Identification of Quantitative Trait Nucleotides and Candidate Genes for Tuber Yield and Mosaic Virus Tolerance in an Elite Population of White Guinea Yam (*Dioscorea Rotundata*) Using Genome-Wide Association Scan

Paterne AGRE (P.Agre@cgiar.org)
International Institute of Tropical Agriculture IITA Ibadan Nigeria
Prince E. Norman
Sierra Leone Agricultural Research Institute, PMB 1313, Tower Hill, Freetown, Sierra Leone
Robert Asiedu
International Institute of Tropical Agriculture IITA Ibadan Nigeria
Asrat Asfaw
International Institute of Tropical Agriculture IITA Ibadan Nigeria

Research Article

Keywords: Genetic diversity, population structure, GWAS, SNP markers, white Guinea yam

Posted Date: July 6th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-612999/v1

Version of Record: A version of this preprint was published at BMC Plant Biology on November 22nd, 2021. See the published version at https://doi.org/10.1186/s12870-021-03314-w.

Abstract

Background

Improvement of tuber yield and tolerance to viruses are priority objectives in white Guinea yam breeding programs. However, phenotypic selection for these traits is quite challenging due to phenotypic plasticity and cumbersome screening of phenotypic-induced variations. This study assessed quantitative trait nucleotides (QTNs) and the underlying candidate genes related to tuber yield per plant (TYP) and yam mosaic virus (YMV) tolerance in a panel of 406 white Guinea yam (*Dioscorea rotundata*) breeding lines using a genome-wide association study (GWAS).

Results

Population structure analysis using 5,581 SNPs differentiated the 406 genotypes into four distinct sub-groups ($K = 4$). Marker-trait association (MTA) analysis using the generalized linear model identified ten QTN regions significant for TYP and five for YMV. We identified variants responsible for predicting higher yield and low virus severity scores in the breeding panel through the marker-effect prediction. Gene annotation for the significant SNP loci identified several essential putative genes associated with the growth and development of tuber yield and those that code for tolerance to mosaic virus.

Conclusion

Our results provide valuable insight for marker validation and deployment for tuber yield and mosaic virus tolerance in white yam breeding. The information on SNP variants and genes from the present study would fast-track the application of genomics-informed selection decisions in breeding white Guinea yam for rapid introgression of the targeted traits.

Background

Root and tuber crops are significant contributors to global food supply next to cereal crops. Yam is among the principal root and tuber crops, after cassava and potato, that are widely grown and consumed as subsistence staples [1]. Yam is a collective name for the *Dioscorea* species extensively cultivated in the tropics and subtropics by smallholder farmers for its starchy underground tuber and aerial bulbils [2, 3]. The global estimated mean annual yam production and gross values are approximately 73 million tons and 14 billion US dollars, respectively, with West Africa accounting for 92% of the total yam production [4, 5]. There are over 600 *Dioscorea* species, of which 11 are economically significant [6]. White Guinea yam (*D. rotundata*), indigenous to Africa, is the most produced and consumed among cultivated species, supporting the livelihood of over 300 million people [2]. Yam is also important in many key life ceremonies in the major producing areas of West Africa [7].

Despite its socio-economic importance, a significant yield increase has not been achieved over the decades compared to cereal crops [1]. Improved varieties are vital for attaining increased productivity in farmers' fields. The development of improved yam varieties requires a better understanding of the genetic control of traits contributing to the increased yield and acceptable quality by growers and consumers. However, the breeding efforts have not adequately explored the genetic basis of tuber yield and virus resistance traits to fast-track improved cultivar development. Genes controlling key traits such as resistance to pests and diseases, tuber yield, and tuber quality traits exhibit quantitative inheritance. They may not be linked in a preferred direction, making improving these traits challenging using conventional breeding techniques [8]. In QTL mapping studies, the variation in virus resistance is attributed to a single major locus with a modest contribution [9]. Two random amplification of polymorphic DNA (RAPD) markers tightly linked in the coupling phase with Ymv-1 locus on the same linkage group were reported in resistant genotypes of *D. rotundata*.

For tuber yield, limited knowledge exists regarding QTL mapping studies [8]. The QTLs detected for YMV in yam were mainly based on conventional family-based linkage mapping. In contrast, the GWAS strategy using naturally occurring variants is a more robust and efficient method for identifying significant loci and the genes involved in the genetic control of complex traits. 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. However, GWAS mapping for tuber yield and YMV tolerance in yam has not been reported to date.

Supporting yam breeding efforts based on quantitative genetics principles and genomics tools is indispensable to increase the program's effectiveness for increasing productivity. Yam cultivar development using conventional strategies spans at least ten years from crossing to variety release recommendation [4, 6]. The complementation of the traditional breeding techniques with advanced molecular tools has reduced the breeding cycle in crops [10]. In theory, genotypic information from molecular markers, when associated with phenotypic traits of interest, may be extensively used to select individuals with higher genetic value through marker-assisted selection (MAS) [11].

This study's objective was to dissect the genetic control of tuber yield and YMV tolerance in white Guinea yam.

Material And Methods

Plant Materials

The study panel comprised 406 white Guinea yam clones comprising 36 trait progenitors, 49 elite clones, and 321 early generation breeding lines from the IITA's yam breeding program (Supplementary Table 1). All the genotypes are from the International Institute of Tropical Agriculture, IITA Ibadan Nigeria and are maintained by the Yam Breeding Improvement Unit.

Phenotyping

Phenotypic data on tuber yield and yam mosaic virus (YMV) severity were recorded on the plant materials assessed at different breeding stages by IITA in Nigeria. The tuber yield and mosaic virus severity were recorded on plants in the field using the procedure described in yam ontology (http://www.cropontology.org/ontology/CO_343/Yam) and yam standard operation protocol [12]. Tuber yield was recorded in kilogram on a plant basis at harvest (eight months after planting). The YMV severity score was assessed at 30-day intervals from 2 to 6 months after planting based on a visual assessment of the relative area of plant leaf surfaces affected by the mosaic virus disease using a five-ordinal scale of 1–5. A score of 1 represented no visible symptoms of virus infection, 2 for mild mosaic, vein-banding, green spotting or flecking, curling and mottling on few leaves but no leaf distortion, 3 for low incidence (25 - 50%) of the mosaic virus on the entire plant, 4 for the severe mosaic on most leaves and leaf distortion, and 5 for severe mosaic and bleaching with severe leaf distortion and stunting. The virus severity score values were converted to percentages and then used to estimate the area under disease progress curve (AUDPC) values as described by Forbes et al. [13]:

$$\mathbf{AUDPC} = \sum_{i=1}^{n-1} \left(\frac{\boldsymbol{y_i} + \boldsymbol{y_{i+1}}}{\mathbf{2}} \right) (\boldsymbol{t_{i+1}} - \boldsymbol{t_i})$$

where y_i = disease severity at the i^{th} observation, t_i= time (days) at the i^{th} observation, and n = total number of observations.

Phenotypic data analysis

We applied a one-step linear mixed model that used G-matrix to compute the best linear unbiased predictor (BLUP) values of an individual clone for a trait from the best fit model using the average information criterion (AIC) in restricted maximum likelihood (REML) algorithm [14] in the ASReml-R version 4 package [15]. The model used was:

$$\mathbf{y} = \mathbf{X_\beta} + \mathbf{Zuu} + \mathbf{Zgua} + \mathbf{Zgu\tilde{a}} + \boldsymbol{\varepsilon}$$

where is the data vector of the response variable across *p* locations with *Nj* plots per location *j*. Each location was treated as a separate trial so that *p*=4; β is the vector of fixed effects associated with the corresponding design matrix (X), including the location main and location-specific design-based replication effects. The term is a vector of random effects associated with location-specific field blocking structures (block within replication) used to capture extraneous variation with the corresponding design matrix ; and are the vectors of random additive and non-additive genetic within location effects, respectively, with corresponding design matrix Zg. Accordingly, the genetic variance was partitioned into the additive effects, which were associated with a covariance structure proportional to genetic relationships derived from the molecular markers and the non-additive genetic effect. This variance is explained by individual identity rather than the genomic relationship matrix following the approach

described by Borgognone et al. [16] and Ovenden et al. [17]. is the vector of residuals modeled as random within each trial.

Broad sense heritability (H^2) estimates for the traits were calculated from phenotypic variance (σ^2p) and the genotypic variance (σ^2g). The BLUP values of the genotypes for the traits extracted from the best fit model were used as input for the GWAS model.

Genotyping and SNP Data Analysis

For each genotype, total genomic DNA was isolated from lyophilized young and fully expanded healthy leaves. Deoxyribonucleic acid (DNA) was extracted from the leaf samples using the CTAB procedure with slight modification [18]. DNA quality and concentration were assessed using agarose gel and nanodrop, respectively, following the methods described in Aljanabi and Martinez [19]. High-throughput genotyping was conducted in 96 plex DArTseq protocol, and SNPs were called using the DArT's proprietary software, DArTSoft, as described by Killian et al. [20]. Reads and tags found in each sequencing result were aligned to the *Dioscorea rotundata* reference genome version 2 (https://drive.google.com/drive/folders/1H5T4xjKAEl9LliR-4qK_lR6TypCDe8nj) with Hisat2 [21]. 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%; minor allele frequency (MAF) <0.05; genotype quality <20 and heterozygosity >50. Of the 16,242 SNP markers subjected to the filtering quality criteria, 5,581 good-quality SNPs were retained for various analyses.

Population Genetic Analysis

Various population genetic analysis methods were conducted to explore the structure and level of genetic diversity in the study material. The SNP distribution and the density were estimated using the '*Cmplot*' function implemented in the CMplot R package [22]. About the SNP mutation from the reference to the alternative, SNPlay open website was used to estimate the rate of the transition and transversion across the retained SNP. Statistics such as the minor allele frequency (MAF), the observed and the expected heterozygosity, and the polymorphism information content were estimated using the function "–freq" and "–hardy" using PLINK [23].

The exploratory Discriminant Analysis of Principal Components (DAPC) was applied using the adegenet package [24]. The number of clusters was assessed using the "*find.clusters*" function, which runs successive K-means clustering with an increasing number of clusters (k).

