

Full Length Research Paper

Identification and analysis of cassava genotype TME3 bacteria artificial chromosome libraries for characterization of the cassava mosaic disease

Paul Kuria^{1,2,3*}, Elijah Ateka¹, Douglas Miano⁴, Justus Onguso¹ and Nigel J. Taylor³

¹Jomo Kenyatta University of Agriculture and Technology, P. O. Box 62000-00200, Nairobi, Kenya.

²Kenya Agricultural and Livestock Research Organization, P. O. Box 57811-00200, Nairobi, Kenya.

³Donald Danforth Plant Science Center, 975 North Warson Road, St Louis, MO, 63132 USA.

⁴University of Nairobi, P. O. Box 30197, 00100, Nairobi, Kenya.

Received 26 February, 2016; Accepted 10 June, 2016

Cassava is an economically important crop in sub-Saharan Africa; however, its yield potential is constrained by cassava mosaic disease (CMD) infection. Classical genetics and biotechnology are being harnessed to overcome the disease and secure yields for farmers. The CMD2 resistance locus flanked by three simple sequence repeats (SSR) markers and one sequence characterized amplified region (SCAR) marker were mapped in West African genotypes and shown to impart qualitative resistant to all species of CMGs. However, gene(s) associated with the CMD2 locus and their mode of actions remains unknown. In an effort to discover gene(s) located in CMD2 locus region, TME3 BAC collections were screened for the presence of CMD2 flanking markers. CMD susceptible and resistant cassava genotypes were found to contain 100% of the markers flanking CMD2 locus. SNPs and nucleotide deletions were identified within the marker sequences but there was no evidence of trait and marker association. All the SSR markers flanking CMD2, and the more recently characterized CMD3 loci were to be located on chromosome 12. Through BAC pools library hybridization with marker probes, 130 BACs were identified, but only 23 BACs contained at least CMD2 specific two markers. Whole BAC sequencing identified five clones that mapped to the marker regions. BAC29 assembled into a 100 kb contig and encoded tandem repeats of three full length R genes (3.5 kb) and two partial repeats. These R genes were conserved and highly expressed in CMD susceptible and CMD resistant cassava genotypes. Promoter sequences derived from R genes showed similar transient expression of GUS as 35S promoter. On cassava genome V6.1 BAC29 sequences were mapped to chromosome 16, eliminating their potential role in CMD resistant.

Key words: Bacteria artificial chromosome, CMD2, cassava, cassava mosaic disease.

INTRODUCTION

Cassava genetic improvement through classical breeding and biotechnology targets various traits such as disease

and pest resistance, enhanced nutritional and reduced cyanogenic content, improved starch quality as well as

decreased post-harvest deterioration (Legg et al., 2015; Akinbo et al., 2012; Welsch et al., 2010; Jørgensen et al., 2005; Lokko et al., 2005; Siritunga et al., 2004). In sub-Saharan Africa, cassava production is constrained by two viral diseases namely Cassava mosaic disease (CMD) caused by multiple species of single stranded DNA viruses in the family geminiviridae, and Cassava brown streak disease (CBSD) caused by two species of RNA viruses in the family Potyviridae (Legg et al., 2011, 2015). CMD is endemic to all cassava production regions of sub-Saharan Africa (Harimalala et al., 2015; Legg et al., 2015; Patil and Fauquet, 2009). As a result, classical breeding programs have focused significant attention on mitigating the negative effects caused by CMD. Through marker assisted breeding, three quantitative trait loci (QTL) namely CMD1, CMD2 and CMD3 have been independently mapped in cassava genotypes namely TMS30572, TME3 and TMS97/2205 (Rabbi et al., 2014; Okogbenin et al., 2012; Lokko et al., 2005; Akano et al., 2002; Fregene and Puonti-Kaerlas 2002). During the 1980s and 1990s, a CMD epidemic occurred in East Africa. A successful approach in managing the crisis was achieved through deployment of the West African landraces TME14 and TME204 which carry the CMD2 locus (Legg, 1999). These cultivars are now among the most popular genotypes in Kenya and Uganda due to their inherent resistance to CMD, good cropping characteristics, cooking qualities and taste. However, they are highly susceptible to CBSD. Others traditional East African cultivars, such as Ebwanatereka, possess desirable organoleptic traits but are highly susceptible to both CMD and CBSD (Kaweesi et al., 2014). Transgenic technologies allow the introduction of multiple traits into an individual cultivar (Taylor et al., 2012). Various breeding programs and research groups are embarking on the use of such techniques to stack multiple resistance traits onto genetic backgrounds with inherent resistance to CMD and improving the popular but CMD susceptible genotypes (Chauhan et al., 2015).

