

# Genetic Structure and Diversity Study of Cassava (*Manihot esculenta*) Germplasm for African Cassava Mosaic Disease and Fresh Storage Root Yield

Janatu Veronica Sesay<sup>1,2</sup>, Aiah Lebbie<sup>2</sup>, Richard Wadsworth<sup>2</sup>, Ephraim Nuwamanya<sup>3</sup>, Souleymane Bado<sup>4</sup>, Prince Emmanuel Norman<sup>1\*</sup>

<sup>1</sup>Sierra Leone Agricultural Research Institute, PMB 1313, Tower Hill, Freetown, Sierra Leone

<sup>2</sup>Department of Biological Sciences, School of Environmental Sciences, Njala University, Njala, Sierra Leone

<sup>3</sup>Makerere University Kampala, Kampala, Uganda

<sup>4</sup>Department of Plant Science, Phytoprise GmbH, DI Rudolf Schober-StraBe 4 7400 Oberwart, Oberwart, Austria

Email: \*pnorman@wacci.ug.edu.gh

**How to cite this paper:** Sesay, J.V., Lebbie, A., Wadsworth, R., Nuwamanya, E., Bado, S. and Norman, P.E. (2023) Genetic Structure and Diversity Study of Cassava (*Manihot esculenta*) Germplasm for African Cassava Mosaic Disease and Fresh Storage Root Yield. *Open Journal of Genetics*, 13, 23-47.

<https://doi.org/10.4236/ojgen.2023.131002>

**Received:** December 26, 2022

**Accepted:** March 11, 2023

**Published:** March 14, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

A better understanding of population structure and genetic diversity among cassava germplasm for African cassava mosaic disease and fresh root yield traits is useful for cassava improvement programme. Phenotype-based selection for these traits is cumbersome due to phenotypic plasticity and difficulty in screening of phenotypic-induced variations. This study assessed quantitative trait loci (QTL) regions associated with African cassava mosaic disease (ACMD) and fresh storage root yield (FSRY) in 131 cassava (*Manihot esculenta*) genotypes using a genome-wide association study (GWAS). The single nucleotide polymorphism (SNP) loci and associated candidate genes, when validated, would be a valuable resource for marker-assisted selection in the breeding process for development of new cassava genotypes with improved resistance to ACMD and desirable high root yield. Population structure analysis using 12,500 SNPs differentiated the 131 genotypes into five distinct sub-groups (K = 5). Marker-trait association (MTA) analysis using the generalized linear model identified two QTL regions significant for ACMD and three for FSRY. This study demonstrated that DArTseq markers are useful genomic resources for genome-wide association studies of ACMD and FSRY traits in cassava for the acceleration of varietal development and release.

## Keywords

Population Structure, Genetic Variability, GWAS, SNP Markers, Cassava

## 1. Introduction

Cassava (*Manihot esculenta* Crantz) is an important starchy root crop used for food, feed and various industrial applications [1]. The starchy storage roots of cassava are important source of dietary energy in sub-Saharan Africa (SSA) as they provide more returns per unit of input than any other staple crop [2] [3]. Cassava serves as food security and income generation crop for resource poor farmers due to its tolerance to erratic rainfall and poor soils compared to other root and tuber crops. In Sierra Leone, cassava ranks as the second most important staple crop after rice. The fresh storage root production of the crop in the country has increased from 82,500 tons in 1970 to 4.59 million tons (MT) in 2019, growing at an average annual rate of 12.08% [4].

However, on-farm cassava yields are significantly lower than the potential yields of improved varieties, which have been estimated at  $\geq 25$  tons per hectare [5]. For instance, in 2019, an estimated 59,660 ha were cultivated to cassava by 101,021 households, producing 817,342 MT. A wide yield variability ranging from 6.5  $\text{MT}\cdot\text{ha}^{-1}$  to 33.9  $\text{MT}\cdot\text{ha}^{-1}$  exists among genotypes, with an average yield (14.5  $\text{MT}\cdot\text{ha}^{-1}$ ) below 50% relative to yields obtained under good agronomic practices [5]. Cassava is grown extensively across the country because it is easily propagated, reliable, adaptable to different soils and climate, as well as its high food productivity potential. Cassava is utilized as food for human consumption, animal feed and agro-industries [6].

Assembling of cassava germplasm is crucial for the conservation and population improvement targeting key traits such as high yield, high starch quality, disease and drought tolerance, early maturity, easy cooking and good flavor. The introduction of genes from diverse populations broadens the genetic base and serves a potent source for delivering of novel genes or quantitative trait loci (QTLs) for important agronomic traits. Genetic diversity is the amount of existing divergence present in genotypes or populations or species [7]. In cassava, genetic divergence results in variations in the DNA sequence, morphological (above and below ground), biochemical (protein structure or isoenzyme), and physiological (abiotic and biotic stress resistance or growth rate) traits of cassava [8]. The genetic diversity of cassava in Sierra Leone is mainly maintained in work collections, in-situ gene bank at the Sierra Leone Agricultural Research Institute and farmers' fields, represented mainly by landraces and improved/introduced varieties selected by farmers [9].

Genetic markers are often used to assess the genetic divergence and population structure in living organisms [10]. For instance, molecular markers have been applied in several crop species to determine the population structure and genetic diversity of genotypes [11]. In cassava, Okogbenin *et al.* [12] reported the useful application of molecular markers in robustly determining population structure and genetic divergence in the crop. Some of the molecular markers utilized in cassava are diversity array technology (DArT), single nucleotide polymorphisms (SNPs), restriction fragment length polymorphism (RFLPs), random

amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), and simple sequence repeats (SSRs) [13].

Advances in microarray-based marker technology reveals the diversity arrays technology (DArT) markers as genetic markers of choice for the construction of high-density maps, mapping quantitative trait loci (QTL) and genetic diversity analysis based on their efficiency and low cost [14]. The QTL mapping methods based on bi-parental mapping populations identify the genomic regions with low resolution, whereas, genome-wide association studies (GWAS), based on linkage disequilibrium (LD), utilizes populations from diverse genetic backgrounds for dissection of the genetic architecture of complex traits with high resolution. The GWAS strategy has increasingly been utilized in many crops, including root and tuber crops, to dissect the underlying genetic control mechanism in complex traits [15]. Moreover, with the decreased genotyping cost and improved statistical methods, GWAS is considered as one of the powerful tools to overcome limitations in traditional QTL mapping [16]. The combination of the complexity reduction of the DArT method with high-throughput next-generation sequencing (NGS) technologies, led to the advent of the DArTseq platform with the merit of sequencing of complexity reduced representations [17]. The DArTseq markers based on genotyping-by-sequencing (GBS) technology have been successfully applied for linkage mapping, QTL identification in bi-parental mapping population, GWAS, genetic diversity, as well as in marker-assisted and genomic selection [18]. This technique is rapidly gaining popularity as a preferred method of genotyping by sequencing [18].

In the last four decades, cassava breeding programs in Africa, Asia, and Latin America, have developed genotypes to ameliorate various production constraints such as biotic and abiotic stresses with improved yield and starch content [19] [20]. While phenotype-based recurrent selection has contributed significant progress, the rate of genetic gain is low due to several breeding complexities associated with the biology of crop typified by the lengthy breeding cycle of the crop due to its inherent poor and asynchronous flowering, insufficient seed production, heterozygosity, slow multiplication rate of planting materials, and a long annual growing cycle of 12 months [21]. These problems limit conventional breeding techniques resulting into inefficient and a low level of genetic gain. The complementation of the conventional crop improvement techniques with advanced molecular tools contribute to reduce the breeding cycle in crops [15] [22]. Moreover, understanding the genetic basis of variation in key traits of interest is critical for increasing their selection efficiency, shortening the breeding cycle and the rate of genetic gain.

Modern crop improvement techniques such as marker-assisted selection (MAS) and genomic selection (GS) can be used to accelerate genetic improvement particularly by reducing generational interval and increasing selection intensity [21] [23]. However, integration of molecular markers as part of MAS in breeding pipelines requires an initial investment in discovery research to identify major effect loci that serve as the targets of selection. With the advances in next-ge-

neration sequencing (NGS) technologies, it is now possible to generate genome-wide marker data in targeted populations. The combination of this molecular information from this technique and phenotype data, makes it possible to identify and map locations of agriculturally important genes and quantitative trait loci (QTL) at the whole genome level [24]. The marker trait association obtained is a useful technique that guides the selection of individuals with higher genetic value through marker-assisted selection (MAS) [25]. Thus, the main objectives of this study were to: 1) investigate genetic diversity and genome-wide association studies (GWAS) in cassava germplasm from Sierra Leone; 2) identify SNP markers associated with African cassava mosaic disease and fresh storage root yield of cassava *via* GWAS. The identified markers are anticipated to facilitate marker assisted selection (MAS) of the studied traits in cassava, and the cassava accessions with high fresh storage root yield and tolerance to ACMMD will be potential parents for cassava breeding.

