Genome–Wide Association Study of Resistance to Cassava Green Mite Pest and Related Traits in Cassava

Lydia Ezenwaka, Dunia Pino Del Carpio, Jean–Luc Jannink, Ismail Rabbi, Eric Danquah, Isaac Asante, Agyemang Danquah, Essie Blay, and Chiedozie Egesi⋆

ABSTRACT

Cassava green mite [CGM, *Mononychellus tanajoa* (Bondar)] is a dry-season pest that usually feeds on the underside of young leaves causing leaf chlorosis, stunted growth, and root yield reduction by 80%. Since cassava (*Manihot esculenta* Crantz) leaves and roots serve as a primary staple food source, a decline in cassava yield can lead to household food, nutrition, and income insecurity. To evaluate the existence of CGM resistance alleles in the available germplasm, a diversity panel of 845 advanced breeding lines obtained from IITA, CIAT, and the National Root Crops Research Institute (NRCRI) were evaluated for CGM severity (CGMS), leaf pubescence (LP), leaf retention (LR), stay green, shoot tip compactness, and shoot tip size. A genome-wide association mapping detected 35 single-nucleotide polymorphisms (SNPs) markers significantly associated with CGMS, LP, and LR on chromosome 8. Colocalization of the most significant SNP associated with CGMS, LP, and LR on chromosome 8 is possibly an indication of pleiotropy or the presence of closely linked genes that regulate these traits. Seventeen candidate genes were found to be associated to CGM resistance. These candidate genes were subdivided into seven categories according to their protein structure namely, Zn finger, pentatricopeptide, MYB, MADS, homeodomain, trichome birefringence-related protein, and ethylene-responsive transcription factor genes. This study revealed significant loci associated with CGM, not previously reported, which together represent potential sources for the ongoing effort to develop multiple pest- and disease-resistant cassava cultivars.

L. Ezenwaka and C. Egesi, National Root Crops Research Institute (NRCRI), P.M.B, 7006 Umudike, Nigeria; L. Ezenwaka, E. Danquah, I. Asante, A. Danquah, and E. Blay, West Africa Center for Crop Improvement (WACCI), Univ. of Ghana, Legon, Accra, Ghana; D.P.D. Carpio, J.-L. Jannink, and C. Egesi, Dep. of Plant Breeding and Genetics, Cornell Univ., Ithaca, NY 14850; J.-L. Jannink, USDA-ARS, Ithaca, NY 14850; I. Rabbi and C. Egesi, IITA, Ibadan, Nigeria; E. Danquah, A. Danquah, and E. Blay, Dep. of Crop Science, Univ. of Ghana, Legon, Accra, Ghana; I. Asante, Dep. of Botany, Univ. of Ghana, Legon, Accra, Ghana. Received 11 Jan. 2018. Accepted 25 May 2018. ⋆Corresponding author (cne22@cornell.edu). Assigned to Associate Editor Yiqun Weng.

Abbreviations: BLUP, best linear unbiased prediction; CGM, cassava green mite; CGMS, cassava green mite severity; GBS, genotyping-by-sequencing; GLM, general linear model; GWAS, genome-wide association study; H^2, broad-sense heritability; LP, leaf pubescence; LR, leaf retention; MAF, minor allele frequency; MLM, mixed linear model; NRCRI, National Roots Crop Research Institute; PC, principal component; PCA, principal component analysis; QTL, quantitative trait locus; SG, stay green; SNP, single-nucleotide polymorphism; SSR, simple sequence repeat; STC, shoot tip compactness; STS, shoot tip size.

Cassava (*Manihot esculenta* Crantz) is the primary source of carbohydrates for >1 billion people, including the poorest on the African continent. Cassava production is particularly vital to the economy of Nigeria, the world's leading producer of cassava, since a decline of cassava yield can lead to household food and income insecurity.

Most often, the yields on smallholder cassava farms are relatively low due to viral diseases and cassava green mite [CGM, *Mononychellus tanajoa* (Bondar)] pest infestation. Cassava green mite is a dry-season pest that feeds by inserting their piercing and sucking mouthparts (chelicerae) into the abaxial surface of cassava

leaves extracting the cell contents (Yaninek and Hanna, 2003). Feeding damage by CGM causes leaf chlorosis, total defoliation, stunted growth, and root yield reduction up to 80%, which can eventually lead, in severely infected plants, to a candlestick appearance and low accumulation of starch in the storage roots.

Important control measures against CGM include cultural practices such as early planting, biological control using natural predators of *M. tanajoa* like *Typhlodromalus aripo* De Leon (Yaninek and Hanna, 2003), and host plant resistance (Leuschner, 1982). The use of resistant cultivars has been, so far, the most effective solution in mitigating the negative effect of CGM in farmers' fields, not only because this approach reduces yield losses, but since it also decreases the levels of pest inoculum in the farming system.

Clones with pubescent leaves, large compact shoot apices, and enhanced leaf retention (LR) and stay green (SG) offer higher levels of resistance to CGM than cultivars that lack these characteristics (Bellotti, 2002). Previous studies have reported the preference of *T. aripo*, the natural predator against CGM, for pubescent vs. glabrous cassava cultivars infested with *M. tanajoa* (Chalwe et al., 2015). Ultimately, leaf pubescence (LP) provides shelter for the predatory mite and enhances the ability of the predator to find the prey (CGM) due to the production of herbivore-induced plant volatiles (Bellotti, 2002; Onzo et al., 2012).

There has been limited work on the genetic control of CGM to determine the gene(s) conferring resistance. Most studies on CGM have focused on conventional breeding with limited work on molecular breeding. Previous studies identified two simple sequence repeat (SSR) markers, NS 1099 and NS 346, which showed high association with CGM resistance (Ceballos et al., 2010; Choperena et al., 2012). Recently, using an F_1 cross developed between the Tanzanian landrace Kiroba and the breeding line AR37-80, two quantitative trait loci (QTLs), qCGMc5Ar and qCGMc10Ar, were reported to be linked to CGM resistance (Nzuki et al., 2017). However, the genetic mechanism underlying CGM resistance and its phenotypically associated traits (LP, LR, SG, shoot tip size [STS], and shoot tip compactness [STC]) has not been studied in diversity panels using a genome-wide association mapping approach.

Conventional phenotype-based recurrent selection to breed CGM resistant cultivars is lengthy and resource intensive owing to several biological aspects associated with cassava breeding, including low seed set, slow multiplication rate of planting materials, and a 12-mo growing cycle (Ceballos et al., 2012). Moreover, the identification of CGM-resistant cultivars through phenotypic selection requires dry environmental conditions that favor *M. tanajoa* infestation. When and where such environmental conditions occur irregularly or nonuniformly, screening for the severity of the pest damage can be difficult. These challenges can be overcome through the use of genomic-assisted and marker-assisted breeding tools to facilitate indirect selection (Wolfe et al., 2016). Therefore, we performed a genome-wide association study (GWAS) to uncover genomic regions associated with natural resistance to *M. tanajoa*. A GWAS is an efficient and effective method for identifying new genes in complex phenotypic traits (Altshuler et al., 2008). This approach has been particularly useful for mapping major loci for various other traits in cassava. Some of the studied traits include genetic architecture of cassava mosaic disease, which is caused by cassava mosaic geminiviruses in the family Geminiviridae and the genus *Begomovirus* (Wolfe et al., 2016), β-carotene and dry matter content (Esuma et al., 2016; Rabbi et al., 2017), shoot weight, fresh root yield, starch fraction amylose content, dry matter content, and starch yield (de Oliveira et al., 2012). The current study was conducted aiming to identify genomic regions and single-nucleotide polymorphism (SNPs) linked to CGM resistance and other CGM-related traits in a GWAS panel of 845 advanced breeding lines.