The Admixture was performed through the Bayesian Information Criterion (BIC). The number and nature of clusters were done to assess the best-supported model recommended by Jombart et al. [24]. The optimal number of clusters was inferred using k-means analysis after varying the possible number of clusters from 2 to 20, and the posterior probability was used to define the cluster membership. A hierarchical cluster was constructed utilizing a kinship relation matrix implemented in GAPIT [25].

Genome Wide-Association Analysis

The GWAS analysis was performed using the compressed mixed linear model (CMLM) [26] implemented in the GAPIT R package and the Manhattan and the QQ plot were visualized in CMplot [22].

In the GAPIT analysis, we accounted for population structure (Q) through a principal component (PC) analysis and for relationships among individuals through a kinship (K) matrix [26] generated from marker data. For each trait, the optimal number of PCs/covariates to include in GWAS models was determined through model selection using the Bayesian information criterion (BIC), with a maximum of four tested PCs. For the GWAS study, the following formula was used: $Y = X_{\beta} + W\alpha + Qv + Zu + \varepsilon;$ where Y was treated as the observed vector of BLUP; β as the fixed effect vector (p × 1) other than molecular markers effects and population structure; α 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 β, α, v, and u parameters; and ε as the residual effect vector. The significance threshold for marker-trait associations (MTA) was set to p = 0.05 after applying the Bonferroni threshold correction (-log10(P)) with P the associated probability assigned to each marker. The percentage of variation explained by the associated marker (R^2) was calculated through a stepwise regression implemented in GAPIT.

Circular Manhattan and Quantile-Quantile (QQ) plots were generated by plotting 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 [22]. The Manhattan plot was created for visualizing GWAS on the entire genome, and zoom mapping was performed on a particular chromosome after identifying a significant SNP marker.

Gene annotation and marker effect prediction

The possible candidate genes within the significant QTL region were searched in the defined range window of 1 MB at 500 Kb (downstream and upstream) from the yam Generic File Format (GFF3) file. Linkage disequilibrium (LD) was assessed between the significant SNPs using the LDheatmap library [27]. The yam generic feature format (GFF3) of the reference genome was used to identify the main gene in the inter-genic region using the SNPReff. The functions of the various genes associated with the SNPs identified were determined using the public database Interpro, European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI) [28]. All pairwise LD (r^2) estimation between loci within a window of 1MB of each chromosome was done in TASSEL 5.0 [29]. The Pairwise LD estimates in the region of interest for significantly associated markers were investigated using Haploview 4.2 [30]. Finally, LD plotting was done based on base pairs (bp) distance, using "*ggplot2"* package in R [31].

Haplotype estimation and in vivo marker prediction

Haplotype associated with significant QTL was developed using "*rstatix"* package implemented in R, and the sequence of each haplotype was defined based on the 406 genetic material considered as testing and or identification population. The variant effect prediction was evaluated through the adjusted posterior probability, and the major QTL was identified, and their effects were predicted in unrelated diversity. Marker effects were then plotted, and SNP effects were compared in both populations.

Results

Phenotypic data of the white yam

Table 1 presents summary statistics for the phenotypic traits assessed. Broad-sense heritability estimates were high, 0.708 for tuber yield per plant and 0.903 for yam mosaic virus. The phenotypic value for the tuber yield ranged from 0.93 to 1.47 kg plant^{-1} with an average of 1.19 kg. The area under the disease progress curve for YMV ranged from 100.56 to 2900.45 with an average of 936.16. (Supplementary Table 2).

Table 1. Descriptive statistics of tuber yield per plant (TYP) and yam mosaic virus (YMV)

Traits	Minimum	Maximum	Mean	Standard Deviation	Broad sense heritability (H^2)
TYP	0.93	1.47	1.19	0.11	0.708
YMV	100.56	2900.45	936.16	481.19	0.903

TYP: Tuber yield per plant; YMV: yam mosaic virus (AUDPC value)

Genetic diversity, population structure and Linkage Disequilibrium

The DArT genotyping of 406 white Guinea yam clones detected the highest number of SNPs (637) mapped on chromosome 5 and the lowest of 123 on chromosome 11 (Supplementary Figure 1A). Transition SNPs (60.13%, 3,356 SNPs) were more frequent than transversions (39.87%, 2225 SNPs) (Supplementary Figure 1B). The observed heterozygosity value ranged from 0.029 to 0.622, with an average of 0.336 (Supplementary Figure 1C). The expected heterozygosity value ranged from 0.09 to 0.5, with an average of 0.331 (Supplementary Figure 1D). The minor allele frequency ranged from 0.05 to 0.5, with a mean of 0.24 (Supplementary Figure 1E). The polymorphic information content (PIC) ranged from 0.087 to 0.335, with an average value of 0.267 (Supplementary Figure 1F).

The DAPC analysis grouped the 406 clones into four main clusters with 0.99% of group assignment probability (Figure 1A). The highest membership was recorded in cluster 4 (154 clones) and the lowest in cluster 1 (70 clones). Through the DAPC, Group 1 comprises 70 individuals constituting early generation breeding lines and the landraces (clones collected from farmers and maintained at the Genetic Resources Center, International Institute of Tropical Agriculture, Nigeria). Group 2 consisted of 96 individuals, including early generation breeding lines, the elite clones and the landraces. Groups 3 (86 members) and

4 (154 members) comprised mainly half-sibs from natural intermating among parents established in poly-cross block (Supplementary Table 3). Phylogeny tree analysis through the kinship matrix revealed four clusters as well (Figure 1C). Cluster membership displayed through the phylogeny tree was in perfect alignment with the DAPC cluster membership (Supplementary Table 3). Using a cut-off of 0.5, about 233 clones were assigned to a respective groups, while many clones (173) were considered as admixed (Figure 1D) with an assigned probability less than 0.5.

Genome-Wide Scan for Traits

Tuber yield

We found ten SNPs significantly associated with tuber yield (kg plant^{-1}), based on p-values ($p \leq 0.005$) across the whole yam genome scan after correction for the false discovery rate (FDR). The evidence for SNP association was also observed in the quantile-quantile (QQ) plot of the observed *p* values of the association analysis (Figure 2). The LOD value for these SNPs ranged from 3.95 (chr08_13478676) to 4.71 (chr14_14300070) with minor allele frequency (MAF) ranging from 0.069 to 0.376. The ten significant SNPs were located on five chromosomes (Table 2; Figure 2). Of the 10 SNP markers associated with tuber yield, five were mapped on chromosome 14, two on chromosome 8, and a single SNP each on chromosomes 6, 15, and 19 (Table 2). Each of the significant SNPs explained over 62% of the observed variation in tuber yield.

Yam mosaic virus resistance

We found five SNP loci that showed a significant association with the reaction to mosaic virus infection (Table 2, Figure 3A). Of the significant SNPs associated with YMV, two markers named chr04_23336802 and chr10_1116193 displayed negative marker effect (Table 2). The SNP markers associated with YMV were explained ~12% of the total phenotypic variance and displayed LOD ³ 4 (Table 2). Although the population structure matrix was included in the models, the QQ-plot remained inflated for some SNPs (Figure 3B). The minor allele frequency (MAF) of the associated SNP marker ranged from 0.069 to 0.455.

Table 2. SNP markers associated with the tuber yield per plant and yam mosaic virus severity score.

Trait	Marker name	Chr	Position (Bp)	PVE (%)	Variants	MAEF	LOD	MAF
TYP	chr06_965642	6	96,5642	62.13	TC	-0.02	4.12	0.376
	chr08_13478676	8	13,478,676	62.08	AT	0.04	3.95	0.095
	chr08_21789522	8	21,789,522	62.19	AT	0.05	4.35	0.096
	chr14_8143018	14	8,143,018	62.13	TA	0.03	4.06	0.155
	chr14_11128124	14	11,1281,24	62.74	GC	0.05	4.70	0.154
	chr14_14292173	14	14,292,173	62.62	GA	-0.07	4.44	0.069
	chr14_14300070	14	14,300,070	62.74	AC	-0.08	4.71	0.069
	chr14_15507116	14	15,507,116	62.53	AT	0.05	4.24	0.071
	chr15_4138634	15	4,138,634	62.36	AT	-0.03	4.00	0.181
	chr19_9446619	19	9,446,619	62.03	AG	0.03	4.00	0.091
YMV	chr03_6338751	3	6,338,751	12.42	GT	169.09	4.28	0.455
	chr04_23336802	4	23,336,802	12.95	AG	-225.25	4.10	0.254
	chr10_613339	10	613,339	12.38	AT	138.99	4.36	0.388
	chr10_1116193	10	1,116,193	12.85	GA	-240.87	4.00	0.260
	chr16_1325272	16	1,325,272	13.52	CT	295.41	4.12	0.069

TYP: tuber yield (kg plant^{-1}), YMV: Yam mosaic virus severity score (AUDPC value); LOD: Logarithm of odds; Chr: chromosomes; bp: base-pair; MAF: Minor allele frequency; PVE: Phenotypic variance explained and MAEF: Marker effect

3.3. Putative gene prediction for the traits

3.3.1 Tuber yield

We explored the association of the identified QTN regions on the physical map with the potential candidate genes and their functions using the white Guinea yam genome sequence. The LD heatmap of the significant SNPs on chromosomes 6, 8, 14, 15 and 19 displayed a high genetic correlation (0.4 to 0.8) between the specific SNPs in the vicinity of the peak adjacent to the putative gene (Figure 4). On chromosome 6, the significant SNP for tuber yield is located on the genomic regions harboring three putative genes (AUX/IAA protein, AP2/ERF domain and Phloem protein 2-like) with known functions. On chromosome 8, we identified two putative genes (AUX/IAA protein; Glycine-rich protein) (Supplementary Table 4). Several putative genes were identified on chromosome 14 (Supplementary Table S4). On chromosome 15, which displayed average correlation through the LDheatmap, five genes were identified

in the vicinity of the targeted SNP marker. The LD heatmap for the SNP found in association with tuber yield on chromosome 19 revealed the presence of 9 putative genes (ABC transporter-like, Exportin-1/Importin-beta-like, Sodium/calcium exchanger membrane region, AUX/IAA protein, Geminivirus AL3 coat protein, AP2/ERF domain, Major facilitator, sugar transporter-like, and Expansin).