## 2. Materials and Methods

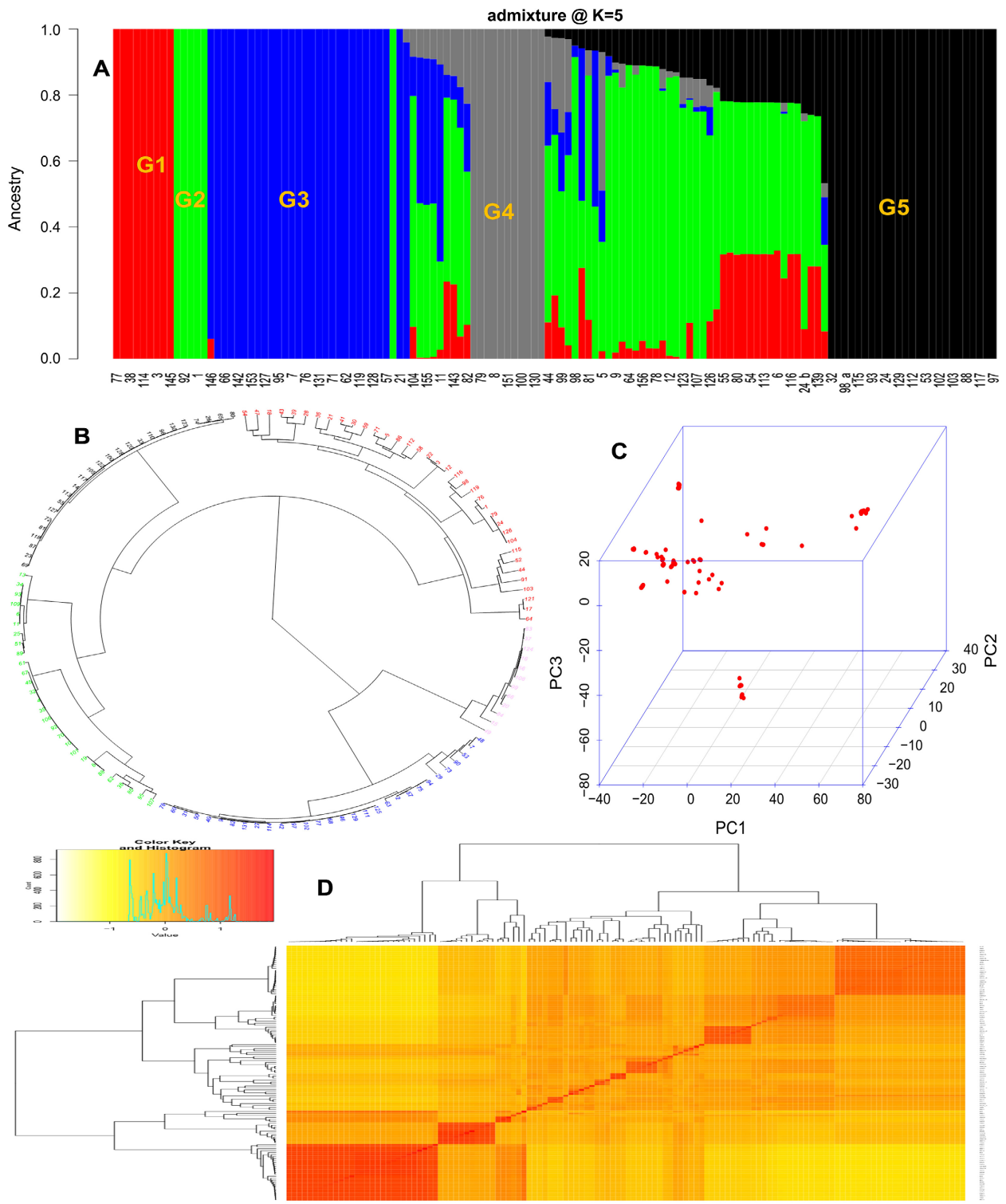
### 2.1. Plant Materials

The study panel comprised of 131 cassava varieties collected from different geographical areas (three agro-ecological zones) of Sierra Leone (**Figure 1; Supplementary Table S1**). The varieties were collected using the multistage sampling approach. Districts considered as the cassava belt areas within the country were selected and three communities per district were also selected at random based on the list of communities used by Statistics Sierra Leone during 2018. The varieties were collected from farmers' fields with the aid of International Institute of Tropical Agriculture (IITA) cassava descriptors, farmers' knowledge, preferences and special attributes of the varieties. The different collection sites were located using global positioning system (GPS). A minimum of 10 km separated the two nearest sites of collection. The planting materials collected were stem cuttings.

### 2.2. Field Experiment and Phenotyping

The experiment was conducted at the Biological Sciences Experimental Farm, Njala University (08° 14'S, 12° 1'W), southern Sierra Leone. The experiment was laid out in an augmented randomized complete block design with three replicates. The stem cuttings of each variety were cut into 20 cm long each and planted in holes made on the crest of single ridge plots measuring 10 m<sup>2</sup>. The cuttings were planted at a planting distance of 1 m × 1 m between and within rows giving a plant population of 10,000 plants ha<sup>-1</sup>. The field trial management was done whenever necessary, following the technical recommendations and standard agricultural practices for cassava [26]. Phenotypic data on fresh storage root yield and African cassava mosaic disease (ACMD) severity were recorded. Detailed trait description and ontologies are available (<https://cassavabase.org/search/traits>). The fresh storage root yield was recorded





**Figure 1.** Population structure and diversity analysis of current GWAS panel. (A) Population structure based on STRUCTURE when K = 5. (B) Neighbor-joining based clustering observed in the study panel using 12,500 SNP markers. (C) Three-dimensional plot of the first three principal components, and (D) heat map of pairwise kinship matrix of 131 cassava genotypes.

in  $t \cdot ha^{-1}$  at-harvest (eight months after planting). The ACMD severity score was done based on a visual assessment of the relative area of plant surface affected by

the mosaic virus disease using a five-ordinal scale of 1 - 5 at 6 months after planting. The 1 - 5 disease rating scales used to record the proportion of a plant surface in a plot affected by ACMD represented the following: 1 for no visible virus symptom, 2 for mild symptoms on few leaves but no leaf distortion, 3 for low symptoms of the mosaic virus on leaves, 4 for the severe mosaic on most leaves and leaf distortion, and 5 for severe mosaic and bleaching with severe leaf distortion and stunting.

### 2.3. Phenotypic Data Analysis

The phenotypic data were analyzed using the one-step linear mixed model that utilizes G-matrix to compute the best linear unbiased predictor (BLUP) values of each genotype for a trait from the best fit model; and the average information criterion (AIC) in restricted maximum likelihood (REML) algorithm [27] was done in the ASReml-R version 4 package [28]. Accordingly, the genetic variance was partitioned into the additive genetic effect (*i.e.* the proportion that associated with a covariance structure proportional to genetic relationships derived from the molecular markers) and the non-additive genetic effect. The non-additive genetic variance is explained by individual identity rather than the genomic relationship matrix [29] [30]. Broad sense heritability ( $H^2$ ) estimates for fresh root yield and ACMD were estimated from phenotypic variance ( $\sigma^2_p$ ) and the genotypic variance ( $\sigma^2_g$ ). The BLUP values of the genotypes for the traits extracted from the best fit model were used as input for the GWAS model.

### 2.4. Genotyping and SNP Data Analysis

Total genomic deoxyribonucleic acid (DNA) was isolated from lyophilized young and fully expanded healthy leaves of each varieties studied. The DNA was extracted using the CTAB procedure with slight modification [31]. The quality and concentration of the DNA were assessed using agarose gel and nanodrop, respectively [32]. High-throughput genotyping was done using 96 plex DArTseq protocol, and SNPs were called using DArTSoft as described by Kilian *et al.* [33]. The raw HapMap file generated was first converted to a Variant call format (VCF) and filtered for missing value and polymorphic SNPs using quality control criteria of low sequence depth < 5; SNP markers with missing values > 20%; genotype quality < 20; minor allele frequency (MAF) < 0.05 and heterozygosity > 50 [15] [34]. Of the 61,214 SNPs subjected to the filtering quality criteria, 12,500 good-quality SNPs were retained for further analyses.

### 2.5. Population Genetic Analysis

Different population genetic analysis techniques were used to explore the structure and level of genetic diversity in the germplasm. The SNP distribution and density were determined using the “*Cmplot*” function implemented in the CMplot R package [35]. The SNPlay open website was used to estimate the rate of transition and transversion across the retained SNP. Summary statistics on the minor

allele frequency (MAF), polymorphism information content (PIC), the observed and the expected heterozygosity were estimated based on the function “--freq” and “--hardy” using PLINK V1.90 [36].

The genetic relationship among the cassava germplasm was explored using the principal component analysis (PCA) in FactorMiner R package V3.3 [37]. For the PCA, the origin of the varieties was used as factor.

The cluster samples were put into populations using structure software version 2.3.3 [38] [39]. The structure simulations were done using the admixture model with a burn in period of 20,000 iterations and a Markov chain Monte Carlo (MCMC) set at 20,000. The simulations were repeated 3 times for K-values of 1 to 10. The optimal subpopulation model was determined using the following procedures: 1) by applying the informal pointers (*i.e.* geographical origin) proposed by Pritchard *et al.* [38] and Falush *et al.* [39]; 2) by considering  $\Delta K$ , a second order rate change with respect to K, as defined in Evanno *et al.* [40], and implemented in STRUCTURE HARVESTER V0.6.94 [41]. Structure population was then plotted using barplot function implemented in R. The phylogeny tree was constructed using ape version 5.0 implemented in R [42]. The exploratory Discriminant Analysis of Principal Components (DAPC) was applied using the adegenet package V2.1.8 [40]. The Admixture was performed through the Bayesian Information Criterion (BIC). A hierarchical cluster was constructed utilizing a kinship relation matrix implemented in GAPIT V3 [41].

## 2.6. Genome Wide-Association Study (GWAS) Analysis

The GWAS analysis was done using the compressed mixed linear model (CMLM) implemented in the GAPIT R package V3 and the Manhattan and the QQ plots were visualized in CMplot [35]. For the GAPIT analysis, population structure (Q) and the relationships among individuals were accounted for through a principal component (PC) analysis and a kinship (K) matrix generated from marker data, respectively [36]. For each trait, the optimal number of PCs/covariates included in the GWAS models was determined through model selection using the Bayesian information criterion (BIC), with a maximum of four tested PCs. The GWAS study was done using the following formula:  $Y = X\beta + W\alpha + Qv + Zu + \varepsilon$ ; where  $Y$  was treated as the observed vector of BLUP;  $\beta$  as the fixed effect vector ( $p \times 1$ ) other than molecular markers effects and population structure;  $\alpha$  as the fixed effect vector of the molecular markers;  $v$  as the fixed effect vector from the population structure;  $u$  as the random effect vector from the polygenic background effect;  $X$ ,  $W$ , and  $Z$  are the incidence matrixes from the associated  $\beta$ ,  $\alpha$ ,  $v$ , and  $u$  parameters; and  $\varepsilon$  as the residual effect vector. The significance threshold for the marker-trait associations (MTA) was set to  $p = 0.05$  after assigning the associated probability (P) to each marker using the Bonferroni threshold correction ( $-\log_{10}(P)$ ). The percentage of variation explained by the associated marker ( $R^2$ ) was determined using a step regression implemented in GAPIT.