MATERIALS AND METHODS
Germplasm Collection
A diverse panel of cassava germplasm maintained by the National Root Crop Research Institute (NRCRI), Umudike, Nigeria, was used for GWAS. The diversity panel consists of 845 advanced breeding lines that had been selected and clonally maintained. Three hundred and seventy-two clones were obtained from the cassava breeding program at IITA in Nigeria, 82 clones from CIAT, Colombia, and 391 clones from the breeding program of NRCRI, Nigeria (Supplemental Table S1).

Experimental Sites
The GWAS panel clones were evaluated in three locations in Umudike (5°29′ N, 7°24′ E; annual rainfall of 2200 mm; altitude of 120 m; mean annual temperature of 22 to 31°C; Dystric Luvisol soils; humid forest), Kano (12°0′0″ N, 8°31′0″ E; annual rainfall of 830 mm; altitude of 470 m; mean annual temperature of 29 to 38°C; coordinates; Calcic Luvisol soils; Sudan savanna), and Otobi (8°41′ N, 7°20′ E; annual rainfall of 1500 mm; altitude of 319 m; mean annual temperature of 24 to 35°C; Ferric Luvisol soils; southern Guinea savanna) in Nigeria. These sites represent the hot spots of CGM in the country.

Field Layout and Experimental Design
The trials were laid out in a randomized incomplete block design with three replications. Each plot consisted of five plants per row on ridges. There was 1 m of spacing between ridges and also between plants within the row, giving a total population of 10,000 plants ha^{-1}. The trials were evaluated in the 2013–2014, 2014–2015, and 2015–2016 cropping seasons. The populations were systematically phenotyped for CGM severity (CGMS), LP, LR, SG, STC, and STS.

Agronomic Practices
Fertilizer (15:15:15 N–P–K) was applied at the rate of 600 kg ha^{-1}. Fertilizer was applied at 8 wk after planting using the ring

method, where a round small channel is made around the plants to input the fertilizers after planting. The field was weeded three times during the first 4 mo. Weeding was done by hoe.

Agronomic Data
The traits evaluated were as follows:

1. Cassava green mite severity was evaluated at the visual rating of the damage caused by CGM at 6 mo after planting in January (peak of dry season). A 1-to-5 scale was used to rate symptoms where 1 = highly resistant (no symptoms observed), 2 = resistant (moderate damage, no reduction in leaf size, scattered chlorotic spots on young leaves), 3 = moderately resistant (severe chlorotic symptoms, slight reduction in leaf size), 4 = susceptible (severe chlorotic symptoms and severe reduction in leaf size of young shoot), 5 = highly susceptible (very severe chlorosis, extensive defoliation, candlestick appearance of young shoots).
2. Leaf pubescence was characterized visually for the degree of hairiness on the young leaf with 0 = glabrous, 3 = little pubescence, 5 = moderate pubescence, and 7 = high pubescence.
3. Leaf retention was evaluated at the visual rating of leaf longevity using a scale of 1 = very poor retention, 2 = poor retention, 3 = fair retention, 4 = good retention, 5 = outstanding retention.
4. Stay green was scored visually based on a 1-to-3 scale where 1 = poor (<50% of the leaves are live and green), 2 = moderately good (50–35% of the leaves are live and green), and 3 = very good (≥75% of the leaves are live and green). Leaf longevity was assessed by scoring for LR and SG.
5. Shoot tip size was a visual assessment of shoot apices based on how large or small the shoot apices are and categorized on a scale of 1 to 3 where 1 = small, 2 = medium, and 3 = large.
6. Shoot tip compactness was also based on visual assessment of the compactness of shoot apices according to how closely the shoot apices are, where 1 = loose, 2 = moderately compact, and 3 = compact.

All the traits were evaluated during the peak of dry season (January) at 6 mo after planting.

Phenotypic Data Analyses
Phenotypic data obtained from three experimental years were initially analyzed separately and subsequently pooled across years, using lme4 (linear mixed effects models using Eigen and S4) package implemented in R software (R Core Team, 2014).

The mixed models were fitted considering the location and year effect using the "lmer" function of the lme4 r package (Bates et al., 2015):

$$Y_{ijk} = \mu + c_i + \beta_j + r_{k(i)} + \varepsilon_{ijk}$$

where Y_{ijk} represents raw phenotypic observations; μ is the grand mean; c_i is a random effects term for clone; β_j is a fixed effect for the combination of location and year harvested; $r_{k(i)}$ is a random effect for replication nested within a location–year combination; and ε_{ijk} is the residual variance.

For the clone effect c_i, the best linear unbiased prediction (BLUP) represents an estimate of the total genetic value for each individual. To calculate the predictor error variance (PEV), the BLUPs were deregressed using the equation

$$\text{deregressed BLUP} = \text{BLUP} \Big/ \left[1 - \left(\text{PEV} / \sigma_i^2 \right) \right]$$

where PEV is the prediction error variance for each clone (Wolfe et al.; 2016, 2017). The deregressed BLUP was used for the association analysis to help reduce noise variation in GWAS. Variance components extracted from the lmer output and broad-sense heritability (H^2), based on clone means, was estimated according to (Hallauer et al., 2010).

Pearson correlation was calculated between traits using trait means of 845 accessions, which was performed using the *cor* function implemented in R.

DNA Extraction and SNP Genotyping
DNA extraction was performed following the DNeasy Plant Mini Kit (Qiagen) protocol with slight modification. The young fresh leaf samples were harvested from the apical part of cassava plant in the field. About three to five tender leaves, weighing ∼500 to 900 mg, were inserted in well-labelled extraction tubes arranged in a labelled 96-well box and placed on ice to maintain DNA integrity. From the field, the leaf samples were transferred to the NRCRI molecular laboratory and stored in a −80°C freezer. Before the commencement of the extraction process, the stored samples were lyophilized for 24 to 48 h. With the use of a tissuelyser running with a 1× speed at a rate of 1500 strokes min⁻¹, the samples were ground to a fine powder. Genomic DNA was extracted and quantified using a NanoDrop 1000 (Thermo Scientific), whereas the molecular weight was assessed by agarose gel electrophoresis. After successful DNA extraction, the samples were sent to the Institute of Genomic Diversity, Cornell University, Ithaca, NY, for SNP genotyping. There, SNP genotyping was performed following the genotyping-by-sequencing (GBS) protocol of (Elshire et al., 2011) using the ApeK1 restriction enzyme recommended by (Hamblin and Rabbi, 2014). Genotypes were called using the TASSEL 5.0 GBS pipeline version 2 (Glaubitz et al., 2014) and aligned to the cassava reference genome, version 6 (http://phytozome.jgi.doe.gov). The GBS data were filtered so that clones with >80% missing and markers with >60% missing genotype calls were removed. The SNPs with low quality (i.e., markers with extreme deviation from Hardy–Weinberg equilibrium, $\chi^2 > 20$) across all samples were removed from the dataset. Only biallelic SNP markers with a call rate >70% and minor allele frequency (MAF) of 0.05 were used for the analyses. Finally, imputation of the filtered dataset with 84,585 SNP markers >0.01 MAF was performed using Beagle version 4.0 (Browning and Browning, 2009).

Population Structure and Genome-Wide Association Analysis
A dataset of 61,307 SNP markers with MAF ≥0.05 was used for the assessment of population structure and GWAS. Principal component analysis (PCA) on SNP markers was used to identify major patterns of relatedness within and among the

breeding populations (NRCRI, IITA, and CIAT) using the *prcomp* function in R.

Association analysis was performed using three different statistical models comprising both general linear models (GLMs) and mixed linear models (MLMs) using TASSEL 5.0 (Bradbury et al., 2007). The following models were tested:

1. Naïve model: GLM without any correction for population structure.
2. PK model: MLM with two principal components (PCs) from PCA and **K** matrix as a correction for population structure.
3. K model: MLM with K matrix as a correction for population structure.

