Sequence aware genotyping improves the accuracy of molecular markers associated with cassava mosaic disease resistance loci

Keelan Krinsky
University of the Witwatersrand
Bulelani Sizani
Agricultural Research Council of South Africa
Marie Emma Rey
chrissie.rey@wits.ac.za

University of the Witwatersrand

Article

Keywords: Cassava mosaic disease, Marker-assisted breeding, SACMV, Cassava, SSR markers, Size homoplasy

Posted Date: March 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8948966/v1

Additional Declarations: No competing interests reported.

Abstract

Cassava (*Manihot esculenta*) is a critical food crop for sub-Saharan Africa. Cassava mosaic disease (CMD) caused by cassava geminiviruses, can severely reduce crop yield. Improved host resistance presents the most viable strategy for CMD management, but long breeding cycles constitute a major stumbling block. To ameliorate this challenge, contemporary CMD resistance breeding is highly reliant on marker-assisted selection for the rapid identification of CMD resistance loci in new germplasm. High rates of false discovery (> 30%) limit the efficiency of CMD resistance loci associated molecular markers. Marker-locus recombination and genetic background epistasis are known to exacerbate false discovery, but the contribution of size homoplasy remains understudied. Further, markers loci have been identified in intronic regions of quantitative trail locus (QTL) genes, but the direct impact of marker sequence polymorphism on CMD response phenotype remains undetermined. In this study, sequencing of marker amplicons derived from CMD resistant, tolerant, and susceptible germplasm, revealed the frequent occurrence of size homoplasy associated ambiguities in marker screening. Single nucleotide polymorphisms (SNPs) in marker amplicons were identified which correlated with CMD response phenotype. Elimination of susceptibility associated marker amplicons based on these SNPs reduced false discovery rates by between 15–22%, offering a substantial improvement in CMD resistance breeding efficiency.

Introduction

Food shortages, and food security, present a serious global challenge with approximately 2.33 billion people still food insecure and over 750 million chronically undernourished [1]. The development of resilient, high-yield varieties of staple crops such as cassava is vital to tackling this challenge. Cassava is cultivated extensively throughout Africa, South America, and Asia, serving as a staple crop for over a billion people [2]. Cassava cultivation provides both a direct food source, with tuberous roots high in carbohydrates, calcium, phosphorus, and vitamin C, and leaves high in protein [3, 4], and a means of poverty alleviation for small scale and subsistence farmers [1, 3]. Cassava is also cultivated commercial for the production of starch and biofuel from cassava tubers [4, 5]. Over the last all 30 years, cassava has shown the greatest relative increase in land usage and crop yield [2, 4]. This increase may be attributed to the acceptable yield obtained from cassava cultivated, without inorganic fertilisers, under adverse abiotic conditions including variable precipitation, periods of protracted drought, variable soil pH, and nutrient-poor soil [4].

Despite the crops general resilience, cassava cultivation is still severely challenged by a multitude of biotic and abiotic stresses, exacerbated by the increasing abundance and continued geographic spread of diseases vectors such as aphids and whitefly [6, 7]. Breeding for improved disease resistance provides an effective and efficient means to alleviate the burden of biotic stress in cassava cultivation [4, 7–9, 11, 12]. Due to low seed set, and long breeding cycles, the development of new cassava germplasm requires at least eight years [11]. Markers assisted selection (MAS), can improve breeding efficiency by genomic tagging of quantitative trait loci (QTLs) [8, 14]. MAS-QTL breeding has allowed rapid introgression and pyramiding of critical agronomic and stress tolerance traits, for example, cold tolerance, draught tolerance [16, 17], cynogenic potential [15], and nutritional value [16].

On the African continent, the introgression of cassava mosaic disease (CMD) resistance into farmer preferred/commercially important cassava germplasm is a longstanding and high priority breeding objective [8–10]. CMD is caused by 11 species of cassava begomoviruses (CBV), including *South African cassava mosaic virus* (SACMV) [5], which impose harsh tuber and leaf crop yield losses in cassava cultivation. Contemporary cassava breeding, most notably the extensive programs co-ordinated by the International Institute of Tropical Agriculture (IITA), rely on three major CMD resistance/tolerance associated QTLs, namely CMD1, CMD2 and CMD3 [8, 9, 11, 12]. The CMD1 associated simple sequence repeat (SSR) marker SSRY40 [13], the CMD2 marker SSRY28 [14], and the CMD3 marker NS198 [15], explained 48, 68, and 11% of phenotypic variance, respectively, in initial mapping populations. Considerable reductions in yield loss were achieved through introgression of CMD2 into farmer preferred germplasm, and additional CMD2 markers, NS158 [16], NS169, and RME1 [17, 18] were developed to this end. In breeding populations derived from elite IITA clones, these six markers, typically display a combined PVE of ~ 70–80, with a false discovery rate (FDR) of ~ 20%. FDR varies widely between breeding families (4%-50%)[11, 16], obstructing MAS for CMD resistance in phylogenetically diverse germplasm collections. As all CMD susceptible progeny are typically discarded, FDR constitutes a significant challenge to breeding efficiency.

Most attention given to FDR reduction has focussed on the identification of additional QTLs, additional markers, and markers with greater genomic proximity to CMD1, 2 and 3. Notably CMD SNP markers were identified from high density SNP maps generated from Genotyping by Sequencing (GBS) [19, 20]. However, in practise SNP markers have largely failed to replace the pre-existing SSR and SCAR markers, [21–24] due to the higher costs, more complex analysis, and lower reproducibility associated with SNP sets relative to SSR. Elimination of size homoplasy induced screening errors provides a potential method for FDR reduction without marker replacement. While instances of size homoplasy between susceptibility and resistance associated marker alleles/amplicons have previously been noted for CMD resistance markers [11, 25], the contribution of size homoplasy to marker FDR remains understudied.

MAS studies typically do not assess homoplasy directly; FDR is treated as synonymous to the rate of recombination between markers and their respective QTLs [25, 26]. SSR/SCAR marker primers used for resistance breeding frequently generate multiple amplicons; the target amplicon which correspond to the resistance associated marker allele and isolocus is distinguished from the susceptibility associated marker allele and from any off-target amplicons (derived from replicate isoloci or off-target primer binding sites) on the basis of amplicon length [11, 27]. The highly repetitive nature of SSR regions favours frequent polymerase slippage and recombination events [28], thus SSRs can track minor phylogenic divergence between closely related individuals/lineages. However, SSRs are also prone to homoplasy, as any polymorphism observed could be derived from the cumulative effect of very few, or very many mutation events resulting in distant relatives with identical SSR sequence [29]. At the intra-specific level the probability of complete SSR homoplasy is considered minimal, but size homoplasy has been observed for cassava SSRs [30–32]. This study sought to investigate the prevalence of size homoplasy and the impact of size homoplasy induced error in CMD marker screening.

Results

1.Frequency of occurrence of molecular marker associated with CMD response QTLs

In this study, we assessed the general prevalence of size homoplasy and determined the FDR due to size homoplasy using six CMD resistance associated molecular markers, namely, SSRY40 associated with the CMD1 resistance QTL, SSRY28, NS158, NS169, and RME1, associated with the CMD2 QTL, and NS198 associated with the CMD3 QTL. Twenty-two cassava germplasm including seven CMD-resistant germplasm; seven CMD-tolerant germplasm, and eight CMD-susceptible germplasm [33] (**Supplementary Table S1**), where screened with these six markers (**Supplementary Fig. S1, Supplementary Table S6**). These germplasm were also subjected to viral challenge with SACMV, and symptoms severity, was determined at 32 days post infection, (dpi) when symptoms when both tolerant and susceptible germplasm may show severe symptoms, and at 67 dpi when recovery/reversion is expected for tolerant germplasm Fig. 1. Marker presence was then correlated directly against response phenotype via a $\chi 2$ independence test and, symptom severity via a kruskal wallis test, and false discovery rates were calculated for each marker Table 1. For each of the markers assessed, amplicons in the range expected for the resistance linked marker allele were also observed in susceptible germplasm (**Supplementary Fig. S1, Supplementary Table S6**). Further several germplasm (both susceptible and tolerant/resistant) produced multiple amplicons in the expected range leading to considerable ambiguity in marker scoring (**Supplementary Fig. S2, Supplementary Table S7**). A substantial FDR was also observed across all markers, ranging from 18% for RME1 to 69% for SSRY40, and with the exception of SSRY28 marker presence was not significantly associated with response phenotype or leaf symptom severity. The low frequency of SSRY40 marker in germplasm derived from the tropical manihot series (TMS) (**Supplementary Table S6**). The poor association observed between the CMD1, 2 and 3 markers assessed and CMD response, may have resulted from false positives due to marker-locus recombination, false negatives due to alternative resistance QTLs and epistatic effects, false negatives due to null alleles, and/or false positives due to size homoplasy between resistance associated and susceptibility associated marker amplicons [27, 34]. To quantify the frequency and establish the potential impact of null alleles and marker size homoplasy, (between isoloci) selected marker bands were sequenced and used to query the 24 collected cassava whole genome assemblies (**Supplementary Table S1**) available from the genomic database hosted by the National Center for Biotechnology Information (NCBI), taxon ID 3982. (Section 2).

Table 1 Statistical evaluation of the associations between CMD response associated molecular markers and CMD response.

QTL	Marker	False discovery Rate	Rank correlation of marker presence against CMD symptom severity		Test of independence between marker presence and CMD response phenotype	
			p	τ	p	χ^2
CMD1	SSRY40	0.69	0.859	0.033	0.337	2.176
CMD2	SSRY28	0.33	0.023	0.428	0.337	1.480
CMD2	NS158	0.29	0.308	0.192	0.085	3.123
CMD2	NS169	0.30	0.308	0.192	0.085	3.123
CMD2	RME1	0.18	0.054	0.536	0.210	5.731
CMD3	NS198	0.26	0.365	0.170	0.115	6.940

Section 2: Genomic mapping of molecular markers linked to major CMD resistance loci

Genomic mapping of QTL linked molecular markers can assist in the identification of QTL genes, and additional QTL markers. Previous efforts towards mapping CMD1-3 associated SSR/SCAR markers produced conflicting results. Rabbi *et al.*, (2014) mapped NS158, NS169, SSRY28, and NS198 to three scaffolds of the reference assembly AM560-2 v4.1 (S05214, S06906, and S04175) all located on the same linkage group based on GBS-SNP markers [35]. Conversely, Wolfe *et al.*, (2016) mapped NS158, NS169 and SSRY28 to chromosome 12, and NS198 to chromosome 8, of the AM560-2 v5.1 assembly, while Kuria *et al.*, (2016) mapped RME1 and NS158 to chromosome 16 of the AM560-2 v6.1 assembly [19, 36]. A number of high quality genome assemblies, including an improved reference assembly (AM560-2 v8.1) are now available [37–40]. We used a combination of *in silico* PCR, and BLASTn queries with the marker amplicons sequenced (SRA: PRJNA1420052) to identify SSR/SCAR marker loci in the 24 *Manihot* genomic assemblies available from the NCBI (taxa ID 3982); Two alternative assemblies were available for the TME3 and TMS60444 germplasm. As the assemblies constructed by Cornet et al., (2025) improved the confidence and contiguity of the CMD2 region, but achieved a lower overall genomic coverage than the assemblies constructed by Kuon *et al.*, (2019), both variants were included in the analysis [37, 38]. A subset of the mapped loci are presented in Fig. 2, (for unabridged results see **Supplementary Table S5**). As the NS158, and NS169 markers were derived from overlapping regions in all cases, only NS169 is shown.