Circular Manhattan and Quantile-Quantile (QQ) plots were generated by plot-

ting the negative logarithms ( $-\log_{10}$ ) of the P-values against their expected P-values to fit the appropriateness of the GWAS model with the null hypothesis of no association and to determine how well the models accounted for population structure using CMplot [35]. The Manhattan plot was created for visualization of GWAS on the entire genome, and zoom mapping was performed on a particular chromosome after identifying a significant SNP marker.

### 3. Results

#### 3.1. Phenotypic Trait Assessment

Broad-sense heritability estimates was high, 0.99 for African cassava mosaic disease and intermediate, 0.51 for fresh root yield. The phenotypic value for ACMD ranged from 1.0 to 5.0 with an average of 1.76. The phenotypic value for the fresh root yield ranged from 0.20 to 21.40 t·ha<sup>-1</sup> with an average of 7.08 t·ha<sup>-1</sup> (Table 1). The best linear unbiased predictor (BLUP) values of African cassava mosaic disease (ACMD) and fresh storage root yield (FSRY) varied among 131 cassava varieties studied (Supplementary Table S2).

#### 3.2. Genetic Diversity, Population Structure and Linkage Disequilibrium

The DArT genotyping of 131 cassava varieties detected the highest number of SNPs (1434) mapped on chromosome 1 and the lowest of 457 on chromosome 13 (Supplementary Figure S1(A)). Transition SNPs (60.76%, 4,470 SNPs) were more frequent than transversions (39.24%, 2887 SNPs) (Supplementary Figure S1(B)). The observed heterozygosity value ranged from 0.0 to 0.802, with an average of 0.233 (Supplementary Figure S1(C)). The expected heterozygosity value ranged from 0.023 to 0.5, with an average of 0.263 (Supplementary Figure S1(D)). The minor allele frequency ranged from 0.012 to 0.500, with a mean of 0.184 (Supplementary Figure S1(E)). The polymorphic information content (PIC) ranged from 0.02 to 0.38, with an average value of 0.22 (Supplementary Figure S1(F)).

The population structure of a diverse panel of 131 cassava varieties was investigated on the basis of a 1K method of model-based Bayesian clustering using 12,500 SNP markers. The population structure analysis revealed the existence of five distinct subpopulations in the cassava panel, which was found consistent with the results of the phylogeny tree analysis through the kinship matrix, and

**Table 1.** Descriptive statistics of African cassava mosaic disease (ACMD) and fresh root yield (FRY) of cassava.

Traits	Minimum	Maximum	Mean	Standard Deviation	Broad sense heritability (H <sup>2</sup> )
ACMD	1.00	5.00	1.76	1.12	0.99
SFRY	0.20	21.40	7.08	3.28	0.51

ACMD = African cassava mosaic virus, SFRY = fresh root yield (FRY) of cassava.

PCA (**Figure 1**). The highest membership was recorded in group 1 (36 varieties) and the lowest reported in group 5 (12 varieties).

Group 1 had 36 varieties comprising 34 landraces (collected from farmers and maintained at the Njala Agricultural Research Center, Njala, Sierra Leone) and two improved varieties. One of the two improved varieties, SLICASS6 (SC6\_6) is a released variety. Group 2 consisted of 30 varieties, mainly comprising 29 local landraces and one improved variety (IP6). Group 3 had 28 varieties, comprising of 23 local and five improved varieties. Group 4 had 25 varieties, comprising 21 local and four improved varieties. SLICASS6 (SC6\_1) and SLICASS4 (SC4\_2) are released varieties. Groups 5 had 12 members, comprising 10 local and two improved varieties (**Supplementary Table S3**). The Cluster membership displayed through the phylogeny tree was in perfect alignment with the Discriminant analysis of principal components (DAPC) cluster membership (**Supplementary Table S3**). The genetic relationship based on the principal component analysis revealed that the first two PCs account for 85.8% of the total variation (**Figure 2**). Both the local and improved varieties were distributed along PC1 and the PC2 (**Figure 2**).

### 3.3. Genome-Wide Scan for African Cassava Mosaic Disease Resistance

Two SNP loci exhibited significant association with the reaction to African cassava mosaic disease infection (**Table 2, Figure 3**). The two SNP loci (chr2\_10463640 and chr2\_11944909) that associated with ACMD had marker effects of 2.47 and 2.93, respectively, and explained 22.6% of the total phenotypic variance (**Table 2**). These SNPs were mapped on chromosome 2 at 10463640 and 11944909 bp physical positions, respectively. The Quantile–Quantile (QQ) plot corroborated with reducing  $-\log_{10}$  (p-value) towards the expected level for the ACMD (**Figure 3**). The LOD values for SNPs chr2\_10463640 and chr2\_11944909 were 4.83 and 4.06 with minor allele frequency (MAF) of 0.012 and 0.018, respectively.

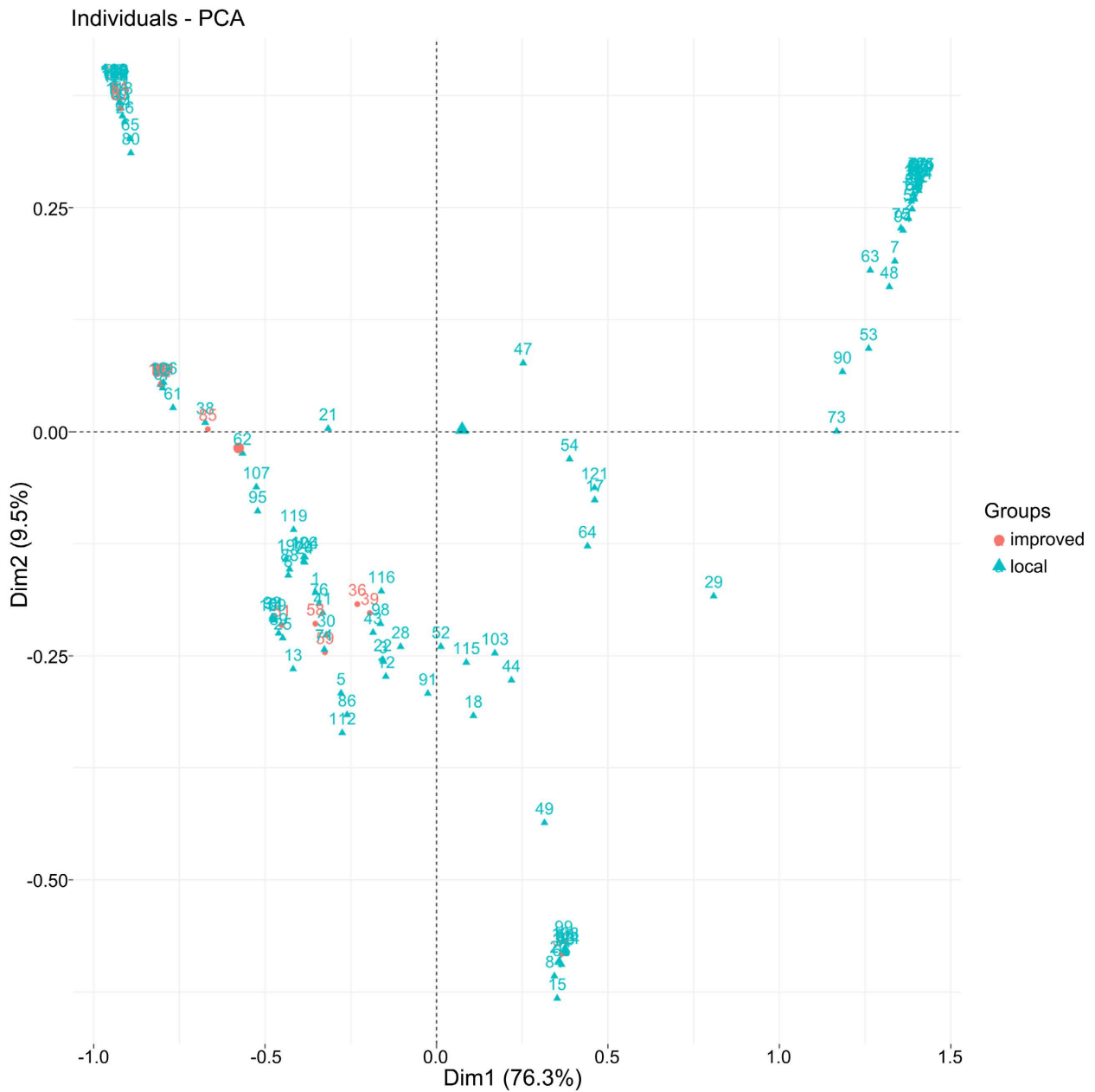
### 3.4. Genome-Wide Scan for Fresh Storage Root Yield

Three SNPs significantly associated with fresh storage root yield ( $t\text{-ha}^{-1}$ ), based

**Table 2.** Summary of significant single nucleotide polymorphism (SNPs) describing different genomic regions associated with African cassava mosaic disease (ACMD) and fresh root yield (FRY) in a panel of 32 cassava varieties.

