viscosities of the wild-type cassava starches (56–99 RVU) were slightly higher than those of the waxy cassava starches (36–55 RVU).

Table 2 Pasting properties of waxy cassava starch compared to wild-type cassava starches

Starch	Pasting parameter						
	Peak viscosity	Trough viscosity	Breakdown viscosity	Final viscosity	Setback viscosity	Pasting temperature	Peak temperature
	RVU	RVU	RVU	RVU	RVU	°C	°C
Wild-type cassava (9.21%)	327–527	112–175	191–379	179–256	56–99	68.45–75.25	79.85–91.15
Waxy cassava (9.21%)	200–243	88–109	102–140	133–151	36–55	67.75–73.36	75–86.3
Waxy cassava (5%)	85–107	40–50	40–60	47–56	7–10	69.35–75.15	76.55–87.95

Population structure and linkage disequilibrium

Sequence reads were aligned to the cassava V6 reference genome (Bredeson et al. 2016), and the SNPs were then called, imputed in regions of low coverage, and filtered to give a total of 327,919 SNPs detected across the population (an average of 1.62 SNP per 1 Kbp) (Fig. 4a).

A total of 327,919 SNPs were filtered with minor allele frequencies less than 1% (> 0.01 , missing data per site $< 90\%$), resulting in 88,934 SNPs. Population structure analysis of the 276 accessions genotyped across the 88,934 SNPs using PC analysis detected subtle genetic differentiation in the germplasm panel, with the first 10 PCs explaining about 25.25% of the genetic variation. The first two PCs accounted for 8.38% and 6.25% of the genetic variation, respectively (Fig. 4b). The average LD decay was not uniform across all chromosomes (Fig. 5). The lowest per-chromosome LD decay was observed on chromosome 11 (24 Kbp) and the largest on chromosome 5 (180 Kbp) ($r^2 < 0.2$).

GWAS analysis

Yield-related traits

We analyzed genotype–phenotype associations in 276 cassava genotypes and 12 traits related to yield, starch pasting properties and waxy and wild-type starch with 89,943 SNPs using CMLM MLMM FarmCPU and BLINK for waxy and wild-type starch analysis in GAPIT. Three different models were used to compare the marker–trait association tests. These three models identified different numbers of significant markers associated with FRW, SC, HI, and RP traits for cassava when the same significance threshold was used (Table 3). Based on the Q–Q plots for all models, FarmCPU model was used for yield-related traits (Fig. 6 and Supplementary Figures S1, S2).

The GWAS analysis on starch pasting property traits was analyzed using only the CLMM model and dichotomous trait of waxy and wild-type starch traits was analyzed using BLINK. A total of eight SNPs were significantly associated with yield-related traits (Fig. 6) with the contribution of phenotypic variance explained (R^2) by a single QLT ranging from 2.54 to 25.32% (Table 3). For FRW, one SNP marker (S1_32060063) on chromosome 1 did not reach the significant threshold but showed a $-\log_{10} p = 5$, and the variance explained 11.82%. Three SNPs were significantly associated with SC and were distributed on chromosomes 1 (S1_23945110), 6 (S1_172308681), 18 (S1_502313244). The significant markers associated with the trait were mostly concentrated in a single region

on chromosome 18. The most significant SNP marker (S1_502313244) had a $-\log_{10} (p \text{ value})$ of 7.66. The most significant marker explained 10.61% of the phenotypic variance. Five SNPs were significantly associated with HI and were distributed on chromosomes 1, 4, 7, 14, and 17. The most significant SNP marker (S1_478075251) had a $-\log_{10} (p \text{ value})$ of 7.35. The marker explained 25.32% of the phenotypic variance. A marker on Chromosome 4 did not reach the significant threshold for RP but had a $-\log_{10} (p \text{ value})$ of 4.96.

Starch type and starch pasting properties

A total of 46 QTLs were associated with wild-type and waxy starch (Table 3 and Fig. 7). Most QTLs were associated starch type and distributed on chromosome 2. The top significant SNP marker (S1_36202626) had a $-\log_{10} (p \text{ value})$ of 12.03. The top significant SNP marker explained 56.89% of the phenotypic variance between waxy and wild-type starch. The second most significant SNP marker (S1_35043061) on chromosome 2 identified three genotypes of cassava, waxy (*wxwx*), heterozygous genotype (*Wxwx*), and wild-type (*WxWx*) genotype, that explained 47.04% of the phenotypic variance (Figure 3b).

A total of 21 significant QTLs of $-\log_{10} (p \text{ value}) > 6.25$ were identified for six starch properties: PV, TV, BD, FV, SB, and PT with the contribution of the variance explained (R^2) ranging from 11.55 to 78.9% for each SNP (Table 3). Two SNPs were associated with PV on chromosomes 2 and 3; SNP S1_35043061 had the lowest p value ($-\log_{10} (p \text{ value}) = 6.89$) on chromosome 2 having the highest R^2 of 77.69% (Figs. 7 and 8). Three SNPs with $-\log_{10} (p \text{ values}) > 6.25$ in a Manhattan plot were strong signals of association between TV on chromosomes 2 and 4 with the lowest p value recorded for S1_35043061 on chromosome 2 ($-\log_{10} (p \text{ value}) = 6.36$). The top significant marker explained 66.05% of the phenotypic variance. For BD, three SNPs were identified on chromosomes 2, 3 and 4, with the lowest p value found for SNP S1_76323649 on chromosome 3 ($-\log_{10} (p \text{ value}) = 6.77$). The top significant marker explained 78.9% of the phenotypic variance. For FV, nine SNPs were identified. The one on chromosomes 2 had the lowest p value recorded for SNP S1_35043061 ($-\log_{10} (p \text{ value}) = 7.23$). The top significant marker explained 75.8% of the phenotypic variance. For SB, two SNPs were detected on chromosomes 2 with the lowest p value of SNP S1_35761681 ($-\log_{10} (p \text{ value}) = 6.43$). The top significant marker explained 76.58% of the phenotypic variance. Two significant SNPs, distributed on chromosomes 2 and 3, were associated with PeT, with the lowest p value recorded for SNP S1_35043061 on chromosome 2 ($-\log_{10} (p \text{ value}) = 6.89$). The top significant marker explained 78.9% of the phenotypic variance. Two SNPs

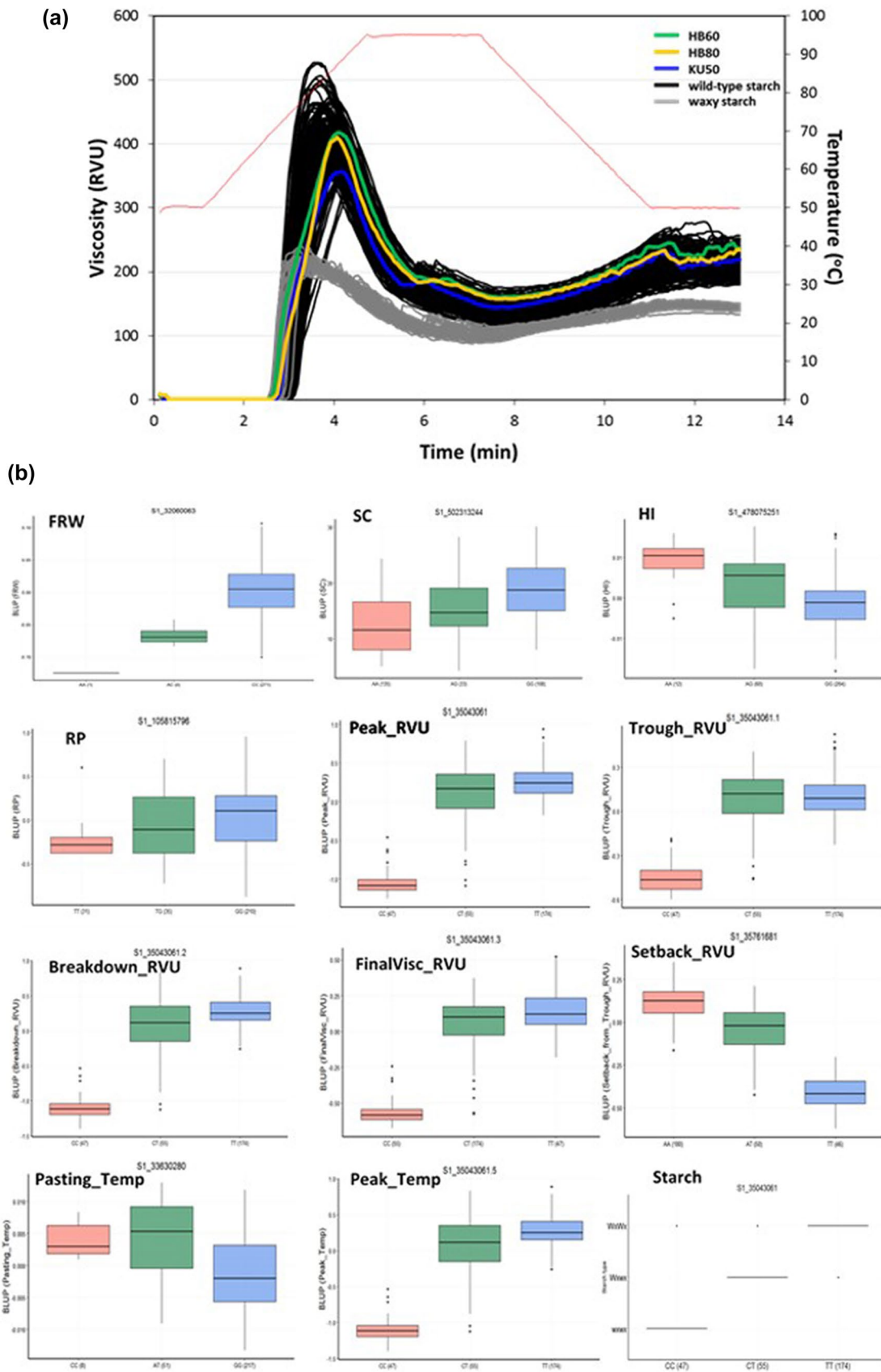


Fig. 3 a Starch paste viscosity profiles of wild-type and waxy starches of 276 cassava were dispersed in distilled water to prepare 9.21% starch suspension. A genome-wide association study identified loci affecting starch pasting properties in the cassava panels. **b** Box plots show the estimated genotypic effect on the phenotypes in the study. The significantly associated SNPs and each trait have been correlated to show the impact each genotype has on the average of the overall traits

S1_35043061 on chromosome 2 and S1_76323649 on chromosomes 3 were associated with PV, TV, BD, BV, FV, SB, and PeT, respectively. The favorable SNP alleles exhibited significantly different starch pasting properties traits and are shown in Fig. 3b.

Genomic prediction

Genomic prediction accuracies for the yield-related traits and starch properties were evaluated with all eight models. Table 4 summarizes rrBLUP, BL, BRR, BayesA, BayesB, BayesC, RKHS, and RF. The prediction accuracies for RP using tenfold cross-validation were 0.64–0.71 using the BL model (Fig. 9). The prediction accuracy of HI was 0.59 using the rrBLUP model. All models had almost the same prediction accuracy of around 0.40 to 0.55 for SC. FRW had low prediction accuracies (– 0.001–0.09) for all models. The genomic prediction accuracies for starch pasting properties were 0.7–0.8 for peak RVU, trough RVU, breakdown RVU, final viscosity RVU, and setback RVU for all models (Fig. 10). Pasting temperature and peak temperature had prediction accuracies between 0.5 and 0.7 in all models. The RFC model had the lowest predicted accuracy values compared to all model and starch property traits.