The West African cassava landraces carrying the CMD2 resistant locus are popular targets for genetic engineering (Beyene et al., 2015; Vanderschuren et al., 2012; Taylor et al., 2012) and as parents for cassava genetic improvement programs (Rabbi et al., 2014; Okogbenin et al., 2012; Akano et al., 2002). Qualitative CMD resistance conferred by CMD2 is reported to be stably inherited as a single locus and to confer broad specificity to cassava-infecting geminiviruses (Rabbi et al., 2014; Okogbenin et al., 2007). CMD2 has been mapped on the genetically similar cassava landraces TME1, TME7 and TME14 (Rabbi et al., 2014) using molecular markers designated as sequence characterized amplified region (SCAR) marker RME1,

simple sequence repeats (SSR) markers; SSRY28, NS158 and NS169 and single nucleotide polymorphism (SNP) markers (Rabbi et al., 2014; Lokko et al., 2005; Akano et al., 2002). SSR and SNP markers linked with the CMD2 locus are collocated on chromosome 12 of cassava genome version 6.1 (International Cassava Genetic Map Consortium [ICGMC] 2015). The current version of cassava genome anchor 72% (382 MB) of cassava sequences on genetic map (ICGMC, 2015) and genomic scan identified gaps within the chromosomal region carrying CMD2 locus. This indicates that there are a significant number of genes missing from the existing cassava reference genome, especially within highly repetitive regions. Therefore, in its current form, the cassava reference genome is not sufficiently reliable to unravel all sequences coding for genes harbored in CMD2 locus. In addition, the sequenced cassava genome was derived from the partial inbred line AM560-2, a progeny of the Latin-American cassava cultivar MCOL1505 (Prochnik et al., 2012).

According to Okogbenin et al. (2007), all cassava germplasm originating from South America are susceptible to CMD. This means that alleles specific to CMD resistance are most likely not present in the AM560-2 genome. In the absence of a fully sequenced and annotated cassava genome from a CMD resistant genotype, bacterial artificial chromosomes (BACs) were adopted to construct large insert libraries for positional cloning of the CMD2 locus in cassava landrace TME3. This covered 10.1X of the haploid genome, indicating 99% chance of finding any given sequence (Tomkins et al., 2004). BACs are preferred as they enabled stable cloning of large genome insert sizes of 100-200 kb and can be propagated easily in *E. coli* cells without undergoing rearrangements (Zhang et al., 2012; Shi et al., 2011). In genomics, large-insert DNA libraries are utilized in the development of physical maps of the genome (Tomkins et al., 2004), for high throughput sequencing of genomes (Prochnik et al., 2012), and in map-based cloning of agronomically important genes such as those imparting disease resistance (Ragupathy et al., 2011). Using BAC by BAC sequencing, resistance gene analogues (RGAs) were isolated and mapped from TMS30001, a CMD resistant genotype derived from *Manihot glaziovii*, and from cultivar MECU72 carrying whitefly resistant loci (Tomkins et al., 2004). Recently, high quality draft-genomes have been assembled through integration of BAC-based physical maps and BAC-end sequences (Wang et al., 2014; International Barley Genome Sequencing Consortium, 2012), indicating the power of the resources and associated technologies for elucidating genomic information in crop plants.

Map based cloning of cassava was initiated through

*Corresponding author. E-mail: kuriapk@gmail.com. Tel: +254736333294.

collaborative efforts by the International Center for Tropical Agriculture (CIAT) and Clemson University Genomics Institute (CUGI) in an effort to identify genes coding for important traits. Through this project, TME3 BAC libraries were constructed and markers in the vicinity of CMD2 locus identified (CIAT, 2006). BAC end sequencing of clones carrying markers BAC33a, BAC33b, BAC36b and SBAC33c were constructed into 13 contigs spanning across the CMD2 region based on the position of markers RME1 and BAC33b located on either side of the CMD2 locus (CIAT, 2007).