Trait	SNP marker	Chromosome	Position (bp)	p-value	MAF	LOD score	Marker effect	R <sup>2</sup> (%)
ACMD	chr2_10463640	2	10463640	1.46 e <sup>-0.5</sup>	0.012	4.83	2.47	11.6
	chr2_11944909	2	11944909	8.68 e <sup>-0.5</sup>	0.018	4.06	2.93	11.0
FRY	chr1_31172650	1	31172650	5.28 e <sup>-0.5</sup>	0.012	4.28	6.08	11.4
	chr18_1103992	18	1103992	4.71 e <sup>-0.5</sup>	0.223	4.33	-2.33	10.1
	chr15_10392402	15	10392402	7.71 e <sup>-0.5</sup>	0.065	4.11	3.00	10.0

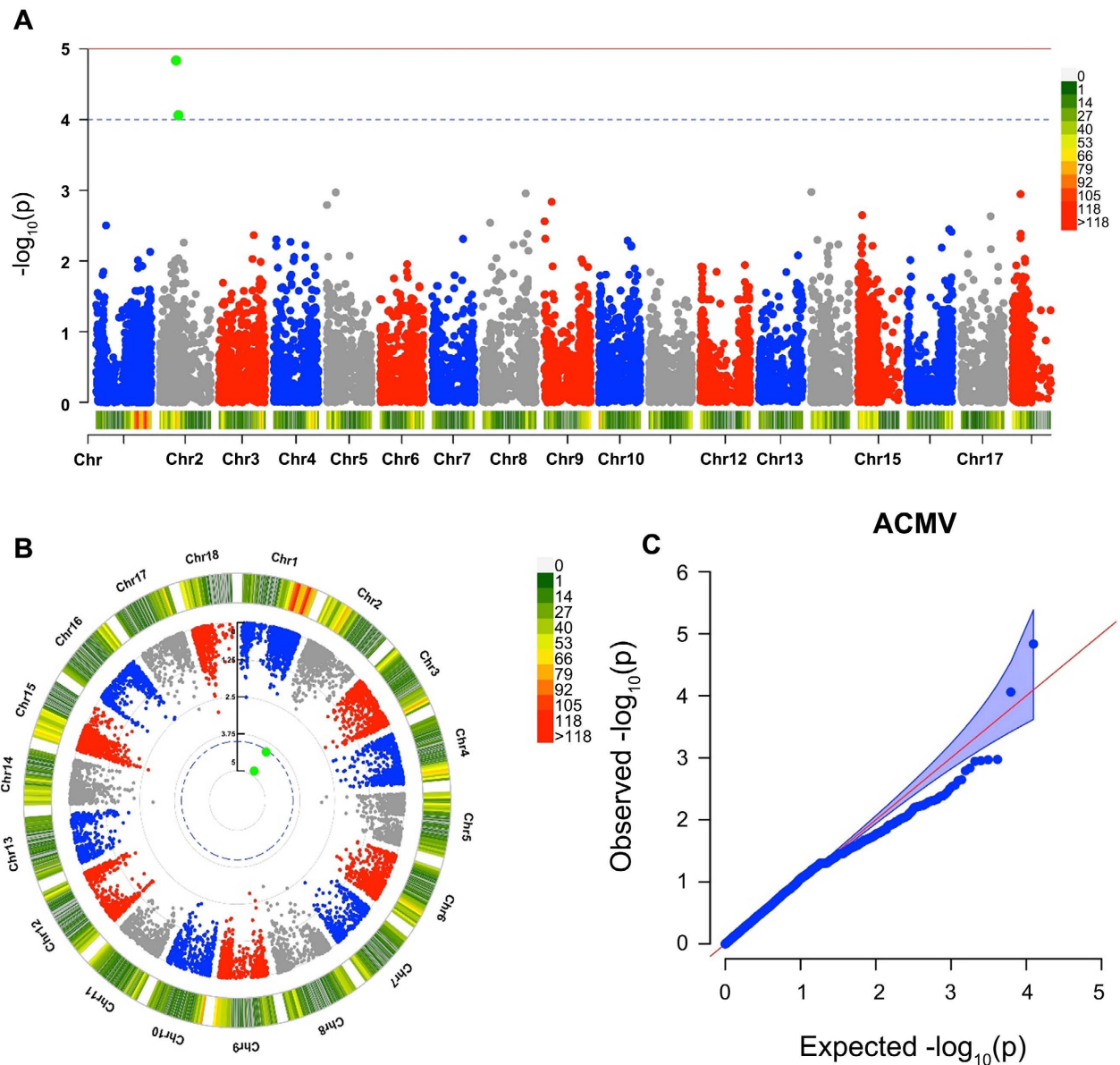
p represents the analysis of variance probability value associated with the variation across variants, SNP = single nucleotide polymorphism, MAF = minor allele frequency, LOD = logarithm of odds.



**Figure 2.** Principal component displaying the relationship between and among the local and improved varieties of cassava used in the study.

on  $p$ -values ( $p \leq 0.005$ ) across the whole cassava genome scan (**Table 2, Figure 4**). Of the three significant markers, one SNP locus (chr18\_1103992) had marker effect of  $-2.33$ , and explained 10.1% of the total phenotypic variation. The other two SNPs chr1\_31172650 and chr15\_10392402 exhibited marker effects of 6.08 and 3.00, respectively, and explained 21.4% of the total phenotypic variation. The SNPs chr1\_31172650, chr15\_10392402 and chr18\_1103992 were mapped on chromosomes 1, 15 and 18 at 31172650, 10392402 and 1103992 bp physical positions, respectively. Evidence of the SNP association was also found in the quantile-quantile (QQ) plot of the observed  $p$ -values of the association analysis for





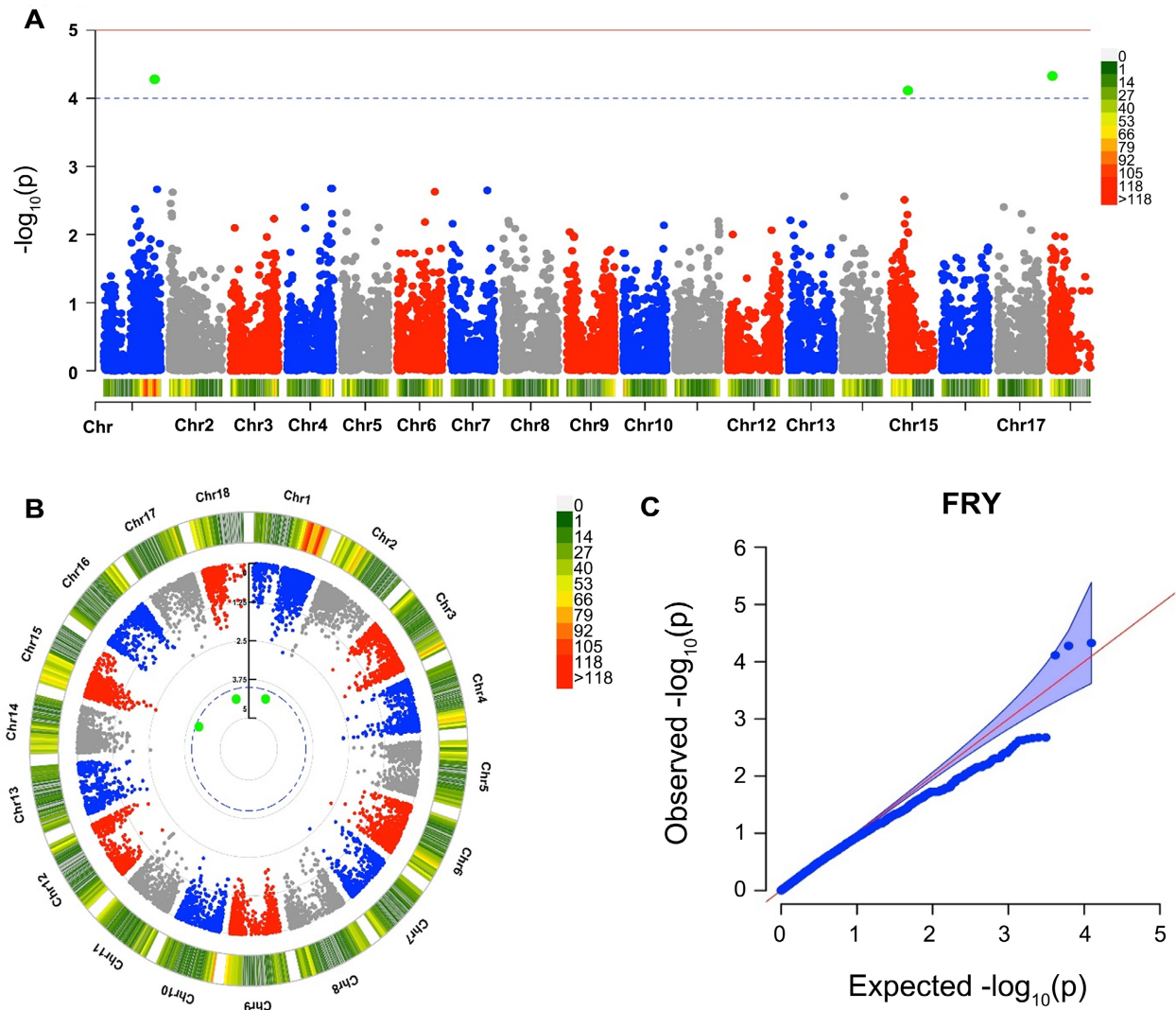
**Figure 3.** Genome-wide association analysis of African cassava mosaic disease (ACMD). (A) Rectangular Manhattan plot of 131 cassava varieties indicating the genomic regions significantly associated with the ACMD. The dashed lines on the Manhattan plot represent the significant threshold. (B) Circular Manhattan plot; and (C) Quantile-quantile (QQ) plot of the observed p-values of the association analysis that is expected in a null association for the phenotype.

the FSRY (**Figure 4**). The LOD values for SNPs chr1\_31172650, chr15\_10392402 and chr18\_1103992 were 4.28, 4.11 and 4.33 with minor allele frequency (MAF) ranging from 0.012, 0.065 and 0.223, respectively.

## 4. Discussion

### 4.1. Phenotypic Variation

The natural variation among the cassava accessions for traits was very informative. High broad-sense heritability of 0.99 for African cassava mosaic disease (ACMD) and intermediate broad-sense heritability of 0.51 for fresh root yield



**Figure 4.** Genome-wide association analysis of fresh root yield (FRY) of cassava. Rectangular Manhattan plot of 131 cassava varieties indicating the genomic regions significantly associated with the FRY. The dashed lines on the Manhattan plot represent the significant threshold. (B) Circular Manhattan plot; and (C) Quantile-quantile (QQ) plot of the observed p-values of the association analysis that is expected in a null association for the phenotype.

demonstrated substantial genetic variation in the studied traits among the different varieties. These findings indicated that the studied traits are amenable to genetic improvement through selection [43]. Furthermore, the observed natural genetic variation in the studied cassava germplasm signifies their relevance for genetic studies.

#### 4.2. Population Differentiation

A good understanding of the existing population structure within cassava breeding population is necessary for determination of its effects on the ability of GWAS to infer marker-trait association. In this study, all the three clustering methods utilized, DAPC, Kinship relationship matrix and structure, revealed five sub-populations that are imperative for preventing sham associations in GWAS

[44]. Thus, the sample size, marker density and diversity demonstrated that the cassava breeding panel used for this study is sufficiently powered to capture allelic variations for ACMD and FSRY studied.