Results were compared to determine the best model for the analysis.

The statistical formula for the GLM is $\mathbf{y} = \mathbf{Xb} + e$ and for the MLM is $\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + e$ (Yu et al., 2006) where **y** is the vector of the phenotypic observations, **X** and **Z** are the known design matrices, **b** is a vector containing the fixed effects (genetic marker information and the population structure [the **Q** matrix]), **u** is a vector containing random additive genetic effects, and **e** is the vector of random residues. In MLM, the variance structure of random vectors **u** and **e** is

$$\mathrm{Var}\begin{bmatrix}\mathbf{u}\\\mathbf{e}\end{bmatrix}=\begin{bmatrix}\mathbf{G} & 0\\0 & \mathbf{R}\end{bmatrix}$$

where $\mathbf{G} = \sigma_a^2\mathbf{K}$ with **K** as the kinship matrix and σ_a^2 as the additive genetic variance, and $\mathbf{R} = \mathbf{I}\sigma_e^2$ with **I** as an identity matrix and σ_e^2 as the residual variance. In MLM, PCs from PCA and kinship (**K**) matrix analyses were used as the covariate (Bradbury et al., 2007; Hao et al., 2012). For the GLM analyses, a permutation test was run, and the number of permutations was 1000. The correcting ability for false positives of these models was tested through the evaluation of the quantile-quantile plots of the observed $-\log_{10}(p$ values) vs. expected $-\log_{10}(p$ values). Manhattan and quantile-quantile plots were generated using the R package qqman (Turner, 2014). To avoid false positives, a Bonferroni correction was used to set the significance cutoff at $-\log_{10}(\alpha/n)$. Where α is 0.05, which is the standard significance threshold, and n is the number of SNPs. In this study, the Bonferroni correction significance cutoff was $-\log_{10}(0.05/61,307) = 6.23$.

Candidate Genes Identification

Significant SNPs from the GWAS results were selected for candidate gene identification. In QTL regions, SNP markers above the Bonferroni threshold ($-\log_{10} > 6.23$) were mapped onto genes within a 4- to 8-Mb interval using the SNP location and gene description from the *M.esculenta*_305_v6.1.gene.gff3 available in Phytozome 11 (Goodstein et al., 2011) and the intersect function from bedtools (Quinlan and Hall, 2010).

Gene ontology annotation for each time point and combining all the datasets was done using Panther (http://go.pantherdb.org/).

Data Availability

The phenotypic and genotypic data generated and analyzed during this study are available in the Cassavabase repository (https://www.cassavabase.org/).

RESULTS

Phenotypic Evaluations

Descriptive Statistics and Broad-Sense Heritability across Three Locations for Three Years

Summary statistics of phenotypic data obtained for three growing seasons in the three sites (Umudike, Otobi, and Kano) are presented in Table 1. The highest mean (3.22 of a maximum of 5) was obtained for CGM at Kano, and the lowest (2.51 of a maximum 5) was obtained for CGM at Umudike. The highest means for LP and LR were obtained in Umudike, whereas the lowest were found in Kano. This implies that the lower the LP and LR, the higher the CGMS. Estimates of H^2 ranged from low to moderate, with the highest H^2 for CGM (0.30) and the lowest H^2 for LR (0.15). Trait variability estimated by a CV ranging from 30.6% for STC to 82.63% for LP.

Distribution and Correlation of the Traits

Correlation values among the different traits evaluated in the three environments are presented in Table 2. Results showed that LP ($r = -0.94$), LR ($r = -0.63$), SG ($r = -0.61$), STS ($r = -0.52$), and STC ($r = -0.65$) significantly and negatively correlated with CGMS ($p < 0.01$), whereas LR ($r = 0.65$), SG ($r = 0.55$), STS ($r = 0.58$), and STC ($r = 0.68$) significantly and positively correlated with LP. The distribution of deregressed BLUPs used as response variables in GWAS can be seen in Fig. 1. A wide range of variation was observed among the genotypes for all traits measured, and most traits exhibited a near-normal distribution (Fig. 1). The results also showed that the performance of plants that were moderately resistant to CGMS were more in number than the resistant and susceptible plants. Moderately pubescent plants tend to show good LR and SG.

Population Structure and Genetic Diversity

A total of 61,307 polymorphic SNPs with MAF >5% were used in this study. Using PCA to summarize genetic variation in the accessions, only a subtle differentiation was observed in the collection of germplasm (Fig. 2). Principal component analysis of the accessions revealed that the first, second, third and fourth PCs accounted for 0.31, 0.17, 0.06 and 0.05, respectively, of the genotypic variability (Supplemental Table S2). The top two PC loadings accounted for 48% of the variation and were included in the subsequent genome-wide association analyses (Fig. 2).

Table 1. Descriptive statistics (mean ± SD), CV, and broad-sense heritability (H^2 ± approximate SEM) for all the traits.

Traits†	Umudike Mean ± SD	CV	H^2 ± SEM	Otobi Mean ± SD	CV	H^2 ± SEM	Kano Mean ± SD	CV	H^2 ± SEM	Pooled Mean ± SD	CV	H^2 ± SEM
		%			%			%			%	
CGMS	2.81 ± 0.93	33.10	0.30 ± 0.01	2.81 ± 0.89	31.53	0.16 ± 0.01	3.22 ± 1.15	35.66	0.08 ± 0.01	2.91 ± 1.00	40.00	0.30 ± 0.01
LP	2.58 ± 2.61	101.00	0.29 ± 0.03	2.39 ± 2.50	104.5	0.15 ± 0.03	2.62 ± 2.56	97.44	0.12 ± 0.04	2.53 ± 2.60	101.30	0.29 ± 0.01
LR	3.02 ± 0.86	28.50	0.13 ± 0.01	3.05 ± 0.86	28.26	0.05 ± 0.01	2.93 ± 0.80	27.31	0.05 ± 0.01	3.01 ± 0.80	28.20	0.15 ± 0.01
SG	2.24 ± 0.61	27.08	0.15 ± 0.01	2.23 ± 0.63	28.37	0.07 ± 0.01	2.22 ± 0.57	25.68	0.06 ± 0.01	2.23 ± 0.60	27.20	0.17 ± 0.01
STC	2.05 ± 0.63	30.69	0.22 ± 0.01	2.03 ± 0.62	30.69	0.11 ± 0.01	2.00 ± 0.60	30.20	0.12 ± 0.01	2.03 ± 0.60	30.60	0.17 ± 0.01
STS	2.22 ± 0.71	31.86	0.16 ± 0.01	2.18 ± 0.72	33.19	0.08 ± 0.01	2.19 ± 0.68	31.00	0.06 ± 0.01	2.20 ± 0.70	32.10	0.23 ± 0.01

† CGMS, cassava green mite; LP, leaf pubescence; LR, leaf retention; SG, stay green; STC, shoot tip compactness; STS, shoot tip size.

Association Analysis
Model Selection

To identify the most suitable model to conduct GWAS for the dataset, three different standard GWAS models were evaluated for CGMS (Fig. 3). As depicted in the quantile–quantile plots, false-positive associations were observed using the naïve model for association analysis. When the K and PK model effects were incorporated, potential spurious associations were filtered out, and the p values were closer to the expected distributions for the trait analyzed. The K and PK models showed a good fit for p values, whereas the naïve model was characterized by an excess of small p values that correspond to the abundance of spurious associations (Fig. 3). This indicated that the K and PK were consistent for reducing $-\log_{10}(p$ values) toward the expected level, thereby controlling for false positives. On the other hand, the K model performed similarly to the PK model, displaying a highly uniform distribution of p values. Therefore, all subsequent results were based on the K model.