In all of the chromosome level assemblies (including AM560-2 v8.1) the SSRY28, NS169, and NS158 CMD2 markers clustered together on chromosome 12. The NS198 CMD3 marker also mapped to chromosome 12 (exclusively), but substantially distanced from the CMD2 markers. These results support the initial characterisation of CMD3 by Okogbenin *et al.*, (2012)[15] as an additional distinct locus, upstream of CMD2. The SSRY40 marker was consistently mapped to chromosome 1, contradicting early mappings to chromosomes 9, and 10 [19, 41]. One SSRY40 isolocus was identified on chromosome 7 of the SC205 assembly, in addition to the locus on chromosome 1; as the SC205 assembly is not haplotype phased it is unclear whether this represents a duplication, or a translocation. In several instances, e.g. SSRY28 in TMEB117 (Fig. 2) a null haplotype was observed whereby the marker locus was present but primer binding regions contained multiple SNPs/INDELs, or were absent entirely. The frequency of null alleles was difficult to confirm as cross-referencing of gels and genomic loci for TME3, TMS60444, TMEB117 and *M. glaziovii* and revealed that some marker loci did produce amplicons despite primer mismatches (**Supplementary Table S7**). Notably, a conserved G/T SNP was identified in the SSRY40 reverse primer binding motif in all assemblies except for a single SSRY40 isolocus in the TME3 Kuon *et al.*, (2018) assemblies (https://doi.org/10.6084/m9.figshare.31287703*)*, but SSRY40 amplicons for M. glaziovii and TMEB117 were still observed from during in vitro screening **Supplementary Figure S1, Supplementary Table S7**). The precise incidence of null alleles is likely to vary with the stringency of the PCR screening reaction employed. Instances of multiple isoloci were confirmed for NS158, NS169, NS198 and SSRY28; although replicate isoloci remained genomically proximal in all cases (**Supplementary Table S5**). Two distinct RME1 alleles previously described by Nzuki *et al.*, (2017), RME1-A2, a short (~ 500 bp) CMD susceptibility associated allele, and RME1-A1 a longer (~ 700–750 bp) CMD resistance associated allele were identified [53, 54]. Several germplasm did contain a RME1-A1 locus that collocated with the SSR CMD2 markers, although RME1-A1 isoloci were also identified on chromosomes 2, 3, 6, 16, 17, and 18. A large number of RME1-A2 isoloci were identified across all assemblies, but none mapped to the CMD2/CMD3 region (**Supplementary Table S5**).

3. Marker polymorphism and allelic variation in CMD1, 2 and 3 loci

As the FDR rate observed is likely results from the combined impact of many different causative factors (see Section 1) to establish the degree, and proportion of FDR specifically derived from homoplasy induced error we sought to identify sequence polymorphisms in marker amplicons that could differentiate susceptibility associated, and resistance associated marker alleles. Allelic variation of this form could improve the accuracy of marker screening through sequence aware marker genotyping h SSR [42–44]. Alignment of the sequenced marker amplicons, and the marker allele sequences obtained from the cassava genomic assemblies (see Section 2) revealed SNPs in the conserved regions surrounding the di/tri-nucleotide repeats, which segregated with response phenotype (https://doi.org/10.6084/m9.figshare.31287703*)* To cross validate these SNPs, and to increases the size of the dataset to improve the confidence of statistical assessment, genotyping by sequencing (GBS) SNP datasets from CMD resistant and CMD susceptible germplasm were called against marker loci in the reference germplasm (AM560-2 v8.1) with Phytozome, Jbrowse (**Supplementary Table S2**). GBS data was available for 35 germplasm including eight of the germplasm screened *in vitro* (**Supplementary Table S1***).* Both GBS-SNPs data, and whole genome assemblies were available for four of the germplasm assessed *in vitro*, namely TME3, TMS60444, TMEB117, and *M. glaziovii* (Mull.Arg.) Allem (accession unspecified). Based on these four germplasm a consolidated set of 45 cross-validated SNP loci, spread across the six CMD response linked markers assessed, were scored in 55 germplasm (**Supplementary Tables S1 & S2**). SNPs that were significantly correlated to response phenotype were identified by a chi-independence test (**Table 2**).

Table 3
SNP markers mined from within CMD-resistance associated molecular markers

Marker	SNP	χ2 test statistic	p-value		All expected Frequencies > 5	Selected for elimination of susceptibility associated amplicons.
			Derived from χ2 distribution	Derived from Monte-Carlo simulation		
NS169	Chr12.7731809	13.297	0.001	0.000		
	Chr12.7731815	20.540	0.000	0.000	X	
	Chr12.7731850	13.297	0.001	0.002		
	Chr12.7731856	8.689	0.009	0.006	X	
	Chr12.7731941	0.059	1.000	1.000		
	Chr12.7731945	8.689	0.009	0.008		
NS158	Chr12.7731809	13.297	0.001	0.002		
	Chr12.7731815	20.540	0.000	0.000	X	X
SSRY28	Chr12.7541095	16.331	0.000	0.000	X	X
	Chr12.7541180	0.013	1.000	1.000		
SSRY40	Chr1.32767646	0.705	0.649	0.452		X
	Chr1.32767693	1.177	0.468	0.291		
	Chr1.32767695	4.415	0.089	0.050		X
	Chr1.32767731	2.763	0.236	0.175		
	Chr1.32767743	3.610	0.111	0.070		
	Chr1.32767748	1.177	0.468	0.315		
	Chr1.32767772	0.705	0.649	0.451		
	Chr1.32767785	4.935	0.055	0.039	X	X
	Chr1.32767791	6.363	0.031	0.015		
	Chr1.32767827	7.651	0.022	0.032		
	Chr1.32767849	3.779	0.111	0.083		
NS198	Chr12.1353228	7.185	0.018	0.015	X	X
	Chr12.1353243	1.984	0.323	0.209		X
	Chr12.1353305	22.290	0.000	0.000		X

Following, elimination of insignificant SNPs and SNPs with expected frequencies too low for accurate statistical assessment, one SNP was selected per marker to differentiate resistance associated and susceptibility associated marker amplicons (**Table 2**), and germplasm were re-scored on this basis (**Supplementary Table S4**). The FDR rate from scoring based on the presence of amplicons in the expected size range for the resistance associated molecular marker, was compared to the FDR rate based only on amplicons in the expected size range which also contained the resistance associated SNP were compared, and the size range for marker alleles with the susceptibility associated, and the resistance associated SNP were compared for each marker Table 3. Suitable SNPs could not be identified for RME1, but for the remaining five markers FDR reductions of 15–20% were achieved, more than halving the total rate of FDR in all cases. Extensive overlaps in the size range of susceptibility associated and resistance associated alleles were apparent for all markers.

Table 3
Relative accuracy of size and sequence based marker genotyping for cassava mosaic disease response associated molecular markers

Marker	Amplicon Size Range (bp)		By marker sequence		By marker presence	FDR Reduction	FDR Reduction Ratio
	Susceptible Germplasm	Resistant and Tolerant Germplasm	FDR	FNR	FDR		
SSRY40	331–341	321–341	0.14	0.42	0.29	0.16	0.526
NS169	298–333	291–331	0.11	0.30	0.29	0.18	0.645
NS158	135–176	136–176	0.11	0.30	0.29	0.18	0.645
SSRY28	157–186	156–182	0.1	0.37	0.29	0.19	0.624
NS198	181–198	181–200	0.08	0.250	0.29	0.21	0.735

4. QTL gene candidates identified from CMD1-3 associated SSR/SCAR markers

In early applications all SSR marker were assumed to be phenotypically neutral, as SSRs were thought to be occur exclusively in non-coding genomic regions. In subsequent SSR studies this assumption has been thoroughly disproved, and assessment of SSR neutrality has become routine. The five SSR markers under study (SSRY40, SSRY28, NS169, NS158, and NS198), were all identified prior to 2005 and applied directly (presumably) without a neutrality assessment, and the neutrality of these markers remains undetermined to date. Subsequently, isoloci for of these markers have been identified within protein coding genes figure, and also reported by Nzuki et al., [45].

Previously, Kuria *et al.*, (2017), mapped NS158 and NS169 to a *HIGH MOBILITY GROUP B PROTEIN 1* (*HMGB1*) gene (JGI: Manes.12G074900), and NS198 to a *PROTEIN PHOSPHATASE TYPE 2A* (*PP2A*) *SUBUNIT B'δ* gene (JGI: Manes.12G016800). Our results confirmed this mapping for the updated AM560-2 v8.1 reference assembly, and additionally mapped SSRY40 to a *PROGRAMMED CELL DEATH PROTEIN 7* (*PDCD7*) gene (JGI: Manes.01G251600). The reference assembly does not contain the CMD2 associated RME1-A1 isolocus on chromosome 12, but sequence homology was identified between RME1-A1 amplicons and long non-coding (lnc) cassava RNAs, with the closest match to the lncRNA genes *LOC128695772* located on chromosome 3 of the AM560-2 v8.1 reference assembly (**Supplementary Table S6**). Interestingly, despite the high frequency of RME1-A2 isoloci observed, RME1-A2 alleles did not match any previously characterized cassava genetic elements. It has not been established whether SSRY40 and NS198 are located proximal to, or are included within the CMD1 and CMD3 loci, respectively [8, 15]. Thus *PDCD7* and *PP2A-B'δ* constitute candidate genes for their respective loci. Conversely, CMD2 markers are generally considered to fall outside of the CMD2 locus [19, 46], however, as the borders identified for the CMD2 locus differ considerably between studies [19, 35, 37, 38, 45, 47], and the number and location of CMD2 associated NS158/NS169 isoloci differs considerably between germplasm (Fig. 2), the a possibility of a minor contribution by *LOC128695772* and/or *HMBG1* to CMD response remains open. To assess the potential activity of *PDCD7*, *HMBG1*, *PP2A-B'δ*, and *LOC128695772* in CMD response, gene ontologies were examined for possible connections to anti-viral immunity, and immunity related gene orthologs were identified from the literature (Table 4).

Table 4

Table 4: Genes nesting CMD1-3 molecular markers, and their immunity related orthologs

QTL	Marker	Marker-linked gene			Immunity related ortholog	
		Gene ID	Gene Name	Ontology	Gene ID	Immune Activity
CMD1	SSRY40	Manes.01G251600	*PDCD7*	Apoptotic Cell dead and Alternative Splicing [48]	*AT2G46200.1* *(U11-59K)*	Alternative splicing and transcriptional regulation in *Arabidopsis* [49]
CMD2	NS158/NS169	Manes.12G074900	*HMBG1*	Transcriptional regulation and chromatin binding [50]	*Q9LG02*	Transcriptional suppression of *pepper golden mosaic virus* [51]
CMD2	RME1-A1	*LOC128695772*		Viral RNA Silencing [52]	*miR168*	Regulation of SACMV recovery in TME3 [53]
CMD3	NS198	Manes.12G016800	*PP2A-B'δ*	Immune regulation and signal transduction [54]	*GC01P212285*	Cofactor recruited by viral replicase [55]

To further characterised the putative role of *PDCD7*, *HMBG1*, *PP2A-B'δ*, and *LOC128695772* in CMD response, differential gene expression post challenge with SACMV was determined by RT-qPCR, in the T200 (CMD-susceptible), TME3 (CMD-tolerant) and CR43.3 (CMD-resistant) germplasm Fig. 3.

PDCD7 was upregulation in all three germplasm at 12 dpi (early response), with significantly greater upregulation observed for TME3 and CR43.3 comparative to T200. At 32 dpi suppression was observed for both TME3 and T200, but appeared much greater in T200. By contrast upregulation increased significantly at 32 dpi in CR43.3. In T200 *PCDC7* expression remained downregulated at 67 dpi (chronic infection), while upregulation was observed for TME3 and CR43.3 (recovery phenotype). *HMGB1* expression, closely resembled that of *PDCD7* with the only qualitative exception being the significant upregulation observed at 32 dpi in T200 (Fig. 3). By contrast, A significant up-regulation was observed for *PP2A-B'δ* in both T200 and TME3 at 12 and 32 dpi. In T200 this upregulation was sustained to 67 dpi, while in TME3 a significant down-regulation was observed at 67 dpi. No significant difference in *PP2A-B'δ* expression was observed between mock and SACMV infected CR43.3 germplasm. *LOC128695772* showed minimal expression in TME3 at 12 and 32 dpi, but a significant upregulation was observed upon recovery at 67 dpi. CR43.3 showed upregulation at 32 dpi which decrease significantly, but remained upregulated at 67 dpi (recovery) Fig. 3. In T200 no amplification was detected, this result was expected as T200 does not possess the RME1 marker (**Supplementary Table S6**).