3.3.2. Yam mosaic virus resistance

We identified four candidate genes, namely AP2/ERF domain, Major facilitator, sugar transporter-like, and AUX/IAA protein on chromosome 3 near the SNP found in association with the YMV. The four identified candidate genes, AP2/ERF domain and AUX/IAA protein, were reported to confer essential gene functions related to plant defense and growth. The pairwise LD between the SNP of chromosome 3, 4 10 and 16 situated in genomic regions associated with YMV displayed a higher correlation with the three main haplotypes block (Figure 5). On chromosome 10, fifteen different putative genes were identified near the significant SNPs as being associated with the YMV resistance, namely SNF2-related domain, Geminivirus AL3 coat protein, SANT/Myb domain, Geminivirus AL1 replication-associated protein, CLV type, Chlorophyll A-B binding protein, AP2/ERF domain, Gdt1 family, NB-ARC, Probable transposase, Ptta/En/Spm plant, Geminivirus AL1 replication-associated protein, catalytic domain, Kinesin-like protein and Geminivirus Rep catalytic domain.

Haplotype SNP distribution and prediction

The frequencies and marker prediction effects of various haplotypes associated with tuber yield and resistance to yam mosaic virus in white Guinea yam are presented in Table 3. Of the ten SNP markers associated with the tuber yield, seven (chr06_965642, chr14_11128124, chr14_14292173, chr14_14300070, chr14_15507116, chr15_4138634 and chr19_9446619) displayed high haplotype segregation among the different variants. For the SNP markers located on chromosome 6, variants CC and CT were identified to be associated with genotypes with higher tuber yield. In contrast, TT was found to be associated with lower tuber yield (Figure 6A). For the two SNP markers located on chromosome 8, variant AA seems to be linked with lower tuber yield, while TT and AT were linked to higher tuber yielding genotypes (Figure 6B, C). For chromosome 14, homozygote variant AA was identified to predict higher tuber yield (Figure 6G, H, I), and the same trend was observed on SNP located on chromosome 15 (Figure 6D) and chromosome 19 (Figure 6E). Of the five SNP markers associated with the YMV, three (chr04_23336802, chr10_1116193 and Chr16_1325272) were found to have high significant haplotype variation (Table 3). For the SNP marker located on chromosome 3, no variation was observed for the different variants (Figure 7A). On chromosome 4, SNP markers associated with the YMV located at 23,336,802 bp showed that variants AG and AA were linked to lower predicted YMV value. For the marker chr10_613339 associated with YMV, no difference was observed among the different allele variants (Figure 7 D). In contrast, variants GA and GG of marker chr10_1116193 predicted lower YMV severity scores (Figure 7E).

Table 3. Frequencies and marker prediction effects of various haplotypes associated with tuber yield (kg plant^{-1}) and reaction to yam mosaic virus infestation (AUDPC value)

Trait	Marker name	Haplotype	Sequence	Frequency	Adjusted Probability	Prob. adj. significance
TYP	chr06_965642	Hap1	CCTC	0.27	0.035	*
		Hap2	CCTT	0.33	1.16 e^{-07}	****
		Hap3	TCTT	0.40	0.001	***
	chr8_21789522	Hap1	AAAT	0.47	0.132	Ns
		Hap2	AATT	0.47	0.319	Ns
		Hap3	ATTT	0.07	0.110	Ns
	chr8_13478676	Hap1	AAAT	0.47	0.272	Ns
		Hap2	AATT	0.47	0.272	Ns
		Hap3	ATTT	0.06	0.110	Ns
	chr14_8143018	Hap1	AATA	0.11	1.000	Ns
		Hap2	AATT	0.44	1.000	Ns
		Hap3	TATT	0.45	1.000	Ns
	chr14_11128124	Hap1	CCCG	0.45	1.90 e^{-07}	****
		Hap2	CCGG	0.44	4.71 e^{-14}	****
		Hap3	CGGG	0.11	0.043	*
	chr14_14292173	Hap1	GAGG	1.00	0.007	**
	chr14_14300070	Hap1	AAAC	1.00	0.008	**
	chr14_15507116	Hap1	AAAT	0.48	0.001	***
		Hap2	AATT	0.47	0.003	**
		Hap3	ATTT	0.05	0.534	Ns
	chr15_4138634	Hap1	AAAT	0.46	0.004	**
		Hap2	AATT	0.40	0.004	**
		Hap3	ATTT	0.14	0.169	Ns
	chr19_9446619	HAP1	AAAG	0.47	0.002	**
		HAP2	AAGG	0.48	1.99 e^{-06}	****
		HAP3	AGGG	0.05	0.971	Ns
YMV	chr03_6338751	Hap1	GGGT	0.4	1.000	Ns

	Hap2	GGTT	0.24	1.000	Ns
	Hap3	GTTT	0.36	1.000	Ns
chr04_23336802	Hap1	AAAG	0.47	0.013	*
	Hap2	AAGG	0.3	0.000	***
	Hap3	AGGG	0.23	0.002	**
chr10_613339	HAP1	AAAT	0.44	0.190	Ns
	HAP2	AATT	0.24	0.023	*
	HAP3	ATTT	0.33	0.190	Ns
chr10_1116193	Hap1	AAGA	0.23	0.239	Ns
	Hap2	AAGG	0.31	0.003	**
	Hap3	GAGG	0.47	$5.25e^{-07}$	****
chr16_1325272	Hap1	CCCT	0.49	0.576	Ns
	Hap2	CCTT	0.45	$2.08\ e^{-06}$	****
	Hap3	CTTT	0.06	0.029	*

Ns=non-significant, *, **, ***, and **** indicate significant association between haplotypes and markers

In vivo marker prediction

In vivo validation for significant markers for tuber yield per plant and YMV revealed high prediction accuracy in both populations (identification and validation population) (Figure 8). We evaluated the segregation of the allele variants of the associated SNP markers to the tuber yield and the YMV across the testing and the validation population. For the tuber yield, marker chr06_965642, predicted high (CC) to low (TT and TC) yield performance in the breeding panel (Figure 8) as well as in the validation panel. One marker, Chr14_1128124 located on chromosome 14, homozygote CC predicted high yield per plant while CG and GG were linked to low yield performance per plant in both populations (Figure 8). On the SNP marker Chr15_4138634 associated with the tuber yield, homozygote AA was identified to predict high tuber yield per plant in the two populations (Figure 8). For the yam mosaic virus and considering the three markers with high haplotype segregation (Table 3), allele variants were identified to predict the same disease tolerance and susceptibility in the two populations (Figure 9).

Discussion

3.1. Phenotypic variation

The natural variation among the studied traits was high and very informative. High broad-sense heritability of 0.708 for tuber yield per plant and 0.903 for yam mosaic virus severity score demonstrated substantial genetic variation in traits between the different clones. Therefore, the studied traits are amenable to genetic improvement through selection [32]. Furthermore, the observed natural genetic variation in the study materials signifies their relevance for genetic studies.

3.2. Population differentiation

Understanding population structure within the study material is imperative to determine how it affects the ability of GWAS to infer marker-trait association. The four clustering methods, BIC, DAPC, Kinship relationship matrix and structure, showed four sub-populations that are imperative for preventing sham associations in GWAS in this study [33, 34]. Thus, the marker density, diversity, and sample size demonstrated that the yam breeding panel used for this study is sufficiently powered to capture allelic variations for the studied traits.

3.3. Genome-wide association studies

The whole-genome scan for phenotypic and allelic variation in tuber yield and yam mosaic virus resistance identified genome regions on nine chromosomes (chromosomes 3, 4, 6, 8, 10, 14, 15, 16, and 19) with significant −log10 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 tuber yield and tolerance to yam mosaic virus 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 [35]. Genome-wide association mapping has been used in exploring the elite alleles of many agronomic traits such as tuber dry matter and oxidative browning [34] in water yam (*Dioscorea alata*). In the present study, the phenotypic effect values of the favorable alleles of TYP and YMV were evaluated and inferred to positively and negatively affect the individual traits. Based on the stringent criterion of −log10, we identified 22 significant markers trait associations ranging between 4.71 e-14 and 0.043 for tuber yield per plant and between 5.25 e-14 and 0.029 for yam mosaic virus. The information on SNP variants from the present study would fast-track the application of genomics-informed selection decisions in breeding white Guinea yam for higher tuber yield and resistance to mosaic virus. Such great potential of GWAS has been reported for some root and tuber crops such as cassava [36], potatoes [37] and water yam [34].