The average minor allele frequency of the risk allele tested for varieties was greater than 10% indicating its significance in detecting genetic effects in the studied populations. The findings support the view that loci with high minor allele frequency have a higher power to detect weak genetic effects compared with those with lower minor allele frequency values [45] [46]. The numbers and patterns of SNP mutations in this study indicate a bias in chloroplast genome evolution in varieties of cassava. The diversity and patterns of SNP mutations in the cassava varieties were possibly due to the function of genes as previously suggested by Cao *et al.* [47].

### 4.3. Genome Wide Association Study

The whole-genome scan for phenotypic and allelic variation in African cassava mosaic disease resistance and fresh root yield identified genome regions on four chromosomes (chromosomes 1, 2, 15, and 18) with significant  $-\log_{10}$  values. Both Q matrix (population structure) and K matrix (Admixture) were considered covariates in a mixed linear model for the association analysis to reduce false-positive associations. The Q-Q plots for tolerance to African cassava mosaic disease and fresh root yield showed no inflation of p-values indicating that the structure of relationships was well accounted for in the GWAS analysis. These findings are consistent with the view that traits with no inflation of p-values show that the structural relationship is adequate for GWAS analysis [15] [48]. Genome-wide association mapping has been used to explore the elite alleles of many agronomic traits such as green mite infestation [49], cassava brown streak disease [50], cassava mosaic disease [51] and provitamin A and dry matter content [52] in cassava (*Manihot esculenta*). The genetic basis of ACMD resistance has been reported to be caused by a single major gene, CMD2 locus, detected on chromosome 12 [53] and chromosome 14 [54]. This gene was discovered using bi-parental linkage mapping [55] and GWAS [51]. In the present study, we detected two additional loci on chromosome 2 that indicate the existence of the ACMD resistance gene.

The phenotypic effect values of the favorable alleles of the studied traits were assessed and inferred to positively affect ACMD; and positively and negatively affect FSRY. Based on the stringent criterion of  $-\log_{10}$ , two significant markers trait associations exhibiting  $1.46 e^{-0.5}$  and  $8.68 e^{-0.5}$  were identified for ACMD and three ranging between  $4.71 e^{-0.5}$  and  $7.71 e^{-0.5}$  were identified for FSRY. The information on SNP variants from this study would fast-track the application of genomics-informed selection decisions in breeding cassava for resistance to African cassava mosaic disease and higher root yield. Similar studies revealing the great potential of GWAS in contributing to genomics-informed selection decisions have been reported for some root and tuber crops such as cassava [56], water yam [44], white yam [15] and potatoes [57].

## 5. Conclusion

The present study demonstrates the potential of highly informative and selective SNP markers for genetic diversity analysis and genome wide association studies in the 131 cassava varieties. This study also provides a direction for breeding efforts in the selection of parents from the current collection with potential for novel genes or QTLs for important agronomic traits: high root storage root yield and tolerance to African cassava mosaic virus (ACMD). Genome-wide association studies successfully identified and tagged five single nucleotide polymorphism (SNP) loci significantly associated with the studied traits. The information from our study could contribute to the design of new breeding strategies to hoard superior alleles for tuber yield per plant and yam mosaic virus in future marker-based breeding. Findings also contribute to a better understanding of the genetic architecture of ACMD and fresh root yield traits in cassava. The chromosomal regions controlling these studied traits could be exploited for selection and effective pyramiding of favorable alleles in white yam population improvement.

## Acknowledgements

This research was supported by the International Atomic Energy Agency (IAEA) TC project (FS-SIL5020-1801894) to Janatu V. Sesay (PhD student), Njala University (NU), Sierra Leone. The authors are thankful to National Crops Resources Research Institute (NaCRRRI), Kampala, Uganda, for facilitation with DNA extraction and analysis. The technical support from the cassava breeding team, Njala, Sierra Leone is dully acknowledged.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this work.

## References

- [1] Sánchez, T., Salcedo, E., Ceballos, H., Dufour, D., Mafla, G., Morante, N., Calle, F., Perez, J.C., Debouck, D., Jaramillo, G. and Moreno, I.X. (2009) Screening of Starch Quality Traits in Cassava (*Manihot esculenta* Crantz). *Starch/Starke*, **61**, 12-19. <https://doi.org/10.1002/star.200800058>
- [2] Scott, G.J., Rosegrant, M.W. and Ringler, C. (2000) Roots and Tubers for the 21st Century: Trends, Projections, and Policy Options. No. 31, 2020 Vision Discussion Papers, International Food Policy Research Institute (IFPRI), Washington DC. <https://econpapers.repec.org/paper/fpr2020dp/31.htm>
- [3] Nassar, N.M.A. (2004) Cassava: Some Ecological and Physiological Aspects Related to Plant Breeding. *Gene Conserve*, **3**, 229-245.
- [4] FAO (2020). <https://knoema.com/atlas/sources/FAO>
- [5] ITC International Trade Centre (2020) The Status of Cassava Production and Markets in Sierra Leone Preliminary Version (October 2020). 42 p. <https://new-staging.intracen.org/media/file/6535>
- [6] Tumuhimbise, R., Melis, R., Shanahan, P. and Kawuki, R. (2014) Genotype × Environment Interaction Effects on Early Fresh Storage Root Yield and Related Traits in Cassava. *Crop Journal*, **2**, 329-337. <https://doi.org/10.1016/j.cj.2014.04.008>

- [7] Brown, W.L. (1983) Genetic Diversity and Genetic Vulnerability—An Appraisal. *Economic Botany*, **37**, 4-12. <https://doi.org/10.1007/BF02859301>
- [8] Rao, V.R. and Hodgkin, T. (2002) Genetic Diversity and Conservation and Utilization of Plant Genetic Resources. *Plant Cell, Tissue and Organ Culture*, **68**, 1-19. <https://doi.org/10.1023/A:1013359015812>
- [9] Ribeiro, M.N.O., Carvalho, S.P., Santos, J.B. and Antonio, R.P. (2011) Genetic Variability among Cassava Accessions Based on SSR Markers. *Crop Breeding and Applied Biotechnology*, **11**, 263-269. <https://doi.org/10.1590/S1984-70332011000300009>
- [10] Andrade, E.K.V., Andrade Júnior, V.C., Laia, M.L., Fernandes, J.S.C., Oliveira, A.J.M. and Azevedo, A.M. (2017) Genetic Dissimilarity among Sweet Potato Genotypes Using Morphological and Molecular Descriptors. *Acta Scientia Agronomy*, **39**, 447-455. <https://doi.org/10.4025/actasciagron.v39i4.32847>
- [11] Melchinger, A. (1999) Genetic Diversity and Heterosis. In: James, G. and Coors, S.P., Eds., *The Genetics and Exploitation of Heterosis in Crops*, American Society of Agronomy, Inc., Madison, 99-118. <https://doi.org/10.2134/1999.geneticsandexploitation.c10>
- [12] Okogbenin, E., Egesi, C.N., Olanmi, B., Ogundapo, O., Kahya, S., Hurtado, P., Marin, J., Akinbo, O., Mba, C., Gomez, H., de Vicente, C., Baiyeri, S., Uguru, M., Ewa, F. and Fregene, M. (2012) Molecular Marker Analysis and Validation of Resistance to Cassava Mosaic Disease in Elite Cassava Genotypes in Nigeria. *Crop Science*, **52**, 2576-2586. <https://doi.org/10.2135/cropsci2011.11.0586>
- [13] Adjebeng-Danquah, J., Manu-Aduening, J., Asante, I.K., Agyare, R.Y., Gracen, V. and Offei, S.K. (2020) Genetic Diversity and Population Structure Analysis of Ghanaian and Exotic Cassava Accessions Using Simple Sequence Repeat (SSR) Markers. *Heliyon*, **6**, e03154. <https://doi.org/10.1016/j.heliyon.2019.e03154>
- [14] Gupta, P., Rustgi, S. and Mir, R. (2008) Array-Based High-Throughput DNA Markers for Crop Improvement. *Heredity*, **101**, 5-18. <https://doi.org/10.1038/hdy.2008.35>
- [15] Agre, P.A., Norman, P.E., Asiedu, R. and Asfaw, A. (2021) Identification of Quantitative Trait Nucleotides and Candidate Gene for Tuber Yield and Mosaic Virus Tolerance in an Elite Population of White Guinea Yam (*Dioscorea rotundata*) Using Genome-Wide Association Scan. *BMC Plant Biology*, **21**, 552. <https://doi.org/10.1186/s12870-021-03314-w>
- [16] Luo, Z., Tomasi, P., Fahlgren, N. and Abdel-Haleem, H. (2019) Genome-Wide Association Study (GWAS) of Leaf Cuticular Wax Components in *Camelina sativa* Identifies Genetic Loci Related to Intracellular Wax Transport. *BMC Plant Biology*, **19**, 187. <https://doi.org/10.1186/s12870-019-1776-0>
- [17] Qiu, X., Gong, R., Tan, Y., Yu, S., *et al.* (2012) Mapping and Characterization of the Major Quantitative Trait Locus qSS7 Associated with Increased Length and Decreased Width of Rice Seeds. *Theoretical and Applied Genetics*, **125**, 1717-1726. <https://doi.org/10.1007/s00122-012-1948-x>
- [18] Sánchez-Sevilla, J.F., Horvath, A., Botella, M.A., Gaston, A., Folta, K., Kilian, A., *et al.* (2015) Diversity Arrays Technology (DArT) Marker Platforms for Diversity Analysis and Linkage Mapping in a Complex Crop, the Octoploid Cultivated Strawberry (*Fragaria × ananassa*). *PLOS ONE*, **10**, e0144960. <https://doi.org/10.1371/journal.pone.0144960>
- [19] Kawano, K. (2003) Thirty Years of Cassava Breeding for Productivity—Biological and Social Factors for Success. *Crop Science*, **43**, 1325-1335. <https://doi.org/10.2135/cropsci2003.1325>