In the present study, attempts were made to assemble sequences and identify genes residing in the CMD2 locus. Thirteen BAC clones were obtained from CIAT courtesy of Dr Luis Augusto, and additional BACs libraries hybridizing to the CMD2 flanking markers were procured from CUGI. Results described here reported presence of CMD2 and CMD3 loci flanking markers in all cassava genotypes irrespective of CMD resistant status. TME3 BAC library hybridization with markers flanking CMD2 locus and whole BAC sequencing identified sequences mapping to markers on either side of CMD2. However, contigs were not long enough to transverse the entire CMD2 locus region. Tandem repeats of five disease resistance genes were identified in one of the clones and their diversity studied in different cassava genotypes. However, they were mapped to chromosome 16 of the current cassava reference genome.

MATERIALS AND METHODS

Selection of candidate BAC libraries spanning around CMD2 locus

Four markers that mapped closest to the CMD2 locus and described by Okogbenin et al. (2012) were PCR amplified from cassava genotype TME3 using the conditions and primer sequences reported by Okogbenin et al. (2012). Amplicons corresponding to the full length marker DNA sequence were cloned into Zero blunt Topo vector (Life Technologies, Carlsbad, CA, USA) and confirmed through sequencing using M13-F (5'-GTAACACGACGGCCAG-3') and M13-R (5'-CAGGAAACAGCTATGAC-3') primers.

Four high density colony filters were hybridized with each probe to cover the entire TME3 BAC libraries. Hybridizations were performed overnight at 60°C, followed by two stringent washes at 60°C with 0.1% SDS and 1X SSC for 1 h. Images of the hybridizations were recorded by phosphor screens and read by a Typhoon 9400 imager (GE Healthcare, Piscataway, NJ, USA). The coordinates of the BAC clones were identified on filters using Hybdecon software (CUGI). Bacterial clones corresponding with positive signals were isolated using a sterile toothpick and grown overnight in 3 ml of liquid LB medium containing 12.5 µg/µl chloramphenicol. Plasmid DNA was extracted from three colonies per BAC using PureLink® Quick Plasmid Miniprep Kit (Life Technologies, Carlsbad, CA, USA). The BAC hybridization results were verified through PCR using CMD2 flanking marker primers as reported by Okogbenin et al. (2012) and described in Supplementary Table 1. Based on PCR results, ten BACs were selected and submitted to Dow Agro Science (Indianapolis, IN, USA) for whole insert sequencing. An additional thirteen BAC clone libraries designated, BAC12, BAC16, BAC21, BAC22, BAC23,

BAC26, BAC29, BAC33, BAC31, BAC38, BAC40, BAC44 and BAC45 reported to have been constructed around the CMD2 locus (CIAT, 2007) were procured from CIAT and submitted to the Genome Technology Access Center at Washington University in St Louis for paired end MiSeq sequencing.

BAC sequence assembly and annotation

Raw sequences obtained from full BAC sequencing were downloaded, demultiplexed by QIIME (Caporaso et al., 2010) and trimmed using Cutadapt (Martin, 2011) to remove adaptor sequences followed by filtering of sequences of bacteria and vector origin by randomized Numerical Aligner (Vezi et al., 2012) Clean reads from each BAC were assembled using Sanger Sequence Assembly Software (DNASTAR) using default settings and all contigs larger than 5 kb used for the present analysis. Comparative analysis of BAC sequences was then performed to identify cassava genomic regions harboring homologous sequences.

Coding regions of BAC sequences were identified through BLASTN using cassava EST sequences (http://cassava.igs.umaryland.edu/blast/db/EST_asmb1_and_single.fasta) as query and the BAC clone sequences as the subjects. The potential coding sequences were blasted against cassava genome V6.1 (<http://phytozome.jgi.doe.gov/>) and homologous genes identified based on functional annotation. To increase the level of confidence, homology searches in NCBI were performed to reveal putative genes based on 95-100% nucleotide identities. For disease resistant genes, conserved protein domains were identified using HMMER software suite (<http://pfam.xfam.org/>). Primers described in Supplementary Table 1 were used to study diversity of full length disease resistant genes in various cassava genotypes.