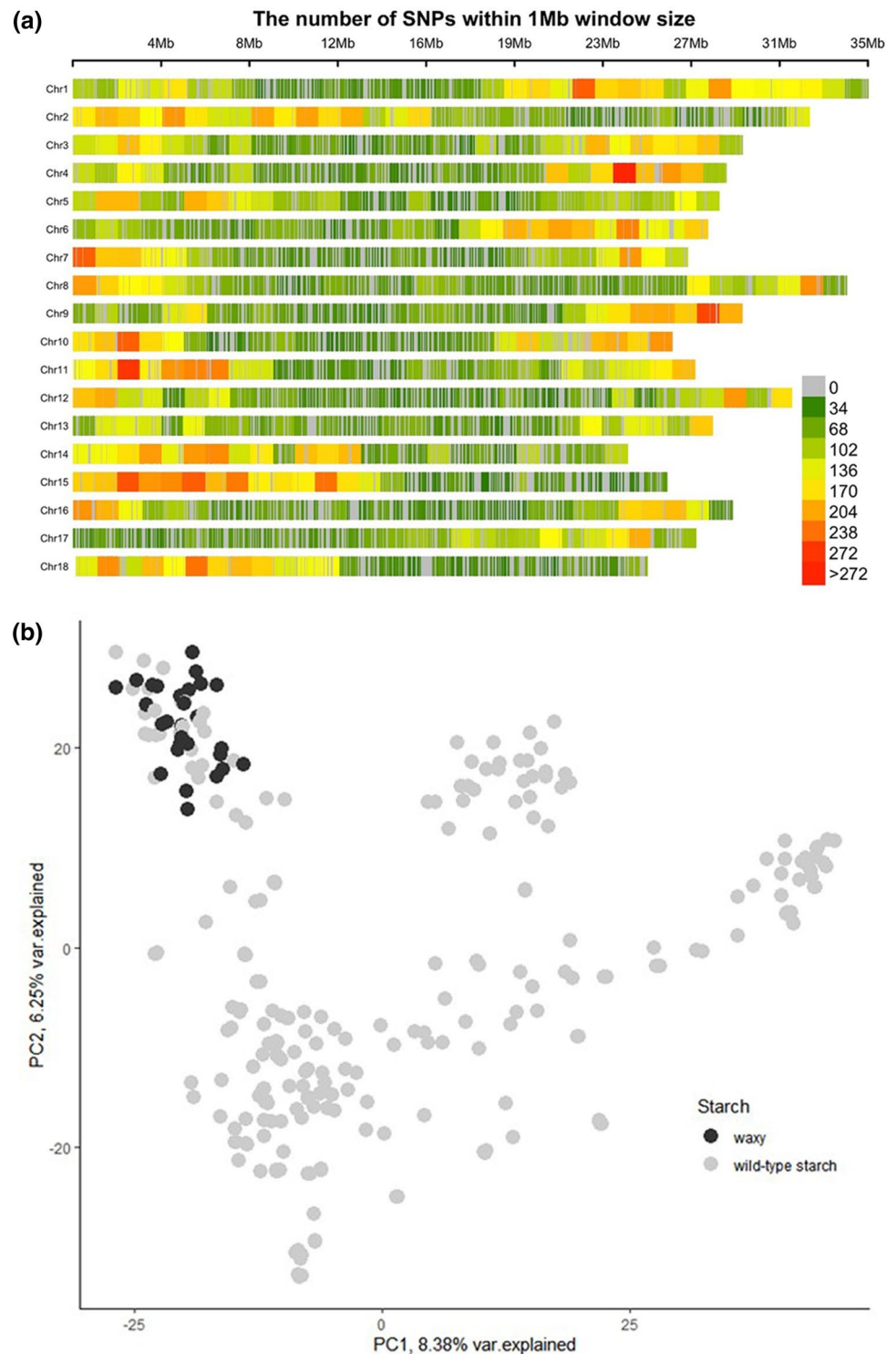
Discussion

Yield-related traits, such as FRW, SC, RP, and HI, were highly variable among the cassava genotypes (Table 1). The heritability estimates were FRW (0.43), SC (0.33), RP roots average per plant (0.40), and HI (0.57) narrow-sense heritability h^2 values ranged from 0.1 to 0.51 (Table 1). These observations are consistent with similar studies concerning yield-related traits in cassava (Wolfe et al. 2017; Andrade et al. 2019). In clonally propagated crops, nonadditive genetic effects can be effectively exploited by identifying superior genotypes as new varieties. For this reason, Wolfe et al. (2016) quantified the amount and nature of the nonadditive genetic variation in key traits in a GS breeding population of cassava. Previous findings in cassava plants were based on diallel crosses and the nonadditive genetic variation was significant, especially for yield traits (Cach et al. 2005; Calle et al. 2005; Jaramillo et al. 2005; Pérez

et al. 2005; Tumuhimbise et al. 2014; Ceballos et al. 2015). Our results indicated nonadditive (mainly epistasis) variance for fresh root yields and mostly additive inheritance for root SC. These findings confirm the conclusions of numerous diallelic studies conducted with cassava germplasms from Latin America (Cach et al. 2005; Calle et al. 2005; Jaramillo et al. 2005; Pérez et al. 2005) and Africa (Zacarias and Labuschagne 2010; Kulembeka et al. 2012; Tumuhimbise et al. 2014; Wolfe et al. 2016a). Previous studies of the general and specific combining abilities of cassava showed that fresh root yield, root number, harvest index, and plant height traits were predominantly controlled by nonadditive effects. Additive effects governed dry matter content and root diameter (Kawano 2003; Ceballos et al. 2016). Thus, the SC or dry matter content is a trait that allows greater predictive ability, even in earlier breeding stages. In contrast, parental selection for fresh root yield is more complex, requiring more refined strategies for analysis and selection (Kawano 2003; Ceballos et al. 2016). Molecular studies have also found strong non-additive genetic effects for complex traits such as fresh root yield (Wolfe et al., 2016b) The heritability of each starch property-related trait was estimated from genetic data only. The SNP-based heritability (narrow-sense heritability) of the starch property traits had high values of 0.58–0.85. There was a significant contribution of additive genetic factors in the expression of these traits. The heritability estimates recorded in our study also indicated good repeatability and reproducibility of our experimental procedures.

The pasting temperatures of the wild-type cassava starches and the waxy cassava (68.45–75.25 °C and 67.75–73.36 °C, respectively) correspond to similar pasting temperatures reported for cassava starches in other studies: 69–73 °C (Toae et al. 2019; Chaengsee et al. 2020) and 67.9–74.4 °C (Aldana and Quintero 2013). The peak viscosity of the wild-type cassava starches (327–527 RVU) was significantly higher than those of the waxy cassava starches (200–243 RVU). Peak viscosity values in this research were different from those in Sánchez et al. (2010) due to the different unit formats. In Sánchez et al (2010), peak viscosity values were reported in centipoise (cP) unit in which 12 cp equivalent to 1 rapid viscosity unit (RVU). Therefore, when data from Sánchez et al. (2010) were transformed into RVU, there were 73–81.58 RVU for normal cassava starch and 93.25 RVU for waxy starch in 5% suspension solution. There were 70–76 RVU for normal cassava starch and 85–107 RVU for waxy starch in 5% suspension solution in this research which were relatively similar to those reported in Sánchez et al. (2010). At the same analysis concentrations (9.21% w/w), Schirmer et al. (2013) reported that wild-type potato starch showed higher peak viscosity than waxy potato starch. However, at the different analysis concentration (5% w/w), Toae et al. (2019) reported that cassava waxy starches (116–131 RVU) had higher peak viscosity than

Fig. 4 **a** SNP density plot on 18 chromosomes of cassava within 1 Mb window size. **b** Scatter plot (PC1 vs PC2) coding by any major genetic groups in the data dark colored by waxy vs. wild-type starch



those of wild-type cassava starches (80–94 RVU). Wang et al. (2017) explained that the lower peak viscosity of rice waxy starch in 9.2% starch suspension compared to wild-type starch was because when the suspension concentration was higher than the close packing concentration of starch, less particle rigidity of partially swollen starch granule of

waxy type than wild type will result, and the reduced rigidity during heating led to lower peak viscosity. Evans and Lips (1992) reported that the close packing concentration for tapioca is 2.27%. Therefore, a 9.2% suspension can be applied to distinguish between waxy and wild-type starch of cassava. Even for 5% starch suspension, Raemakers et al.

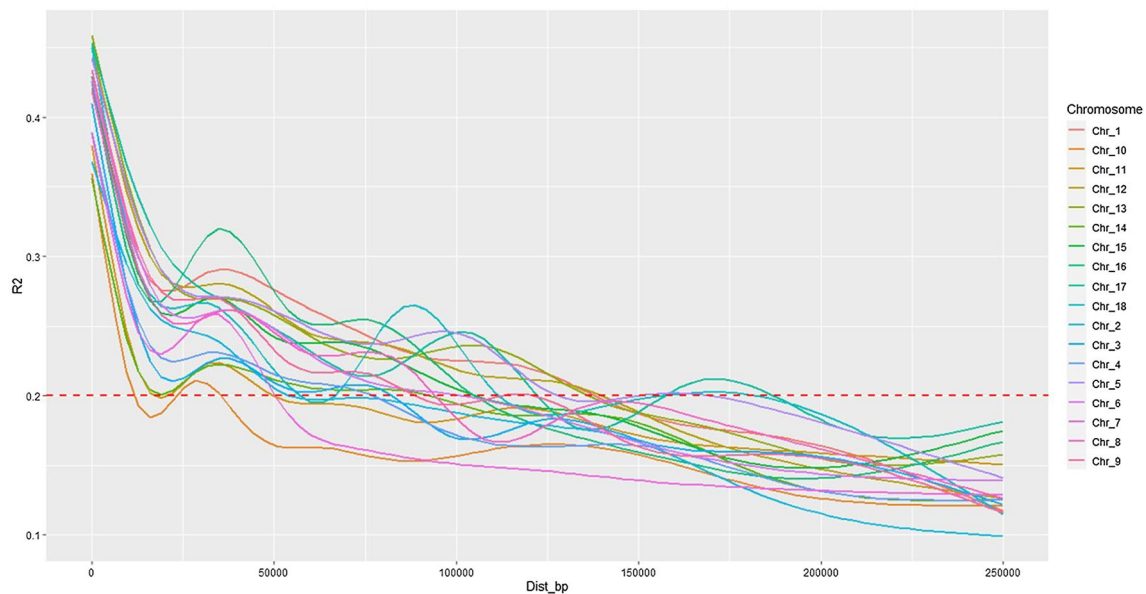


Fig. 5 LD decay by chromosome. LD plots for each chromosome and the mean LD for 18 chromosomes. Inset is the distance (bp) where LD drops below r^2 of 0.2 for each chromosome

(2005) reported that waxy cassava that was mutated by anti-sense inhibition of *GBSSI* had lower peak viscosity than that of wild type. For other pasting properties in our study, the breakdown viscosity of wild-type cassava starches (191–379 RVU) was considerably higher than those of waxy cassava starches (102–140 RVU), indicating their lower resistance to high temperatures and increased sensitivity to shearing stress. After cooling, the final viscosities of the wild-type cassava starches were significantly higher than the waxy cassava starches (179–256 and 133–151 RVU, respectively), indicating that the retrogradation tendency of the wild-type cassava starches was higher than the waxy cassava starches. The final viscosities of the wild-type cassava starches (56–99 RVU) were slightly higher than those of the waxy cassava starches (36–55 RVU). The lower breakdown and setback viscosity values indicate greater starch paste stability and a lower retrogradation tendency of waxy cassava starches. The lower peak viscosity of waxy starch can be explained by the particle rigidity of partially swollen granules. The hypothesis that a pleiotropic effect of waxy gene has a yield penalty was raised and studied in Karlström et al. (2016) in 87 waxy and 87 wild-type counterpart cassava lines from eight full-sib families showing that waxy lines had 0.8% lower dry matter content than the wild-type counterpart. However, there is no evidence of a pleiotropic effect of waxy gene on starch pasting properties. There were reports of higher fat and water soluble carbohydrate content in waxy compared to wild type in wheat (Yasui and Ashida 2011) and rice (Pérez et al. 2019), but starch contents in both studies were reported to be inconsistent and unchanged compared to wild type,

respectively. In this study, there were no significant SNPs detected for SC on spontaneous mutants of *GBSSI* gene (Aiemnaka et al. 2012). However, the preliminary results of waxy cassava had low SC or DMC comparing with commercial varieties. Therefore, it is possibly due to the linkage between the waxy gene and a locus for SC or DMC. Then, the further crosses among waxy cassava (*wxwx*) or improved heterozygous genotypes (*Wxwx*) to select for waxy with higher SC or DMC could help increase the chance to breaking the linkage waxy starch and low desirable SC.

Linkage disequilibrium

LD in cross-pollinated species decays more rapidly than in self-pollinated species due to the more effective recombination that occurs in the former (Flint-Garcia et al. 2003). We observed rapid LD decay in the 276 genotypes with a $r^2 < 0.2$ at approximately 243 kb, which was higher than in Brazilian germplasms (~107 kb; $r^2 < 0.1$; Andrade et al. 2019), Brazilian cassava (~20 kb; $r^2 < 0.2$; Albuquerque et al. 2018), East–West African (~50 kb; $r^2 < 0.2$) datasets (Wolfe et al. 2016), HapMap (3 kb; $r^2 = 0.1$; Ramu et al. 2017), and elite IITA cassava breeding genotypes (10 kb; $r^2 = 0.212$; Rabbi et al. 2020), and lower compared to West African (~2 Mb; $r^2 < 0.1$; Rabbi et al. 2017). Given the allogamous nature of cassava, the LD pattern may reduce prediction accuracy since LD is crucial for GS breeding (Jannink 2010; Yabe et al. 2018); however, the LD observed in this study indicates the potential for improved accuracies when using this population for GS breeding.