Marker-Trait Association

Association tests were performed for all traits in an analysis that combined all years and locations. Thirty-five SNP markers in total passed the Bonferroni significance threshold (Table 3). The GWAS identified 12, 17, 5, and 1 significant

Table 2. Phenotypic correlation coefficient analysis of all traits evaluated across the three locations for 3 yr.

Trait†	CGMS	LP	LR	SG	STS	STC
CGMS	1	−0.94***	−0.63***	−0.61***	−0.52***	−0.65***
LP		1	0.65***	0.64***	0.53***	0.67***
LR			1	0.55***	0.50***	0.53***
SG				1	0.58***	0.58***
STS					1	0.68***
STC						1

*** Significant at the 0.001 probability level.

† CGMS, cassava green mite; LP, leaf pubescence; LR, leaf retention; SG, stay green; STC, shoot tip compactness; STS, shoot tip size.

SNP markers for CGMS, LP, LR, and SG, respectively. The most significant SNP marker (S8_5962253) had a $-\log_{10}(p$ value) of 12 and explained 7% of the observed phenotypic variation. No significant SNP markers were observed for STS and STC in the combined dataset. The contribution of any single SNP to the phenotypic variation ranged from 4 to 7% across the traits (Table 3).

Cassava Green Mite Severity

Twelve markers were significantly associated with CGM resistance. The significant markers associated with the trait mostly concentrated in a single region on chromosome 8. The top significant SNP marker (S8_5962253) explained 7% of the phenotypic variance (Table 3).

Leaf Pubescence

Seventeen markers were found to be significantly associated with LP. The significant markers associated with this trait also lies on chromosome 8. Variance explained by the significant markers ranged from 4 to 7% (Table 3).

Leaf Retention

Five markers displayed significant associations with LR. These markers (S8_5962253, S8_6439483, S8_6439519, S8_6439891, and S8_6439935) were found on chromosome 8. Each of these markers explained 4% of the phenotypic variation of this trait (Table 3).

Stay Green

A marker was found to be significantly associated with SG. This marker (S13_692620) was found on chromosome 13. The marker explains 4% of the phenotypic variation (Table 3).

Genotype × Environment Effects

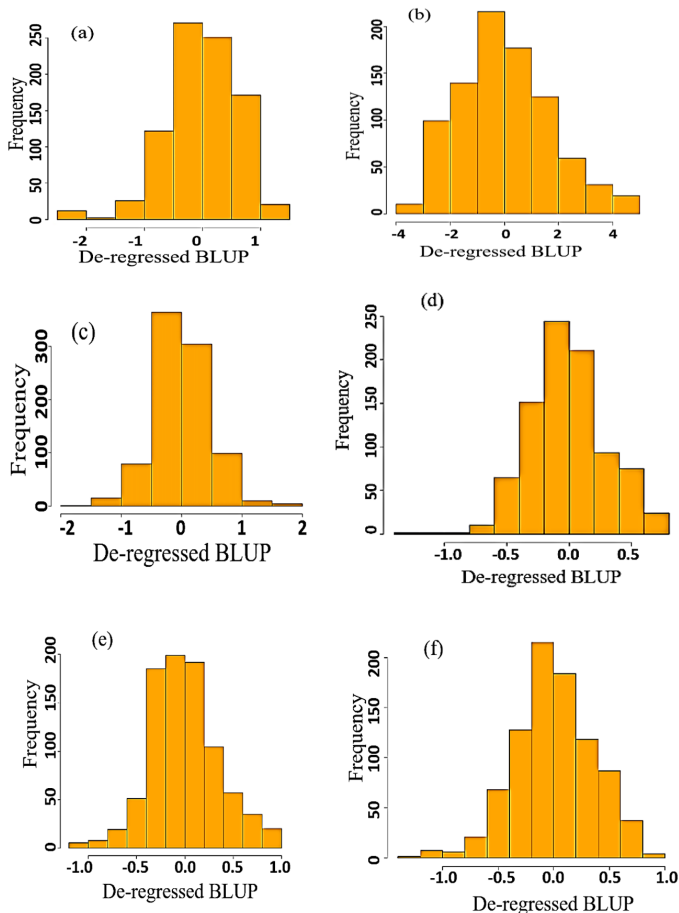The effect of genotype × environment interaction for CGMS, LP, LR, SG, STS, and STC was demonstrated

Fig. 1. Histogram of deregressed best linear unbiased predictions (BLUPs) for all traits—(a) cassava green mite severity, (b) leaf pubescence, (c) leaf retention, (d) stay green, (e) shoot tip compactness, and (f) shoot tip size—evaluated across the three locations for 3 yr.

by comparing GWAS results for the three seasons (2013, 2014, and 2015) across the three locations (Supplemental Table S5). The GWAS QTL regions explained between 4 and 8% of the observed phenotypic variation for all traits in Kano, Otobi, and Umudike, respectively (Supplemental Table S3). As seen in Fig. 4, a QTL with the most significant signal marker (S8_5962253) on chromosome 8 observed for CGMS, LP, and LR across the three locations. Different numbers of significant SNP markers were observed for different traits. In 2013, 34 significant SNP markers were observed for CGMS (11), LP (13), LR (8), SG (1), and STC (1). In 2014, 11 SNPs were found to be associated with two traits (LP [8] and STS [3]) on chromosomes 8 and 9, respectively. In 2015, GWAS identified 60 SNPs on chromosome 8 for five traits (CGMS [15], LP [19], LR [14], SG [8], and STC [3]). The SNP marker (S8_5962253) was the most significant hit for years 2013 and 2015, but not in 2014, although it was significant using the average from the 3 yr. Moreover, the same QTL region on chromosome 8 was identified as significantly associated with CGMS, LP, and LR when mapping across all accessions in the panel.

In addition, ANOVA for all traits evaluated in the three different locations across 3 yr revealed highly significant differences among the genotypes for all the traits ($p < 0.001$), thereby indicating the presence of substantial phenotypic variance (Supplemental Table S4). The results also indicated that the effects of location and year were significant ($p < 0.001$ and 0.05, respectively) for all the traits. Genotype × location interactions were not significant for all the traits. Apart from CGMS, all other traits were significant for genotype × year interactions. Genotype × location × year interactions were significant for CGMS, LP, LR, and STC.
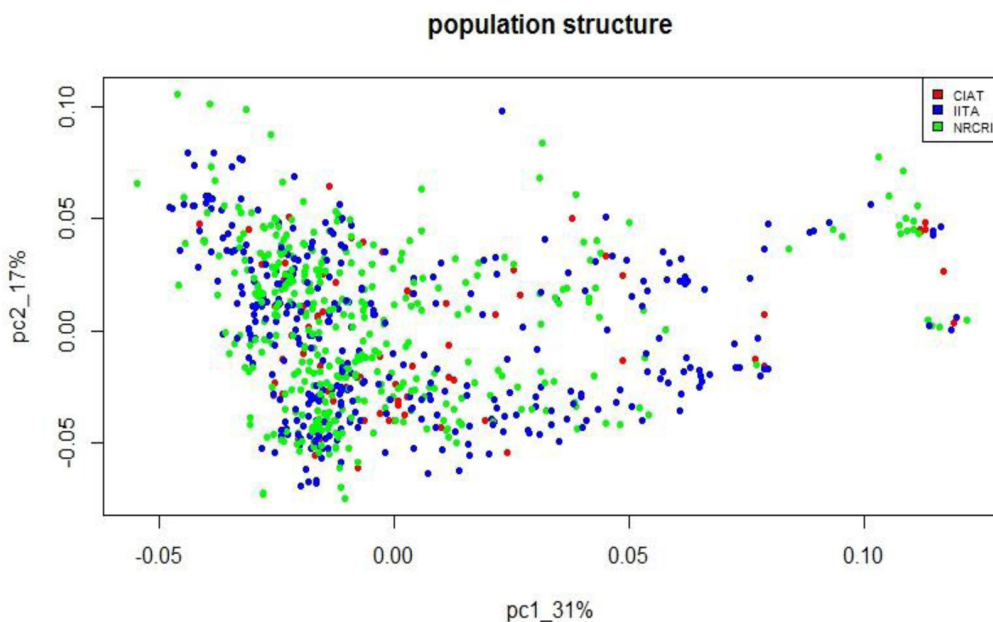