- [20] Okechukwu, R. and Dixon, A.G. (2008) Genetic Gains from 30 Years of Cassava Breeding in Nigeria for Storage Root Yield and Disease Resistance in Elite Cassava Genotypes. *Journal of Crop Improvement*, **22**, 181-208. <https://doi.org/10.1080/15427520802212506>
- [21] Ceballos, H., Kawuki, R.S., Gracen, V.E., Yencho, G.C. and Hershey, C.H. (2015) Conventional Breeding, Marker-Assisted Selection, Genomic Selection and Inbreeding in Clonally Propagated Crops: A Case Study for Cassava. *Theoretical and Applied Genetics*, **128**, 1647-1667. <https://doi.org/10.1007/s00122-015-2555-4>
- [22] Norman, P.E., Asfaw, A., Tongoona, P.B., Danquah, A., Danquah, E.Y. and Koeyer, D.D. (2018) Can Parentage Analysis Facilitate Breeding Activities in Root and Tuber Crops? *Agriculture Journal*, **8**, 1-24. <https://doi.org/10.3390/agriculture8070095>
- [23] García-Ruiz, A., Cole, J.B., VanRaden, P.M., Wiggans, G.R., Ruiz-López, F.J. and Van Tassell, C.P. (2016) Changes in Genetic Selection Differentials and Generation Intervals in US Holstein Dairy Cattle as a Result of Genomic Selection. *Proceedings of National Academy of Science, USA*, **113**, E3995. <https://doi.org/10.1073/pnas.1519061113>
- [24] Varshney, R.K., Terauchi, R. and McCouch, S.R. (2014) Harvesting the Promising Fruits of Genomics: Applying Genome Sequencing Technologies to Crop Breeding. *PLOS Biology*, **12**, e1001883. <https://doi.org/10.1371/journal.pbio.1001883>
- [25] Jiang, G.L. (2013) Molecular Markers and Marker-Assisted Breeding in Plants. In: Andersen, S.B., Ed., *Plant Breeding from Laboratories to Fields*, IntechOpen, London, 45-83. <https://doi.org/10.5772/52583>
- [26] Fukuda, W.M.G., Guevara, C.L., Kawuk, R. and Ferguson, M.E. (2010) Selected Morphological and Agronomic Descriptors for the Characterization of Cassava. International Institute of Tropical Agriculture (IITA), Ibadan, 19 p.
- [27] Gilmour, A.R., Thompson, R. and Cullis, B.R. (1995) Average Information REML: An Efficient Algorithm for Variance Parameter Estimation. *Biometrics*, **51**, 1440-1450. <https://doi.org/10.2307/2533274>
- [28] Butler, D.G., Cullis, B.R., Gilmour, A.A., Gogel, B.J. and Thome, R. (2018) ASReml-R Reference Manual Version 4. VSNi Ltd., Hemel Hempstead.
- [29] Borgognone, M.G., Butler, D.G., Ogonnaya, F.C. and Dreccer, M.F. (2016) Molecular Marker Information in the Analysis of Multi-Environment Trials Helps Differentiate Superior Genotypes from Promising Parents. *Crop Science*, **56**, 2612-2628. <https://doi.org/10.2135/cropsci2016.03.0151>
- [30] Ovenden, B., Milgate, A., Wade, L.J., Rebetzke, G.J. and Holland, J.B. (2018) Accounting for Genotype-by-Environment Interactions and Residual Genetic Variation in Genomic Selection for Water Soluble Carbohydrate Concentration in Wheat. *G3 Genes Genome Genetics*, **8**, 1909-1919. <https://doi.org/10.1534/g3.118.200038>
- [31] Dellaporta, S.L., Wood, J. and Hicks, J.B. (1983) A Plant DNA Miniprep: Version II. *Plant Molecular Biology Report*, **1**, 19-21. <https://doi.org/10.1007/BF02712670>
- [32] Aljanabi, S.M. and Martinez, I. (1997) Universal and Rapid Salt-Extraction of High-Quality Genomic DNA for PCR-Based Techniques. *Nucleic Acids Research*, **25**, 4692-4693. <https://doi.org/10.1093/nar/25.22.4692>
- [33] Kilian, A., Sanewski, G. and Ko, L. (2016) The Application of DArTseq Technology to Pineapple. *Acta Horticulturae*, **1111**, 181-188. <https://doi.org/10.17660/ActaHortic.2016.1111.27>
- [34] Danecek, P., Auton, A., Abecasis, G., Albers, C.A., *et al.* (2011) The Variant Call Format and VCFtools. *Bioinformatics*, **27**, 2156-2158.



- <https://doi.org/10.1093/bioinformatics/btr330>
- [35] Yin, L. (2019) Package “CMplot” 2019. <https://github.com/YinLiLin/CMplot>
- [36] Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J. and Sham, P.C. (2007) PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *American Journal of Human Genetics*, **81**, 559-575. <https://doi.org/10.1086/519795>
- [37] Le, S., Josse, J., Husson, F., *et al.* (2008) FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistics Software*, **25**, 1-18. <https://doi.org/10.18637/jss.v025.i01>
- [38] Pritchard, J.K., Stephens, M. and Donnelly, P. (2000) Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, **155**, 945-959. <https://doi.org/10.1093/genetics/155.2.945>
- [39] Falush, D., Stephens, M. and Pritchard, J.K. (2003) Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. *Genetics*, **164**, 1567-1587. <https://doi.org/10.1093/genetics/164.4.1567>
- [40] Evanno, G., Regnaut, S. and Goudet, J. (2005) Detecting the Number of Clusters of Individuals Using the Software Structure: A Simulation Study. *Molecular Ecology*, **14**, 2611-2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- [41] Earl, D.A. and vonHoldt, B.M. (2012) Structure Harvester: A Website and Program for Visualizing Structure Output and Implementing the Evanno Method. *Conservation and Genetic Resources*, **4**, 359-361. <https://doi.org/10.1007/s12686-011-9548-7>
- [42] Paradis, E. and Schliep, K. (2019) Ape 5.0: An Environment for Modern Phylogenetics and Evolutionary Analyses in R. *Bioinformatics*, **35**, 526-528. <https://doi.org/10.1093/bioinformatics/bty633>
- [43] Piaskowski, J., Hardner, C., Cai, L., Zhao, Y., Iezzoni, A. and Peace, C. (2018) Genomic Heritability Estimates in Sweet Cherry Reveal Non-Additive Genetic Variance Is Relevant for Industry-Prioritized Traits. *BMC Genetics*, **19**, 23. <https://doi.org/10.1186/s12863-018-0609-8>
- [44] Gatarira, C., Agre, P., Matsumoto, R., Edemodu, A., Adetimirin, V., Bhattacharjee, R., Asiedu, R. and Asfaw, A. (2020) Genome-Wide Association Analysis for Tuber Dry Matter and Oxidative Browning in Water Yam (*Dioscorea alata* L.). *Plants*, **9**, 969. <https://doi.org/10.3390/plants9080969>
- [45] Ardlie, K.G., Lunetta, K.L. and Seielstad, M. (2002) Testing for Population Subdivision and Association in Four Case Control Studies. *American Journal of Human Genetics*, **71**, 304-311. <https://doi.org/10.1086/341719>
- [46] Norman, P.E., Paterno, A.A., Danquah, A., Tongoona, P.B., *et al.* (2020) Paternity Assignment in White Guinea Yam (*Dioscorea rotundata*) Half-Sib Progenies from Polycross Mating Design Using SNP Markers. *Plants*, **2020**, 527. <https://doi.org/10.3390/plants9040527>
- [47] Cao, J., Jiang, D., Zhao, Z. *et al.* (2018) Development of Chloroplast Genomic Resources in Chinese Yam (*Dioscorea polystachya*). *BioMedical Research International*, **2018**, Article ID: 6293847. <https://doi.org/10.1155/2018/6293847>
- [48] Yang, J., Zaitlen, N.A., Goddard, M.E., Visscher, P.M. and Price, A.L. (2014) Advantages and Pitfalls in the Application of Mixed Model Association Methods. *Nature Genetics*, **46**, 100-106. <https://doi.org/10.1038/ng.2876>
- [49] Ezenwaka, L., Del Carpio Dunia, P., Jannink, J.L., Rabbi, I., Danquah, E., Asante, I., Danquah, A., Blay, E. and Egesi, C. (2018) Genome-Wide Association Study of Resistance to Cassava Green Mite Pest and Related Traits in Cassava. *Crop Science*,

- 58, 1907-1918. <https://doi.org/10.2135/cropsci2018.01.0024>
- [50] Kayondo, S.I., Pino Del Carpio, D., Lozano, R., Ozimati, A., Wolfe, M., Baguma, Y., Gracen, V., Ofei, S., Ferguson, M., Kawuki, R. and Jannink, J.-L. (2018) Genome-Wide Association Mapping and Genomic Prediction for CBSD Resistance in *Manihot esculenta*. *Scientific Report*, **8**, Article No. 1549. <https://doi.org/10.1038/s41598-018-19696-1>
- [51] Wolfe, M.D., Rabbi, I.Y., Egesi, C., Hamblin, M., Kawuki, R., Kulakow, P., Lozano, R., Carpio, D.P.D., Ramu, P. and Jannink, J.-L. (2016) Genome-Wide Association and Prediction Reveals Genetic Architecture of Cassava Mosaic Disease Resistance and Prospects for Rapid Genetic Improvement. *Plant Genome*, **9**, 1-13. <https://doi.org/10.3835/plantgenome2015.11.0118>
- [52] Ikeogu, U.N., Akdemir, D., Wolfe, M.D., Okeke, U.G., Amaefula, C., Jannink, J.-L. and Egesi, C.N. (2019) Genetic Correlation, Genome-Wide Association and Genomic Prediction of Portable NIRS Predicted Carotenoids in Cassava Roots. *Frontiers in Plant Science*, **10**, Article No. 1570. <https://doi.org/10.3389/fpls.2019.01570>
- [53] Akano A., Dixon, A., Mba, C., Barrera, E. and Fregene, M. (2002) Genetic Mapping of a Dominant Gene Conferring Resistance to Cassava Mosaic Disease. *Theoretical and Applied Genetics*, **105**, 521-525. <https://doi.org/10.1007/s00122-002-0891-7>
- [54] Rabbi, I.Y., Kayondo, S.I., Bauchet, G., Yusuf, M., Aghogho, C.I., Ogunpaimo, K., Uwugiaren, R., Smith, I.A., Peteti, P., Agbona, A., Parkes, E., Ezenwaka, L., Wolfe, M., Jannink, J.-L., Egesi, C. and Kulakow, P. (2020) Genome-Wide Association Analysis Reveals New Insights into the Genetic Architecture of Defensive, Agro-Morphological and Quality-Related Traits in Cassava. *Plant Molecular Biology*, **109**, 195-213. <https://doi.org/10.1101/2020.04.25.061440>
- [55] Rabbi, I.Y., Hamblin, M.T., Kumar, P.L., Gedil, M.A., Ikpan, A.S., Jannink, J.L. and Kulakow, P.A. (2014) High-Resolution Mapping of Resistance to Cassava Mosaic Geminiviruses in Cassava Using Genotyping-by-Sequencing and Its Implications for Breeding. *Virus Research*, **186**, 87-96. <https://doi.org/10.1016/j.virusres.2013.12.028>
- [56] Zhang, S., Chen, X., Lu, C., Ye, J., Zou, M., Lu, K. and Ma, P.A. (2018) Genome-Wide Association Studies of 11 Agronomic Traits in Cassava (*Manihot esculenta* Crantz). *Frontiers in Plant Science*, **9**, 503. <https://doi.org/10.3389/fpls.2018.00503>
- [57] Björn, B., Keizer, P.L., Paulo, M.J., Visser, R.G., Van Eeuwijk, F.A. and Van Eck, H.J. (2014) Identification of Agronomically Important QTL in Tetraploid Potato Cultivars Using a Marker-Trait Association Analysis. *Theoretical and Applied Genetics*, **127**, 731-748. <https://doi.org/10.1007/s00122-013-2254-y>