Amplification of putative CMD resistance genes from cassava genotypes

From one of the BACs sequenced, hereafter referred to as BAC29, three tandem repeats of putative full length resistance (R) genes containing motifs for coiled-coil nucleotide binding site leucine rich repeat (CC-NBS-LRR) and two truncated R genes were identified. To identify nucleotide diversity of the R genes in cassava genotypes, primers were designed from BAC29 sequences to independently amplify 3.5 and 1.6 kb of full length and truncated R genes, respectively (Supplementary Table 1). The promoter regions of putative CMD2 gene(s) was targeted for amplification using primers designed upstream of the genes to cover 1.5-2.0 kb (Table 1). Total DNA was extracted from CMD2 types cassava genotypes TME3, Oko iyawo, TME7, TME204, CMD1 types TMS30001, TMS30572, CMD3 types TMS97/2205 TMS98/0505 and susceptible types TME117, 60444 and Ebwanatereka, using DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The DNA was quantified on NanoDrop (Thermo scientific Waltham, MA, USA) and 100 ng mixed with 22 µl AccuPrime™ Pfx SuperMix (Life Technologies, Carlsbad, CA, USA), 1 µl of each primer (0.5 µM final concentration) and subjected to the following PCR conditions; one cycle of 5 min at 94°C followed by 35 cycles of amplification (35 s at 94°C, 30 s at 52°C and 3 min (R gene ORF), 1.5 min (promoter) at 68°C) and a final cycle of 10 min at 68°C. The PCR amplicons were analyzed by running on a 1% (w/v) agarose gel electrophoresis at 120 V for 30 min. The desired fragments were excised from the gel and purified using PureLink® Quick Gel Extraction Kit (Life Technologies, Carlsbad, CA, USA) as per the manufacturer's instructions.

Cloning putative R genes and their endogenous promoters

The gel purified genes and promoter PCR products were cloned

Table 1. Markers flanking CMD resistant loci. All the markers associated with CMD resistance were located in different genomic regions of chromosome12.

Marker name	Chromosomal location V6.1	Length	Cassava gene carrying marker
SSRY28	Chromosome12:7541033..7541198	166 bp	Manes.12G074000 (5th exon and 4th intron)
SSR NS158	Chromosome12:7731640..7731816	177 bp	Manes.12G074900 (3rd intron)
SSR NS169	Chromosome12:7731640..7731970	331 bp	Manes.12G074900 (3rd intron and 4th exon)
RME1	Scaffold01154:5660..6055 Scaffold01154:14884..15045 Scaffold01154:12726..12851	740 bp	Multiple R genes
NS198	Chromosome12:1353173..1353364	192 bp	Manes.12G016800 (2 nd intron)
GBS-SNP	Chromosome12:7216154..7223867		Manes.12G071900
GBS-SNP	Chromosome12:6910006..6917228		Manes.12G069800
GBS-SNP	Chromosome12:6640962..6649139		Manes.12G068200

into Zero Blunt TOPO cloning vector following the manufacturer's protocol (Life Technologies, Carlsbad, CA, USA). Two microliters of ligation mix was transformed into 50 µl of One Shot TOP10 chemically competent *E. coli* cells (Life Technologies, Carlsbad, CA, USA) as described by the manufacturer. After one hour incubation at 37°C in 250 µl LB medium minus antibiotics, 50 µl of transformed cells were plated on LB agar plates containing 50 µg/ml kanamycin, 100 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) and 20 mg/ml X-gal and incubated overnight at 37°C. Six individual white colonies per clone were grown in liquid LB media containing 50 µg/ml kanamycin for 8 h. Plasmid DNA was purified using Purelink Quick plasmid miniprep kit (Life Technologies, Carlsbad, CA, USA) and confirmed for insert size through double digestion of 100 ng plasmid DNA with *Xma*I and *Xba*I restriction enzymes (New England Biolabs, Ipswich, MA USA). Six clones per cassava genotype were submitted to Genewiz (<http://www.genewiz.com>) for Sanger sequencing using M13-F, M13-R, and additional primers designed in the middle of gene to cover the entire open reading frame (ORF), shown in Supplementary Table 1. The sequence contigs were assembled using SeqMan Pro and aligned with MegAlign Pro of DNASTAR Lasergene 12.2 (DNASTAR, USA). Maximum likelihood phylogenetic trees were constructed to study the diversity of sequences. The sequences of putative promoters were analyzed for promoter elements using program PLACE web Signal Scan (<http://www.dna.affrc.go.jp/PLACE/signalup.html>). To study the promoter activities, clones were constructed into binary vector pCambia5000 using promoter sequences derived from CMD resistant and susceptible genotypes respectively to drive the expression of beta-glucuronidase (GUS) reporter gene whereas 35s promoter was used as the control. The promoter activity was determined through leaf infiltration transient expression in *Nicotiana benthamiana* following the procedures described by Wroblewski et al. (2005).