Fig. 2. Principal component analysis (PCA) plot of Principal Component 1 (PC1) against Principal Component 2 (PC2) illustrating population structure in the *M. esculenta* diversity panel genotyped with the single-nucleotide polymorphism markers. The red, blue, and green rectangles represent *M. esculenta* with CIAT, IITA, and National Root Crops Research Institute (NRCRI) origins.
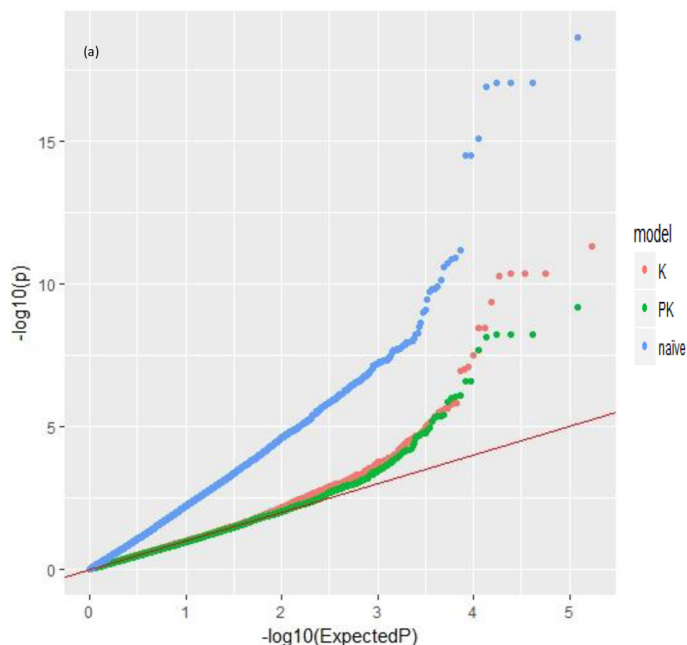
(a)

Fig. 3. Comparison of different genome-wide association study models. Cumulative distribution of p values computed from 61,307 single-nucleotide polymorphisms, and cassava green mite severity phenotype for different association models are presented. The more uniform the distribution of p values, the better the model.

Stay green and STS were not significant for genotype × location × year interactions (Supplemental Table S4).

Candidate Genes

The association results were intersected with the gene annotations within a 4- to 8-MB region on chromosome 8, and this region was the significant QTL for CGMS, LP, and LR. A total of 35 unique genes were identified within this region (Fig. 5, Supplemental Table S6). Most of the annotated genes are classes of membrane proteins that are involved in diverse functions ranging from plant growth and development to stress tolerance.

Among these genes, 17 candidate genes were found in the significant GWAS signal region linked to CGM resistance. The 17 candidate genes were subdivided into seven categories according to their protein structure: Zn finger, pentatricopeptide, MYB, MADS, homeodomain, trichome birefringence–related protein, and ethylene–responsive transcription factor genes (Table 4).

DISCUSSION

In this study, a diversity panel of 845 clones from the NRCRI cassava breeding program was used for genome-wide association mapping to identify QTL regions associated with CGM resistance and other related traits.

Overall, the estimates of H^2 for CGMS, LP, LR, SG, STS, and STC were low to moderate, ranging between 0.15 and 0.30. The H^2 estimates for these traits were lower than the narrow–sense heritability estimates reported by

Table 3. Single-nucleotide polymorphism (SNP) marker loci significantly associated with the traits and their explained proportion of phenotypic variation by marker (R^2).

Trait†	SNP marker	Chr.	SNP position	p value	R^2
LP	S8_5962253	8	5,962,253	8.69×10^{-13}	0.06965
CGMS	S8_5962253	8	5,962,253	5.08×10^{-12}	0.06508
LP	S8_6439483	8	6,439,483	8.25×10^{-12}	0.06383
LP	S8_6439519	8	6,439,519	8.25×10^{-12}	0.06383
LP	S8_6439891	8	6,439,891	8.25×10^{-12}	0.06383
LP	S8_6439935	8	6,439,935	1.50×10^{-11}	0.06228
CGMS	S8_6439483	8	6,439,483	4.61×10^{-11}	0.05940
CGMS	S8_6439519	8	6,439,519	4.61×10^{-11}	0.05940
CGMS	S8_6439891	8	6,439,891	4.61×10^{-11}	0.05940
LP	S8_6125583	8	6,125,583	5.44×10^{-11}	0.05897
CGMS	S8_6439935	8	6,439,935	5.79×10^{-11}	0.05881
CGMS	S8_6125583	8	6,125,583	4.52×10^{-10}	0.05355
LP	S8_6061415	8	6,061,415	1.12×10^{-9}	0.05123
LP	S8_6061421	8	6,061,421	1.12×10^{-9}	0.05123
LP	S8_5962010	8	5,962,010	2.25×10^{-9}	0.04945
CGMS	S8_6061415	8	6,061,415	3.54×10^{-9}	0.04831
CGMS	S8_6061421	8	6,061,421	3.54×10^{-9}	0.04831
LP	S8_5962464	8	5,962,464	5.23×10^{-9}	0.04732
LP	S8_6494548	8	6,494,548	7.37×10^{-9}	0.04644
LP	S8_5389240	8	5,389,240	7.70×10^{-9}	0.04633
LP	S8_5191862	8	5,191,862	2.88×10^{-8}	0.04300
CGMS	S8_6494548	8	6,494,548	3.18×10^{-8}	0.04235
CGMS	S8_5962464	8	5,962,464	8.59×10^{-8}	0.04023
CGMS	S8_5962010	8	5,962,010	1.02×10^{-7}	0.03980
CGMS	S8_5389240	8	5,389,240	1.12×10^{-7}	0.03957
LR	S8_5962253	8	5,962,253	1.19×10^{-7}	0.03941
LP	S8_5389232	8	5,389,232	1.24×10^{-7}	0.03931
LP	S8_5388872	8	5,388,872	1.56×10^{-7}	0.03873
LP	S8_5389151	8	5,389,151	1.56×10^{-7}	0.03873
LR	S8_6439483	8	6,439,483	2.64×10^{-7}	0.03350
LR	S8_6439519	8	6,439,519	2.64×10^{-7}	0.03350
LR	S8_6439891	8	6,439,891	2.64×10^{-7}	0.03350
LP	S8_5355904	8	5,355,904	3.00×10^{-7}	0.03709
SG	S13_692620	13	692,620	3.37×10^{-7}	0.03679
LR	S8_6439935	8	6,439,935	3.83×10^{-7}	0.03647

† CGMS, cassava green mite; LP, leaf pubescence; LR, leaf retention; SG, stay green; STC, shoot tip compactness; STS, shoot tip size.

Chalwe et al. (2015) and were lower than other agronomic traits in cassava (Wolfe et al., 2017). Despite the low heritability of these traits, they are good candidates for genomic selection, which is a breeding approach that is more accurate than traditional selection, especially for low-heritability traits (Calus et al., 2008).