Table 3 Most significant SNPs associated with four yield-related traits and seven starch pasting property traits identified fitting the FarmCPU and BLINK in GAPIT and annotations of candidate genes harboring the significant SNPs using the phytozome 12 portal link to the biomart (<https://phytozome.jgi.doe.gov/biomart/>)

Trait	SNPID	Alleles	Chr	$-\log_{10}(p \text{ value})$	R^2 (%)	Position	Gene name	Functional annotation	Gene ontology
FRW	SI_32060063	C/A	1	5.00	11.8	32,060,063	<i>Manes.01G240800</i>	<i>PF02365 No apical meristem (NAM) protein</i>	
	SI_478075251	G/A	17	7.67	14.0	12,189,438	<i>Manes.17G024400</i>	<i>PF05938 Plant self-incompatibility protein SI</i>	
HI							<i>Manes.17G024500</i>	<i>PTHR15184 ATP SYNTHASE</i>	
							<i>Manes.17G024600</i>	<i>PF09747 Coiled-coil domain-containing protein (DUF2052)</i>	
	SI_184745424	T/C	7	7.26	1.4	2,813,010			
	SI_389866391	C/G	14	6.48	25.3	4,008,933	<i>Manes.14G050300</i>	<i>PF04185 Phosphoesterase family</i>	GO:0,016,788 hydrolase activity, acting on ester bonds
	SI_32891402	A/T	1	7.35	0.1	32,891,402	<i>Manes.01G253200</i>	<i>PTHR11994 60S RIBOSOMAL PROTEIN L11-RELATED; PTHR11994:SF4</i> <i>54S RIBOSOMAL PROTEIN L7, MITOCHONDRIAL</i>	
SC	SI_100762573	A/T	4	7.02	0.1	3,958,753	<i>Manes.04G034200</i>	<i>PF12796 Ankyrin repeats (3 copies)</i>	GO:0,005,515 protein binding;
	SI_502313244	G/A	18	6.96	10.6	9,038,408	<i>Manes.18G102200</i>	<i>PF02225 PA domain; PFI3639 Ring finger domain</i>	GO:0,008,270 zinc ion binding
							<i>Manes.18G102300</i>	<i>PF05695 Plant protein of unknown function (DUF825)</i>	GO:0,005,524 ATP binding;
							<i>Manes.01G121600</i>	<i>PTHR12746 NONSENSE-MEDIATED MRNA DECAY PROTEIN 3</i>	GO:0,009,507 chloroplast;
RP	SI_172308681	T/G	6	6.48	4.8	18,316,327	<i>Manes.06G068300</i>	<i>PTHR31769:SF5 EXPRESSED PROTEIN</i>	
	SI_105815796	G/T	4	4.96	2.5	9,011,976	<i>Manes.04G052900</i>		
Starch	SI_36202626	C/T	2	12.04	56.8	1,242,805	<i>Manes.02G015200</i>	<i>PTHR30540 OSMOTIC STRESS POTASSIUM TRANSPORTER</i>	GO:0,015,079 potassium ion transmembrane transporter activity; GO:0,016,020 membrane;
							<i>Manes.02G015100</i>	<i>PTHR24349 SERINE/THREONINE-PROTEIN KINASE; PTHR24349:SF95 CALCIUM-DEPENDENT PROTEIN KINASE 3</i>	GO:0,071,805 potassium ion transmembrane transport
							<i>Manes.02G015200</i>	<i>PF02705 K + potassium transporter</i>	GO:0,004,672 protein kinase activity; GO:0,005,524 ATP binding; GO:0,006,468 protein phosphorylation
									GO:0,015,079 potassium ion transmembrane transporter activity; GO:0,016,020 membrane;
									GO:0,071,805 A process in which a potassium ion is transported from one side of a membrane to the other

Table 3 (continued)

Trait	SNPID	Alleles	Chr	$-\log_{10}(p \text{ value})$	R^2 (%)	Position	Gene name	Functional annotation	Gene ontology
Starch	SI_35043061	T/C	2	10.44	56.9	83,240	<i>Manes.02G000500</i>	<i>PF00891 O-methyltransferase; PF08100 Dimerization domain</i>	GO:0,008,171 O-methyltransferase activity; GO:0,046,983 protein dimerization activity
								<i>Manes.02G000600</i>	<i>PTHR11562 CATION EFFLUX PROTEIN/ ZINC TRANSPORTER</i>
Peak_RVU	SI_35043061	T/C	2	6.90	77.7	83,240	<i>Manes.02G000700</i>	<i>PF13639 Ring finger domain</i>	GO:0,005,515 protein binding; GO:0,008,270 zinc ion binding;
							<i>Manes.02G000500</i>	<i>PWY-1121 suberin monomers biosynthesis</i>	GO:0,008,171 Catalysis of the transfer of a methyl group to the oxygen atom of an acceptor molecule
Peak_RVU	SI_76323649	C/T	3	6.54	0.4	8,932,332	<i>Manes.02G000600</i>	<i>K14689 solute carrier family 30 (zinc transporter), member 2</i>	GO:0,006,812 The directed movement of cations, atoms or small molecules with a net positive charge, into, out of or within a cell, or between cells
							<i>Manes.02G000700</i>	<i>PTHR14155:SF110 RING/U-BOX DOMAIN-CONTAINING PROTEIN</i>	GO:0,005,515 Interacting selectively and non-covalently with any protein or protein complex
Trough_RVU	SI_35043061	T/C	2	6.37	66.1	83,240	<i>Manes.03G071000</i>	<i>PWY0-1319 CDP-diacylglycerol biosynthesis II</i>	GO:0,008,171 Catalysis of the transfer of a methyl group to the oxygen atom of an acceptor molecule
							<i>Manes.02G000500</i>	<i>PWY-1121 suberin monomers biosynthesis</i>	GO:0,006,812 The directed movement of cations, atoms or small molecules with a net positive charge, into, out of or within a cell, or between cells
SI_105948813	T/C	4	6.24	51.2	9,144,993	<i>Manes.02G000600</i>	<i>K14689 solute carrier family 30 (zinc transporter), member 2</i>	GO:0,005,515 Interacting selectively and non-covalently with any protein or protein complex	
						<i>Manes.04G053300</i>	<i>PF00076 RNA recognition motif</i>	GO:0,003,676 Interacting selectively and non-covalently with any nucleic acid	
SI_105949949	G/A	4	6.24	0.5	9,146,129	<i>Manes.04G053300</i>	<i>PF00076 RNA recognition motif</i>	GO:0,003,676 Interacting selectively and non-covalently with any nucleic acid	

Table 3 (continued)

Trait	SNPID	Alleles	Chr	$-\log_{10}(p \text{ value})$	R^2 (%)	Position	Gene name	Functional annotation	Gene ontology
Breakdown_RVU	S1_76323649	C/T	3	6.77	0.3	8,932,332	<i>Manes.03G071000</i>	<i>PWY0-1319 CDP-diacylglycerol biosynthesis II</i>	
	S1_35043061	T/C	2	6.76	78.9	83,240	<i>Manes.02G000500</i>	<i>PWY-1121 suberin monomers biosynthesis</i>	GO:0,008,171 Catalysis of the transfer of a methyl group to the oxygen atom of an acceptor molecule
							<i>Manes.02G000600</i>	<i>K14689 solute carrier family 30 (zinc transporter), member 2</i>	GO:0,006,812 The directed movement of cations, atoms or small molecules with a net positive charge, into, out of or within a cell, or between cells
FinalVisc_RVU	S1_118281691	G/A	4	6.28	0.6	21,477,871	<i>Manes.02G000700</i>	<i>PTHR14155:SF110 RING/U-BOX DOMAIN-CONTAINING PROTEIN</i>	GO:0,005,515 Interacting selectively and non-covalently with any protein or protein complex
							<i>Manes.04G076800</i>	<i>PTHR11544 COLD SHOCK DOMAIN-CONTAINING PROTEINS</i>	GO:0,003,676 Interacting selectively and non-covalently with any nucleic acid
							<i>Manes.04G076900</i>	<i>PTHR23315 BETA CATENIN-RELATED ARMADILLO REPEAT-CONTAINING</i>	GO:0,005,515 Interacting selectively and non-covalently with any protein or protein complex
FinalVisc_RVU	S1_35043061	T/C	2	7.23	74.1	83,240	<i>Manes.04G077000</i>	<i>PTHR24078:SF231 CHAPERONE PROTEIN DNAJ</i>	
							<i>Manes.02G000500</i>	<i>PWY-1121 suberin monomers biosynthesis</i>	GO:0,008,171 Catalysis of the transfer of a methyl group to the oxygen atom of an acceptor molecule
							<i>Manes.02G000600</i>	<i>K14689 solute carrier family 30 (zinc transporter), member 2</i>	GO:0,006,812 The directed movement of cations, atoms or small molecules with a net positive charge, into, out of or within a cell, or between cells
	S1_35761681	A/T	2	7.01	72.6	801,860	<i>Manes.02G000700</i>	<i>PTHR14155:SF110 RING/U-BOX DOMAIN-CONTAINING PROTEIN</i>	GO:0,005,515 Interacting selectively and non-covalently with any protein or protein complex
							<i>Manes.02G0008100</i>	<i>PTHR13215:SF6 RNA POLYMERASE II TRANSCRIPTIONAL COACTIVATOR KELP</i>	GO:0,003,677 DNA binding; GO:0,003,713 transcription coactivator activity; regulation of transcription, GO:0,006,355 DNA-templated
							<i>Manes.02G0008400</i>	<i>PF13474 SnoaL-like domain</i>	
							<i>Manes.02G0008300</i>	<i>PF08766 DEK C terminal domain</i>	
							<i>Manes.02G0008200</i>	<i>KOG1515 Arylacetamide deacetylase</i>	

Table 3 (continued)

Trait	SNPID	Alleles	Chr	$-\log_{10}(p \text{ value})$	R^2 (%)	Position	Gene name	Functional annotation	Gene ontology
	S1_35761682	G/T	2	7.01	72.6	801,861	<i>Manes.02G008100</i>	<i>PTHR13215:SF6 RNA POLYMERASE II TRANSCRIPTIONAL COACTIVATOR KELP</i>	GO:0,003,677 DNA binding; GO:0,003,713 transcription coactivator activity; regulation of transcription, GO:0,006,355 DNA-templated GO:0,006,629 lipid metabolic process
	S1_35830922	A/G	2	6.64	72.2	871,101	<i>Manes.02G009100</i>	<i>PTHR31479:SF2 ALPHA/BETA-HYDROLASES SUPERFAMILY PROTEIN</i>	GO:0,003,924 GTPase activity; GO:0,005,525 GTP binding GO:0,005,515 protein binding; GO:0,016,773 phosphotransferase activity, alcohol group as acceptor
	S1_35880519	A/T	2	6.64	72.2	920,698	<i>Manes.02G009200</i>	<i>PTHR11566:SF58 DYNAMIN-RELATED PROTEIN 1A</i>	
	S1_35875738	T/C	2	6.46	72.3	915,917	<i>Manes.02G010200</i>	<i>PTHR10666:SF134 PHOSPHATIDYLINOSITOL 4-KINASE GAMMA 2</i>	
							<i>Manes.02G009000</i>	<i>PTHR31852:SF22 LATE EMBRYO-GENESIS ABUNDANT HYDROXY-PROLINE-RICH GLYCOPROTEIN</i>	
	S1_35880519	A/T	2	6.64	72.2	920,698	<i>Manes.02G010200</i>	<i>PF00234</i> Protease inhibitor/seed storage/LTP family	
	S1_35875738	T/C	2	6.46	72.3	915,917	<i>Manes.02G010100</i>	<i>PF13947</i> Wall-associated receptor kinase galacturonan-binding	
	S1_35843804	C/T	2	6.46	72.3	883,983	<i>Manes.02G009200</i>	<i>PTHR11566:SF58 DYNAMIN-RELATED PROTEIN 1A</i>	GO:0,003,924 GTPase activity; GO:0,005,525 GTP binding
	S1_35197419	A/C	2	6.40	75.8	237,598	<i>Manes.02G002300</i>	<i>PF14215 bHLH-MYC and R2R3-MYB transcription factors N-terminal</i>	GO:0,046,983 protein dimerization activity GO:0,046,983 protein dimerization activity
	S1_35836599	C/T	2	6.27	72.3	876,778	<i>Manes.02G002400</i>	<i>PF14215 bHLH-MYC and R2R3-MYB transcription factors N-terminal</i>	GO:0,006,629 lipid metabolic process
							<i>Manes.02G009100</i>	<i>PTHR31479:SF2 ALPHA/BETA-HYDROLASES SUPERFAMILY PROTEIN</i>	
							<i>Manes.02G009200</i>	<i>PTHR11566:SF58 DYNAMIN-RELATED PROTEIN 1A</i>	GO:0,003,924 GTPase activity; GO:0,005,525 GTP binding
							<i>Manes.02G009000</i>	<i>PTHR31852:SF22 LATE EMBRYO-GENESIS ABUNDANT HYDROXY-PROLINE-RICH GLYCOPROTEIN</i>	