Discussion

The six selected CMD resistance markers, namely SSRY40, SSRY28, NS158, NS169, RME1 and NS198 displayed characteristically high rates of false discovery [15, 36, 56–58], in the pathosystem under study. As the impacts of epistatic locus suppression, non-functional locus variants, and marker locus recombination have been addressed elsewhere [15, 26, 56, 59], analysis was focused toward assessment of the prevalence and impact of marker size homoplasy, by gel-electrophoresis based marker genotyping. Further, as marker loci were identified in genic regions (congruent with previous genomic mapping [35, 36, 60]), the potential contributions of four genes containing marker loci, namely *PDCD7*, *HMGB1*, *PP2A-B'δ*, and *LOC128695772*, to CMD response phenotype, were assessed. SNP based elimination of susceptibility marker amplicons, showing size homoplasy with CMD tolerant/resistance associated marker amplicons, confirmed that size homoplasy accounted for at least half of the FDR observed for the SSRY40, SSRY28, NS158, NS169, and NS198 markers in the 55 germplasm assessed (**Supplementary Table S1**). Direct measures of variance/confidence intervals for these FDR reductions would require iterative sampling from a larger test population. While this option was not available, the 55 available germplasm were sufficient for a robust confirmation of the statistical significance (by χ^2 analysis) of these FDR reductions. Further, direct comparison between fractionation patterns observed in this study, and fractionation patterns reported by previous cassava SSR screening studies, revealed a clear set of

common features (supported by genomic mapping of marker loci (**Supplementary Fig. S1, Supplementary Table S7**) that favour a high prevalence of homoplasy.

Firstly, the number of isoloci for a given primer pair differs between, germplasm, null alleles are frequent, and allele/amplicon density is high (multiple amplicons within a narrow size range). [30, 61–63]. Secondly, a heterozygous genotype with two alleles near identical in size is very common (**Supplementary Fig. S1, Supplementary Table S5 and S7**) [58, 62–64]. This pattern cannot be attributed to stutter bands as, bands are of equal intensity [30, 62, 63, 65] and heterozygous individuals carrying a null allele for the locus in question are also common, and show no, or only very faint stutter bands (**Supplementary Fig. S2**) [63, 64, 66]. Thirdly, stutter bands are also frequently present; characterized by asymmetric band intensity and laddering [27, 49, 64]. Finally, band intensity varies dramatically between germplasm of the same genotype (**Supplementary Fig. S2**)[66–69]. This phenomenon is distinct from ghost bands due to lane contamination as faint bands do not only appear in lanes adjacent to stronger bands, they can also be found surrounded by other faint bands or by null alleles [46, 67, 68]. The band intensity of a genotype homologous for a given allele should be more intense than a heterologous genotype including the given allele and a null allele; while this might explain some of the differences seen in this study, variable intensity between heterologous loci (identified by two very close but well-defined bands) is also apparent (**Supplementary Fig. S3)** [46, 61, 64]. The propensity towards homoplasy induced false discovery by these features is difficult to quantify numerically, but it does appear considerable. It should also be noted that the impact of homoplasy induced false discovery can extend beyond a single selection cycle. In the first set of CMD resistant germplasm exchanged between the IITA and the International Center for Tropical Agriculture (CIAT), one of the eighteen genotypes provided, C127, did not contain the CMD2 locus, but was incorrectly genotyped due to homoplasious alleles of the CMD2 markers considered in this study. In the breeding families subsequently generated, a CMD susceptibility rate over 50% was observed; when a rate below 30% was anticipated [11, 25].

The high allelic density observed in this study, and conserved across previous CMD SSR studies [63, 64, 66, 68, 69] is also notable, as screening results indicated a greater risk of false discovery resulting from the decreased resolution of agarose based fractionation comparative to polyacrylamide based fractionation (**Supplementary Fig. S1, Supplementary Table S7)**. For instance, AGE based screening for NS158 resolved two separate amplicons within the expected region, with the majority of germplasm (including susceptible germplasm) showing at least one, and more often, both amplicons *(***Supplementary File S2)**. Enhanced resolution of polyacrylamide gels revealed that the 'single' bands observed in AGE were comprised of multiple amplicons of similar size. This phenomenon has previously been identified for SSR markers in cassava [30, 70], since even high agarose density (5%) can introduce considerable ambiguity [71]. Thus AGE marker genotyping serves to increase the risk of homoplasy derived error.

In this study, the genomic mapping of QTL-linked marker loci lead to the identification of four putative QTL genes, namely *PDCD7* linked to CMD1, *HMBG1* and *LOC128695772* linked to CMD2 and *PP2A-B'δ* linked to CMD3. Although, the CMD1 locus is thought to be polygenic [25], specific CMD1 gene candidates have not been proposed. The CMD3 locus also currently lacks candidate genes, as while the position of NS198 loci in the *PP2A-B'δ* gene has previously been reported, the potential contribution of *PP2A-B'δ* to the CMD3 resistance locus was not been assessed. Conversely, a number of credible candidates have been proposed as the source of the ostensibly monogenic CMD2 locus [19, 72, 73], most recently a DNA polymerase δ subunit 1 gene identified by Lim *et al.*, (2022). The alternative scenario, of a polygenetic CMD2 locus, with complex dominant and epistatic relationships between different CMD2 genes, has also been proposed [38]. Given that at least one instance of an unanticipated break down of previously stable CMD field resistance, has already been observed in Uganda, resulting in mass famine-related fatalities and economic devastation, genes which provide only minor contributions to CMD resistance, may still merit inclusion in the CMD2 locus to re-enforce locus durability.

The main ontology for *PDCD7* is programmed cell death. Cassava does not demonstrate the rapid localised cell death characteristic of hypersensitive response, towards cassava CBVs [74, 75], however, upregulation of genes with ontologies in programmed cell death, has been observed following CBV infection [76, 77]. Further, reactive oxygen species induced apoptosis improves resistance to SLCMV [78], while disruption of caspase-dependant early apoptosis promotes susceptibility to SACMV [79]. In conjunction, these observations suggest a less severe form of early apoptotic cell death, linked to *PDCD7* upregulation, as a possible resistance/tolerance mechanism [76–79]. Aside from cell death, *PDCD7* also functions in alterative splicing as it encodes the 59K subunit of the U11 small nuclear ribonucleoprotein (snRNP), which is an essential component of the minor

spliceosome [49], and additionally a determinant of exon-definition interactions between the minor and major spliceosomes [80]. The U11 snRNP also functions as a promoter of genome-wide alternative splicing by enhancing splice site recognition in the major spliceosome [81, 82]. Alternative splicing can promote resistance or susceptibility in (gemini)viral pathosystems [83–85] through differential alternative splicing (DAS) of host transcripts, typically transcripts from R genes, transcription factors, and RNA silencing/splicing protein genes ([85–87], or alternative splicing of viral transcripts. To date no intronic sequences have been confirmed or predicted for SACMV. Virus induced DAS may proceed via disruption of the transcriptional regulation of pre-mRNA processing genes and/or their endogenous regulators; or via direct interactions between viral effectors and RNA processing complexes e.g. spliceosomes [88, 89] and cajal bodies [90]. Thus far only direct transcriptional regulation of snRNP genes has been confirmed in geminiviral pathosystems [84, 91]. As the *PDCD7* gene is the only gene that encodes the U11-59K protein in the AM560-2 v8.1 reference assembly, the strong transcriptional suppression observed for both T200 and TME3 at 32 dpi (Fig. 3) could potentially induce widespread disruption of endogenous alternative splicing and substantially contribute towards symptom development. This contribution may have a quantitative aspect as transcriptional suppression of *PDCD7* was significantly greater in susceptible T200 than TME3. The up regulation the *PDCD7* gene observed in the resistant germplasm CR43.3 and TME3 at 67 dpi (Fig. 3), could also contribute to resistance/tolerance by DAS induced activation of endogenous defence elements [84]. The higher expression level of this gene observed in TME3 relative to CR43.3 could be due to the higher viral titre in this germplasm (**Supplementary Fig. S3**).

The role of high mobility group B (HMGB) proteins in plant-virus pathosystems remains predominately uncharacterised [50, 92, 93]. Strong transcriptional suppression of *HMBG* genes has been observed in plum pox virus and *bamboo mosaic virus* (BaCV) pathosystems, with increased viral titre and symptom severity in *HMBG* knock-out/down mutants [94]. In BaCV infection transcriptional suppression of *HMGB* was linked to dysregulation of host factors implicated in the systemic spread of the virus [95]. Animal DNA viruses frequently hijack the HMGB1 protein to assist in the replication [96–98] and/or expression [84, 99] of their own genomes, or for transcriptional reprogramming of host genes [100, 101]. The HMGB1 protein plays a central role in the replication, maintenance, and epigenetic regulation of viral dsDNA minichromosomes [97, 102]. In particular, HMGB1 is typically essential for the rolling circle DNA replication mechanism, which is, conserved across all geminiviruses [41, 103]. To date, the only direct instance of HMGB1 activity in a geminiviral pathosystem was observed in *pepper golden mosaic virus* (PepGMV) infection; HMGB1 proteins were identified in complex with PepGMV minichromsomes in recovered/reverted leaf tissue but not in symptomatic leaf tissue, suggesting a possible role in resistance/tolerance via transcriptional suppression, either as a direct repressor or via chromatin remodelling, [51]. A similar phenomenon was apparent for SACMV, whereby strong transcriptional suppression of the *HMBG1* gene was apparent in severely symptomatic leaf tissue (T200: 32 dpi and TME3 32 dpi) while *HMGB1* upregulation was observed in asymptomatic/mildly symptomatic leaf tissue (T200:12 dpi, TME3 12 and 32 dpi, and CR43.3 at 12, 32 and 67 dpi) (Fig. 3). The comparatively minor upregulation in the resistant CR43.3 relative to the tolerant TME3 germplasm may also be related to the higher viral load in the tolerant germplasm. The minor up regulation of *HMGB1* observed in the susceptible T200 germplasm at 67 dpi, despite high symptom severity and viral load, may be associated with allelic variation in the HMBG1 gene. The HMBG1 gene sequence from SACMV-susceptible T200 has a SNP mutation (SNP12-7731945) resulting in a premature stop codon (https://doi.org/10.6084/m9.figshare.31287703), this mutation is not present in the SACMV-tolerant TME3. Thus the up-regulation observed at 67 dpi may be counteracted by the impaired functionality of the truncated HMGB1 isoform.

In plants protein phosphatase 2A (PP2A) is most active in hormonal regulation, such as abcisic acid [54, 104], brassinosteroids [105–107] and auxins [108–111]). It is also involved in abiotic stress [110, 111], immunity linked hypersensitive response (HR) and apoptosis [112, 113]. Functional characterisation of PP2A activity in phytoviral pathosystems remains lacking, but the rapidly expanding set of confirmed interactions between the PP2A-B'δ protein and viral effectors, is indicative of a substantial role in viral pathogenesis. The B'δ subunit has been identified as a key binding partner for viral replicase [114], and integrase [115, 116] as well as a promoter of viral transcription through dephosphorylation of transcriptional activator proteins [117]. Viral recruitment of B'δ is theorised to proceed via a conserved binding site characterized by a LxxIxE motif. This motif, which appears to derive from mimicry of endogenous B'δ binding partners [55, 118, 119], has been identified across a wide range of plant and animal viruses, notably including the movement proteins of two geminiviruses; *Beet curly top virus* and *Beet curly top Iran virus* [118]. A sequence search of the SACMV genome (GCA_000838365.1) revealed an instance of this motif (LNVIRE) within the replication initiation protein (Rep) derived from the *AC1* gene (AAF34895.1). The putative role of PP2A-B'δ as a direct binding partner for SACMV Rep

protein may explain the relatively higher expression levels observed at higher viral loads; the strong upregulation observed in T200 and TME3 at the presymptomatic stage suggest that the while the PP2A-B'δ protein may function as a susceptibility factor in assisting SACMV replication, it is likely not directly involved in symptom development.