Northern blot analysis of potential R genes

To determine the level of putative R gene expression in different cassava genotypes, total RNA was extracted using a modified cetyltrimethylammonium bromide (CTAB) protocol (Doyle and Doyle 1990). Following extraction, RNA was incubated at 37°C for one hour with 4 µl of TURBO DNA-free™ Kit (Life Technologies, Carlsbad, CA, USA) to remove contaminating DNA. The RNA was quantified on NanoDrop (Thermo scientific Waltham, MA, USA) and

10 µg electrophoresed in 1% denaturing agarose gel for 2 h at 80 V before transfer to a positively charged Hybond nylon membrane (Amersham, UK) using DEPC treated 20X SSC (0.3 M trisodium citrate and 3.0 M sodium chloride) for 12 h. Membrane bound RNA was subjected to UV at 120,000 microjoules/cm² using a Stratilinker UV crosslinker 1800 (Stratagene, La Jolla, CA) and pre hybridized in Digoxigenin (DIG) Easy Hyb solution (Roche, Indianapolis, IN, USA) for 1 h and hybridization overnight at 42°C with DIG-labeled probes specific to each R gene.

Cloning of TOM1 gene homolog from cassava

A protein homologous to tobamovirus multiplication protein 1 (TOM1) designated Manes.16G009700.1 was identified in the same genomic region harboring a cluster of NBS-LRR genes that were homologous to BAC29 R genes. Primers TOM1-1 AGAGAATGACCAGAATGCCAGTGC and TOM1-2 TTACCGAATAGGGTGATATTGCGCCG were used to amplify 850 bp transcript of the gene from cassava genotypes TME3, TME204, 60444 and Ebwanatereka. The RT-PCR products were cloned into Zero Blunt TOPO vector and transformed into one shot TOPO ten competent cells. Colonies were screened and positive transformants were sequenced. Sequence contigs were assembled using SeqMan Pro and aligned with MegAlign Pro of DNASTAR Lasergene 12.2 (DNASTAR, USA).

RESULTS

Analysis of CMD2 locus flanking markers in different cassava genotypes

Cassava genotypes possessing resistance or susceptibility to CMD were evaluated for the presence of markers flanking CMD2 and CMD3 resistance loci. Analysis of CMD2 flanking markers revealed presence of two RME1 marker alleles with the upper band corresponding to 740 bp while the shorter fragment was about 500 bp in size (Figure 1A). Two alleles of RME1 were identified in cassava genotypes TME117, Ebwanatereka (both CMD susceptible) and in