3.4. Prediction of candidate genes

Our results identified SNP markers that associate significantly with allelic variation for tuber yield and YMV tolerance in white yam. The detected markers offer good targets for further validation and analysis due to their location in proximity to candidate genes regulating growth, development and disease resistance. The SNP in chromosome 3 is near to AP2/ERF domain, AUX/IAA protein, major facilitator, sugar transporter-like genes. Zarei et al. [38] reported that the AP2/ERF-domain transcription factor

ORA59 acts as the integrator of the jasmonic acid (JA) and ethylene (ET) signaling pathways and is the key regulator of JA- and ET-responsive PLANT DEFENSIN1.2 (PDF1.2) expression. The SNP in chromosome 4 is near to Geminivirus AL1 replication-associated protein, catalytic domain, AP2/ERF domain, NB-ARC, Dirigent protein, and membrane transport protein genes. The NB-ARC domain is noted to play a role in ATPase domain that comprises NB, ARC1, and ARC2 subdomains, which in its nucleotide-binding state regulates the R protein activity or resistance in plants [39]. The plant defense is induced by the R proteins in response to specific pathogen-derived molecules, called avirulence (AVR) proteins, thereby restricting pathogen proliferation [40]. The SNP in chromosome 10 is near to Geminivirus AL1 replication-associated protein, catalytic domain, Geminivirus Rep catalytic domain, Geminivirus AL3 coat protein, AP2/ERF domain, NB-ARC, Chlorophyll A-B binding protein, plant and chromista. Geminivirus AR1/BR1 coat protein, AP2/ERF domain, Geminivirus AL1 replication-associated protein, catalytic domain, Geminivirus AL1 replication-associated protein, central domain, and NB-ARC genes. Geminiviruses have been reported by Sunter and Bisaro [41] to play role in the Transactivation of Geminivirus AR1 and BR1 Gene Expression by the Viral AL2 Gene Product. Chlorophyll A-B binding protein is known as a light receptor that stimulates growth and development in plants [42]. The SNP in chromosome 16 is near to Geminivirus AR1/BR1 coat protein; AP2/ERF domain; Geminivirus AL1 replication-associated protein, catalytic domain; Geminivirus AL1 replication-associated protein, central domain; and NB-ARC genes. The SNP in chromosome 14 is near to expansin, cellulose-binding-like domain; mitochondrial substrate/solute carrier, expansin, root cap; dirigent protein; small auxin-up RNA; major facilitator, sugar transporter-like genes. Expansins or expansin-like proteins (loosenins) were reported to loosen plant cell wall activity and lignocellulose saccharification [43]. Mitochondrial carrier proteins play roles in plant growth and disease resistance [44]. The SNP in chromosome 15 is near to Gibberellin regulated protein; Major facilitator, sugar transporter-like; Senescence regulator S40; ABC transporter-like genes. The gibberellin regulated protein (GRP) has been noted to be up-regulated by gibberellin, and most of these proteins have a role in plant development and some of its members have antimicrobial activity [45, 46]. The SNP in chromosome 19 is near to Exportin-1/Importin-beta-like; Expansin; Sodium/calcium exchanger membrane region; Major facilitator, sugar transporter-like; AUX/IAA protein. The sodium/calcium exchanger has been reported to influence metabolic regulation on ion carrier interactions in living organisms [47]. The SNPs in chromosomes 6 and 8 are near to AUX/IAA protein and Protein ENHANCED DISEASE RESISTANCE 2, C-terminal (EDR2) genes. The Aux/IAA gene has been noted to play cellular and developmental roles in plants' lifespan, such as root development, shoot growth, and fruit ripening [48]. The Protein ENHANCED DISEASE RESISTANCE 2, C-terminal (EDR2) in plants limits cell death initiation and the establishment of hypersensitive response [49]. The identified putative candidate genes and SNPs linked with these important economic traits could help design new breeding strategies to hoard superior alleles for these key traits in future marker-based breeding. The novel regions identified in this study have not been previously detected, possibly due to the limitations of the various marker systems used in earlier studies.

Our findings indicated that multiple loci having unequal effects can influence the variation for TYP, and YMV in white yam. The identified novel candidate genomic regions with growth, development and

disease resistance genes in our study require further validation and testing in yam germplasm. This could be done by converting these MTAs into low cost Kompetitive Allele-Specific PCR (KASP) markers that can efficiently transfer alleles into elite yam genotypes as reported for wheat [50]. These valuable genomic resources and PCR based markers (KASP markers) could greatly support selection initiatives for key traits in yam breeding through marker-assisted selection (MAS). These will also support the systematic study of the genetics, comparative genomics and evolution of yam, aimed at expediting the isolation and characterization of genes that control agronomically important traits, such as tuber yield and yam mosaic virus.

Based on the marker effects of our study, we observed alleles CC and CT on chromosome 6, alleles TT and AT on chromosome 8 and allele AA on chromosomes 14, 15 and 19 to be responsible for high tuber yield per plant in the diversity panel used in the study. For the YMV, we found alleles AG and AA on chromosome 4 and alleles GA and GG on chromosome 10 to be responsible for low YMV disease infection. These findings suggest that data mining of favorable alleles is essential for improving the quantitative trait for tuber yield and YMV in yam using marker-assisted selection. Moreover, the results could be helpful for marker validation and deployment in yam breeding. Our findings agree with the view that information on marker effect based on segregation pattern is fundamental for marker validation and deployment in a breeding program [26, 34]. Association mapping has been utilized to explore elite alleles present in many agronomic traits, including yield and related attributes in bread wheat [51].

Conclusion

Useful genetic variability exists in the 406 genotypes studied. The associated SNP markers could be potentially employed for targeted and accelerated tuber yield per plant and YMV resistance in white yam. The information from our study could help design new breeding strategies to hoard superior alleles for tuber yield per plant and yam mosaic virus in future marker-based breeding. The chromosomal regions controlling these studied traits could be exploited for selection and effective pyramiding of favorable alleles in white yam population improvement.

Declarations

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

The Variant Call Format (VCF) file used in this study for the analysis can be viewed on www.yambase.org under genotypic data. Phenotypic data can be obtained as well upon request from the corresponding author.

COMPETING INTEREST

The authors declare that the research was conducted in the absence of any potential conflict of interest.

FUNDING

The work was financially supported by Bill and Melinda Gates Foundation (BMGF) through the AfricaYam project (OPP1052998). The sequencing activities was fully sponsored by BMGF as well as the charges related to field evaluation. Publication fee will be paid as well the BMGF.

AUTHOR CONTRIBUTIONS

AA conceptualized the study, PAA did the data curation and analysis. PEN managed the phenotypic data. AA acquired the funding for the research. PEN and PAA wrote the draft manuscript with input from AA. PEN, PAA, RA and AA reviewed and edited the manuscript. All authors have read, made corrections, and approved the final manuscript.

ACKNOWLEDGMENTS

We are grateful to the Yam Improvement Program team for their assistance during the research implementation we also acknowledge Afolabi Agbona, Kayondo Siraj Ismael for their critical suggestion during the analysis.

References

1. FAO Food and Agriculture Organization of the United Nations Statistics database, FAOSTAT. 2020. http://www.fao.org/faostat/en/#data/ QC
2. Asiedu R, Sartie A. Crops that feed the world 1. Yams: Yams for income and food security. Food Security 2010; 2: 305–315, https://doi.org/10.1007/s12571-010-0085-0.
3. Cormier F, Lawac F, Maledon E, Gravillon MC, Nudol E, Mournet P, et al. A reference high-density genetic map of greater yam (*Dioscorea alata* L.). Theor Appl Genet. 2019; 132: 1733–1744. https://doi.org/10.1007/s00122-019-03311-6.
4. Darkwa K, Olasanmi B, Asiedu R, Asfaw A. Review of empirical and emerging methods and tools for yam (*Dioscorea* spp.) improvement: status and prospects. Plant Breed. 2020a; 139(3): 474–497. doi.org/doi.10.1111/PBR.12783
5. Onda Y, Mochida K. Exploring genetic diversity in plants using high-throughput sequencing techniques. Curr Genomic 2016; 17: 358–367.

6. Lebot V. Tropical root and tuber crops: cassava, sweet potato, yams and aroids. CABI, Wallingford, XIX 2009; 413 p.
7. Obidiegwu JE, Akpabio EM. The geography of yam cultivation in southern Nigeria: Exploring its social meanings and cultural functions. J Ethnic Foods 2017; 4: 28–35.
8. Darkwa K, Agre P, Olasanmi B, Iseki K, Matsumoto R, Powell A, et al. Comparative assessment of genetic diversity matrices and clustering methods in white Guinea yam (Dioscorea rotundata) based on morphological and molecular markers. Sci Reports 2020b; 10: 13191 https://doi.org/10.1038/s41598-020-69925-9
9. Mignouna H, Mank R, Ellis T, Van Den Bosch N, Asiedu R, Ng S, Peleman J. A genetic linkage map of Guinea yam (Dioscorea rotundata Poir.) based on AFLP markers. Theoretical and Applied Genetics 2002; 105(5): 716–725. https://doi.org/10.1007/ s00122-002-0911-7
10. Norman PE, Asfaw A, Tongoona PB, Danquah A, Danquah EY, Koeyer DD, et al. Can parentage analysis facilitate breeding activities in root and tuber crops? Agric J. 2018; 8, 1–24.
11. Jiang GL. Molecular Markers and Marker-Assisted Breeding in Plants. In: Plant Breeding from Laboratories to Fields Sven Bode Andersen (ed). IntechOpen 2013; pp 45–83. http://dx.doi.org/10.5772/52583
12. Asfaw A. ed. (2016). Standard operating protocol for yam variety performance evaluation trial. IITA, Ibadan, Nigeria, 27 p.
13. Forbes, G., Pérez, W., Andrade-Piedra, J.L., 2014. Field assessment of resistance in potato to *Phytophthora infestans*: International cooperators guide. Lima (Peru). International Potato Center (CIP). ISSBN 978-92-9060-440-2$4 35p. https://doi.org/10.4160/9789290604402
14. Gilmour AR, Thompson R, Cullis BR. Average information REML: An efficient algorithm for variance parameter estimation. Biometrics 1995; 51: 1440–1450. https://doi.org/10.2307/ 2533274
15. Butler DG, Cullis BR, Gilmour AA, Gogel BJ, Thome R. *ASReml-R Reference manual version 4*. VSNi Ltd, Hemel Hempstead, HP1IES, UK. 2018.
16. Borgognone MG, Butler DG, Ogbonnaya FC, Dreccer MF. Molecular marker information in the analysis of multi-environment trials helps differentiate superior genotypes from promising parents. Crop Sci. 2016; 56: 2612–2628.
17. Ovenden B, Milgate A, Wade lJ, Rebetzke GJ, Holland JB. Accounting for genotype-by-environment interactions and residual genetic variation in genomic selection for water soluble carbohydrate concentration in wheat. G3 Genes Genome Genet. 2018; 8: 1909–1919.
18. Dellaporta SL, Wood J, Hicks JB. A plant DNA minipreparation: version II. Plant Mol Biol Rep. 1983; 1: 19–21.
19. Aljanabi SM, Martinez I. Universal and rapid salt-extraction of high-quality genomic DNA for PCR-based techniques. Nucleic Acids Res. 1997; 25: 4692–4693. https://doi.org/10.1093/nar/25.22.4692
20. Kilian A, Sanewski G, Ko L. The application of DArTseq technology to pineapple. Acta Hortic. 2016; 1111: 181–188.