Table 3 (continued)

Trait	SNPID	Alleles	Chr	Position	Gene name	Functional annotation	Gene ontology
Setback_RVU	S1_35761681	A/T	2	801,860	<i>Manes.02G008100</i>	<i>PTHR13215:SF6 RNA POLYMERASE II TRANSCRIPTIONAL COACTIVATOR KELP</i>	GO:0,003,677 DNA binding; GO:0,003,713 transcription coactivator activity; regulation of transcription, GO:0,006,355 DNA-templated
					<i>Manes.02G008400</i>	<i>PF13474 Snoal-like domain</i>	
					<i>Manes.02G008300</i>	<i>PF08766 DEK C terminal domain</i>	
					<i>Manes.02G008200</i>	<i>KOG1515 Arylacetylase</i>	
Pasting_Temp	S1_33630280	G/C	1	33,630,280	<i>Manes.01G264200</i>	<i>KOG1515 Arylacetylase</i>	
					<i>Manes.01G264300</i>	<i>KOG1515 Arylacetylase</i>	GO:0,008,152 metabolic process; GO:0016787hydrolase activity
Peak_Temp	S1_76323649	C/T	3	8,932,332	<i>Manes.03G071000</i>	<i>PWY0-1319 CDP-diacylglycerol biosynthesis II</i>	

GWAS and candidate genes

We scanned for associations between SNPs using CMLM, BLINK, and FarmCPU to identify genomic regions associated with yield-related traits, starch type of waxy, and wild-type starch and starch pasting property traits in cassava, respectively. The MLM model controls the P value inflation well but leads to false negatives, which weakens the identification of true associations (Zhang et al. 2010). To deal with this problem, we used the compressed MLM model (CMLM), which clusters individuals into groups and fits the genetic values of groups as random effects in the model (Zhang et al. 2010). The CMLM method improves statistical power compared to regular MLM methods (Zhang et al. 2010). FarmCPU is a multi-locus model developed to control false positives without false negatives (Liu et al. 2016). The FarmCPU model uses a modified MLM method, MLMM, and incorporates multiple markers simultaneously as covariates in a stepwise MLM to partially remove the confusion between testing markers and kinship. To eliminate this confusion, MLMM is divided into two parts: a fixed-effect model (FEM) and a random effect model (REM) and uses them iteratively. The FEM uses single marker testing and multiple associated markers as covariates to control false positives. The associated markers are estimated in the REM to define kinship to avoid model overfitting in the FEM. We evaluated these three models for false positives and false negatives based on the Q–Q plots. A sharp deviation from the expected P value distribution in the tail area would indicate that a model appropriately controlled both false positives and false negatives. Q–Q plots of complex models including CMLM, MLMM, and FarmCPU had a straight line with a slightly deviated tail, indicating that these models reduced false positives. We found that the FarmCPU model controlled both false positives and false negatives more effectively than the CMLM and MLMM models for yield-related traits.

Usually, many QTLs for a given trait are detected in different populations from different environments and the effect of QTLs by environment interaction is inconsistent. In previous studies, Andrade et al. (2019), Ozimati et al. (2019), and Yonis et al. (2020) detected root yield marker–trait associations. As Bonferroni correction method may cause more false-negative result (Kaler and Purcell, 2019), one QTL on chromosome 4 associated with RP and a QTL on chromosome 1 for FRW were mentioned in this study, as the $-\log_{10} p$ value was quite close to the threshold value, for further marker validation with the caution that this SNP did not pass the Bonferroni threshold in the FarmCPU analysis. Wolfe et al. (2017) and Zhang et al. (2018) used a compressed MLM approach to significantly decrease the number of false positives for yield components (number of

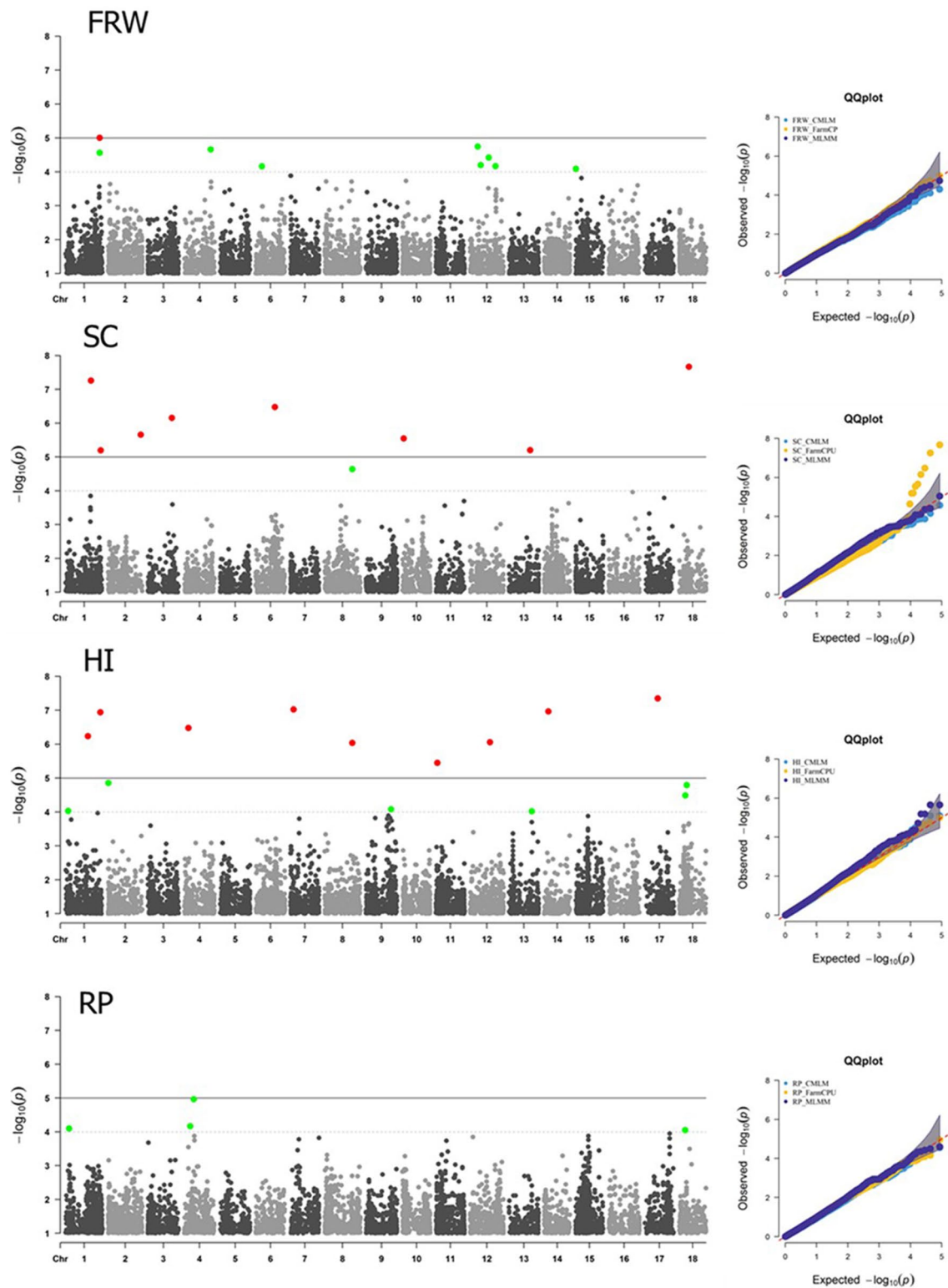


Fig. 6 Manhattan plot of SNP markers associated with FRW, SC, HI and RP the quantile–quantile (Q–Q) plots in cassava using the fixed and random model circulating probability unification (FarmCPU; yel-

low dot), compressed mixed linear model (CMLM; navy blue), and the multi-locus mixed model (MLMM; blue dot) (color figure online)

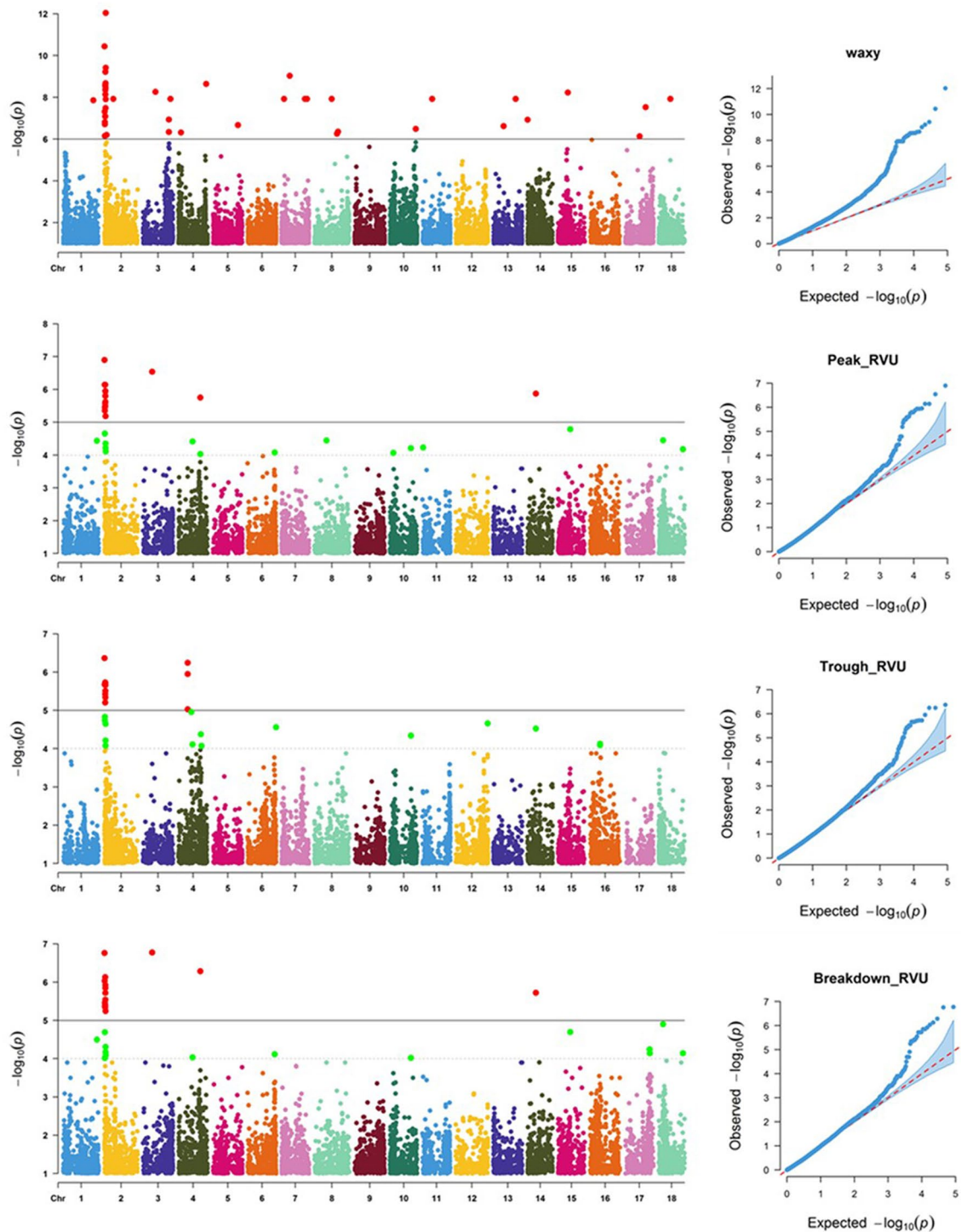


Fig. 7 Manhattan plot of SNP markers associated with starch (waxy and wild-type) implemented with BLINK method, peak RVU. Trough RVU, breakdown RVU were implemented with CMLM method and quantile–quantile (Q–Q) plots in cassava found with the fixed and

random model circulating probability unification (FarmCPU; blue dot). A gray line represents the significance threshold. $-\log_{10}(p \text{ values})=6.25$ represents the threshold from the Bonferroni correction method (color figure online)

storage roots, storage roots weight, and dry mass weight), and storage root quality (dry matter content and starch content), and found one SNP locus located on chromosome 2

that was associated with fresh weight and dry mass weight. MLM-based GWAS for harvest index, the ratio of fresh root weight to total plant weight, uncovered two genomic regions