## Supplementary

**Supplementary Table S1.** Description of cassava varieties utilized for the study.

SN	Genotype	Status	SN	Genotype	Status
1	Cooksoon	local	43	Yangay	local
2	CKSHW101 R	local	44	Cocoa	local
3	Bandaquee	local	45	Kandagboe	local
4	SLICASS6	improved	46	Milikit	local
5	Marie	local	47	Blueboat	local
6	Kandae	local	48	Blueboat	local
7	Bomba	local	49	Cooksoon	local
8	Madawei	local	50	Nyon	local
9	Madawei	local	51	Kandae	local
10	Tangayawa	local	52	Cooksoon	local
11	Diamonyamawoi	local	53	Cocoa	local
12	S110 R	local	54	SLICASS6	improved
13	Rubber	local	55	Milikit	local
14	Guawae	local	56	Cocoa	local
15	IDA	local	57	Worldvision	local
16	Blueman	local	58	IDA	local
17	Diamonyamawoi	local	59	Pesae	local
18	Kadija	local	60	Kandabendu	local
19	Improved	improved	61	Munafa	local
20	Munafa	local	62	Farannaka	local
21	Tangayawa	local	63	Cooksoon	local
22	Rubber	local	64	Cocoa	local
23	Bongoman	local	65	Warima	local
24	Warima	local	66	SLICASS6	improved
25	Warima	local	67	SLICASS4	improved
26	Warima	local	68	Blueboat	local
27	Ya-Kanu	local	69	Blueboat	local
28	Warima	local	70	Warima	local
29	Jet	local	71	Bayamala	local
30	Improved	improved	72	Konko	local
31	Pajero	local	73	Guawai	local
32	Yarima	local	74	Warima	local
33	Warima	local	75	Guawae	local
34	Kickbox	local	76	Warima	local

**Continued**

35	Warima	local	77	Munafa	local
36	Munafa	local	78	Guineanca	local
37	Butter cassava	local	79	Cooksoon	local
38	Care	local	80	Munafa	local
39	Kanda-bendu	local	81	Cocoa	local
40	Munafa	local	82	Unknown	local
41	Diamonyamawoi	local	83	White	local
42	Potho E	local	84	SLICASS6	improved
85	Improved	improved	109	Marie	local
86	TME419	improved	110	Morovia	local
87	Blueboat	local	111	Nyon	local
88	Diamonyawawoi	local	112	Blueman	local
89	Cocoa	local	113	Improved	improved
90	Improved	improved	114	Cooksoon	local
91	Warima	local	115	Blueman	local
92	Marie	local	116	Warima	local
93	SLICASS4	improved	117	Gbagbanya	local
94	Tapioca	local	118	SLICASS6	improved
95	Cocoa	local	119	Tesae	local
96	Munafa	local	120	SLICASS4	improved
97	Njagae	local	121	Kamawai	local
98	Ya-kanu	local	122	Blueboat	local
99	Konko	local	123	Samuyana	local
100	Slicass6	improved	124	Diamonyamawoi	local
101	Diamonyamawoi	local	125	CHW151 R	local
102	Mamoudukuma	local	126	Rubber	local
103	Project	local	127	Shortman	local
104	Kendemeh	local	128	Cocoa	local
105	Tapioca	local	129	Korpoigibagie	local
106	Butter	local	130	Potho (wc)	local
107	Butter	local	131	Warima	local
108	Tangagboi	local			

**Supplementary Table S2.** Best linear unbiased predictor (BLUP) values of African cassava mosaic disease (ACMD) and fresh root yield (FRY) among 131 cassava varieties.

Entry	ACMV	FRY	Entry	ACMV	FRY
BA1	1.0	15.1	GE2	2.0	7.5
BB1	1.0	4.0	GW1	5.0	5.7

**Continued**

BC1	1.0	8.3	IDA	1.0	8.8
BM2	1.0	10.0	IDA2	1.0	3.8
BM3	1.0	7.7	IP1	2.0	6.6
BM4	1.0	4.2	IP2	1.0	8.5
BN1	1.0	2.1	IP4	1.0	3.9
BQ1	1.0	1.8	IP5	3.0	6.6
BT1	1.0	8.3	IP6	1.0	6.1
BT2	1.0	7.2	JET1	1.0	9.9
BT3	1.0	10.3	KA1	3.0	12.5
BT4	1.0	6.8	KB1	4.0	9.9
BT5	3.0	11.2	KD1	1.0	3.3
BT6	3.0	5.5	KD2	1.0	3.2
BU	1.0	12.9	KE1	0.0	7.3
BU2	2.0	5.6	KG2	3.0	6.8
CA1	3.0	1.8	KJ1	4.0	7.2
CH	2.0	5.1	KK2	2.0	7.8
CHS5	3.0	3.5	KM2	1.0	7.5
CK	1.0	7.1	Konko	1.0	5.3
CKS1	1.0	4.7	KU1	1.0	7.2
CKS2	3.0	12.8	KU2	1.0	5.8
CKS3	4.0	3.6	MA2	1.0	5.5
CKS4	1.0	3.0	MF1	2.0	6.7
CKS6	1.0	4.1	MF4	1.0	7.7
CO10	1.0	4.9	MF5	1.0	10.3
CO2	5.0	2.5	MF6	1.0	9.8
CO4	4.0	2.3	MF7	3.0	10.0
CO5	1.0	1.2	MF8	1.0	10.2
CO6	3.0	7.8	MF9	4.0	10.3
CO7	1.0	13.5	MK3	1.0	4.5
CO8	3.0	5.3	MK4	2.0	8.5
CO9	1.0	2.4	MM1	4.0	9.6
D1	1.0	7.5	MR1	1.0	6.8
D2	1.0	12.1	MV1	1.0	7.3
D3	0.0	7.3	MW1	2.0	5.7
D4	1.0	9.0	MW2	1.0	7.3
D5	1.0	5.3	NE1	1.0	9.2
D6	1.0	5.0	NY1	4.0	8.5

## Continued

FK1	1.0	7.4	NY2	1.0	10.3
GA1	4.0	1.5	PE1	3.0	3.9
GC2	1.0	6.1	PE2	1.0	6.6
GE1	1.0	4.1	PJ1	1.0	7.3
PO1	3.0	6.5	TS1	1.0	4.0
RB1	3.0	6.1	UN1	3.0	3.6
RU3	1.0	4.4	WA10	1.0	8.4
RU5	1.0	4.0	WA11	1.0	8.0
SC4_2	1.0	4.9	WA12	1.0	8.7
SC4_3	3.0	3.8	WA14	1.0	6.3
SC4_4	2.0	7.5	WA15	1.0	4.6
SC6_1	1.0	6.2	WA16	1.0	9.3
SC6_5	1.0	11.0	WA17	3.0	7.7
SC6_6	0.0	2.4	WA18	1.0	7.2
SC6_7	0.0	0.6	WA3	1.0	6.0
SC62	3.0	5.0	WA4	1.0	9.0
SC63	3.0	6.3	WA5	3.0	7.1
SIR	3.0	9.8	WA8	2.0	12.0
SM1	4.0	8.0	WA9	1.0	7.2
SY2	1.0	10.2	WC2	1.0	6.5
TA1	1.0	6.9	WH1	3.0	1.2
TA2	0.0	3.0	WV1	1.0	10.5
TC1	4.0	5.6	YA1	1.0	5.7
TC2	1.0	10.6	YK2	1.0	10.3
TME419	1.0	4.1	YK3	1.0	2.9
TO1	4.0	5.1	YY1	0.0	0.0

**Supplementary Table S3.** Cluster membership of 131 varieties of cassava based on Discriminant analysis of principal components (DAPC).