In plant immunity lncRNAs have been characterised as post transcriptional regulators of defence related protein coding genes [120]; decoys for RNA-binding proteins, substrates for rapid sRNA biogenesis, sponges for the sequestration of miRNAs, and scaffolds for RNA dependant DNA methylation (RdDM) [121, 122]. Histone modification and chromatin remodelling through recruitment and regulation of remodellor complexes have also been observed [123, 124]. Alternatively, lncRNAs can contribute to susceptibility, via viral silencing suppression and general host reprogramming [125, 126]. The low homology observed between RME1 and full miR genes does not support a role in miRNA biogenesis, but the high sequence homology between RME1 and mature miRNAs was suggestive of a role in miRNA sequestration. Notably, the strong upregulation observed at recovery in TME3 is consistent with putative activity in the suppression of miR167. Bizabani *et al.*, (2021) identified suppression of miR167 and the subsequent activation of auxin signalling as an active component of TME3 recovery [53].

In conclusion, sequencing and genomic mapping of five SSR markers; SSRY40, SSRY28, NS158, NS169, and NS198 associated with three major CMD resistance loci (CMD1-3), identified size homoplasy between susceptibility associated, and resistance associated marker amplicons as a major source of false discovery. Homoplasy induced false discovery accounted for over half of the total FDR associated with each marker in 55 germplasm screened, and SNP based elimination of susceptibility associated marker amplicons reduced FDR by between 15 and 22%. These results support the adoption of sequence aware marker genotyping methods and the use of PAGE based fractionation in place of AGE fractionation, which can easily obscure minor deviations in amplicon size, or the presence of multiple amplicons within a narrow size range. Four QTL gene candidates were identified from genomic mapping of marker loci. Based on functional annotation and gene expression post challenge with SACMV: the CMD1 candidate gene *PDCD7* was characterised as a putative antiviral CMD response gene active in alternative splicing induced activation of endogenous defence elements; the CMD2 candidate gene *HMBG1* was characterised as a putative antiviral CMD response gene active in transcriptional silencing of viral genes; the CMD2 gene candidate *LOC128695772* was characterised as a putative positive regulator of immune response to SACMV; and the CMD3 candidate gene *PP2A-B'δ* was characterised as a putative CMD susceptibility factor active in viral replication, as a Rep cofactor.

Methods

1. Plant material, growth conditions and inoculation

A collection of 21 germplasm was sourced from the IITA (http://www.iita.org) and from the International Center for Tropical Agriculture (CIAT) (https://www.ifpri.org/partnership/international-center-tropical-agriculture-ciat/). Additionally, a local CMD-susceptible landrace T200 was sourced from the University of the Witwatersrand, Johannesburg, South Africa. For each germplasm, plantlets were cultivated from tissue culture on Murashige and Skoog 2 media (4.4 g/L MS salts, 20 g/L sucrose, solidified with 6.8 g/L plant agar, pH 5.8)[127]. Plantlets were sustained on MS media for approximately 2 weeks, with a 16/8 photo-period (28°C light, 25°C dark, with photosynthetic photon flux density of 150 $\mu Em^{-2} s^{-1}$) to allow for root development. Subsequently, plantlets were transferred to peat jiffy (Jiffy Group™, Zwijndrecht, Netherlands) growth bags and placed within plastic planter trays. These trays were then sealed with plastic wrap to maintain humidity, and sealed trays were placed in a phytotron facility with environmental conditions of 28^0C, and 60% humidity, under 16:8 hour light:dark photoperiod, with a light intensity of 8000–10000 lux. Over a 4 week period, the plastic wrap covering was successively perforated to allow for acclimatisation of the plantlets.

Agroinfection was used for inoculation of cassava plants, due to its low cost, and high infection rate [128]. Agroinfectious clones comprised SACMV DNA-A and DNA-B dimers in pCambia 1300 vectors (GenBank: AF234296.1) were transformed (separately) into *Agrobacterium tumefaciens* hyper-infective strain AGL1 [129]. *Agrobacterium* transformed with only the pCambia 1300 vector was included as a negative control (henceforth referred to as mock treatment). Agrobacteria for mock treatment were cultured in yeast extract peptone (YEP) broth liquid growth media containing 50 mg/mL carbenicillin Agrobacteria containing infectious clones were cultured in YEP broth containing 50 mg/mL carbenicillin and 50 mg/mL kanamycin. Flasks containing broth cultures were incubated in a shaking incubator at 30°C and 250 rpm. Growth was monitored by optical density measured by

spectrophotometry at 600 nm. Once an optical density of 1.8-2.0 was achieved, *Agrobacterium* cells were retrieved by centrifugation at 4200 rpm for 30 min. The supernatant was discarded and the agrobacterium pellet was rinsed in 40 mL of autoclaved distilled water, and re-suspended in 5 mL of infiltration buffer (YEP liquid media supplemented with 200 µM acetosyringone). SACMV-A and SACMV-B infectious clones were combined to obtain the final inoculum [129]. Approximately 100 µL of inoculum was injected directly into the petiole just below each apical leaf using a 0.3 mL needle tip syringe. For each germplasm three biological replicates were employed, each composed of 4–6 plants (experimental replicates) per treatment group. Once inoculated plants were placed in deep containers, covered with cling wrap (to prevent desiccation of wounded tissue) and placed back into the phytotron. Approximately one week after inoculation plants were transferred from jiffy grow bags to pots containing a 3:1 mix of autoclaved soil and sterilised vermiculite, where they were maintained at 28^{O} C and 60% humidity, under 16:8 hour light:dark photoperiod, with a light intensity of 8000–10000 lux for the remainder of the trial.

2. Symptomology assessment

Symptom scoring and tissue sampling was performed at 12, 32, and 67 days post infection (dpi). The second and third uppermost apical leaves of each plant were observed for symptom scoring before harvesting with a sterilised metal tweezers and flash-freezing in liquid nitrogen. Leaf tissue was stored at minus 80°C until use. Symptom scoring was performed following via a disease severity index (DSI) scores [130]. The DSI scores for each germplasm at a given time-point were obtained from an arithmetic mean of the sample means corresponding to each biological replicate. Confidence intervals were obtained from the arithmetic mean of the unbiased population variance estimated by Eq. (1).

3. Determination of viral load.

To calculate viral load qPCR was performed on the extracted DNA with primers designed to target the core coat protein *AC1* gene (**Supplementary Table S9**). Prior to qPCR, the total DNA extracts of experimental replicates were pooled for each biological replicate, and standardised to a final concentration of 100 ng/µL. A qPCR reaction volume of 10 µL was utilised, composed of 5 µL of 2x SYBR Green master mix (Thermofisher, Maxima SYBR Green/ROX, Thermofisher Scientific Ballics UAB), 0.5 µL each of forward and reverse primers (10 µM), 1 µL (100 ng) DNA, and 3 µL of PCR grade water. Three analytical replicates where prepared for each biological replicate and run on a CFX96 T100 Lightcycler (BIO-RAD, CFX96™ Optics Module, Bio-Rad Laboratories, Singapore), was used for qPCR, with a thermocycling regime composed of initial denaturation at 95°C for 10 min, followed by 35 cycles of denaturation at 95°C for 10 s, annealing at 60°C for 10 s and extension at 72°C for 15 s (with a plate read at the end of each extension step), followed by a melting curving from 65°C to 95°C with a ramp rate of 0.1°C/s. The double standard-curve method advised by Gullup & Acherman (2008)[131] was employed for target quantification. A five-parameter sigmodal model was selected as optimal for amplification curve fitting based on second-order corrected Akaike Information Criterion, and target concentrations were calculated based on the second derivate crossing point, using the *qpcR* R package [132].

4. Detection of marker amplicons

Screening for CMD resistance markers was performed as follows: 25 µL reaction mix was prepared composed of 5 µL of 10X DreamTaq buffer (Thermofisher Scientific, USA), 0.5 µL dNTP mix (10 µM), 1 µL each of forward and reverse primers (10 µM) (**Supplementary Table S9**), 0.25 µL of Taq enzyme, 1 µL of DNA template (100 ng/µL) and 1.25 µL of $MgCl_2$ (2.5 µL in the case of NS198) made up to 25 µL with PCR grade water, Thermocycling was performed with an initial denaturation at 95°C for 3 min, followed by 35 cycles of 95˚C for 30 s, annealing for 30 s (1 min for NS169 and RME1), and extension at 72°C for 1 min (5 min for NS169 and RME1), with a final extension at 72°C for 5 min. Annealing temperatures of 60, 60, 65, 54, 54, and 48˚C were utilised for SSRY40, NS158, NS169, SSRY28, RME1, and NS198, respectively.

PCR amplicons were resolved on 3% agarose with 20 ppm EtBr gels, run at 80 V, 400 mA, and visualised after 90 min. SSR amplicons were additionally separated on 8% polyacrylamide gels, run at 110 V, 400 mA, for 70 min, and visualised using the acidic silver staining technique developed by Huang *et al*., (2018) [133]. To gain insight into the extent of marker sequence polymorphism between diverse accession, For each marker in each germplasm, all amplicons within the expected size range

were excised from the electrophoresis gel and retrieved by silica column based DNA purification utilising the Zygomo-Spin PCR clean-up kit (Macherey-Nagel, Duren, Germany) in accordance with kit recommendations. Briefly, amplicon bands were excised from the gel and dissolved in Extraction Buffer (100 µL Buffer per mg of gel fragment). The dissolved gel-buffer solution was then loaded onto Zygomo-Spin microcentrifuge tube silica columns and centrifuged at 12,500 rpm for 30 s. Column bound DNA was washed by the addition of 700 µL of an ethanol based wash buffer and centrifugation at 12,500 rpm for 30 s. Finally, DNA was eluted from the column with 100 µL of autoclaved distilled water, and concentrated down to 20 µL utilising a vacuum spin (Eppendorf). Purified amplicons were then utilized as a template for a second round of PCR (to increase amplicon concentration), and sequenced via Sanger sequencing, at Inqaba Biotech (Inqaba Biotechnical Industries (Pty) Ltd, Menlo Park, South Africa) with a 3730xl DNA Analyzer. Sanger traces obtained were resolved with the Pearl (Patching references via trace assemblies) online application available from the Gear-Genomics online platform (https://www.gear-genomics.com/ accessed 06/2025), and aligned in MEGAX (https://www.megasoftware.net/), utilising the Clustal-W function with default parameters [134].

5. Polymorphism in marker amplicons

NCBI primer-BLAST [135] was employed to perform *in silico* PCR on the twenty four available *Manihot* genomic assemblies, available from the NCBI database (https://www.ncbi.nlm.nih.gov/datasets/genome/?taxon=3982) **Supplementary Table S1.** Marker amplicon sequences obtained from Sanger sequencing (see Section 4) were used also used to query each genome assembly with a BLASTn search. Results were cross referenced to identify: target marker loci with 100% sequence homology with marker primers and > 90% sequence homology with the respective marker amplicon consensus sequence; null alleles with > 90% sequence homology marker amplicon consensus sequence and SNPs or INDELs in primer binding regions; and off target amplicons derived from regions containing primer binding motifs, but lacking homology (< 80%) which the consensus sequence for the respective marker. Marker allele sequences obtained from the marker loci identified in each genomic assembly, were align with MEGAX (as in Section 4) to identify SNPs/INDELs between marker alleles. Additionally, Genotyping-by-Sequencing dataset or germplasm of known CMD response phenotype [136] were called against the AM560-2 v8.1 reference assembly via phytozome-JBrowse (https://phytozome-next.jgi.doe.gov/jbrowse/ accessed 06/2025) [137] to identify SNPs in marker loci.

SNPs derived from Sanger sequencing, *in silico* PCR and GBS-SNP alignment were cross-referenced and a consolidated set of 45 SNP loci were scored in 55 germplasm (**Supplementary Table S1**). To determine the capacity of the SNPs identified to correct false discovery, a X^2 independence test was performed against CMD response phenotype, a minimum expected frequency above 5 is generally considered sufficiently high that the the X^2 test statistic will conform to the X^2 -*distribution.* Below this frequency, p-values derived from the X^2 -distribution may not be accurate; for SNPs within this category p-values were calculated by Monte-Carlo simulation (n = 2000), using the *chiq.test* function from the *r-base* package. Significant SNPs were selected for marker to differentiate susceptibility and resistance associated markers, and false discovery rates were calculated before and after the elimination of germplasm with susceptibility associated marker alleles.