21. Kim D, Langmead B, Salzberg SL. HISAT: A fast spliced aligner with low memory requirements. Nat Methods 2015; 12: 357.
22. Yin L. *Package "CMplot"*. 2019. URL https://github.com/YinLiLin/R-CMplot/blob/master/CMplot.r
23. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum. Genet. 2007; 81: 559–575. https://doi.org/10.1086/519795.
24. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics 2010; 11: 1–15. https://doi.org/10.1186/1471-2156-11-94.
25. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, et al. GAPIT: genome association and prediction integrated tool. Bioinformatics 2012; 28: 2397–2399. doi:10.1093/bioinformatics/bts444
26. Li L, Tacke E, Hofferbert HR, Lübeck J, Strahwald J, Draffehn AM, Gebhardt C. Validation of candidate gene markers for marker-assisted selection of potato cultivars with improved tuber quality. Theor Appl Genet. 2013; 126: 1039–1052.
27. Shin JH, Blay S, McNeney B, Graham J. LDheatmap: An R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. J. Stat. Softw. 2006; 16: 1–10.
28. Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, Yong SY. InterPro new developments in the family and domain prediction database. Nucleic Acids Res. 2011; 40: 306–312.
29. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 2007; 23: 2633–2635. https://doi.org/10.1093/ bioinformatics/btm308
30. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005; 21: 263–265. https://doi.org/10.1093/bioinformatics/bth457
31. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York 2016 ISBN 978-3-319-24277-4 https://ggplot2.tidyverse.org.
32. Piaskowski J, Hardner C, Cai L,Zhao Y, Iezzoni A, Peace C. Genomic heritability estimates in sweet cherry reveal non-additive genetic variance is relevant for industry-prioritized traits. BMC Genetics 2018; 19(1): 23. https://doi.org/10.1186/s12863-018-0609-8.
33. Yu D, Lane SN. Urban fluvial flood modelling using a two-dimensional diffusion-wave treatment, part 2: Development of a sub-grid-scale treatment. Hydrol Proccess. 2006; 20: 1567–1583.
34. Gatarira C, Agre P, Matsumoto R, Edemodu A, Adetimirin V, Bhattacharjee R, Asiedu R, Asfaw A. Genome-Wide Association Analysis for Tuber Dry Matter and Oxidative Browning in Water Yam (*Dioscorea alata* L.). Plants 2020; 9: 969; https://doi.org/10.3390/plants9080969
35. Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. Advantages and pitfalls in the application of mixed model association methods. Nat Genet. 2014; 46: 100–106.
36. Zhang S, Chen X, Lu C, Ye J, Zou M, Lu K, Ma PA. Genome-wide association studies of 11 agronomic traits in cassava (Manihot esculenta Crantz). Front Plant Sci. 2018; 9: 503.

37. Björn B, Keizer PL, Paulo MJ, Visser RG, Van Eeuwijk FA, Van Eck HJ. Identification of agronomically important QTL in tetraploid potato cultivars using a marker–trait association analysis. Theor Appl Genet. 2014; 127: 731–748.
38. Zarei A, Körbes PA, Younessi P, Montiel G, Champion A, Memelink J. Two GCC boxes and AP2/ERF-domain transcription factor ORA59 in jasmonate/ethylene-mediated activation of the *PDF1.2* promoter in Arabidopsis. Plant Mol Biol. 2011; 75: 321–331. https://doi.org/10.1007/s11103-010-9728-y.
39. van Ooijen G, Mayr G, Kasiem MMA, Albrecht M, Cornelissen BJC, Takken FLW. Structure–function analysis of the NB-ARC domain of plant disease resistance proteins. J Exp Bot. 2008; 59(6): 1383–1397, doi:10.1093/jxb/ern045
40. DeYoung and Innes, 2006
41. Sunter G, Bisaro DM. Transactivation of Geminivirus AR1 and BR1 gene Expression by the Viral AL2 Gene Product occurs at the Level of Transcription. The Plant Cell 1992; 4(10): 1321–1331.
42. Yan-Hong X, Rui L, Lu Y, Zhi-Qiang L, Shang-Chuan J, Yuan-Yue S, Xiao-Fang W, Da-Peng Z. Light-harvesting chlorophyll *a*/*b*-binding proteins are required for stomatal response to abscisic acid in *Arabidopsis*, *J Exp Bot*. 2012; 63(3): 1095–1106, https://doi.org/10.1093/jxb/err315
43. Ríos-Fránquez FJ, Rojas-Rejón ÓA, Escamilla-Alvarado C. Microbial Enzyme Applications in Bioethanol Producing Biorefineries: Overview. In: Bioethanol Production from Food Crops Sustainable Sources, Interventions, and Challenges, Academic Press, Ray, R.C. Ramachandran, S. (eds.). 2019; 249–266 p.
44. Palmieri F. Mitochondrial carrier proteins. FEBS Lett. 1994; 346: 48–54. https://doi.org/10.1016/0014-5793(94)00329-7
45. Berrocal-Lobo M, Segura A, Moreno M, Lopez G, Garcia-Olmedo F, Molina A. Snakin-2, an antimicrobial peptide from potato whose gene is locally induced by wounding and responds to pathogen infection. Plant Physiol 2002; 128: 951–961.
46. Inomata N. Gibberellin-regulated protein allergy: Clinical features and cross-reactivity. Allergol Int 2020; 69: 11–18.
47. DiPolo R, Beaugé L. Sodium/calcium exchanger: influence of metabolic regulation on ion carrier interactions. Physiol Rev. 2006; 86: 155–203. doi:10.1152/physrev.00018.2005.
48. Luo J, Zhou JJ, Zhang JZ. Aux/IAA Gene Family in Plants: Molecular Structure, Regulation, and Function. Int J Mol Sci. 2018; 19: 259. Published 2018 Jan 16. doi:10.3390/ijms19010259
49. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. 2019; 47(W1), W636–W641. https://doi.org/10.1093/nar/gkz268
50. Rasheed A, Wen W, Gao F, Zhai S. Jin H, Liu J, et al. (2016). Development and validation of KASP assays for genes underpinning key economic traits in bread wheat. Theor. Appl. Genet. 129, 1843–1860. https://doi.org/10.1007/s00122-016-2743-x

51. Sun C, Zhang F, Yan X, Zhang X, Dong Z, Cui D, et al. Genome-wide association study for 13 agronomic traits reveals distribution of superior alleles in bread wheat from the Yellow and Huai Valley of China. Plant Biotechnol J. 2017; 15: 953–969. https://doi.org/10.1111/pbi.12690

Figures

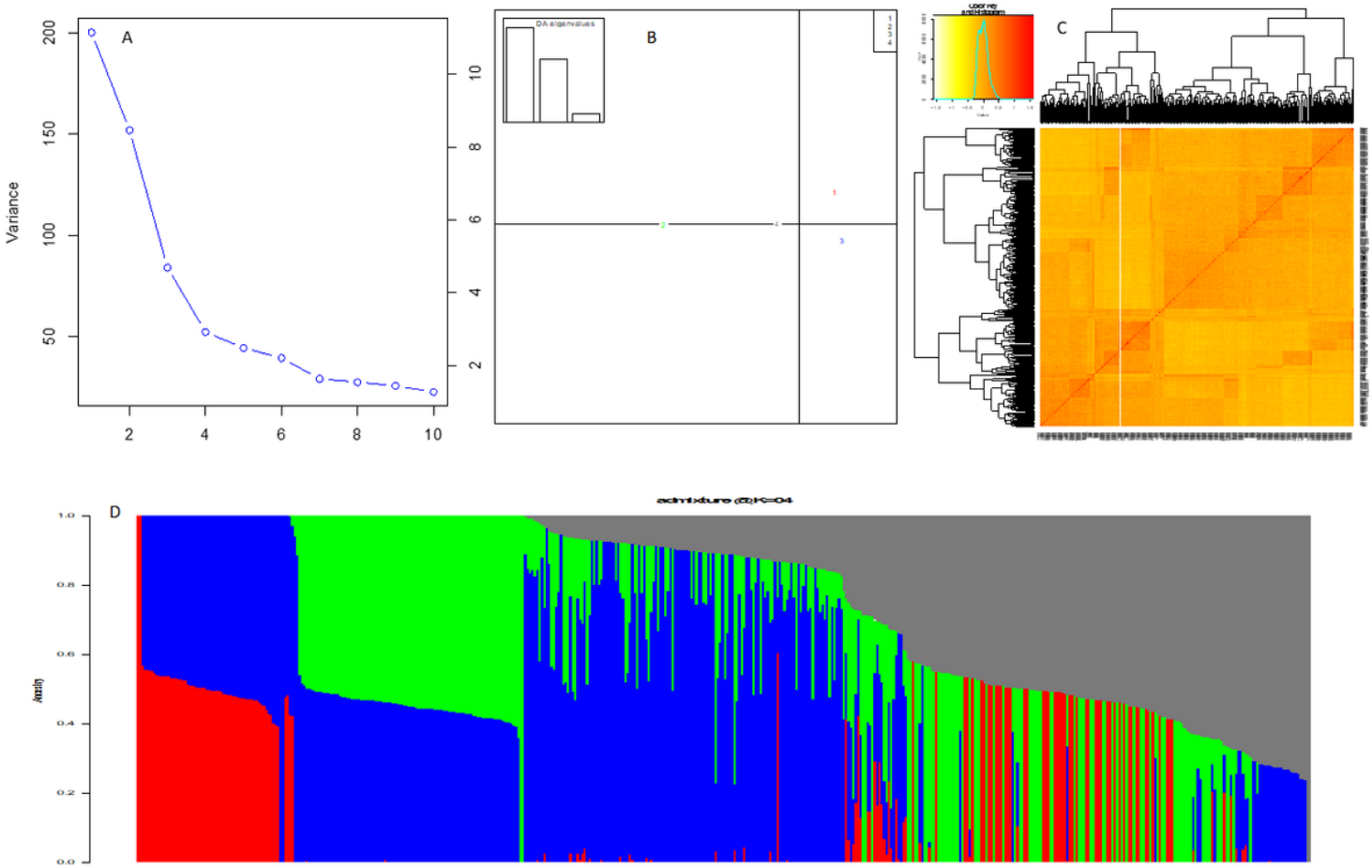

Figure 1

Population structure and genetic diversity assessment. (A) Bayesian information criteria showing a rapid elbow defined 4 as number of clusters; (B) Population structure based on the discriminant analysis of the principal component; (C) hierarchical grouping based on the kinship relationship matrix and (D) Population structure based on STRUCTURE at K = 4.