Correlation estimates in crop breeding help to determine the success of indirectly selecting one trait for another (Falconer and Mackay, 1996). The correlation between leaf and shoot pubescence and CGM was reported by Byrne et al. (1983). Pubescence, therefore, acts to protect the most susceptible part of the plant from the *Mononychellus* mites (Hershey, 1987). Our correlation results showed that LP, LR, SG, STS, and STC are significantly and negatively correlated with CGMS. Thus, the observed correlation between CGM and other traits indicates that genotypes with pubescent leaves, outstanding LR, SG ability, compacted shoot tip, and large STS exhibited resistance to tolerance to CGM infestations.
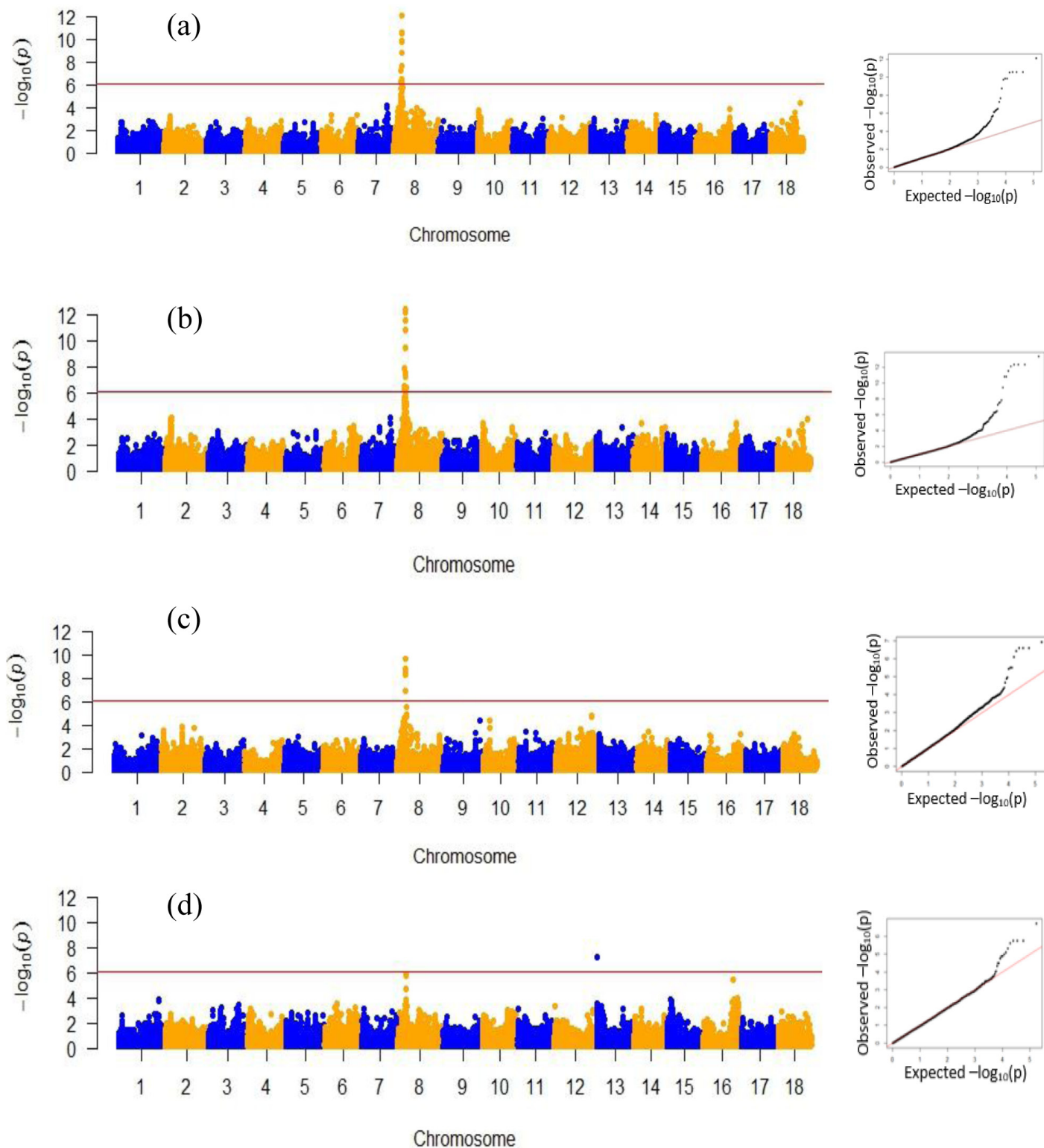
Fig. 4. Manhattan plots summarizing genome-wide association results for (a) cassava green mite severity, (b) leaf pubescence, (c) leaf retention, and (d) stay green across the three locations. The Bonferroni significance threshold is shown in red. A quantile-quantile plot if the *p* values is shown to access the number and magnitude of observed associations between genotyped single-nucleotide polymorphisms and the traits.

In Africa, the history of a cassava has followed complex domestication and breeding processes (Bredeson et al., 2016), with germplasm sharing and recurrent use of elite parents among the breeding programs (Wolfe et al., 2016). Conceivably in cassava, relatedness among cultivars can lead to population structure, which can confound the results of GWAS when relatedness is not included in the GWAS model (Yu et al., 2006). However, the population structure analysis of the GWAS accessions showed only a subtle differentiation between accessions,
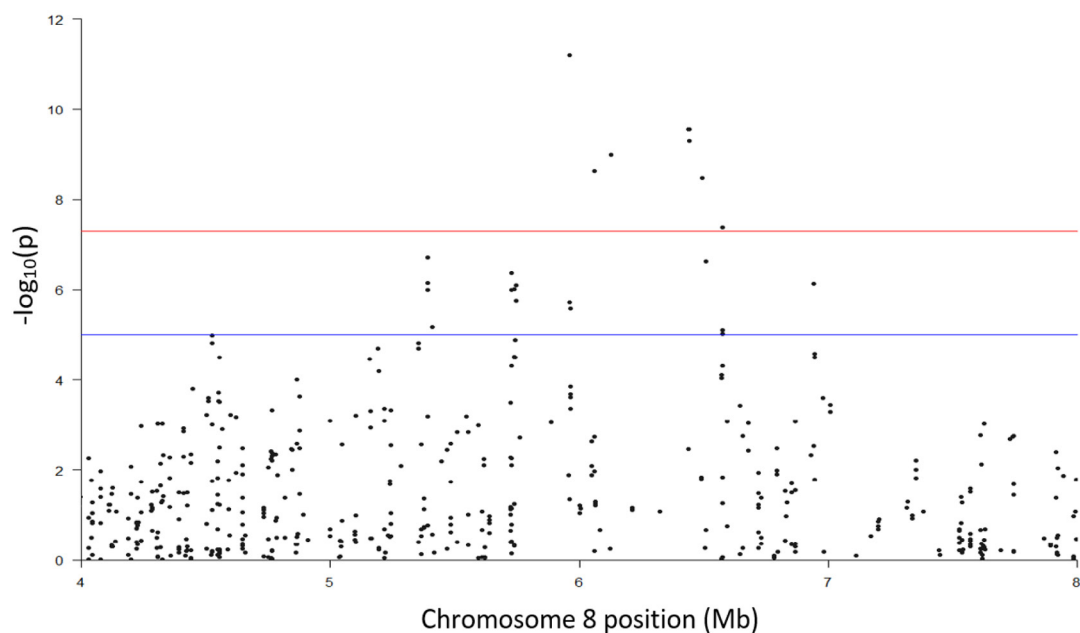
Fig. 5. Manhattan plot for genome-wide diagnosis of association for chromosome 8, zooming out the region with association signal for cassava green mite. Horizontal blue and red lines are the suggestive and Bonferroni threshold lines, respectively.

similar to those previously reported in cassava (Wolfe et al., 2016).

The incorporation of both K and PK matrices into a MLM has been successfully used in other species that exhibited significant population structure and relatedness (Gajardo et al., 2015; Zhou et al., 2016). Furthermore, the K model computational time is faster, and no additional steps like identifying population structure and PCs in the germplasm are required (Pasam et al., 2012).

Using three different standard GWAS models, we observed a stringent reduction in the number of significant markers when either the K or PK model was incorporated in GWAS, whereas in comparison, the naïve GLM model exhibited a high rate of false–positive rates.