6. QTL gene candidates with nested isoloci of CMD-resistance QTL linked markers.

Genomic mapping of markers associated with the CMD1,2 and 3 loci lead to the identification of four genes containing nested marker loci, namely: Manes.01G251600 a *PROGRAMMED CELL DEATH PROTEIN 7 (PDCD7)* gene with a SSRY40 marker locus in the 2nd intron of the gene sequence derived from the AM560-2 v8.1 reference genome (GenBank: GCA_001659605); Manes.12G074900 a *HIGH MOBILITY GROUP B PROTEIN 1 (HMGB1)* gene with a NS169 locus beginning in the 4th intron and ending in the 5th exon of the gene sequence derived from the AM560-2 v8.1 reference genome; and the *PROTEIN PHOSPHATASE TYPE 2A (PP2A) SUBUNIT B'δ* gene with a NS198 marker locus in the 1st intron of the gene sequence derived from the AM560-2 v8.1 reference genome. As the reference assembly does not contain the CMD2 associated RME1-A1 isolocus on chromosome 12. The RME1-A1 amplicons consensus sequence was queried against the microRNA (miRNA) database miRBase (https://www.mirbase.org/accessed06/2025) and the long non-coding RNA database GreenNC (http://greenc.sequentiabiotech.com/accessed06/2025) [140, 141]. The closely sequence match was to the lncRNA gene *LOC128695772* located on chromosome 3 of the AM560-2 v8.1 reference assembly. To assess the possible activity of *PDCD7*, *HMBG1*, *PP2A-B'δ* and *LOC128695772* in CMD response, gene ontologies were obtained from the TAIR

(https://www.arabidopsis.orgaccessed07/2025) and Panther v19.0 (http://go.pantherdb.orgaccessed07/2025) [138] databases accessed via Phytozome [139]. Further, the expression of these genes, post SACMV viral challenge, was determined via RT-qPCR as in Section 3. SACMV viral challenge was performed for the CMD-susceptible germplasm T200, the CMD-tolerant germplasm TME3, and CMD-resistant germplasm CR43.3. RNA extraction and cDNA conversion were performed on the leaf tissue collected for each germplasm at each 12, 32, and 67 dpi, with SACMV DNA-A and DNA-B infectious clones. Separate extraction and conversion reactions were prepared for each experimental replicate, and cDNA samples were pooled for each biological replicate. Quantitative PCR, was performed as previously for each QTL gene candidate and each reference gene, (**Supplementary Table S9)**, and gene expression ratios were calculated with the *qpcR* R package [132].

Declarations

Competing Interests

The authors declare no competing interests.

Funding

The specific research in this manuscript was partly funded by the French National Research Agency (ANR) through the grant agreement ANR-18-LEAP-0004-03. Further funding was provided by the Carnegie Diversifying the Academy (CDTA) programme, ID RKSBS21. Research work performed by K. Krinsky was funded by a post-graduate bursary awarded by the National Research Foundation (NRF).

Author Contribution

All wet lab and bio-informatics work presented was performed by Mx K. Krinsky in pursuit of a doctoral degree from the University of The Witwatersrand, under the supervision of Professor M.E.C Rey, and Doctor B. L. Sizani. Manuscript figures and text were prepared by Mx. K. Krinsky, and all authors contributed extensively to manuscript review.

Data Availability

All Sanger sequencing files used for this study are available from the NCBI Sequence Read Archive under the bio-project number PRJNA1420052. Marker Sequence Alignments along with the R code used for all statistical calculation, qPCR data for gene expression and viral load, and any remaining data related to this study but not included in the supplementary materials. are available at https://doi.org/10.6084/m9.figshare.31287703.

References

1. FAO, IFAD, UNICEF, WFP & WHO. The State of Food Security and Nutrition in the World 2024. (FAO (Food and Agricultural Organization), IFAD (International Fund for Agricultural Development), UNICEF (United Nations Children's Fund), WFP (World Food Programme), WHO (World Health Organisation), Rome, (2024).
2. FAO. *Agricultural Production Statistics 2010–2023* (FAO, 2024).
3. Parmar, A., Sturm, B. & Hensel, O. Crops that feed the world: Production and improvement of cassava for food, feed, and industrial uses. *Food Sec.* **9**, 907–927 (2017).
4. Immanuel, S., Jaganathan, D., Prakash, P. & Sivakumar, P. S. Cassava for food security, poverty reduction and climate resilience: A review. *Indian J. Ecol.* **51**, 21–31 (2024).
5. Jha, P. & Schmidt, S. State of biofuel development in sub-Saharan Africa: How far sustainable? *Renew. Sustain. Energy Rev.* **150**, 111432 (2021).

6. Kriticos, D. J. et al. Improving climate suitability for Bemisia tabaci in East Africa is correlated with increased prevalence of whiteflies and cassava diseases. *Sci. Rep.* **10**, 22049 (2020).
7. Chikoti, P. C. & Tembo, M. Expansion and impact of cassava brown streak and cassava mosaic diseases in Africa: A review. *Front Sustain. Food Syst* **6**, (2022).
8. Ceballos, H. et al. Excellence in cassava breeding: perspectives for the future. *Crop Breed. Genet. Genomics* **2**, (2020).
9. Ceballos, H., Hershey, C., Iglesias, C. & Zhang, X. Fifty years of a public cassava breeding program: evolution of breeding objectives, methods, and decision-making processes. *Theor. Appl. Genet.* **134**, 2335–2353 (2021).
10. Hahn, S. K., Terry, E. R. & Leuschner, K. Breeding cassava for resistance to cassava mosaic disease. *Euphytica* **29**, 673–683 (1980).
11. Hershey, C. H. Cassava genetic improvement: theory and practice. (2020). https://cgspace.cgiar.org/items/48ad4cd4-84f9-4393-8bbc-2c9b7ee2e56c
12. Mutoni, C. K. *Development of Dual Virus Resistant Cultivars and Rapid Propagation Methods for Cassava in Kenya* (University of Nairobi, 2024).
13. Fregene, M. & Puonti-Kaerlas, J. Cassava biotechnology. in Cassava: biology, production and utilization 179–207 (CABI Publishing, UK, doi:10.1079/9780851995243.0179. (2001).
14. Akano, A., Dixon, A., Mba, C., Barrera, E. & Fregene, M. Genetic mapping of a dominant gene conferring resistance to cassava mosaic disease. *Theor. Appl. Genet.* **105**, 521–525 (2002).
15. Okogbenin, E. et al. Molecular marker analysis and validation of resistance to cassava mosaic disease in elite cassava genotypes in Nigeria. *Crop Sci.* **52**, 2576–2586 (2012).
16. Annual Report 2002-Project SB-2: Assessing and utilizing agrobiodiversity through biotechnology.
17. Moreno Cabrera, M. I., Tomkins, J., Okogbenin, E. & Fregene, M. A. Positional cloning of CMD2 the gene that confers high level of resistance to the cassava mosaic disease (CMD). (2004).
18. CIAT_Publications.
19. Wolfe, M. D. et al. Genome-Wide Association and Prediction Reveals Genetic Architecture of Cassava Mosaic Disease Resistance and Prospects for Rapid Genetic Improvement. *The Plant. Genome* **9**, (2016). plantgenome2015.11.0118.
20. Somo, M. et al. Genomic prediction and quantitative trait locus discovery in a cassava training population constructed from multiple breeding stages. *Crop Sci.* (2020).
21. Codjia, E. D. et al. Selection for resistance to cassava mosaic disease in African cassava germplasm using single nucleotide polymorphism markers. *S Afr. J. Sci.* **118**, 1–7 (2022).
22. Thuy, C. T. L. et al. Identifying new resistance to Cassava Mosaic Disease and validating markers for the CMD2 locus. *Agriculture* **11**, 829 (2021).
23. Soro, M. et al. Evaluation of Ten Cassava Varieties for Resistance to Cassava Mosaic Disease in Burkina Faso. *ujar* 9, 266–276 (2021).
24. Olasanmi, B., Kyallo, M. & Yao, N. Marker-assisted selection complements phenotypic screening at seedling stage to identify cassava mosaic disease-resistant genotypes in African cassava populations. *Sci. Rep.* **11**, 2850 (2021).
25. Fregene, M. et al. Molecular markers for Introgression of Useful Traits from Wild Manihot Relatives of Cassava. *Marker-Assisted Selection (MAS) Disease Root Qual. Traits Root Crop J.* **32**, 1–31 (2006).
26. Okogbenin, E. et al. Marker-Assisted Introgression of Resistance to Cassava Mosaic Disease into Latin American Germplasm for the Genetic Improvement of Cassava in Africa. *Crop Sci.* **47**, 1895–1904 (2007).
27. Ferguson, M. et al. Molecular Markers and Their Application to Cassava Breeding: Past, Present and Future. *Trop. Plant. Biol.* **5**, 95–109 (2012).
28. Li, Y. C., Korol, A. B., Fahima, T. & Nevo, E. Microsatellites Within Genes: Structure, Function, and Evolution. *Mol. Biol. Evol.* **21**, 991–1007 (2004).
29. Kirk, H. & Freeland, J. R. Applications and Implications of Neutral versus Non-neutral Markers in Molecular Ecology. *Int. J. Mol. Sci.* **12**, 3966–3988 (2011).

30. Raji, A. A. et al. Gene-based microsatellites for cassava (Manihot esculentaCrantz): prevalence, polymorphisms, and cross-taxa utility. *BMC Plant. Biol.* **9**, 118 (2009).
31. Olsen, K. M. *Evolution in a recently arisen species complex: Phylogeography of Manihot esculenta Crantz (Euphorbiaceae)* (Washington University in St. Louis, United States -- Missouri, 2000).
32. Olsen, K. M. & SNPs SSRs and inferences on cassava's origin. *Plant. Mol. Biol.* **56**, 517–526 (2004).
33. Sizani, B. L., Krinsky, K., Mokoka, O. A. & Rey, M. E. C. Comparative analysis of pattern-triggered and effector-triggered immunity gene expression in susceptible and tolerant cassava genotypes following begomovirus infection. *PLOS ONE*. **20**, e0318442 (2025).
34. Jones, N., Ougham, H. & Thomas, H. Markers and mapping: we are all geneticists now. *New Phytol.* **137**, 165–177 (1997).
35. Rabbi, I. Y. et al. High-resolution mapping of resistance to cassava mosaic geminiviruses in cassava using genotyping-by-sequencing and its implications for breeding. *Virus Res.* **186**, 87–96 (2014).
36. Kuria, P., Ateka, E., Miano, D., Onguso, J. & Taylor, N. J. Identification and analysis of cassava genotype TME3 bacteria artificial chromosome libraries for characterization of the cassava mosaic disease. *AJB* **15**, 1575–1596 (2016).
37. Kuon, J. E. et al. Haplotype-resolved genomes of geminivirus-resistant and geminivirus-susceptible African cassava cultivars. *BMC Biol.* **17**, 1–15 (2019).
38. Cornet, L. et al. A BAC-guided haplotype assembly pipeline increases the resolution of the virus resistance locus CMD2 in cassava. *Genome Biol.* **26**, 185 (2025).
39. Xu, X. D. et al. Telomere-to-telomere assembly of cassava genome reveals the evolution of cassava and divergence of allelic expression. *Hortic. Res.* **10**, uhad200 (2023).
40. Landi, M. et al. Haplotype-resolved genome of heterozygous African cassava cultivar TMEB117 (Manihot esculenta). *Sci. Data*. **10**, 887 (2023).
41. Mohan, C. et al. Mapping New Genetic Markers Associated with CMD Resistance in Cassava (Manihot esculenta Crantz) Using Simple Sequence Repeat Markers. *JAS* 5, p57 (2013).
42. Hoffman, J. I. & Nichols, H. J. A Novel Approach for Mining Polymorphic Microsatellite Markers In Silico. *PLoS One*. **6**, e23283 (2011).
43. Šarhanová, P., Pfanzelt, S., Brandt, R., Himmelbach, A. & Blattner, F. R. SSR-seq: Genotyping of microsatellites using next-generation sequencing reveals higher level of polymorphism as compared to traditional fragment size scoring. *Ecol. Evol.* **8**, 10817–10833 (2018).
44. Yang, J. et al. Target SSR-Seq: A Novel SSR Genotyping Technology Associate With Perfect SSRs in Genetic Analysis of Cucumber Varieties. *Front Plant. Sci* **10**, (2019).
45. Nzuki, I. et al. QTL mapping for pest and disease resistance in cassava and coincidence of some QTL with introgression regions derived from Manihot glaziovii. *Front. Plant Sci.* **8**, 1168 (2017).
46. Akano, A. O., Dixon, A. G. O., Mba, C., Barrera, E. & Fregene, M. Genetic mapping of a dominant gene conferring resistance to cassava mosaic disease. *Theor. Appl. Genet.* **105**, 521–525 (2002).
47. Ige, A. D. et al. Conversion and validation of uniplex SNP markers for selection of resistance to cassava mosaic disease in cassava breeding programs. *Agronomy* **11**, 420 (2021).
48. Peng, S. Y., Chang, K. W., Lin, S. C. & Tu, H. F. AmiR-134 targets programmed cell death 7 (PDCD7) gene to modulate the pathogenesis of head and neck carcinoma. *Cancer Res.* **75**, 181 (2015).
49. Zhao, J., Peter, D., Brandina, I., Liu, X. & Galej, W. P. Structural basis of 5' splice site recognition by the minor spliceosome. *Mol. Cell.* **85**, 652–664e4 (2025).
50. Grasser, K. D., Launholt, D. & Grasser, M. High mobility group proteins of the plant HMGB family: Dynamic chromatin modulators. *Biochim. Biophys. Acta*. **1769**, 346–357 (2007).
51. Ceniceros Ojeda, E. A. Estudio del minicromosoma de PepGMV durante su ciclo infectivo en Capsicum annuum. (Tesis (D.C.)--Centro de Investigación y de Estudios Avanzados del I.P.N. Unidad Irapuato. Departamento de Ingeniería Genética., (2016).
52. Bouba, I., Kang, Q., Luan, Y. S. & Meng, J. Predicting miRNA-lncRNA interactions and recognizing their regulatory roles in stress response of plants. *Math. Biosci.* **312**, 67–76 (2019).