SN	Clone	Code	Status	Cl	SN	Clone	Code	Status	Cl
1	Cooksoon	CKS2	local	1	57	Worldvision	WV1	local	2
3	Bandaquee	BQ1	local	1	60	Kandabendu	KU2	local	2
5	Marie	MA2	local	1	63	Cooksoon	CHS5	local	2
12	S110 R	S1R	local	1	68	Blueboat	BT5	local	2
17	Diamonyamawoi	D3	local	1	73	Guawai	GE2	local	2
18	Kadija	KJ1	local	1	75	Guawae	GW1	local	2
21	Tangayawa	TA1	local	1	77	Munafa	MF4	local	2



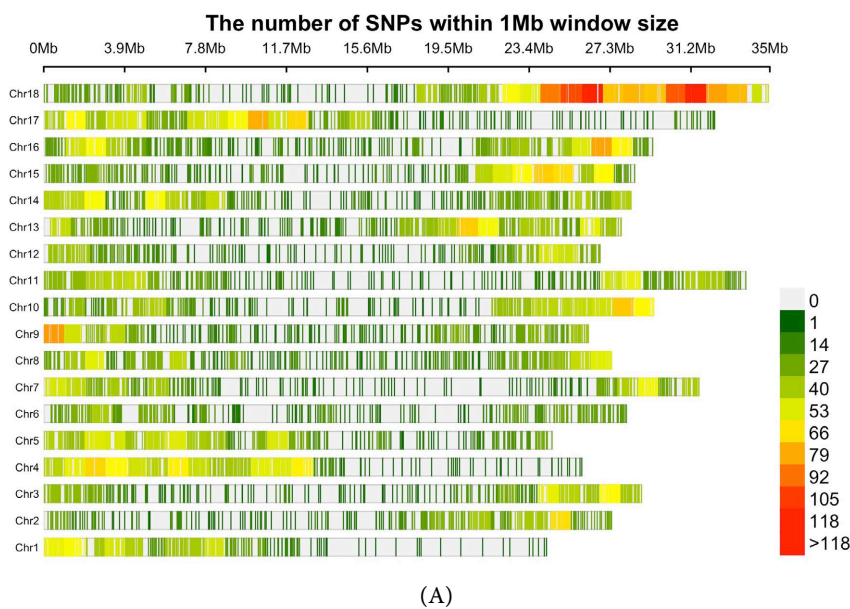
## Continued

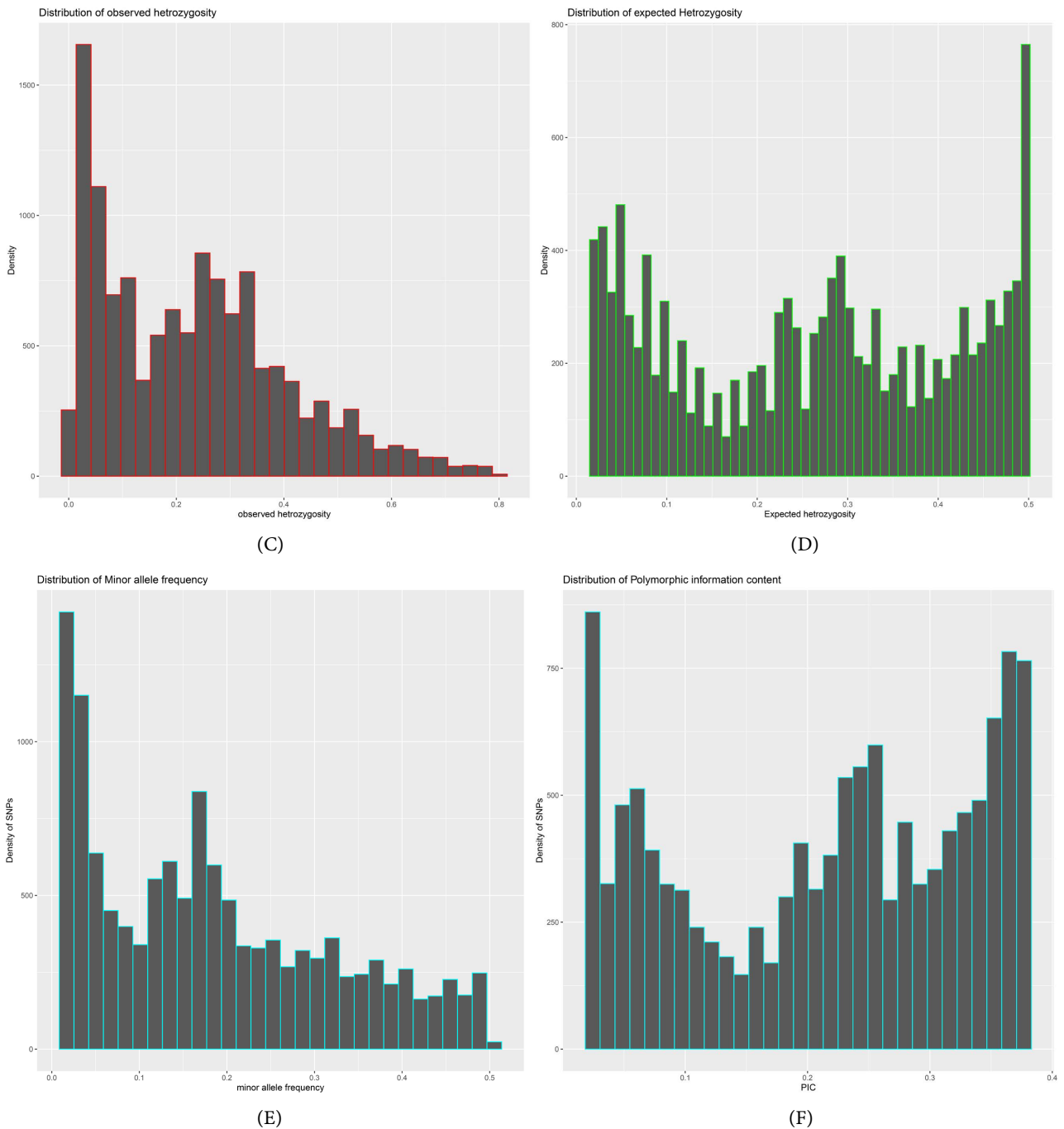
22	Rubber	RB1	local	1	78	Guineanca	GC2	local	2
24	Warima	WA17	local	1	82	Unknown	UN1	local	2
28	Warima	WA9	local	1	90	Improved	IP6	improved	2
30	Improved	IP1	improved	1	94	Tapioca	TC1	local	2
36	Munafa	MF8	local	1	97	Njagae	NE1	local	2
39	Kanda-bendu	KU1	local	1	102	Mamoudukuma	MM1	local	2
41	Diamonyamawoi	D2	local	1	111	Nyon	NY2	local	2
43	Yangay	YY1	local	1	114	Cooksoon	CKS3	local	2
44	Cocoa	CO7	local	1	125	CHW151 R	CH	local	2
47	Blueboat	BT3	local	1	129	Korpoigibagie	KE1	local	2
52	Cooksoon	CKS6	local	1	131	Warima	WA4	local	2
54	SLICASS6	SC6_6	improved	1	4	Slicass6	SC62	improved	3
58	IDA	IDA	local	1	6	Kandae	KD2	local	3
59	Pesae	PE2	local	1	8	Madawei	MW2	local	3
64	Cocoa	CO5	local	1	10	Tangayawa	TA2	local	3
71	Bayamala	BA1	local	1	11	Diamonyamawoi	D6	local	3
76	Warima	WA15	local	1	13	Rubber	RU3	local	3
79	Cooksoon	CKS1	local	1	19	Improved	IP4	improved	3
86	TME419	TME419	improved	1	25	Warima	WA11	local	3
91	Warima	WA18	local	1	32	Yarima	YA1	local	3
98	Ya-kanu	YK3	local	1	34	Kickbox	KB1	local	3
103	Project	PJ1	local	1	35	Warima	WA8	local	3
104	Kendemeh	KM2	local	1	38	Care	CA1	local	3
112	Blueman	BM3	local	1	45	Kandagboe	KG2	local	3
115	Blueman	BM4	local	1	51	Kandae	KD1	local	3
116	Warima	WA14	local	1	61	Munafa	MF6	local	3
119	Tesae	TS1	local	1	62	Farannaka	FK1	local	3
121	Kamawai	KA1	local	1	67	SLICASS4	SC4_4	improved	3
126	Rubber	RU5	local	1	70	Warima	WA3	local	3
2	CKSHW101 R	CK	local	2	85	Improved	IP5	improved	3
7	Bomba	BB1	local	2	88	Diamonyawawoi	D1	local	3
9	Madawei	MW1	local	2	89	Cocoa	CO4	local	3
27	Ya-Kanu	YK2	local	2	92	Marie	MR1	local	3
29	Jet	JET1	local	2	93	SLICASS4	SC4_3	improved	3
31	Pajero	PO1	local	2	95	Cocoa	CO6	local	3
40	Munafa	MF1	local	2	101	Diamonyamawoi	D5	local	3
42	Potho E	PE1	local	2	106	Butter	BU	local	3

Continued

46	Milikit	MK3	local	2	107	Butter	BU2	local	3
48	Blueboat	BT6	local	2	109	Marie	MA2	local	3
50	Nyon	NY1	local	2	14	Guawae	GE1	local	4
53	Cocoa	CO9	local	2	23	Bongoman	BN1	local	4
26	Warima	WA12	local	4	122	Blueboat	BT2	local	4
33	Warima	WA5	local	4	123	Samuyana	SY2	local	4
55	Milikit	MK4	local	4	127	Shortman	SM1	local	4
65	Warima	WA10	local	4	128	Cocoa	CO2	local	4
69	Blueboat	BT1	local	4	130	Potho (wc)	WC2	local	4
72	Konko	KK1	local	4	15	IDA	IDA2	local	5
74	Warima	WA16	local	4	16	Blueman	BM2	local	5
80	Munafa	MF5	local	4	20	Munafa	MF7	local	5
81	Cocoa	CO10	local	4	37	Butter cassava	BC1	local	5
87	Blueboat	BT4	local	4	49	Cooksoon	CKS4	local	5
96	Munafa	MF9	local	4	56	Cocoa	CO8	local	5
100	Slicass6	SC63	improved	4	66	SLICASS6	SC6_7	improved	5
105	Tapioca	TC2	local	4	83	White	WH1	local	5
110	Morovia	MV1	local	4	84	SLICASS6	SC6_5	improved	5
113	Improved	IP2	improved	4	99	Konko	KK2	local	5
117	Gbagbanya	GA1	local	4	108	Tangagboi	TO1	local	5
118	SLICASS6	SC6_1	improved	4	124	Diamonyamawoi	D4	local	5
120	SLICASS4	SC4_2	improved	4					

Cl = cluster.





**Supplementary Figure S1.** Overview of SNP genotyping data. (A) Density of SNPs on the 18 chromosomes of the cassava association mapping panel, (B) number of transition and transversion SNPs, (C) histogram of observed heterozygosity, (D) histogram of expected heterozygosity, (E) histogram of minor allele frequency distribution and (F) histogram of polymorphic information content across the *Manihot exculenta* genome.