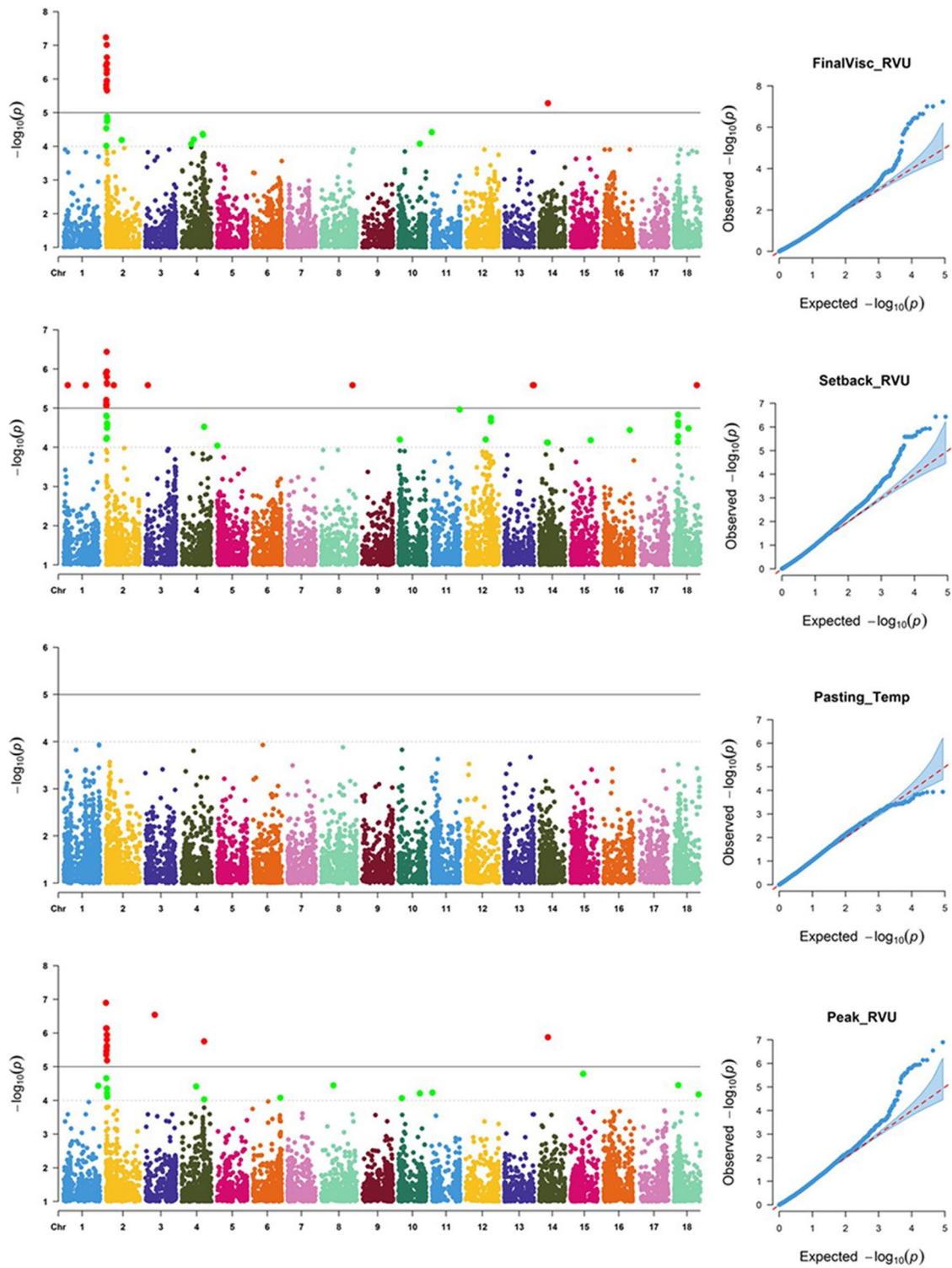


Fig. 8 Manhattan plot of SNP markers associated with final viscosity RVU, setback RVU, pasting temp RVU, peak temp RVU, and quantile–quantile (Q–Q) plots in cassava found with the CMLM method.

A gray line represents the significance threshold. $-\text{Log}_{10}(p \text{ values})=6.25$ represents the threshold from the Bonferroni correction method

Table 4 Average predictabilities and their standard error (SE) for 11 traits using eight prediction models

Model	FRW	SC	HI	RP	Peak viscosity (RVU)	Trough (RVU)	Breakdown (RVU)	Final viscosity (RVU)	Set back (RVU)	Pasting temperature (°C)	Peak time (minutes)
rrBLUP	0.0917 ± 0.017	0.5238 ± 0.09	0.5941 ± 0.10	0.7142 ± 0.13	0.7654 ± 0.03	0.7556 ± 0.02	0.7587 ± 0.03	0.8232 ± 0.02	0.8078 ± 0.02	0.6079 ± 0.03	0.7587 ± 0.03
BayesA	− 0.001 ± 0.000	0.4905 ± 0.08	0.5775 ± 0.10	0.6963 ± 0.12	0.7664 ± 0.03	0.7511 ± 0.03	0.7531 ± 0.03	0.8233 ± 0.02	0.8016 ± 0.02	0.5838 ± 0.03	0.7531 ± 0.03
BayesB	0.0284 ± 0.005	0.5037 ± 0.09	0.5827 ± 0.10	0.7050 ± 0.12	0.7754 ± 0.03	0.7601 ± 0.02	0.7619 ± 0.03	0.8259 ± 0.02	0.8067 ± 0.02	0.5983 ± 0.03	0.7619 ± 0.03
BayesC	0.0178 ± 0.003	0.5062 ± 0.09	0.5842 ± 0.10	0.7076 ± 0.12	0.7689 ± 0.03	0.7591 ± 0.02	0.7615 ± 0.03	0.8264 ± 0.02	0.807 ± 0.02	0.5903 ± 0.03	0.7615 ± 0.03
BL	0.0372 ± 0.007	0.5018 ± 0.09	0.5896 ± 0.10	0.7189 ± 0.13	0.7822 ± 0.02	0.7669 ± 0.02	0.7774 ± 0.03	0.8292 ± 0.02	0.8181 ± 0.02	0.6118 ± 0.03	0.7774 ± 0.03
BRR	0.0397 ± 0.007	0.5139 ± 0.09	0.5884 ± 0.10	0.7108 ± 0.12	0.7757 ± 0.03	0.7629 ± 0.02	0.7675 ± 0.03	0.8278 ± 0.02	0.8082 ± 0.02	0.6006 ± 0.03	0.7675 ± 0.03
RFC	0.0024 ± 0.001	0.4005 ± 0.07	0.5085 ± 0.09	0.6460 ± 0.11	0.6459 ± 0.03	0.5999 ± 0.03	0.5993 ± 0.03	0.6544 ± 0.03	0.6038 ± 0.03	0.3312 ± 0.03	0.5993 ± 0.03
RKHS	0.0556 ± 0.01	0.5577 ± 0.10	0.5913 ± 0.10	0.7166 ± 0.13	0.7756 ± 0.03	0.7616 ± 0.02	0.7679 ± 0.03	0.8277 ± 0.02	0.8075 ± 0.02	0.6019 ± 0.03	0.7656 ± 0.03

FRW Fresh root weight, SC Starch content, HI Harvest index, RP Number of roots per plant. Predictabilities are averaged over 30 cross-validation runs

that were significantly associated with the trait. The first peak is in chromosome 2, tagged by SNP S2_2809137 (p value = 3×10^{-8}). The second locus occurred on chromosome 12, with SNP S12_6055806 showing the strongest association with the trait (p value = 5.4×10^{-24}). MLM analysis revealed two major loci, of which only one was previously reported (Rabbi et al. 2017). The most significant locus occurred on chromosome 1 in the 24.64 Mb region, and the second locus was tagged on chromosome 12. This study used FarmCPU in GWAS for yield-related traits, instead of CMLM or MLMM (Fig. 6), which could exclude the false-positive associations. The Q–Q plots also suggested that the false-positive associations in this study were well controlled for the GWAS of the four yield-related traits.

GWAS has been applied to dissect the genetic architecture of several complex traits in cassava (Zhang et al. 2018; Rabbi et al. 2020). However, no previous studies have focused on GWAS for starch pasting properties in cassava. Here, we performed GWAS for seven pasting properties in a 276 wild-type and waxy cassava panel with 89,934 SNPs and identified 21 significant QTLs using CMLM model. The QTLs detected in this study were compared with those in previous studies that used F_1 mapping population (Thanyasiriwat et al. 2014). As a result, the major genes of interest that encode the family of glycosyl or glucosyl transferases and hydrolases corresponded to the QTLs identified in previous studies. The remaining could possibly be novel QTLs (Table 3).

Notably, the detected loci may not be the actual causative loci due to false positives caused by LD or population structure. To understand the molecular basis of starch pasting properties, we further investigated the locations of associated QTLs for possible candidate genes by examining the candidate genes within 50 kb downstream and upstream of the identified QTLs (Table 3). According to functional annotations, the candidate genes involving in yield-related traits which were fresh root weight, starch content, and harvest index were in the group of DNA-binding protein such as histone H2A and WRKY DNA-binding domain and in the group of protein and ATP binding gene such as ring finger domain and F-box domain. Some specific protein function such as no apical meristem (NAM) protein, protein phosphatase 2A regulatory B subunit, LYCOPENE BETA/EPSILON CYCLASE, stress-responsive A/B barrel, plant self-incompatibility protein S1, ATP SYNTHASE, phosphoesterase, diacylglycerol kinase, and protein kinase were linked with the yield-related traits suggesting the complex of plant interaction with environmental stress signal involving with cassava yield. Some of the candidate genes or their homologs are known to be linked to starch properties. Another example is the chemical reactions and pathways involving carbohydrates and carbohydrate metabolic

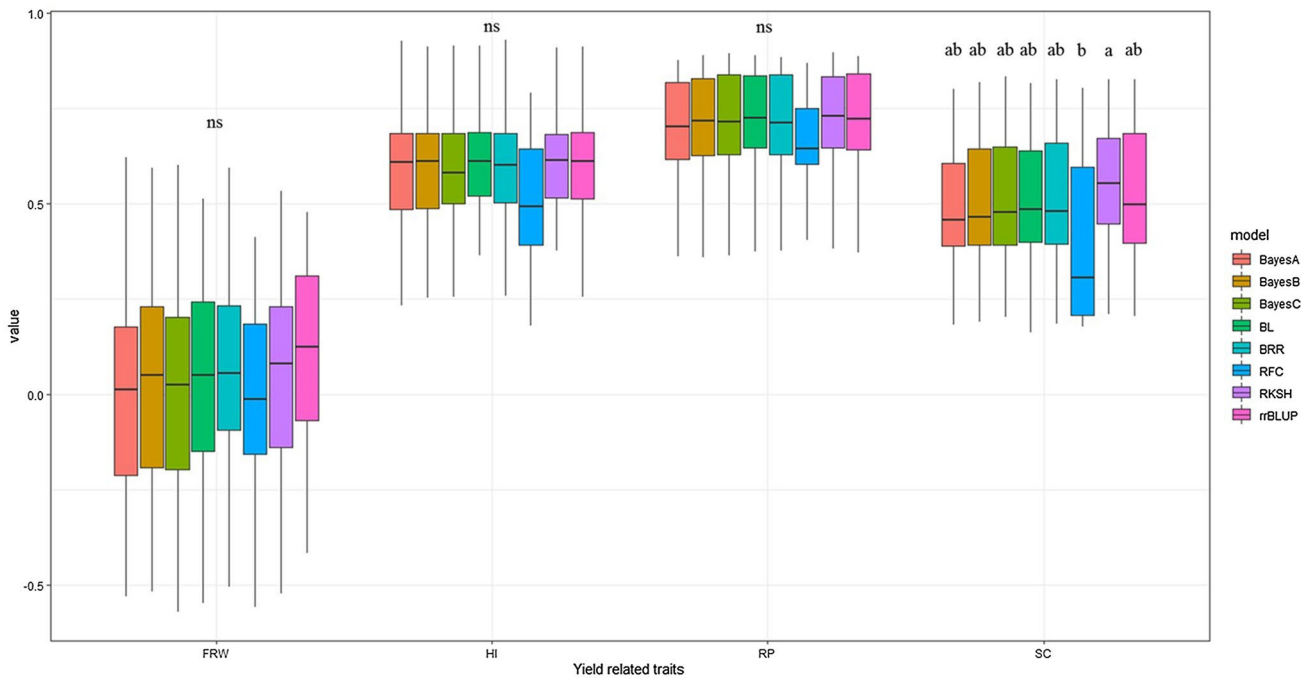


Fig. 9 Predictive yield-related traits based on the Pearson correlations between GEBVs and true breeding values from eight statistical models using tenfold cross-validation with 9:1 for the training set and a

validation set with 30 replications. Different capital letters above the group labels indicate significant differences between groups