53. Bizabani, C., Rogans, S. J. & Rey, M. E. C. Differential miRNA profiles in South African cassava mosaic virus-infected cassava landraces reveal clues to susceptibility and tolerance to cassava mosaic disease. *Virus Res.* **303**, 198400 (2021).
54. Yoon, H. S., Fujino, K., Liu, S., Takano, T. & Tsugama, D. Characterizing the role of PP2A B" family subunits in mechanical stress response and plant development through calcium and ABA signaling in Arabidopsis thaliana. *PLoS One*. **19**, e0313590 (2024).
55. Zhang, Y., Jiang, H., Yin, H., Zhao, X. & Zhang, Y. Emerging Roles of B56 Phosphorylation and Binding Motif in PP2A-B56 Holoenzyme Biological Function. *Int. J. Mol. Sci.* **25**, 3185 (2024).
56. Ferguson, M. et al. Molecular Markers and Their Application to Cassava Breeding: Past, Present and Future. *Trop. Plant. Biol.* **5**, 95–109 (2012).
57. Fernandez, L. A., Martin-Mayor, V. & Yllanes, D. Phase transition in the computational complexity of the shortest common superstring and genome assembly. *Phys. Rev. E*. **109**, 014133 (2024).
58. Olasanmi, B. et al. Bulked segregant analysis identifies molecular markers associated with early bulking in cassava (Manihot esculenta Crantz). *Euphytica* **195**, 235–244 (2014).
59. Rabbi, I. Y. et al. Genome-wide association analysis reveals new insights into the genetic architecture of defensive, agro-morphological and quality-related traits in cassava. *Plant. Mol. Biol.* **109**, 195–213 (2020).
60. Kuria, P. et al. Differential response of cassava genotypes to infection by cassava mosaic geminiviruses. *Virus Res.* **227**, 69–81 (2017).
61. Beovides, Y. et al. Molecular diversity of Cuban cassava (Manihot esculenta Crantz) cultivars assessed by simple sequences repeats (SSR). *Biotechnol. Agron. Soc. Environ.* **19**, 364–377 (2015).
62. Hurtado, P. et al. Assessment of the diversity in global cassava genetic resources based on simple sequence repeat (SSR) markers. in First Scientific Meeting of the Global Cassava Partnership GCP-1 (2008).
63. Soro, M. et al. Estimation of Genetic Diversity and Number of Unique Genotypes of Cassava Germplasm from Burkina Faso Using Microsatellite Markers. *Genes* **15**, 73 (2024).
64. Acquah, E. W., Quain, M. D., Twumasi, P. & Parkes, E. Y. Genetic relationships revealed by simple sequence repeat (SSR) markers among Ghanaian cassava cultivars released by different research groups. *J. Ghana. Sci. Assoc.* **48**, 99–109 (2014).
65. Sraphet, S. et al. SSR and EST-SSR-based genetic linkage map of cassava (Manihot esculenta Crantz). *Theor. Appl. Genet.* **122**, 1161–1170 (2011).
66. Ndung'u, J. N. et al. Genetic diversity study of Kenyan cassava germplasm using simple sequence repeats. *AJB* **13**, 926–935 (2014).
67. Bi, H., Aileni, M. & Zhang, P. Evaluation of cassava varieties for cassava mosaic disease resistance jointly by agro-inoculation screening and molecular markers. *Afr. J. Plant. Sci.* **4**, 330–338 (2010).
68. Asare, P., Galyuon, I., Asare-bediako, E. & Tetteh, J. Phenotypic and molecular screening of cassava (Manihot esculenta Crantz) genotypes for resistance to cassava mosaic disease. *J. Gen. Mol. Virol.* **6**, 6–18 (2014).
69. Varshney, R. K. & Tuberosa, R. Translational Genomics in Crop Breeding for Biotic Stress Resistance: An Introduction. in Translational Genomics for Crop Breeding 1–9 (Wiley, doi:10.1002/9781118728475.ch1. (2013).
70. Mba, R. E. C. et al. Simple sequence repeat (SSR) markers survey of the cassava (Manihot esculenta Crantz) genome: towards an SSR-based molecular genetic map of cassava. *Theor. Appl. Genet.* **102**, 21–31 (2001).
71. Scotti, I., Magni, F., Paglia, G. & Morgante, M. Trinucleotide microsatellites in Norway spruce (Piceaabies): their features and the development of molecular markers. *Theor. Appl. Genet.* **106**, 40–50 (2002).
72. Chatukuta, P. & Rey, M. E. C. Susceptibility to South African cassava mosaic virus is Associated with a RING Finger E3 Ubiquitin Ligase. (2020). https://scholar.archive.org/work/ngfvb5plbnbb3k26uoizgxb6bi/access/wayback/https://assets.researchsquare.com/files/rs-41420/v1/40208db0-f562-40e5-96e1-9b518e4c2306.pdf
73. Lim, Y. W. et al. Mutations in DNA polymerase δ subunit 1 mediate CMD2-type resistance to Cassava Mosaic Geminiviruses. *Nat. Commun.* **13**, 3933 (2022).

74. Dalio, R. J. D. et al. Hypersensitive response: From NLR pathogen recognition to cell death response. *Ann. Appl. Biol.* **178**, 268–280 (2021).
75. Freeborough, W., Gentle, N. & Rey, M. E. C. WRKY Transcription Factors in Cassava Contribute to Regulation of Tolerance and Susceptibility to Cassava Mosaic Disease through Stress Responses. *Viruses* **13**, 1820 (2021).
76. Allie, F., Pierce, E. J., Okoniewski, M. J. & Rey, C. Transcriptional analysis of South African cassava mosaic virus-infected susceptible and tolerant landraces of cassava highlights differences in resistance, basal defense and cell wall associated genes during infection. *BMC Genom.* **15**, 1006 (2014).
77. Chaowongdee, S. et al. Comparative transcriptomics analysis reveals defense mechanisms of Manihot esculenta Crantz against Sri Lanka Cassava Mosaic Virus. *BMC Genom.* **25**, 436 (2024).
78. Siriwan, W. et al. Integrated Proteomic and Metabolomic Analysis of Cassava cv. Kasetsart 50 Infected with Sri Lankan Cassava Mosaic Virus. *Agronomy* **13**, 945 (2023).
79. Chatukuta, P. & Rey, M. E. C. A cassava protoplast system for screening genes associated with the response to South African cassava mosaic virus. *Virol. J.* **17**, 1–15 (2020).
80. Olthof, A. M. et al. Disruption of exon-bridging interactions between the minor and major spliceosomes results in alternative splicing around minor introns. *Nucleic Acids Res.* **49**, 3524–3545 (2021).
81. Niemelä, E. H., Verbeeren, J., Singha, P., Nurmi, V. & Frilander, M. J. Evolutionarily conserved exon definition interactions with U11 snRNP mediate alternative splicing regulation on U11–48K and U11/U12–65K genes. *RNA Biol.* **12**, 1256–1264 (2015).
82. Verbeeren, J. et al. An Ancient Mechanism for Splicing Control: U11 snRNP as an Activator of Alternative Splicing. *Mol. Cell.* **37**, 821–833 (2010).
83. Farooq, T. et al. Enhanced antiviral defense against begomoviral infection in Nicotiana benthamiana through strategic utilization of fluorescent carbon quantum dots to activate plant immunity. *J. Nanobiotechnol.* **22**, 707 (2024).
84. Laskar, P., Hazra, A., Pal, A. & Kundu, A. Deciphering the role of alternative splicing as modulators of defense response in the MYMIV–Vigna mungo pathosystem. *Physiol. Plant.* **175**, e13922 (2023).
85. Yang, S., Tang, F. & Zhu, H. Alternative splicing in plant immunity. *Int. J. Mol. Sci.* **15**, 10424–10445 (2014).
86. Godinho, D. P., Yanez, R. J. R. & Duque, P. Pathogen-responsive alternative splicing in plant immunity. *Front Plant. Sci* **0**, (2024).
87. Mandadi, K. K. & Scholthof, K. B. G. Genome-Wide Analysis of Alternative Splicing Landscapes Modulated during Plant-Virus Interactions in Brachypodium distachyon. *Plant. Cell.* **27**, 71–85 (2015).
88. Maio, F. et al. Identification of Tomato Proteins That Interact With Replication Initiator Protein (Rep) of the Geminivirus TYLCV. *Front Plant. Sci* **11**, (2020).
89. Ncube Kanyika, B. Protein-protein interactions in a geminivirus-cassava system. (University of the Witwatersrand, Johannesburg (South Africa), South Africa, (2022).
90. Lettin, L., Erbay, B. & Blair, G. E. Viruses and Cajal Bodies: A Critical Cellular Target in Virus Infection? *Viruses* **15**, 2311 (2023).
91. Miozzi, L., Napoli, C., Sardo, L. & Accotto, G. P. Transcriptomics of the Interaction between the Monopartite Phloem-Limited Geminivirus Tomato Yellow Leaf Curl Sardinia Virus and Solanum lycopersicum Highlights a Role for Plant Hormones, Autophagy and Plant Immune System Fine Tuning during Infection. *PLoS One*. **9**, e89951 (2014).
92. Lotze, M. T. & Tracey, K. J. High-mobility group box 1 protein (HMGB1): nuclear weapon in the immune arsenal. *Nat. Rev. Immunol.* **5**, 331–342 (2005).
93. Tang, D., Kang, R., Zeh, H. J. & Lotze, M. T. The multifunctional protein HMGB1: 50 years of discovery. *Nat. Rev. Immunol.* **23**, 824–841 (2023).
94. Vozárová, Z., Žilová, M. & Šubr, Z. Differentially expressed genes in healthy and plum pox virus-infected Nicotiana benthamiana plants. *Acta virol.* **59**, 389–397 (2015).
95. Alazem, M., He, M. H., Chang, C. H., Cheng, N. & Lin, N. S. Disrupting the Homeostasis of High Mobility Group Protein Promotes the Systemic Movement of Bamboo mosaic virus. *Front Plant. Sci* **11**, (2020).
96. Guo, Y. et al. miR-505 inhibits replication of Borna disease virus 1 via inhibition of HMGB1-mediated autophagy. *J Gen. Virol* **103**, (2022).