To gain insight into the genetics of CGM resistance and its correlated traits, we selected the results from the K model to annotate the significant markers into candidate genes. We identified 35 unique genes within the SNPs

Table 4. genes annotated as candidates for association with cassava green mite resistance.

Gene	Chromosome	Gene description
Zn finger transcription factor genes		
Manes.08G058500	8	C2H2-like Zn finger protein
Manes.08G048200	8	C2H2-type Zn finger family protein
Manes.08G048800	8	CCCH-type Zn finger protein with ARM repeat domain
Manes.08G034200	8	Dof-type Zn finger DNA-binding family protein
Manes.08G046400	8	K homology-domain-containing protein-/-Zn finger (CCCH-type) family protein
Manes.08G041900	8	Zn finger protein 8
Pentatricopeptide repeat family		
Manes.08G026500	8	Pentatricopeptide repeat (PPR) superfamily protein
Manes.08G053900	8	Pentatricopeptide repeat (PPR-like) superfamily protein
Manes.08G060500	8	PENTATRICOPEPTIDE-REPEAT-596
MYB-domain transcription factor genes		
Manes.08G058000	8	myb-domain protein 106
Manes.08G045400	8	myb-like helix-turn-helix (HTH) transcriptional regulator family protein
MADS-domain transcription factor genes		
Manes.08G035100	8	AGAMOUS-like 80
Homeodomain transcription factor genes		
Manes.08G043900	8	Homeobox from *Arabidopsis thaliana*
Manes.08G024700	8	Basic leucine zipper (bZIP) transcription factor family protein
Manes.08G046700	8	bZIP transcription factor family protein
Trichome birefringence-related protein		
Manes.08G044000	8	PROTEIN TRICHOME BIREFRINGENCE-RELATED
Ethylene-responsive transcription factors		
Manes.08G026900	8	SAUR-like auxin-responsive protein family

associated with CGM resistance, LP, and LR in a 4-MB interval located on chromosome 8. Previous studies in a backcross population detected two SSR markers (NS1009 and NS346) that showed high association with CGM resistance (Ceballos et al., 2012; Choperena et al., 2012). In addition, a QTL mapping study in a biparental population identified two QTLs linked to CGM resistance (qCGMc5Ar and qCGMc10Ar) on chromosomes V and X respectively, with a maximum logarithm of odds of 20.19 at C1 (with phenotypic explained variance of 6.48) and 24.03 at C2 (with phenotypic explained variance of 4.11) (Nzuki et al., 2017).

We indirectly assessed the effect of genotype × environment interaction for all traits over 3 yr by comparing GWAS results across locations. Here, the GWAS peak associated with the candidate genes was consistently significant in 2013 and 2015, but not in 2014. These differences in significant QTLs are likely due to differences in weather patterns experienced during the growing seasons throughout years, which can have an impact on trait scores. In addition, for CGMS, LP, and LR, the same QTL was detected across locations, suggesting the lack of genotype × environment interaction for all traits across locations.

The fact that CGMS, LP, and LR show a stable QTL across locations indicates that these traits can be under the same genetic control or that the genes controlling each of these traits are linked and located on the chromosome 8 QTL region. The presence of pleiotropic effects may be beneficial, as genes conferring resistance to CGM are linked to LP and LR. These desirable traits (LP and LR) may be cointroduced along with the pest resistance into susceptible, glabrous cassava cultivars. Thus, we focused on the annotation of candidate genes within the 4- to 8-Mb region containing the CGM GWAS peak. Within the QTL region, we found 17 candidate genes that have strong homology with genes that are related to plant defense against insects. Among the most promising candidates, we identified protein trichome birefringence–related genes, Homeodomain-leucine zipper genes, myb-like helix-turn-helix (HTH) family protein and Zn finger protein 8 (ZFP8). The protein trichome birefringence–related genes are involved in increased levels of crystalline secondary wall cellulose in trichomes and stem development (Bischoff et al., 2010). Trichome birefringence–related genes are also involved in many developmental processes including defense to insects, herbivores, microbes, maintenance of leaf temperature, and transpiration regulation (Zhao et al., 2015).

Homeodomain-leucine zipper genes are unique to plant kingdom and participate in organ and vascular development including trichome development (Ariel et al., 2007; Zhao et al., 2015), and MYB domain transcription factor genes like myb-like HTH are involved in the specification of the leaf proximal distal axis and also regulates trichome differentiation in tobacco (*Nicotiana tabacum* L.)

(Payne et al., 1999). Finally, Zn finger protein 8 (ZFP8) is required for the initiation of inflorescence trichomes in response to gibberellin and cytokinin (Gan et al., 2007), and in *Cucumis sativus* L., ethylene treatment increased the number of cells per trichomes (Kazama et al., 2004).

Here, we identified a number of candidate genes in a GWAS panel representative of African cassava diversity. Earlier studies identified QTL regions for CGM resistance in other chromosomes than the ones reported here, suggesting loci specificity dependent on germplasm origin. Therefore, to fully dissect the genetic architecture of CGM resistance, additional clones should be tested over several years and at multiple locations. After validation, our results can be translated into marker-assisted breeding strategies that will serve as a complement to conventional breeding approaches to improve cassava cultivars for resistance to CGM.

CONCLUSION

This study is the first one to follow a genome-wide association mapping approach to dissect the genetic basis of CGM resistance and related traits in an African cassava germplasm. The genome-wide association analysis led to the identification of 17 candidate genes associated with CGM resistance, LP, and LR in cassava. This approach provides to plant breeders a useful tool to identify genes and discover valuable alleles for cultivar improvement. Validation studies involving fine mapping, joint linkage mapping, and genetic evaluation of candidate genes are required to understand the relationship between these candidate genes and the phenotypes evaluated in this GWAS analysis.

Conflict of Interest

The authors declare that there is no conflict of interest.

Supplemental Information Available

Supplemental Table S1. Origin of the training population.

Supplemental Table S2. PCA of training populations genotyped with SNP markers.

Supplemental Table S3. Summary of genome-wide association tests passing the Bonferroni significance threshold for Kano, Otobi, and Umudike locations.

Supplemental Table S4. ANOVA for all traits evaluated in the three locations across three years.

Supplemental Table S5. Summary of genome-wide association tests passing the Bonferroni significance threshold for years 2013, 2014, and 2015.

Supplemental Table S6. Genes annotated as candidates for association with CGM resistance, LP, and LR.

References

Altshuler, D., M.J. Daly, and E.S. Lander. 2008. Genetic mapping in human disease. Science 322:881–888. doi:10.1126/science.1156409

Ariel, F.D., P.A. Manavella, C.A. Dezar, and R.L. Chan. 2007. The true story of the HD-Zip family. Trends Plant Sci. 12:419–426. doi:10.1016/j.tplants.2007.08.003

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67(1). doi:10.18637/jss.v067.i01

Bellotti, A.C. 2002. Arthropod pests. In: R.J. Hillocks and J.M. Thresh, editors, Cassava: Biology, production and utilization. CABI, Wallingford, UK. p. 209–235. doi:10.1079/9780851995243.0209

Bischoff, V., J. Selbig, and W.-R. Scheible. 2010. Involvement of TBL/DUF231 proteins into cell wall biology. Plant Signaling Behav. 5:1057–1059.