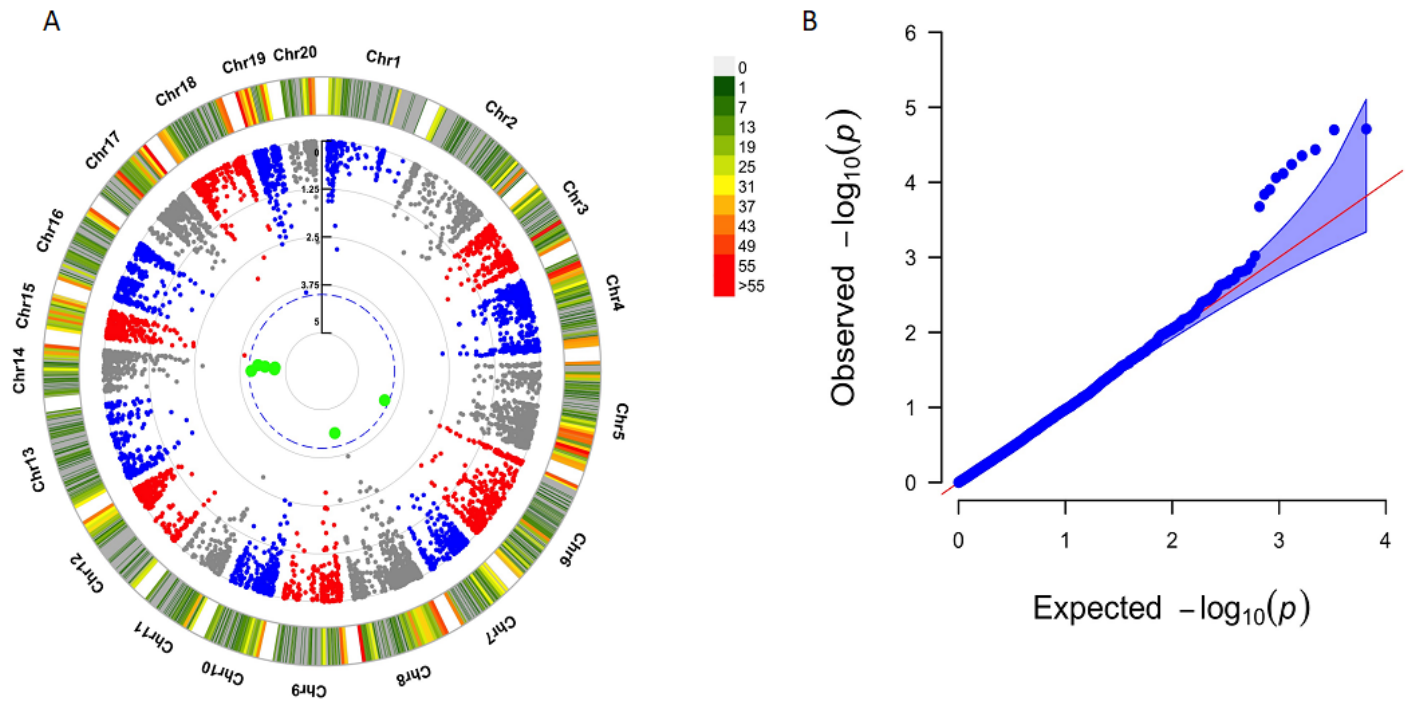

Figure 2

Genome-wide association analysis of tuber yield per plant. (A) Manhattan plot indicating three SNP markers located on chromosome 6, 8, 14, 15 and 19 associated with the tuber yield; and (B) Quantile-quantile (QQ) plot of the observed p values of the association analysis that is expected in a null association for the phenotype.

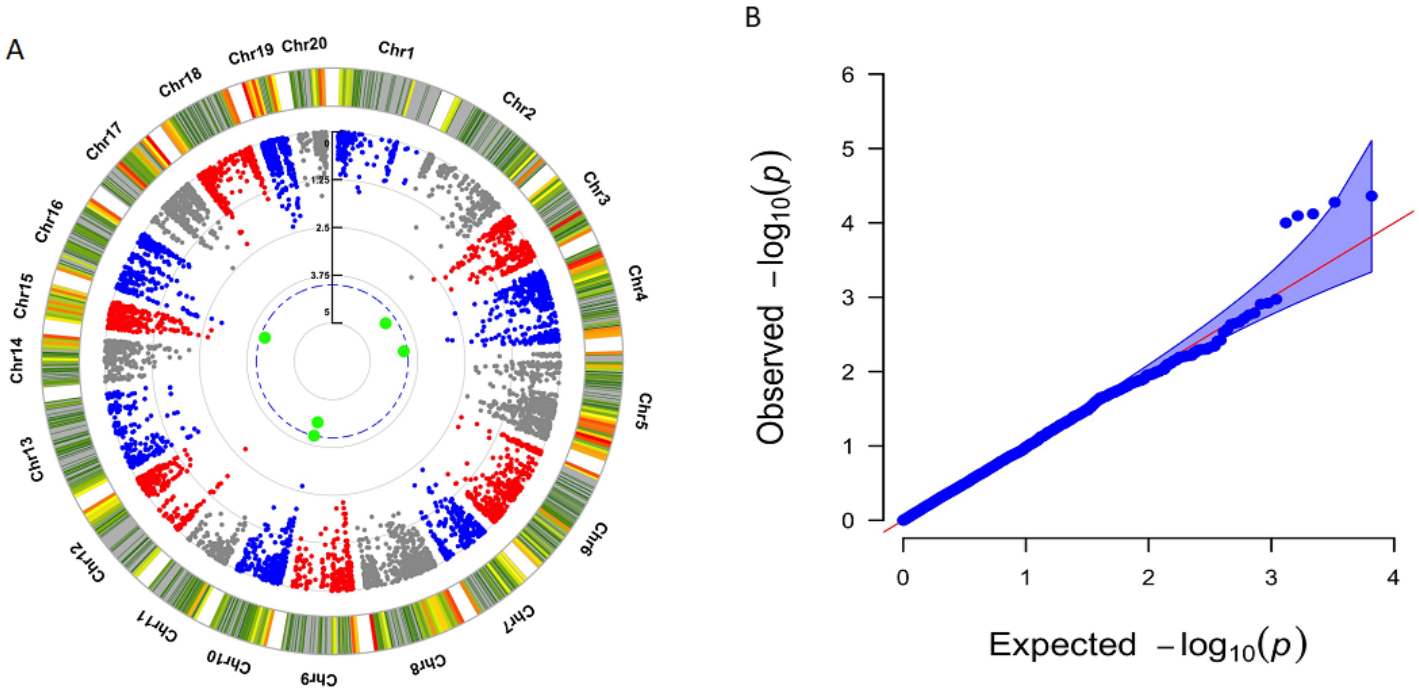

Figure 3

Genome-wide association analysis of yam mosaic virus. (A) Manhattan plot indicating SNPs associated with the YMV. The y-axis represents the p-value of the marker-trait association on a − log10 scale. (B) quantile-quantile (QQ) plot of the observed p values of the association analysis that is expected in a null association for the phenotype

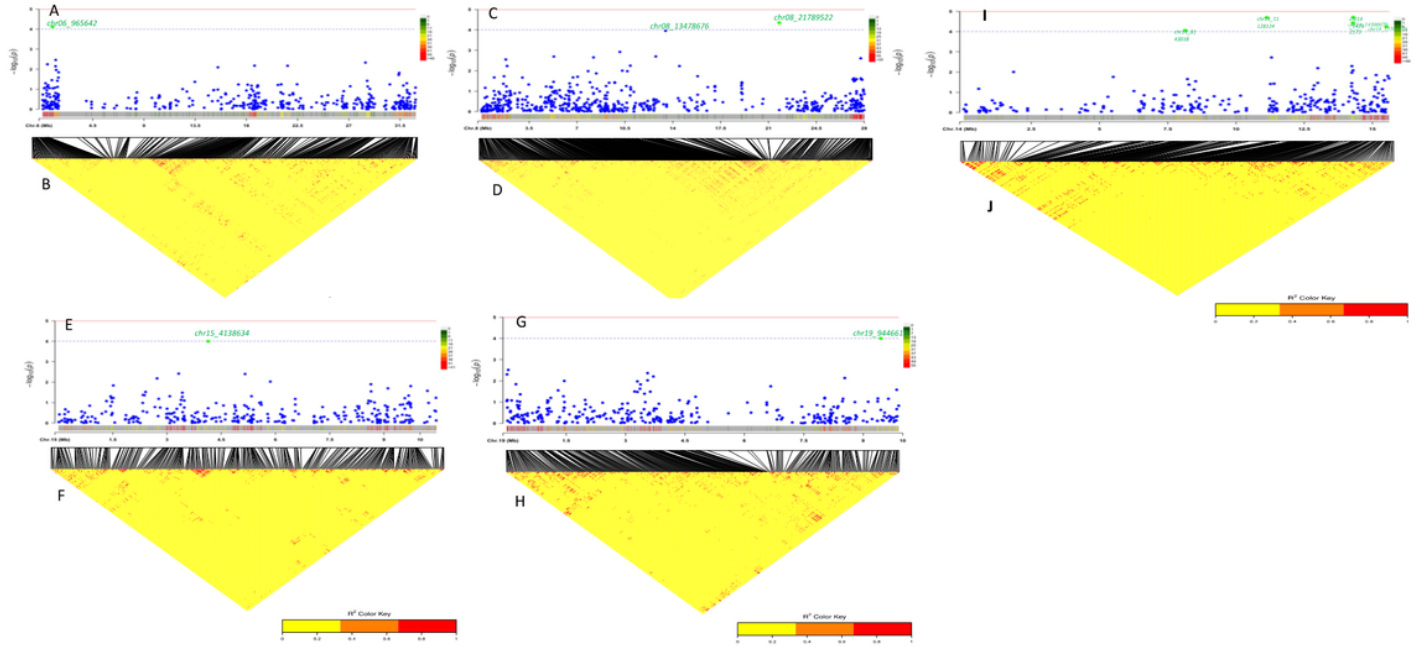

Figure 4

Heatmap LD haplotype blocks for different SNP markers located on different chromosomes (A & B) chromosome 6; (C & D) chromosome 8; (I & J) chromosome 14; (E and F) chromosome 15 and (G & H) chromosome 19. The R2 color key indicates the degree of significant association with the putative genes.