97. Hidalgo, P. et al. The protein composition of human adenovirus replication compartments. *mBio* **16**, e02144–e02124 (2025).
98. Yu, R. et al. HMGB1 Promotes Hepatitis C Virus Replication by Interaction with Stem-Loop 4 in the Viral 5' Untranslated Region. *J. Virol.* **90**, 2332–2344 (2016).
99. Jung Song, M. et al. The DNA Architectural Protein HMGB1 Facilitates RTA-Mediated Viral Gene Expression in Gamma-2 Herpesviruses. *J. Virol.* **78**, 12940–12950 (2004).
100. Kim, E. S. et al. Hepatitis B virus X protein counteracts high mobility group box 1 protein-mediated epigenetic silencing of covalently closed circular DNA. *PLoS Pathog.* **18**, e1010576 (2022).
101. Lynch, K. L. et al. A viral histone-like protein exploits antagonism between linker histones and HMGB proteins to obstruct the cell cycle. *Curr. Biol.* **31**, 5227–5237e7 (2021).
102. Bächer, J., Allweiss, L. & Dandri, M. SMC5/6-Mediated Transcriptional Regulation of Hepatitis B Virus and Its Therapeutic Potential. *Viruses* **16**, 1667 (2024).
103. Manti, S., Harford, T. J., Salpietro, C., Rezaee, F. & Piedimonte, G. Induction of high-mobility group Box-1 in vitro and in vivo by respiratory syncytial virus. *Pediatr. Res.* **83**, 1049–1056 (2018).
104. Manalansan, B. *Type 2A Protein Phosphatases mediate Abscisic Acid Responses by interacting with the protein kinase Open Stomata 1* (University of California, 2013).
105. Rubbo, S. D., Niloufer, G., Irani & Russinova, E. PP2A phosphatases: the 'on-off' regulatory switches of brassinosteroid signaling. *Sci. Signal.* **4**, pe25 (2011).
106. Tang, W. et al. PP2A activates brassinosteroid-responsive gene expression and plant growth by dephosphorylating BZR1. *Nat. Cell. Biol.* **13**, 124–131 (2011).
107. Wang, J. et al. Crystal structure of a PP2A B56-BubR1 complex and its implications for PP2A substrate recruitment and localization. *Protein Cell.* **7**, 516–526 (2016).
108. Ballesteros, I. et al. Specialized functions of the PP2A subfamily II catalytic subunits PP2A-C3 and PP2A-C4 in the distribution of auxin fluxes and development in Arabidopsis. *Plant J.* **73**, 862–872 (2013).
109. Lillo, C. et al. Protein phosphatases PP2A, PP4 and PP6: mediators and regulators in development and responses to environmental cues. *Plant. Cell. Environ.* **37**, 2631–2648 (2014).
110. Máthé, C., Freytag, C., Kelemen, A., M-Hamvas, M. & Garda, T. B Regulatory Subunits of PP2A: Their Roles in Plant Development and Stress Reactions. *Int. J. Mol. Sci.* **24**, 5147 (2023).
111. Máthé, C., Garda, T., Freytag, C. & M-Hamvas, M. The Role of Serine-Threonine Protein Phosphatase PP2A in Plant Oxidative Stress Signaling—Facts and Hypotheses. *Int. J. Mol. Sci.* **20**, 3028 (2019).
112. Barski, M. S., Minnell, J. J. & Maertens, G. N. PP2A Phosphatase as an Emerging Viral Host Factor. *Front. Cell. Infect. Microbiol.* **11**, 725615 (2021).
113. Garcia, A. et al. Serine/threonine protein phosphatases PP1 and PP2A are key players in apoptosis. *Biochimie* **85**, 721–726 (2003).
114. Anwar, M. I. et al. PPP2R5D promotes hepatitis C virus infection by binding to viral NS5B and enhancing viral RNA replication. *Virol. J.* **19**, 118 (2022).
115. Barski, M. S. et al. Cryo-EM structure of the deltaretroviral intasome in complex with the PP2A regulatory subunit B56γ. *Nat. Commun.* **11**, 5043 (2020).
116. Maertens, G. N., Engelman, A. N. & Cherepanov, P. Structure and function of retroviral integrase. *Nat. Rev. Microbiol.* **20**, 20–34 (2022).
117. von Creytz, I., Rohde, C. & Biedenkopf, N. The cellular protein phosphatase 2A is a crucial host factor for Marburg virus transcription. *J. Virol.* **98**, e01047–e01024 (2024).
118. Hertz, E. P. T. et al. A Conserved Motif Provides Binding Specificity to the PP2A-B56 Phosphatase. *Mol. Cell.* **63**, 686–695 (2016).
119. Kruse, T. et al. The Ebola Virus Nucleoprotein Recruits the Host PP2A-B56 Phosphatase to Activate Transcriptional Support Activity of VP30. *Mol. Cell.* **69**, 136–145e6 (2018).
120. Arunkumar, G. LncRNAs: the good, the bad, and the unknown. *Biochem. Cell. Biol.* **102**, 9–27 (2024).

121. Song, L., Fang, Y., Chen, L., Wang, J. & Chen, X. Role of non-coding RNAs in plant immunity. *Plant. Comm.* **2**, 100180 (2021).
122. Teixeira, R. M., Ferreira, M. A., Raimundo, G. A. S. & Fontes, E. P. B. Geminiviral Triggers and Suppressors of Plant Antiviral Immunity. *Microorganisms* **9**, 775 (2021).
123. Ali, S., Tang, Y. & Noncoding RNA-Mediated Regulation of DNA Methylation: Insights into Plant Epigenetic Mechanisms. *J. Plant. Growth Regul.* **44**, 373–388 (2025).
124. Kumar, K. & Chakraborty, S. Roles of long non-coding RNAs in plant virus interactions. *J. Plant. Biochem. Biotechnol.* **30**, 1–14 (2021).
125. Wang, J. et al. Re-analysis of long non-coding RNAs and prediction of circRNAs reveal their novel roles in susceptible tomato following TYLCV infection. *BMC Plant. Biol.* **18**, 104 (2018).
126. Wang, L. & Lozano-Durán, R. Manipulation of plant RNA biology by geminiviruses. *J. Exp. Bot.* **74**, 2311–2322 (2023).
127. Classic Murashige, T. & Skoog, F. A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol. plant.* **15**, 473–497 (1962).
128. Sserubombwe, W. S. et al. Diversity of begomoviruses associated with mosaic disease of cultivated cassava (Manihot esculenta Crantz) and its wild relative (Manihot glaziovii Müll. Arg.) in Uganda. *J. Gen. Virol.* **89**, 1759–1769 (2008).
129. Berrie, L. C., Rybicki, E. P. & Rey, M. E. C. Complete nucleotide sequence and host range of South African cassava mosaic virus: further evidence for recombination amongst begomoviruses. *J. Gen. Virol.* **82**, 53–58 (2001).
130. Louis, B. & Rey, C. Resistance gene analogs involved in tolerant cassava–geminivirus interaction that shows a recovery phenotype. *Virus Genes*. **51**, 393–407 (2015).
131. Gallup, J. M. & Ackermann, M. R. The 'PREXCEL-Q Method' for qPCR. *Int. J. Biomed. Sci.* **4**, 273–293 (2008).
132. Ritz, C. & Spiess, A. N. *qpcR*: an R package for sigmoidal model selection in quantitative real-time polymerase chain reaction analysis. *Bioinformatics* **24**, 1549–1551 (2008).
133. Huang, L. et al. A fast silver staining protocol enabling simple and efficient detection of SSR markers using a non-denaturing polyacrylamide gel. *J. Vis. Exp.* **1**, 57192 (2018).
134. Kumar, S., Nei, M., Dudley, J. & Tamura, K. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief. Bioinform*. **9**, 299–306 (2008).
135. Ye, J. et al. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinform.* **13**, 134 (2012).
136. Tibihika, P. D. et al. Application of microsatellite genotyping by sequencing (SSR-GBS) to measure genetic diversity of the East African Oreochromis niloticus. *Conserv. Genet.* **20**, 357–372 (2019).
137. Buels, R. et al. JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.* **17**, 66 (2016).
138. Mi, H. et al. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res.* **49**, D394–D403 (2021).
139. Berardini, T. Z. et al. The Arabidopsis information resource: making and mining the gold standard annotated reference plant genome. *Nucleic Acids Res.* **53**, 474–485 (2015).
140. Di Marsico, M., Paytuvi Gallart, A. & Sanseverino, W. Aiese Cigliano, R. GreeNC 2.0: a comprehensive database of plant long non-coding RNAs. *Nucleic Acids Res.* **50**, D1442–D1447 (2022).
141. Kozomara, A., Birgaoanu, M. & Griffiths-Jones, S. miRBase: from microRNA sequences to function. *Nucleic Acids Res.* **47**, D155–D162 (2019).

Equations

Eqn1: $\sigma^2 = s^2(n-1)/n$ |

s^2 = biological replicate sample variance

σ^2 = estimated variance for a given germplasm.

Figures

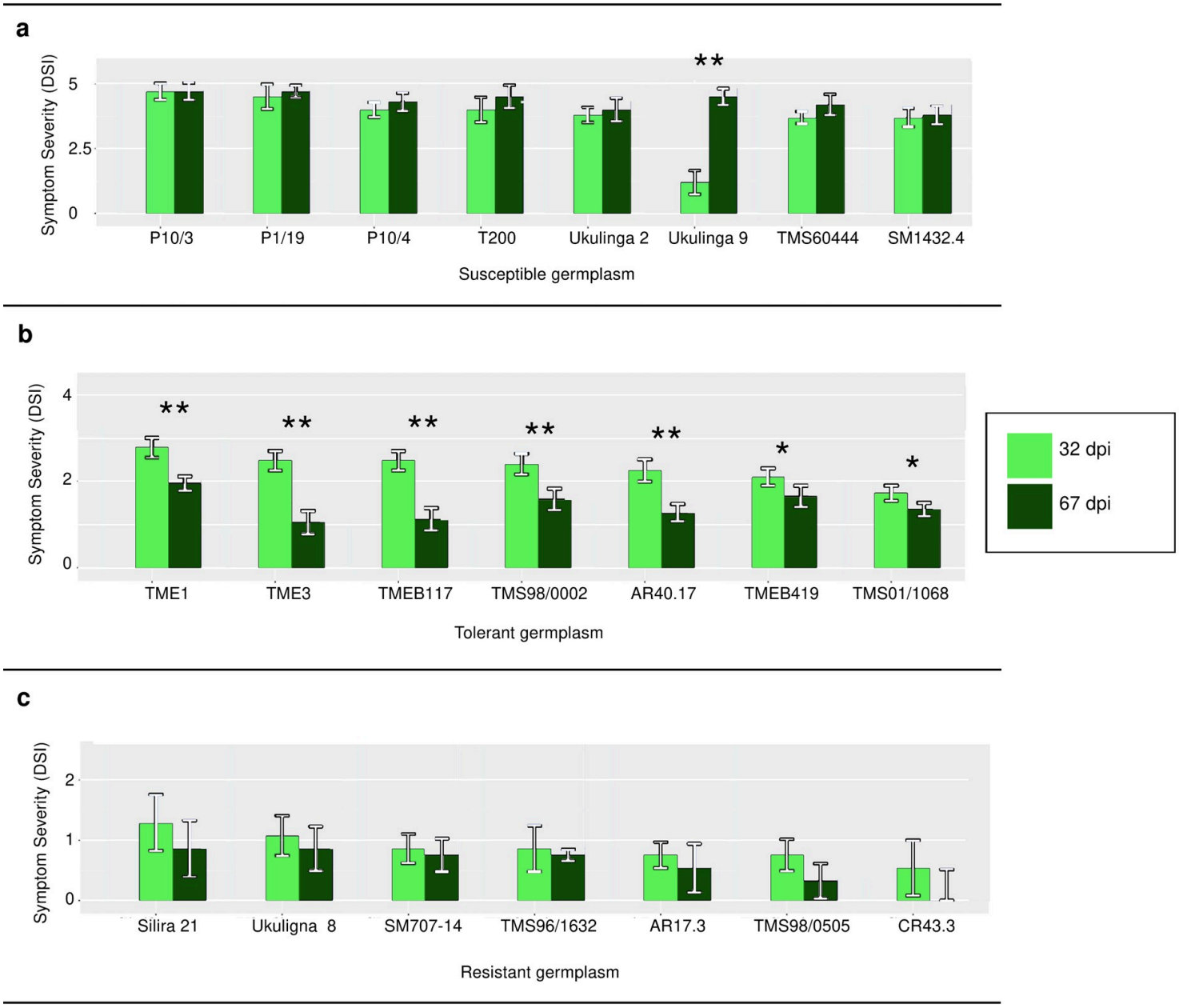

Figure 1

Leaf symptom severity induced by challenge with South African cassava mosaic virus