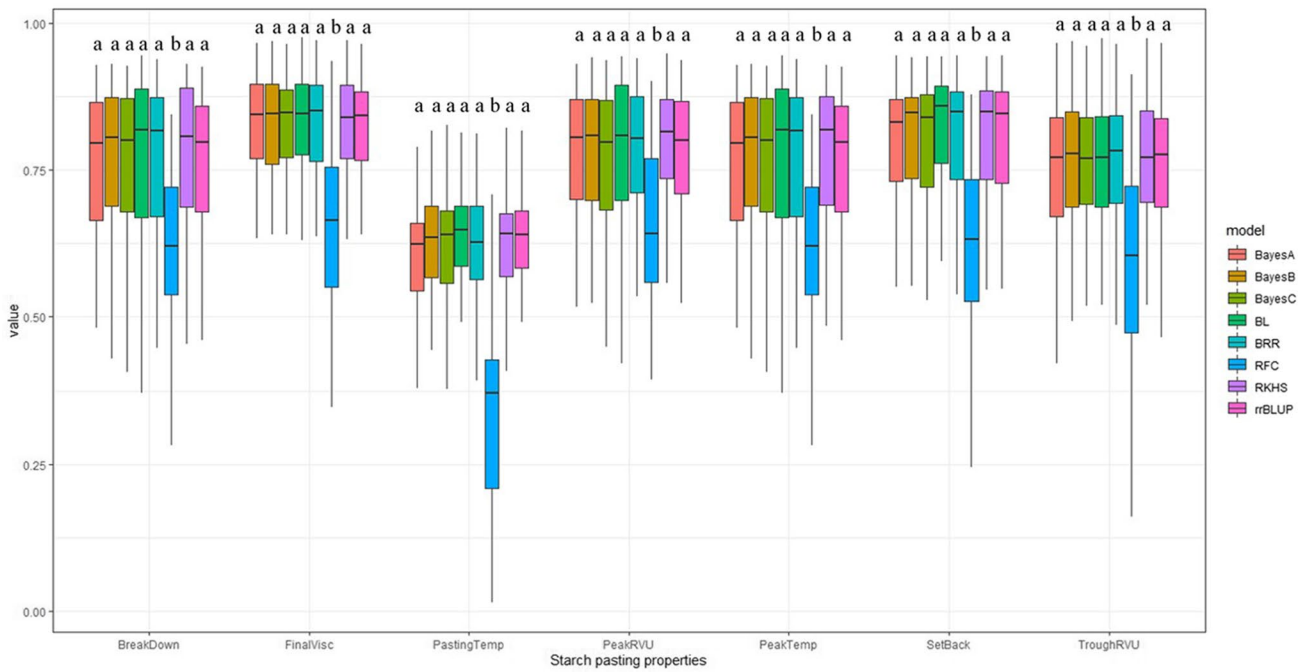


Fig. 10 Predictive ability of starch pasting property traits based on the Pearson correlations between GEBVs and true breeding values using eight statistical models and tenfold cross-validation with 9:1 for the training set and a validation set with 30 replications

processes (*Manes.04G057600.1*). *Manes.02G000500* encodes for suberin monomers biosynthesis enzyme. Suberin is the part of cell wall polymer complex. Interestingly, *Manes.02G000500* located in chromosome 2 (position 83,240) which is near *Manes.02G001000* encoding for *GBSSI* at position 117,671 and this QTL of suberin biosynthesis enzyme encoding gene was related to various traits of starch pasting properties including peak viscosity, trough viscosity, breakdown, final viscosity, and peak temperature. *bHLH-MYC* and *R2R3-MYB* transcription factors (*Manes.02G002300*) which involve in epicuticular wax synthesis were detected to link with peak viscosity and final viscosity as well (Elango et al. 2020). *Manes.02G011700* encodes pectinesterase, one of the most heat-resistant enzymes, leading to gelatinization and precipitation of pectin in puree and juice with subsequent loss of the juice's cloudy appearance (Salas-Tovar et al. 2017). Other genes related to the synthesis of fat were shown to link with starch pasting properties as well such as CDP-diacylglycerol biosynthesis (*Manes.03G071000*) and Beta-ketoacyl synthase (*Manes.02G007700*). The gene encoding DYNAMIN-RELATED PROTEIN which is GTP binding protein (*Manes.02G009200*) was associated with final viscosity. Dynamin protein was reported to involve the late stage of amyloplast division that determines the size of starch granule and affects starch gelatinization peak temperature (Yun and Kawagoe 2009).

Genomic prediction accuracies

The Pearson's correlations (r) between the observed phenotypic values and the genomic estimated breeding values (GEBVs) were based on the tenfold cross-validation scheme to evaluate the efficiency of GS for yield-related traits and starch pasting property traits in wild-type and waxy cassava. Random sampling training and validation sets were repeated 30 times, and the mean of the correlations was defined as the genomic predictive ability (PA; Spindel et al. 2015; Gouy et al. 2013; Heffner et al. 2010). Usually, a large sample of training and test data sets will be generated to estimate PA, which generally results in a normal or nearly normal PA distribution. The mean or median of the distribution represents the GS predictive ability of a trait (Rutkoski et al. 2012). In this study, the mean PA was used to represent the predictive ability under a certain training population size with a combination of cassava panels, GS models, and marker sets. The PA for the yield-related traits ranged -0.001 – 0.71 depending on trait, and 0.33 – 0.82 for the starch property traits, which was comparable to the values implemented for GS in cassava breeding programs (Wolfe et al. 2016, 2017; Okeke et al. 2017; Ozimati et al. 2018; Kayondo et al. 2018; Andrade et al. 2019; Somo et al. 2020; Yonis et al. 2020).

Generally, PA is related to the heritability. In this study, HI trait was moderately heritable (0.51; Table 1) and PA was 0.50 – 0.59 . Whereas the heritability of FRW was low (0.10), FRW had low PA (-0.001 – 0.09). These results correspond to Wolfe (2016b, 2017) and Andrade et al. (2019). Due to the higher heritability (0.58 – 0.85) of starch property traits, their direct phenotypic selection could be sufficient for the selection in a breeding program. In contrast, PA for FRW a higher genomic heritability is needed for predicting yet-to-be-observed phenotypes. As the heritability of the trait decreases, the ratio of environmental variance (residual) to genetic variance increases (Momen et al. 2018). For genomic evaluation programs to succeed, markers should have an acceptable level of LD with QTL so that the marker can express the QTL effects efficiently in the population. The successful implementation of genomic prediction in the process of selection is determined by the ability of the model to predict or estimate the genetic potential of new breeding lines for a specific trait. Training populations have a major impact on model performance, especially its size, genetic composition, heritability, and relatedness to the selection population (Isidro et al. 2015). FRW is an important yield-related trait in cassava. Low heritability traits such as FRW are complex traits that control by many small effects of gene interaction and genotype by environment interaction (Kawano 2003; Ceballos et al. 2016; Wolfe et al. 2016b; 2017). The additional costs of genotyping would be counterbalanced by a higher reliability for predicting FRW. However, high-density markers and analytical skills for genomic prediction for nonadditive gene effects and environmental interaction are required for low heritability traits.

GS models

In GS, the best prediction model is commonly identified by the cross-validation of phenotypic and genotypic data from a training set (Heffner et al. 2009). We tested eight models and found that the lower PA values of FRW were also reported by Wolfe et al. (2017) and Andrade et al. (2019), who reported low PA for fresh root yield (0.18 – 0.37 , and 0.475 , respectively). Overall, previous studies of crops and traits reported only slight differences in performance among the different genomic prediction algorithms (Asoro et al. 2011; Crossa et al. 2013; Heslot et al. 2012; Lorenz et al. 2012). We found that FRW, a more complex and lower H^2 trait, generally had lower PA than a simpler and higher H^2 trait, such as RP. This is consistent with other studies, where complex traits controlled by many loci with small effects had lower prediction accuracy than less complex traits (Hayes et al. 2010; Wolfe et al. 2016a, b). Genomic predictions should be more accurate for traits with higher H^2 (Daetwyler et al. 2010; Lorenz 2013; Combs and Bernardo 2013).

Among the parametric and nonparametric methods, RFC showed the lowest PA, which was not statistically significant ($P > 0.05$) for FRW, HI, and RP. The variation in prediction superiority for RKHS has been observed in previous studies (Wolfe 2017; Andrade et al. 2019). Empirical studies in cassava demonstrated lower predictive abilities for dry matter content of 0.24–0.68 (Wolfe et al. 2017) and 0.5655–0.5670 (Andrade et al. 2019). They did not find any great differences between the GS methods for SC and dry matter content. Although different methods were used to evaluate the starch content, our predictive ability results were higher (i.e., the specific gravity). In this study, the PA of SC was significantly different for the RFC and RKHS models.

Of the starch pasting-related traits, there were no significant differences between the parametric and nonparametric models, except for RFC, which had the lowest performance. Differences in the predictive ability of the nonparametric models could be due to the intrinsic way in which the various prediction models incorporate marker information. While models make no assumptions about gene action, nonlinearity is introduced using specific methods (Morota and Gianola 2014). In general, the performance of predictive models is known to depend on the genetic architecture of the trait under consideration (Daetwyler et al. 2010; Su et al. 2014). Although nonadditive models, including RFC and RKHS, capture dominance and epistasis effects, GBLUP is more suitable for prediction when traits are determined by a large number of unlinked and non-epistatic loci with small effects.

The cross-validation results were mostly consistent across the breeding programs, and the superiority of one prediction method over the others was trait-dependent. RF and RKHS usually predicted phenotypes more accurately for yield-related traits, which are known to have a significant amount of nonadditive genetic variation (Wolfe et al. 2016). Similar findings have been made in wheat (*Triticum aestivum* L.) for grain yield, an additive, and epistatic trait, in which RKHS, radial basis function neural networks, and Bayesian regularized neural network models had a better predictive ability than additive models like BL, Bayesian ridge-regression, BayesA, and BayesB (Perez-Rodriguez et al. 2012).

Conclusions

This study aimed to identify major QTL to facilitate MAS and explore the utility of GS for yield-related and starch pasting property-related traits in cassava panel breeding programs. In this study, GWAS identified eight yield-related, peak starch type of waxy and wild-type starch and 21 starch pasting property-related traits (QTLs), respectively. We performed GWAS and evaluated GS for 12 yield and

starch-related traits by utilizing data collected for 276 cassava-derived lines from TTDI, CIAT and IITA, and other advanced breeding lines. We provided a comparison of eight genomic prediction models for predictive ability. The results showed that predictabilities for the eight models were not significantly different for the yield-related traits, except that the RKHS models were better than the RFC models for SC, and the RFC model had a significantly lower predictability for the starch pasting property-related traits. Given these promising results, the implementation of MAS and GS can enable breeders to better select those quality traits in a larger population at earlier stages, thus enhancing selection efficiency. Opportunity exists to explore the genetics underlying other traits and agronomic characteristics, which will be addressed in future investigations.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00122-021-03956-2>.

Acknowledgements We would like to thank James Tanaka for his assistance in the laboratory of Dr. Mark Sorrells.

Author contribution statement CP and JC designed the experiment. PA, PN, SH, PJK, PK, WW, and CP conducted the field experiments. CR and VV provided germplasm. CK, SC, and KP conducted the starch pasting properties experiments. CP, PF, and PT performed data analysis. CP, SC, and PJK wrote the first draft of the manuscript. CP, MS, JJ, and MW revised the manuscript. All authors read and approved the manuscript.

Funding This work was supported by Kasetsart University Research and Development Institute (KURDI), Grant Number ๖๓๕๘ ๗-๒(๓๕)2.57, and was partially supported by the Center of Excellence on Agricultural Biotechnology, Office of the Permanent Secretary, Ministry of Higher Education, Science, Research and Innovation (AG-BIO/MHESI) Grant Number (60-005-001)J. The authors would also like to thank TTDI for providing supporting plant materials.