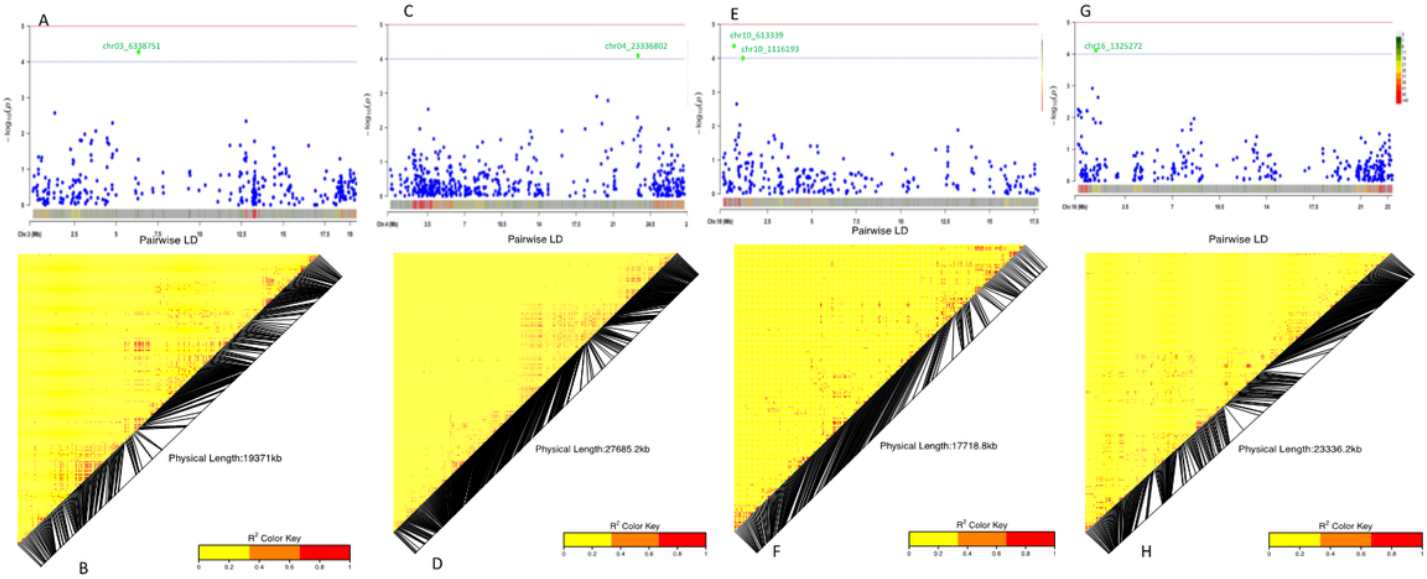

Figure 5

Summary of the local LD and haplotype blocks for different SNP marker located on different chromosome (A) chr3 and (B): chr10. The R2 color key indicates the degree of significant association.

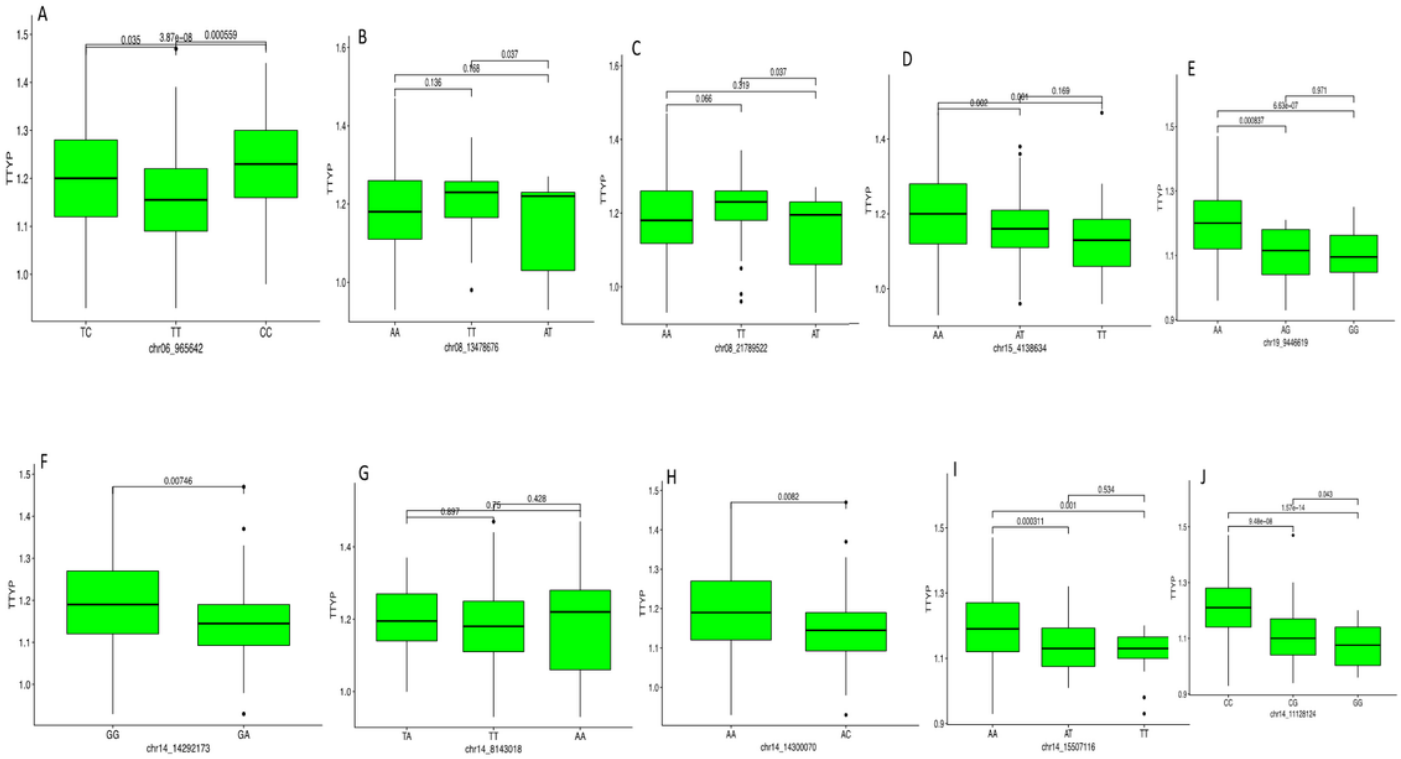

Figure 6

Boxplots showing the effect of the significant markers associated with tuber yield per plant on: (A) chromosome 6 with one SNP chr06_965642; (B & C) chromosome 8 with two SNPs chr08_13478676 and chr08_21789522, respectively; (D) chromosome 15 with one SNP chr15_4138634 (F, G, H, I & J) chromosome 14 with five SNPs chr14_8143018, chr14_11128124, chr14_14292173, chr14_14300070 and chr14_15507116; (E) chromosome 19 with one SNP chr19_9446619. The letters on the X axis represent allele variants; significant codes: * = 0.05 and p represents the analysis of variance probability value associated with the variation across variants.