Leaf symptom severity induced by challenge with South African cassava mosaic virus in 22 cassava germplasm at 32 and 67 days post inoculation (dpi). a: susceptible cassava germplasm, b: tolerant cassava germplasm, c: resistant cassava germplasm. Germplasm are arranged left to right from least to most resistant). Symptom severity is given as average disease severity index (DSI) scores, with a DSI ranging from 0 for asymptomatic germplasm to 5 for severely disease germplasm showing extensive leaf deformation, and chlorosis. Error bars represent 95% confidence intervals determined by error propagation of unbiased variance estimators through second order Taylor expansion. Where significant changes in symptoms occurred between time points the associated p-values as indicated with asterisks

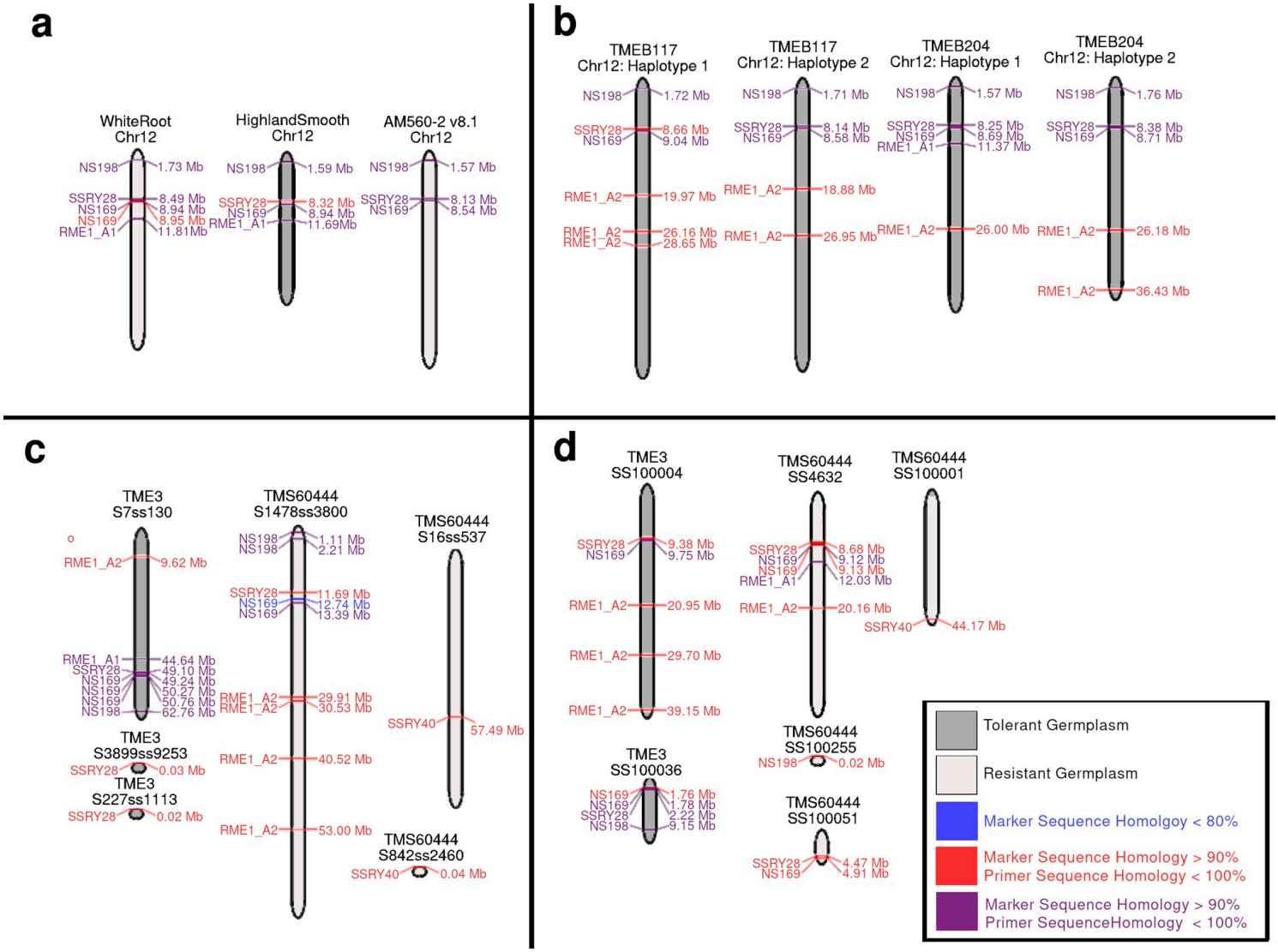

Figure 2

Genomic mapping of molecular markers associated with cassava mosaic disease resistance loci.

Six molecular markers; SSRY40 linked to the cassava mosaic disease (CMD) resistance locus CMD1, SSRY28, NS169, NS158 and RME1 linked to the CMD2 resistance locus, and NS198 linked to the CMD3 resistance locus were mapped to 24 genomic assemblies of *Manihot*germplasm through in silico PCR, and BLASTn queries with marker amplicon sequences. Two alleles (A1 and A2) corresponding to distinct isoloci were identified for RME1. **a**: Marker loci identified on Chromosome 12 for three of the 24 assemblies. The HighlandSmooth germplasm shows the canonical configuration with one CMD3 marker (NS198) upstream of the four clustered CMD2 markers, SSRY28, NS158, NS169 and RME1_A1, (as NS158 and NS169 overlapped only NS169 is shown). The CMD susceptible WhiteRoot germplasm contains two distinct NS169 isoloci, and lacks the CMD2 locus, notably the downstream NS169 isolocus contains a null allele with primer binding site mutations. The CMD susceptible AM560-2 germplasm is the cassava reference genome, which lacks both the RME1_A1 and CMD2 locus. **b**: The CMD tolerant TMEB117 germplasm and the CMD tolerant TMEB204 germplasm show variation in locus position and size between haplotypes. **c**: Variations in number of location of marker isoloci between the CMD tolerant germplasm TME3, and the CMD susceptible germplasm TMS60444. Variations between TME3 and TMS60444 as in part **c,** but based on alternative genomic assemblies, which have a higher confidence/quality in the CMD2 region but lower confidence/quality for the overall assembly.

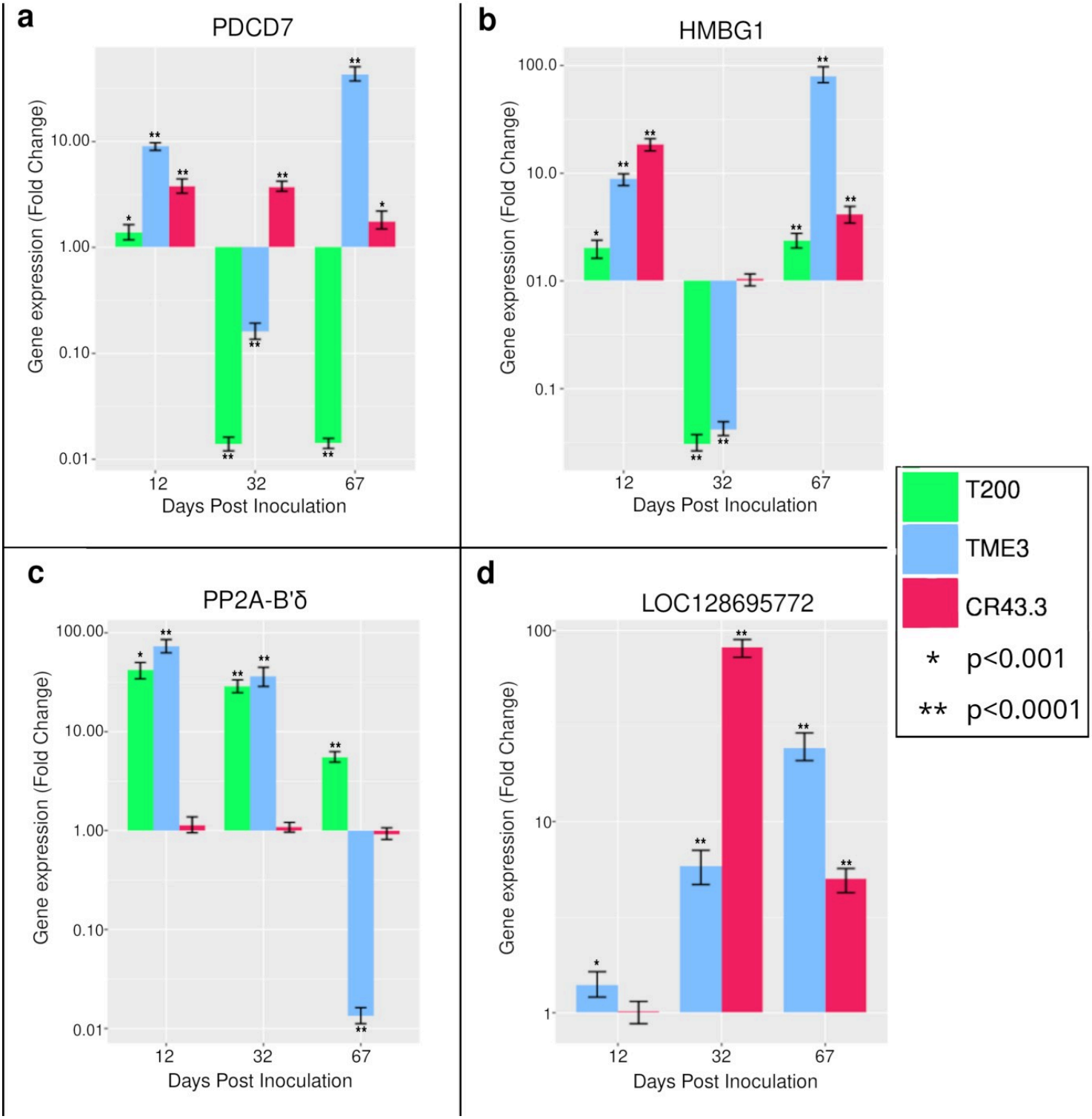

Figure 3

RNA Expression of genes containing cassava mosaic disease resistance markers

Gene expression at 12, 32, and 67 days post inoculation (dpi) with agroinfectious clones of South African cassava mosaic virus (SACMV) is presented for four cassava mosaic disease (CMD) resistant quantitative trait loci (QTL) gene candidates. **a**: *PROGRAMMED CELL DEATH PROTEIN 7* (*PDCD7*); **b**: *HIGH MOBILITY GROUP B PROTEIN 1* (*HMGB1*) gene; **c**: *PROTEIN PHOSPHATASE TYPE 2A*(PP2A) *SUBUNIT* B'δ gene; **d**: lncRNA gene *LOC128695772* for the (green) CMD susceptible T200, the CMD tolerant germplasm TME3, (blue) and the CMD resistant germplasm CR43.3 (pink). Error bars represent 95% confidence

intervals determined by error propagation through second order Taylor expansion. The R package *ggplot2* was used for data visualisation.

Supplementary Files

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

- FiguresandSupplementaryMaterials.docx
- TablesandSupplimentaryMaterials..xlsx