Data availability Phenotypic data used in the analyses and the genotypic data are available in Cassavabase (<http://cassavabase.org/search/traits>). The imputed SNP genotypic data obtained from the 276 genotypes used in this study are available in Cassavabase.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

- Aiemnaka P, Wongkaew A, Chanthaworn J, Nagashima SK, Boonma S, Authapun J et al (2012) Molecular characterization of a spontaneous waxy starch mutation in cassava. *Crop Sci* 52:2121–2130. <https://doi.org/10.2135/cropsci2012.01.0058>
- de Albuquerque HYG, Carmo CDD, Brito AC, Oliveira EJD (2018) Genetic diversity of manihot esculenta crantz germplasm based on single-nucleotide polymorphism markers. *Annals Appl Biol* 173(3):271–284. <https://doi.org/10.1111/aab.12460>

- Aldana AS, Quintero AF (2013) Physicochemical characterization of two cassava (*Manihot esculenta* Crantz) starches and flours. *Rev Sci Agroaliment* 1:19–25
- Andrade LRB, Sousa MBE, Oliveira EJ, Resende MDV, Azevedo CF (2019) Cassava yield traits predicted by genomic selection methods. *PLoS One* 14(11):e0224920. <https://doi.org/10.1371/journal.pone.0224920>
- Asoro FG, Newell MA, Beavis WD, Scott MP, Jannink J-L (2011) Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. *Plant Gen* 4:132–144. <https://doi.org/10.3835/plantgenome2011.02.0007>
- Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. *J Stat Soft* 67(1):1–48. <https://doi.org/10.18637/jss.v067.i01>
- Begum H, Spindel JE, Lalusin A, Borromeo T, Gregorio G, Hernandez J et al (2015) Genome-wide association mapping for yield and other agronomic traits in an elite breeding population of tropical rice (*Oryza sativa*). *PLoS ONE* 10(3):e0119873. <https://doi.org/10.1371/journal.pone.0119873>
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) Tassel: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23(19):2633–2635
- Bredeson J V, Lyons JB, Prochnik SE, Wu GA, Ha CM, 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. <https://doi.org/10.1038/nbt.3535>
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
- Ceballos H, Iglesias CA, Perez JC, Dixon AG (2004) Cassava breeding: opportunities and challenges. *Plant Mol Biol* 56:503–516. <https://doi.org/10.1007/s11103-004-5010-5>
- Cach NT, Perez JC, Lenis JJ, Calle F, Morante N et al (2005) Epistasis in the expression of relevant traits in cassava (*Manihot esculenta* Crantz) for subhumid conditions. *J Hered* 96:586–592
- Calle F, Perez JC, Gaitán W, Morante N, Ceballos H, Llano G et al (2005) Diallel inheritance of relevant traits in cassava (*Manihot esculenta* Crantz) adapted to acid-soil savannas. *Euphytica* 144:177–186
- Ceballos H, Pérez JC, Barandica OJ, Lenis JJ, Morante N, Calle F, Pino L, Hershey CH (2016) Cassava breeding I: the value of breeding value. *Front Plant Sci* 7:1227
- Ceballos H, Kawuki RS, Gracen VE, Yencho GC, Hershey CH (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
- Ceballos H, Rojanaridpiched C, Phumichai C et al (2020) Excellence in cassava breeding: perspectives for the future. *Crop Breed Gene Genom* 2:e200008
- Chaengsee P, Kongsil P, Siritwong N, Kittipadukul P, Piyachomkwan K, Petchpoung K (2020) Potential yield and cyanogenic glucoside content of cassava root and pasting properties of starch and flour from cassava Hanatee var and breeding lines grown under rain-fed condition. *Agri Natural Resour* 54(3):237–244
- Combs E, Bernardo R (2013) Accuracy of genomewide selection for different traits with constant population size, heritability, and number of markers. *Plant Genome* 6:1–7. <https://doi.org/10.3835/plantgenome2012.11.0030>
- Crossa J, Beyene Y, Kassa S, Pérez P, Hickey JM, Chen C, de los Campos G, Burgueño J, Windhausen VS, Buckler E, Jannink J-L, Lopez Cruz MA, Babu R (2013) Genomic prediction in maize breeding populations with genotyping-by-sequencing. *G3 Genes/Genomes/Genetics* 3(11):1903–1926. <https://doi.org/10.1534/g3.113.008227>
- Crossa J et al (2014) Genomic prediction in CIMMYT maize and wheat breeding programs. *Heredity* 112:48–60
- Crossa J et al (2017) Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci* 22:961–975
- Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA (2010) The impact of genetic architecture on genome-wide evaluation methods. *Genetics* 185:1021–1031
- de los Campos G, Naya H, Gianola D, Crossa J, Legarra A, Manfredi E (2009) Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics* 182:375–385. <https://doi.org/10.1534/genetics.109.101501>
- de los Campos G, Gianola D, Rosa GJM, Weigel KA, Cross J (2010) Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. *Genet Res* 92:295–308. <https://doi.org/10.1017/S0016672310000285>
- de los Campos G., Pérez, P. 2013. BGLR: Bayesian generalized linear regression. R package version 1.0.4. <https://cran.r-project.org/web/packages/BGLR>
- Desta ZA, Ortiz R (2014) Genomic selection: genome-wide prediction in plant improvement. *Trends Plant Sci* 19:592–601
- Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus* 12:13–15
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (gbs) approach for high diversity species. *PloS one* 6(5):e19379
- Ehret A, Hochstuhl D, Gianola D, Thaller G (2015) Application of neural networks with back-propagation to genome-enabled prediction of complex traits in Holstein-Friesian and German Fleckvieh cattle. *Genet Sel Evol* 47:22
- Elango D, Xue W, Chopra S (2020) Genome wide association mapping of epi-cuticular wax genes in *Sorghum bicolor*. *Physio Mol Biol Plants* 26(8):1727–1737. <https://doi.org/10.1007/s12298-020-00848-5>
- Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. *The Plant Genome* 4:250–255
- Endelman JB, Jannink JL (2012) Shrinkage estimation of the realized relationship matrix. *G3 Genes/Genomes/Genetics* 2:1405–1413
- Evans ID, Lips A (1992) Viscoelasticity of gelatinized starch dispersions. *J Texture Stud* 23:69–86. <https://doi.org/10.1111/j.1745-4603.1992.tb00512.x>
- Ezenwaka L, Del Carpio DP, Jannink J-L, Rabbi I, Danquah E, Asante I, Danquah A, Blay E, Egesi C (2018) Genome-wide association study of resistance to cassava green mite pest and related traits in cassava. *Crop Sci* 58:1907–1918
- FAO. 2018 Food Outlook - biannual report on global food markets – November 2018. Rome. 104 pp.
- Flint-Garcia SA, Thornsberry JM, Buckler ESIV (2003) Structure of linkage disequilibrium in plants. *Annu Rev Plant Biol* 54:357–374. <https://doi.org/10.1146/annurev.arplant.54.031902.134907>
- Gelandi P, Kowalski BR (1986) Partial least-squares regression: a tutorial. *Anal Chim Acta* 185:1–17
- Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES (2014) Tassel-gbs: a high capacity genotyping by sequencing analysis pipeline. *PLoS One* 9(2):e90346
- González-Camacho JM, Ornella L, Pérez-Rodríguez P, Gianola D, Dreisigacker S, Crossa J (2018) Applications of machine learning methods to genomic selection in breeding wheat for rust resistance. *Plant Genome*. <https://doi.org/10.3835/plantgenome2017.11.0104>
- Gouy M, Rousselle Y, Bastianelli D et al (2013) Experimental assessment of the accuracy of genomic selection in sugarcane. *Theor Appl Genet* 126:2575–2586. <https://doi.org/10.1007/s00122-013-2156-z>
- Gracen VE, Kogsil P, Napasintuwong O, Duangjit J, Phumichai C. The story of Kasetsart 50. The most important cassava variety

- in the world. Bangkok-Kamphaeng Saen (Thailand): Center for Agricultural Biotechnology, Kasetsart University; 2018.
- Guo G, Zhao F, Wang Y, Zhang Y, Du L et al (2014) Comparison of single-trait and multiple-trait genomic prediction models. *BMC Genet* 15:30. <https://doi.org/10.1186/1471-2156-15-30>
- Hayes BJ, Pryce J, Chamberlain AJ, Bowman PJ, Goddard ME (2010) Genetic Architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in holstein cattle as contrasting model traits. *PLoS Genet* 6(9):e1001139. <https://doi.org/10.1371/journal.pgen.1001139>
- Habier D, Fernando R, Kizilkaya K, Garrick D (2011) Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics* 12:186
- Heffner EL, Sorrells ME, Jannink J-L (2009) Genomic selection for crop improvement. *Crop Sci* 49:1–12
- Heffner EL, Lorenz AJ, Jannink J-L, Sorrells ME (2010) Plant breeding with genomic selection: gain per unit time and cost. *Crop Sci* 50:1681–1690
- Heslot N, Yang H, Sorrells M, Jannink J (2012) Genomic selection in plant breeding: a comparison of models. *Crop Sci* 52:146–160. <https://doi.org/10.2135/cropsci2011.09.0297>
- Hickey JM, Chiurugwi T, Mackay I, Powell W (2017) Implementing genomic selection in CBPWP genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nat Genet* 49:1297–1303
- Howard R, Carriquiry AL, Beavis WD (2014) Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. *G3 Genes/Genomes/Genetics* 4:1027–1046
- Ikeogu UN, Akdemir D, Wolfe MD, Okeke UG, Amaefula C, Jannink J-L, Egesi CN (2019) Genetic correlation, genome-wide association and genomic prediction of portable NIRS predicted carotenoids in cassava roots. *Front Plant Sci* 10:1570–1570
- Isidro J, Jannink JL, Akdemir D, Poland J, Heslot N, Sorrells ME (2015) Training set optimization under population structure in genomic selection. *Theor Appl Genet* 128(1):145–158. <https://doi.org/10.1007/s00122-014-2418-4>
- Jaramillo G, Morante N, Pérez JC, Calle F, Ceballos H et al (2005) Diallel analysis in cassava adapted to the midaltitude valleys environment. *Crop Sci* 45:1058–1063
- Jannink JL, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. *Brief Funct Genomics* 9:166–177
- Juliana P, Singh RP, Singh PK, Poland JA, Bergstrom GC, Huerta-Espino J, Bhavani S, Crossa J, Sorrells ME (2018) Genome-wide association mapping for resistance to leaf rust, stripe rust and tan spot in wheat reveals potential candidate genes. *Theor Appl Genet* 131(7):1405–1422. <https://doi.org/10.1007/s00122-018-3086-6>
- Kaler AS, Purcell LC (2019) Estimation of a significance threshold for genome-wide association studies. *BMC Genomics* 20:618. <https://doi.org/10.1186/s12864-019-5992-7>
- Karlström A, Calle F, Salazar S, Morante N, Dufour D, Ceballos H (2016) Biological implications in cassava for the production of amylose-free starch: impact on root yield and related traits. *Front Plant Sci* 7:604. <https://doi.org/10.3389/fpls.2016.00604>
- Kawano K, Fukuda WMG, Cenpukdee U (1987) Genetic and environmental effects on dry matter content of cassava root. *Crop Sci* 27(1):69–74
- Kawano K (2003) Thirty years of cassava breeding for productivity – biological and social factors for success. *Crop Sci* 43:1325–1335. <https://doi.org/10.2135/cropsci2003.1325>
- Kayondo SI, Pino Del Carpio D, Lozano R, Ozimati A, Wolfe M et al (2018) Genome-wide association mapping and genomic prediction for CBSD resistance in *Manihot esculenta*. *Sci Rep* 8:1549. <https://doi.org/10.1038/s41598-018-19696-1>
- Kulembeka HP, Ferguson M, Herselman L, Kanju E, Mkamilo G et al (2012) Diallel analysis of field resistance to brown streak disease in cassava (*Manihot esculenta* Crantz) landraces from Tanzania. *Euphytica* 187:277–288
- Kuon J-E, Qi W, Schläpfer P, Hirsch-Hoffmann M, Rogalla P, von Bieberstein A, Patrignani LP et al (2019) Haplotype-resolved genomes of geminivirus-resistant and geminivirus-susceptible african cassava cultivars. *BMC Biol* 17(1):75
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760
- Liaw A. 2013 Breiman and Cutler's random forests for classification and regression. Available 403 at: <http://cran.r-project.org/web/packages/randomForest/index.html>.
- Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ et al (2012) GAPIT, genome association and prediction integrated tool. *Bioinformatics* 28:2397–2399. <https://doi.org/10.1093/bioinformatics/bts444>
- Liu X, Huang M, Fan B, Buckler ES, Zhang Z (2016) Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet* 12(2):e1005767. <https://doi.org/10.1371/journal.pgen.1005767>
- Lorenz AJ, Smith KP, Jannink J-L (2012) Potential and optimization of genomic selection for Fusarium head blight resistance in six-row barley. *Crop Sci* 52:1609–1621
- Lorenz AJ (2013) Resource allocation for maximizing prediction accuracy and genetic gain of genomic selection in plant breeding: a simulation experiment. *G3 Genes/Genomes/Genetics* 3:481–491
- Maenhout S, De Baets B, Haesaert G, Van Bockstaele E (2007) Support vector machine regression for the prediction of maize hybrid performance. *Theor Appl Genet* 115:1003–1013
- McKey D, Elias M, Pujol B, Duputié A (2010) The evolutionary ecology of clonally propagated domesticated plants. *New Phytol* 186(2):318–332
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829
- Momen M, Mehrgardi AA, Sheikhi A, Kranis A, Tusell L et al (2018) Predictive ability of genome-assisted statistical models under various forms of gene action. *Sci Rep* 8:12309. <https://doi.org/10.1038/s41598-018-30089-2>
- Morota G, Gianola D (2014) Kernel-based whole-genome prediction of complex traits: a review. *Front Gene* 5:363
- Newport Scientific operation manual of the series 4 rapid visco analyzer. Australia. 1995. p 93
- Nwokocho LM, Aviara NA, Senan C, Williams PA (2009) A comparative study of some properties of cassava and cocoyam starches. *Carbohydr Polym* 76:362–367. <https://doi.org/10.1016/j.carbpol.2008.10.034>
- Ogbonna AC, Braatz de Andrade LR, Mueller LA et al (2021) Comprehensive genotyping of a Brazilian cassava (*Manihot esculenta* Crantz) germplasm bank: insights into diversification and domestication. *Theor Appl Genet*. <https://doi.org/10.1007/s00122-021-03775-5>
- Okeke UG, Akdemir D, Rabbi I et al (2017) Accuracies of univariate and multivariate genomic prediction models in African cassava. *Genet Sel Evol* 49:88. <https://doi.org/10.1186/s12711-017-0361-y>
- Ozimati AR, Kawuki W, Esuma IS, Kayondo M, Wolfe et al (2018) Training population optimization for prediction of cassava brown streak disease resistance in west african clones. *G3 Genes/Genomes/Genetics* 8:3903–3913
- Ozimati A, Kawuki R, Esuma W, Kayondo SI, Pariyo A, Wolfe M, Jannink JL (2019) Genetic variation and trait correlations in an