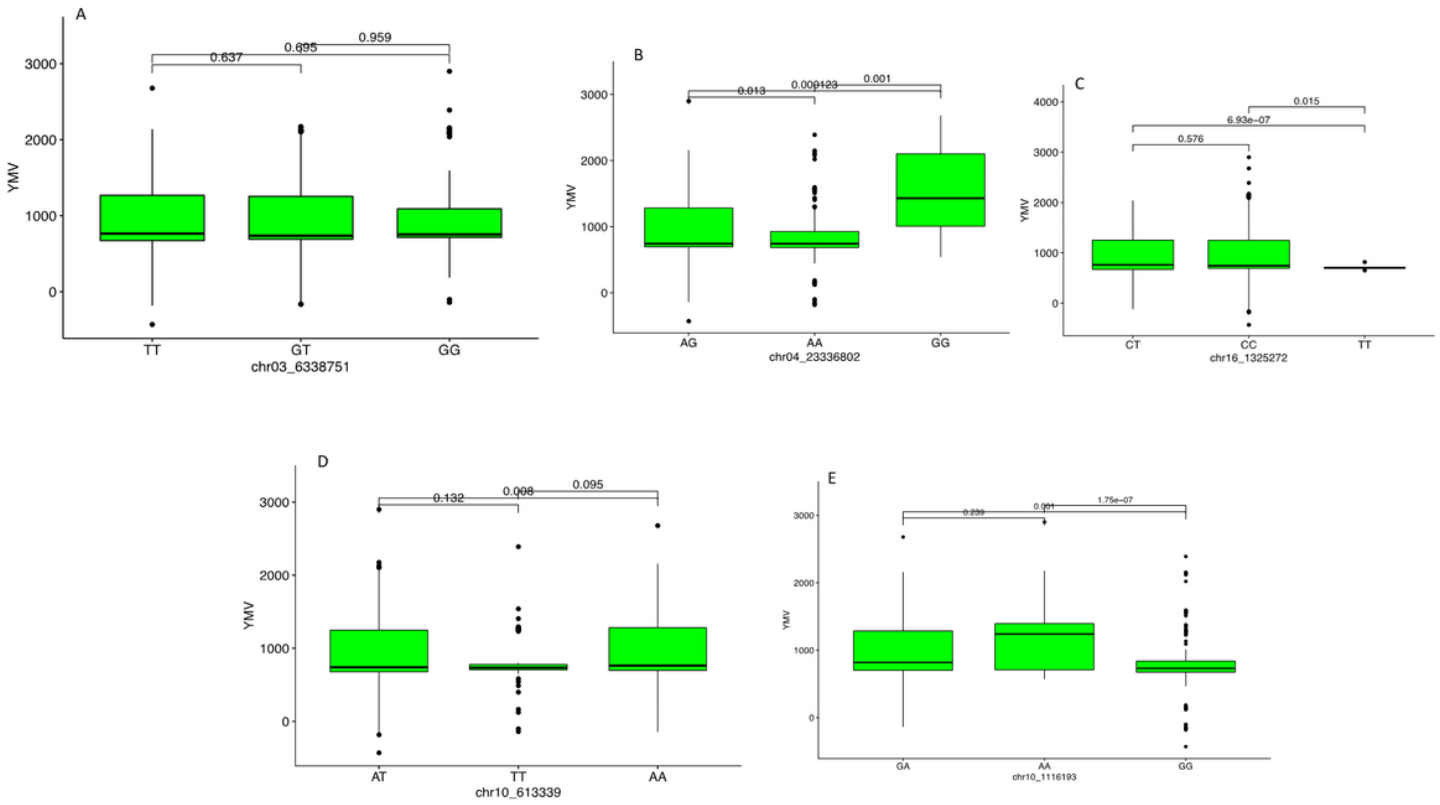

Figure 7

Boxplots showing the effect of the significant markers associated with yam mosaic virus on: (a) chromosome 3 with one SNP chr03_6338751; (b) chromosome 4 with one SNP chr04_23336802; (D & E) chromosome 10 with two SNPs chr10_613339 and chr10_1116193; and (C) chromosome 16 with one SNP chr16_1325272. The letters on the X axis represent allele variants for the different SNP markers; significant codes: * = 0.05 and p represents the analysis of variance probability value associated with the variation across variants

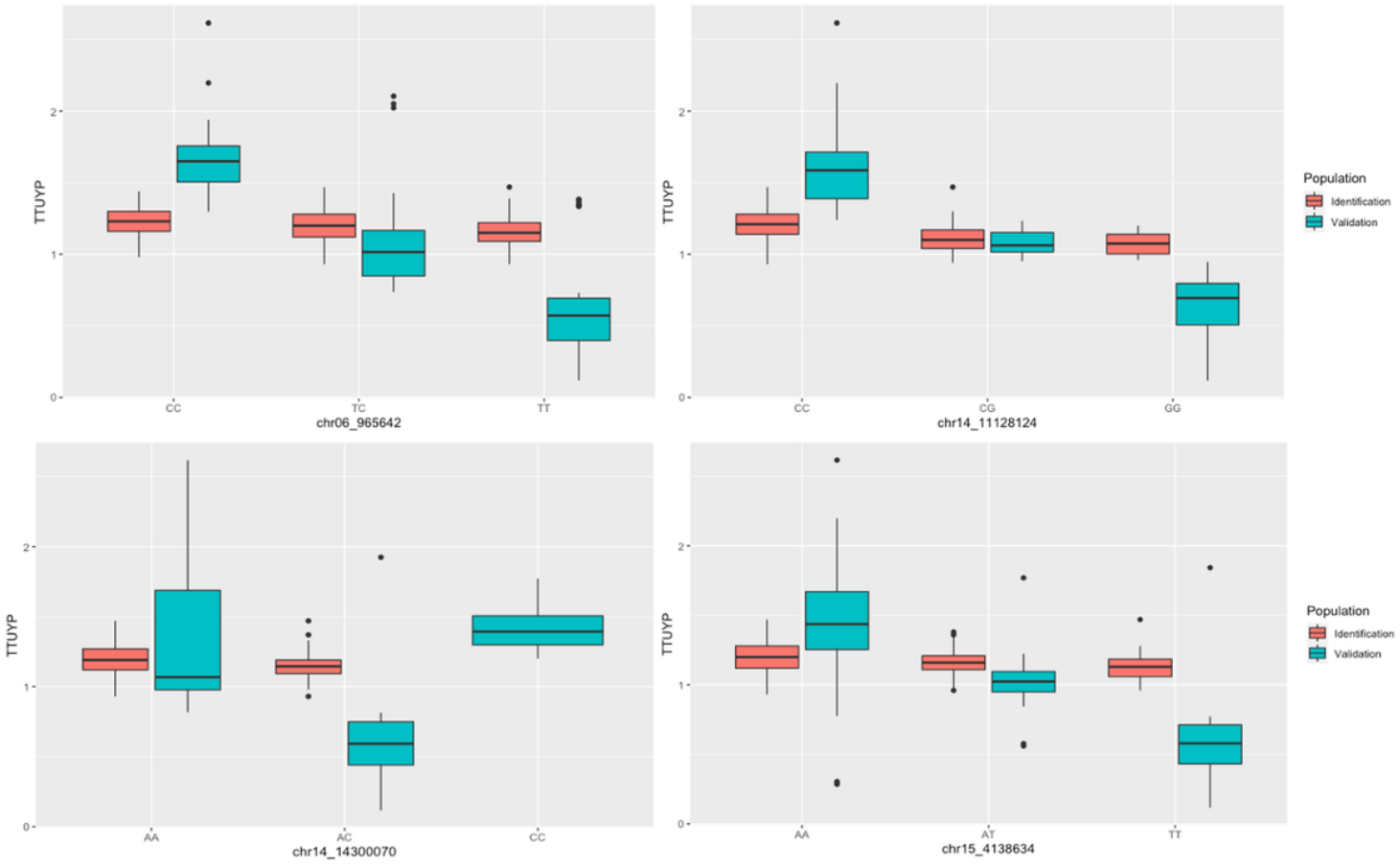

Figure 8

In Vivo marker validation for tuber yield per plant using indentication and panel breeding population

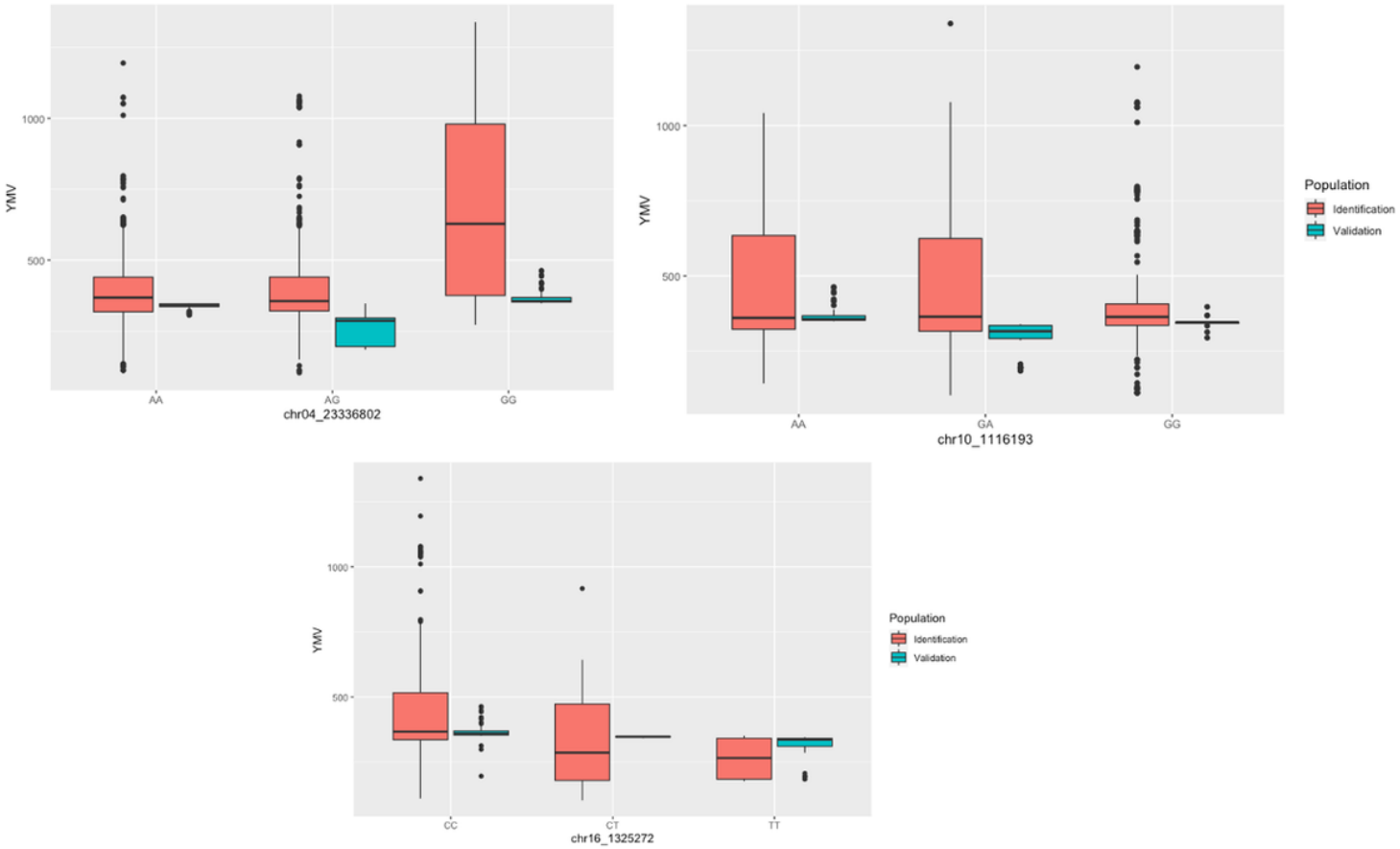

Figure 9

In Vivo marker validation for yam mosaic virus (YMV) using indentication and panel breeding population

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SUPPLEMENTARYTABLESGWASApr21.docx