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, and E.S. Buckler. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635. doi:10.1093/bioinformatics/btm308

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

Browning, B.L., and S.R. Browning. 2009. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am. J. Hum. Genet. 84:210–223. doi:10.1016/j.ajhg.2009.01.005

Byrne, D.H., A.C. Bellotti, and J.M. Guerrero. 1983. The cassava mites. Trop. Pest Manage. 29:378–394. doi:10.1080/09670878309370833

Calus, M.P.L., A.P.W. De Roos, and R.F. Veerkamp. 2008. Accuracy of genomic selection using different methods to define haplotypes. Genetics 178:553–561. doi:10.1534/genetics.107.080838

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

Ceballos, H., E. Okogbenin, J.C. Pérez, L.A.B. López-Valle, and D. Debouck. 2010. Cassava. In: J.E. Bradshaw, editor, Root and tuber crops. Springer, New York. p. 53–96. doi:10.1007/978-0-387-92765-7_2

Chalwe, A., R. Melis, P. Shanahan, and M. Chiona. 2015. Inheritance of resistance to cassava green mite and other useful agronomic traits in cassava grown in Zambia. Euphytica 205:103–119. doi:10.1007/s10681-015-1404-5.

Choperena, E.d.P.M., C. Ospina, M. Fregene, J. Montoya-Lerma, and A.C. Bellotti. 2012. Identificación de microsatélites en yuca asociados con la resistencia al ácaro Mononychellus tanajoa (Acari: Tetranychidae). Rev. Colomb. Entomol. 38:70–75.

de Oliveira, E.J., M.D.V. de Resende, V. da Silva Santos, C.F. Ferreira, G.A.F. Oliveira, M.S. da Silva, et al. 2012. Genomewide selection in cassava. Euphytica 187263–276. doi:10.1007/s10681-012-0722-0

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379. doi:10.1371/journal.pone.0019379

Esuma, W., R.S. Kawuki, L. Herselman, and M.T. Labuschagne. 2016. Stability and genotype by environment interaction of provitamin A carotenoid and dry matter content in cassava in Uganda. Breed. Sci. 66:434–443. doi:10.1270/jsbbs.16004

Falconer, D.S., and T.F.C. Mackay. 1996. Introduction to quantitative genetics. 4th ed. Pearson Educ., Harlow, UK.

Gajardo, H.A., B. Wittkop, B. Soto-Cerda, E.E. Higgins, I.A.P. Parkin, R.J. Snowdon, et al. 2015. Association mapping of seed quality traits in Brassica napus L. using GWAS and candidate QTL approaches. Mol. Breed. 35:143. doi:10.1007/s11032-015-0340-3

Gan, Y., C. Liu, H. Yu, and P. Broun. 2007. Integration of cytokinin and gibberellin signalling by Arabidopsis transcription factors GIS, ZFP8 and GIS2 in the regulation of epidermal cell fate. Development 134:2073–2081. doi:10.1242/dev.005017

Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, and E.S. Buckler. 2014. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS One 9:e90346. doi:10.1371/journal.pone.0090346

Goodstein, D.M., S. Shu, R. Howson, R. Neupane, R.D. Hayes, J. Fazo, et al. 2011. Phytozome: A comparative platform for green plant genomics. Nucleic Acids Res. 40:D1178–D1186. doi:10.1093/nar/gkr944

Hallauer, A.R., M.J. Carena, and J.B. de Miranda Filho. 2010. Testers and combining ability In: M.J. Carena, et al., editors, Quantitative genetics in maize breeding. Springer, New York. p. 383–423. doi:10.1007/978-1-4419-0766-0_8

Hamblin, M.T., and I.Y. Rabbi. 2014. The effects of restriction-enzyme choice on properties of genotyping-by-sequencing libraries: A study in cassava. Crop Sci. 54:2603–2608. doi:10.2135/cropsci2014.02.0160

Hao, D., M. Chao, Z. Yin, and D. Yu. 2012. Genome-wide association analysis detecting significant single nucleotide polymorphisms for chlorophyll and chlorophyll fluorescence parameters in soybean (Glycine max) landraces. Euphytica 186:919–931. doi:10.1007/s10681-012-0697-x

Hershey, C. 1987. Cassava breeding: A multidisciplinary review. Proceedings of a workshop held in the Philippines. 4–7 Mar. 1985. CIAT, Cali, Colombia.

Kazama, H., H. Dan, H. Imaseki, and G.O. Wasteneys. 2004. Transient exposure to ethylene stimulates cell division and alters the fate and polarity of hypocotyl epidermal cells. Plant Physiol. 134:1614–1623. doi:10.1104/pp.103.031088

Leuschner, K. 1982. Pest control for cassava and sweet potato. In: Root crops in eastern Africa: Proceedings of a workshop held in Kigali, Rwanda. 23–27 Nov. 1980. Int. Dev. Res. Ctr., Ottawa, ON.

Nzuki, I., M.S. Katari, J.V. Bredeson, E. Masumba, F. Kapinga, K. Salum, et al. 2017. QTL mapping for pest and disease resistance in cassava and coincidence of some QTL with introgression regions derived from Manihot glaziovii. Front. Plant Sci. 8:1168. doi:10.3389/fpls.2017.01168

Onzo, A., A.F. Houedokoho, and R. Hanna. 2012. Potential of the predatory mite, Amblyseius swirskii to suppress the broad mite, Polyphagotarsonemus latus on the gboma eggplant, Solanum macrocarpon. J. Insect Sci. 12:7. doi:10.1673/031.012.0701

Pasam, R.K., R. Sharma, M. Malosetti, F.A. van Eeuwijk, G. Haseneyer, B. Kilian, and A. Graner. 2012. Genome-wide association studies for agronomical traits in a world wide spring barley collection. BMC Plant Biol. 12:16. doi:10.1186/1471-2229-12-16

Payne, T., J. Clement, D. Arnold, and A. Lloyd. 1999. Heterologous myb genes distinct from GL1 enhance trichome production when overexpressed in Nicotiana tabacum. Development 126(4):671–682.

Quinlan, A.R., and I.M. Hall. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842. doi:10.1093/bioinformatics/btq033

R Core Team. 2014. R: A language and environment for statistical computing. R Found. Stat. Comput., Vienna.

Rabbi, I.Y., L.I. Udoh, M. Wolfe, E.Y. Parkes, M.A. Gedil, A. Dixon, et al. 2017. Genome-wide association mapping of correlated traits in cassava: Dry matter and total carotenoid content. Plant Genome 10(3). doi:10.3835/plantgenome2016.09.0094

Turner, S.D. 2014. qqman: An R package for visualizing GWAS results using QQ and Manhattan plots. (Preprint.) bioRxiv 5165. doi:10.1101/005165

Wolfe, M.D., D.P. Del Carpio, O. Alabi, L.C. Ezenwaka, U.N. Ikeogu, I.S. Kayondo, et al. 2017. Prospects for genomic selection in cassava breeding. Plant Genome 10(3). doi:10.3835/plantgenome2017.03.0015

Wolfe, M.D., I.Y. Rabbi, C. Egesi, M. Hamblin, R. Kawuki, P. Kulakow, 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). doi:10.3835/plantgenome2015.11.0118

Yaninek, S., and R. Hanna. 2003. Cassava green mite in Africa: A unique example of successful classical biological control of a mite pest on a continental scale. In: P. Neuenschwander, et al., editors, Biological control in IPM systems in Africa. CABI, Wallingford, UK. p. 61–76. doi:10.1079/9780851996394.0061

Yu, J., G. Pressoir, W.H. Briggs, I.V. Bi, M. Yamasaki, J.F. Doebley, et al. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38:203–238. doi:10.1038/ng1702

Zhao, J.L., Y.L. Wang, D.Q. Yao, W.Y. Zhu, L. Chen, H. Le He, et al. 2015. Transcriptome profiling of *trichome-less* reveals genes associated with multicellular trichome development in *Cucumis sativus*. Mol. Genet. Genomics 290:2007–2018. doi:10.1007/s00438-015-1057-z

Zhou, G., D. Hao, G. Chen, H. Lu, M. Shi, Y. Mao, et al. 2016. Genome-wide association study of the husk number and weight in maize (*Zea mays* L.). Euphytica 210:195–205. doi:10.1007/s10681-016-1698-y