- East African cassava breeding population for genomic selection. *Crop Sci* 59(2):460–473
- Park T, Casella G (2008) The Bayesian lasso. *J Am Stat Assoc* 103:681–686
- Pérez JC, Ceballos H, Calle F, Morante N, Gaitán W et al (2005) Within-family genetic variation and epistasis in cassava (*Manihot esculenta* Crantz) adapted to the acid-soils environment. *Euphytica* 145:77–85
- Pérez P, de Loscampos G (2014) Genome-wide regression and prediction with the *bgrr* statistical package. *Genetics* 198:483–495. <https://doi.org/10.1534/genetics.114.164442>
- Pérez-Rodríguez P, Gianola D, González-Camacho JM, Crossa J, Manès Y, Dreisigacker S (2012) Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. *G3 Genes/Genomes/Genetics* 2(12):1595–1605. <https://doi.org/10.1534/g3.112.003665>
- Pérez L, Soto E, Farré G, Juanos J, Villorbina G, Bassie L, Medina V, Serrato AJ, Sahrawy M, Rojas JA, Romagosa I, Muñoz P, Zhu C, Christou P (2019) CRISPR/Cas9 mutations in the rice *Waxy*/GBSSI gene induce allele specific and zygosity dependent feedback effects on endosperm starch biosynthesis. *Plant Cell Reports* 38:417–433. <https://doi.org/10.1007/s00299-019-02388-z>
- R Core Team (2017). R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rabbi IY, Udoh LI, Wolfe M, Parkes EY, Gedil MA, Dixon A, Ramu P, Jannink J-L, Kulakow P (2017) Genome wide association mapping of correlated traits in cassava: dry matter and total carotenoid content. *Plant Genome* 10(3):944. <https://doi.org/10.3835/plantgenome2016.09.0094>
- Rabbi IY, Kayondo SI, Bauchet G 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*. <https://doi.org/10.1007/s11103-020-01038-3>
- Raemakers K, Schreuder M, Suurs L et al (2005) Improved cassava starch by antisense inhibition of granule-bound starch synthase I. *Mol Breed* 16:163–172. <https://doi.org/10.1007/s11032-005-7874-8>
- Ramu P, Esuma W, Kawuki R, Rabbi IY, Egesi C, Bredeson JV, Bart RS, Verma J, Buckler ES, Fei Lu (2017) Cassava haplotype map highlights fixation of deleterious mutations during clonal propagation. *Nat Genet*. <https://doi.org/10.1038/ng.3845>
- Riedelshheimer C et al (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nat Genet* 44:217–220
- Rojanaridpiched, C., V. Vichukit, S. Thongsri, O. Boonseng, A. Lim-sila and D. Suparhan (2010) Recent progress in cassava breeding and varietal adoption in Thailand. In: Howeler, R. (ed.) *A new future for cassava in Asia: its use as food and fuel to benefit the poor. Proceedings of the 8th regional workshop*, Vientiane, Lao PDR, pp. 202–210
- Rutkoski J, Benson J, Jia Y, Brown-Guedira G, Jannink JL, Sorrells M (2012) Evaluation of genomic prediction methods for Fusarium head blight resistance in wheat. *Plant Genome* 5:51–61. <https://doi.org/10.3835/plantgenome2012.02.0001>
- Salas-Tovar JA, Flores-Gallegos AC, Contreras-Esquivel JC et al (2017) Analytical methods for pectin methylesterase activity determination: a review. *Food Anal Methods* 10:3634–3646. <https://doi.org/10.1007/s12161-017-0934-y>
- Sánchez T, Dufour D, Moreno IX, Ceballos H (2010) Comparison of pasting and gel stabilities of waxy and normal starches from potato, maize, and rice with those of a novel waxy cassava starch under thermal, chemical, and mechanical stress. *J Agric Food Chem* 58:5093–5099
- Sauvage C, Segura V, Bauchet G, Stevens R, Do PT, Nikoloski Z, Fernie AR, Causse M (2014) Genome-wide association in tomato reveals 44 candidate loci for fruit metabolic traits. *Plant Physiol* 165(3):1120–1132. <https://doi.org/10.1104/pp.114.241521>
- Schirmer M, Höchstötter A, Jekle M, Arendt E, Becker T (2013) Physicochemical and morphological characterization of different starches with variable amylose/amylopectin ratio. *Food Hydrocoll* 32:52–63
- Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, Nordborg M (2012) An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat Genet* 44(7):825–30
- Somo M, Kulembeka H, Mtunda K et al (2020) Genomic prediction and quantitative trait locus discovery in a cassava training population constructed from multiple breeding stages. *Crop Sci* 60:896–913. <https://doi.org/10.1002/csc2.20003>
- Spindel J, Begum H, Akdemir D, Virk P, Collard B, Redoña E et al (2015) Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite. *Tropical Rice Breed Lines Plos Genet* 11(2):e1004982. <https://doi.org/10.1371/journal.pgen.1004982>
- Stapleton G. (2012). Global starch market outlook and competing starch raw materials for by product segment and region. Pricing outlook and cassava growth potential. Cassava Starch World 2010. Centre for Management Technology (CMT), Phnom Penh
- Su G, Christensen OF, Janss L, Lund MS (2014) Comparison of genomic predictions using genomic relationship matrices built with different weighting factors to account for locus-specific variances. *J Dairy Sci* 97(10):6547–6559. <https://doi.org/10.3168/jds.2014-8210>
- Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP, Feuston BP (2003) Random forest: a classification and regression tool for compound classification and QSAR modeling. *J Chem Inf Comput Sci* 43:1947–1958
- Swarts K, Li H, Romero Navarro JA, An D, Romay MC, Hearne S, Acharya C, Glaubitz JC, Mitchell S, Elshire RJ, Buckler ES, Bradbury PJ (2014) Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. *Plant Genome*. <https://doi.org/10.3835/plantgenome2014.05.0023>
- Thanyasiriwat T, Sraphet S, Whankaew S, Boonseng O, Bao J, Lightfoot DA, Tangphatsornruang S, Triwitayakorn K (2014) Quantitative trait loci and candidate genes associated with starch pasting viscosity characteristics in cassava (*Manihot esculenta* Crantz). *Plant Biol (stuttg)* 16(1):197–207. <https://doi.org/10.1111/plb.12022> (Epub 2013 Apr 24 PMID: 23614826)
- Toae R, Sriroth K, Rojanaridpiched C, Vichukit V, Chotineerant S, Wansuksri R, Chatakanonda P, Piyachomkwan K (2019) Outstanding characteristics of Thai non-gm bred waxy cassava starches compared with normal cassava starch, waxy cereal starches and stabilized cassava starches. *Plants* 8(11):447. <https://doi.org/10.3390/plants8110447>
- Tibshirani R (1996) Regression shrinkage and selection via the Lasso. *J R Stat Soc Ser B Methodol* 58:267–288
- Tumuhimbise R, Melis R, Shanahan P (2014) Diallel analysis of early storage root yield and disease resistance traits in cassava (*Manihot esculenta* Crantz). *F Crop Res* 167:86–93
- VanRaden PM (2008) Efficient methods to compute genomic predictions. *J Dairy Sci* 91:4414–4423
- Wang X, Yang Z, Xu C (2015) A comparison of genomic selection methods for breeding value prediction. *Sci Bull* 60:925–935
- Wang X, Li L, Yang Z, Zheng X, Yu S, Xu C, Hu Z (2017) Predicting rice hybrid performance using univariate and multivariate

- GBLUP models based on North Carolina mating design II. *Heredity* 118:302–310
- Wickham H (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. ISBN 978–3–319–24277–4, <https://ggplot2.tidyverse.org>.
- Wolfe MD, Rabbi IY, Egesi C, Hamblin M, Kawuki R, Kulakow P et al (2016) Genome-wide association and prediction reveals the genetic architecture of cassava mosaic disease resistance and prospects for rapid genetic improvement. *Plant Genome* 9(2):1–13. <https://doi.org/10.3835/plantgenome2015.11.0118>
- Wolfe MD, Kulakow P, Rabbi IY, Jannink J-L (2016b) Marker-based estimates reveal significant nonadditive effects in clonally propagated cassava (*Manihot esculenta*): implications for the prediction of total genetic value and the selection of varieties. *G3 Genes/Genomes/Genetics* 6:3497. <https://doi.org/10.1534/g3.116.033332>
- Wolfe MD, Del Carpio DP, Alabi O, Ezenwaka LC, Ikeogu UM, Kayondo IS et al (2017) Prospects for genomic selection in cassava breeding. *Plant Genome* 10(3):1–19
- Wolfe MD, Bauchet GJ, Chan AW, Lozano R, Ramu P et al (2019) Historical introgressions from a wild relative of modern cassava improved important traits and may be under balancing selection. *Genetics* 213:1237–1253
- Xu S (2017) Predicted residual error sum of squares of mixed models: an application for genomic prediction. *G3 Genes/Genomes/Genetics* 7:895–909
- Xu S, Zhu D, Zhang Q (2014) Predicting hybrid performance in rice using genomic best linear unbiased prediction. *Proc Natl Acad Sci USA* 111:12456–12461
- Xu Y, Wang X, Ding X, Zheng X, Yang Z, Xu C, Hu Z (2018) Genomic selection of agronomic traits in hybrid rice using an NCH population. *Rice* 10, 11(1):32. <https://doi.org/10.1186/s12284-018-0223-4>. **PMID: 29748895; PMCID:PMC5945574**
- Yabe S, Hara T, Ueno M, Enoki H, Kimura T, Nishimura S, Yasui Y, Ohsawa R, Iwata H (2018) Potential of genomic selection in mass selection breeding of an allogamous crop: an empirical study to increase yield of common buckwheat. *Front Plant Sci*. <https://doi.org/10.3389/fpls.2018.00276>
- Yasui T, Ashida K (2011) Waxy endosperm accompanies increased fat and saccharide contents in bread wheat (*Triticum aestivum* L.) grain. *Cereal Sci* 53:104–111. <https://doi.org/10.1016/j.jcs.2010.10.004>
- Yang N, Lu Y, Yang X, Huang J, Zhou Y, Ali F et al (2014) Genome wide association studies using a new nonparametric model reveal the genetic architecture of 17 agronomic traits in an enlarged maize association panel. *PLoS Genet* 10(9):e1004573. <https://doi.org/10.1371/journal.pgen.1004573>
- Yin L, Zhang H, Tang Z, Xu J, Yin D, Zhang Z, Yuan X, Zhu M, Zhao S, Li X (2021) rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study, genomics. *Proteom Bioinform*. <https://doi.org/10.1016/j.gpb.2020.10.007>
- Yonis BO, Pino del Carpio D, Wolfe M et al (2020) Improving root characterisation for genomic prediction in cassava. *Sci Rep* 10:8003. <https://doi.org/10.1038/s41598-020-64963-9>
- Yun MS, Kawagoe Y (2009) Amyloplast division progresses simultaneously at multiple sites in the endosperm of rice. *Plant Cell Physiol* 50(9):1617–1626. <https://doi.org/10.1093/pcp/pcp104>
- Zacarias AM, Labuschagne MT (2010) Diallel analysis of cassava brown streak disease, yield and yield related characteristics in Mozambique. *Euphytica* 176:309–320
- Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordoñas JM (2010) Buckler ES. *Nat Genet* 42(4):355–360
- Zhang N et al (2015) Genome-wide association of carbon and nitrogen metabolism in the maize nested association mapping population. *Plant Physiol* 168:575–583
- Zhang S, Chen X, Lu C, Ye J, Zou M, Lu K, Feng S, Pei J, Liu C, Zhou X, Ma P, Li Z, Liu C, Liao Q, Xia Z, Wang W (2018) Genome-wide association studies of 11 agronomic traits in cassava (*Manihot esculenta* Crantz). *Front Plant Sci* 9:503. <https://doi.org/10.3389/fpls.2018.00503>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.