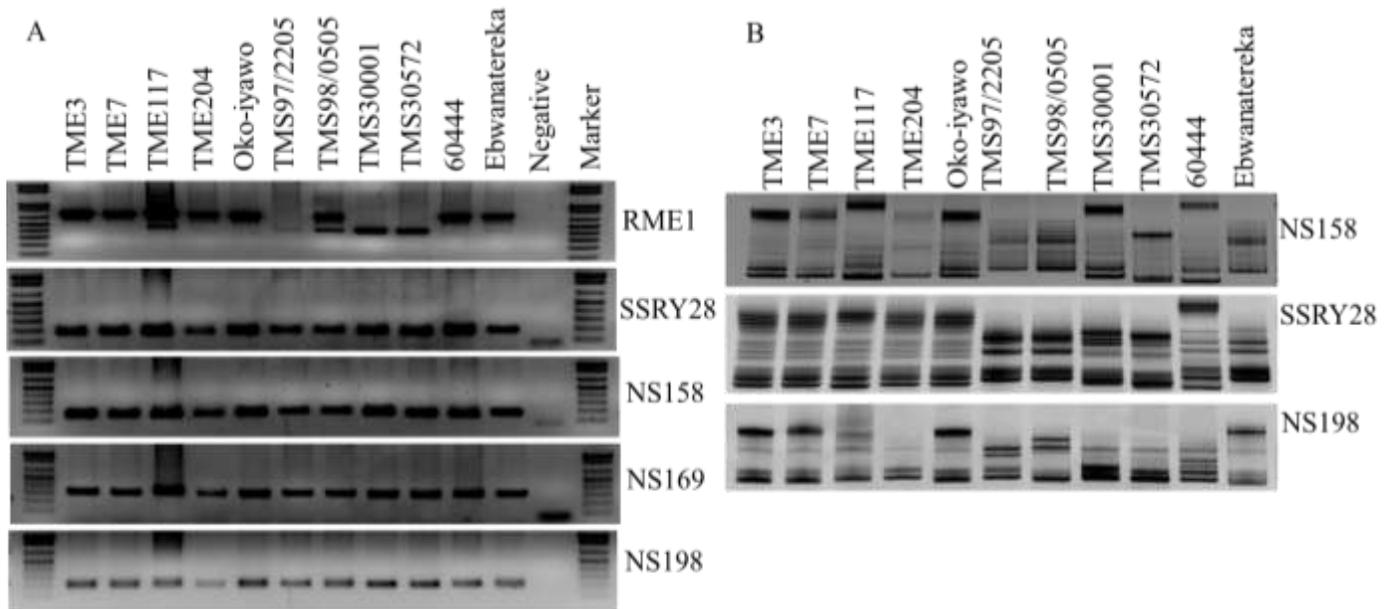


Figure 1. Analysis of cassava genotypes for the presence of markers flanking CMD resistance locus. (A) The PCR products were resolved on 1% (w/v) agarose gel for 1 h at 100 V. (B) Single-strand conformation polymorphism of SSR markers analyzed on 8% non-denaturation polyacrylamide gel electrophoresis.

TMS98/0505 (CMD resistant). One allele of RME1 corresponding to 740 nt was observed in cassava genotypes TME3, TME204, Oko-iyawo, (all CMD resistant), TME7 and 60444 (both CMD susceptible). On the other hand, genotypes conferring quantitative CMD resistance mediated by CMD1 (TMS30572) and CMD3 (TMS97/2205) respectively contained only the shorter, 500 bp sized fragment (Figure 1A). Sequence analysis of the two RME1 alleles revealed that RME1 sequences encode for a partial nucleotide binding motif of a disease resistance gene. Homology match of RME1 sequences on AM560-2 genome revealed imperfect matches, an indication of highly repetitive sequence (Figure 1A).

The SSR markers flanking the CMD2 and CMD3 loci were found to be present in all cassava genotypes (Figures 1A and B). However, analysis of sequences using single stranded conformation polymorphism (SSCP) showed differences across the genotypes (Figure 1B). Sequence analysis of SSRY28 derived from TME3 and 60444 respectively revealed a 12 nucleotide deletion in TME3 starting from position 114. The position of CMD2 flanking markers was identified as follows: SSRY28 on chromosome12:7541033 - 7541198, NS169 on chromosome12:7731640 - 7731970, NS158 on chromosome12:7731640 - 7731816 whereas CMD3 flanking marker NS198 was located on chromosome12:1353345 - 1353192 of cassava genome v6.1 (Table 1). Based on the positions of the markers flanking CMD2 and CMD3 loci, the CMD3 locus is located upstream of the CMD2 locus in the same cassava genomic region.

High-throughput TME3- BAC pools library hybridization with CMD2 marker probes

The four cassava genetic markers RME1, SSRY28, NS169 and NS158 were PCR amplified from cassava genotype TME3, cloned into zero blunt topo and confirmed with *EcoRI* restriction digestion (Figure 2A). Each marker probe was independently hybridized to two replicates of high density colony filters containing 70,000 TME3 BAC clones and images recorded by phosphor screens and read by a Typhoon 9400 imager (Figure 2B). Eighteen BAC clones hybridized with markers NS169 and NS158 respectively while 23 clones were positive with SSRY28 and 89 with RME1 (Table 2). Cross hybridization was detected only for SSRY markers whereby 23 BACs positively hybridized with at least two marker probes (Supplementary Table 2).

Amplification of CMD2 markers from candidate BAC clones using PCR

The 23 BAC clones identified to carry more than one marker were subjected to PCR analysis using primers specific to each marker (Figure 3). Ten clones were positively identified with RME1, two clones with SSRY28, six clones with NS158 (Figure 3B and C). None of the clones were positive with NS169 (Figure 3D). The two strategies of BAC screening gave conflicting results, whereby some BAC clones that positively hybridized with marker probe were scored as negative using PCR

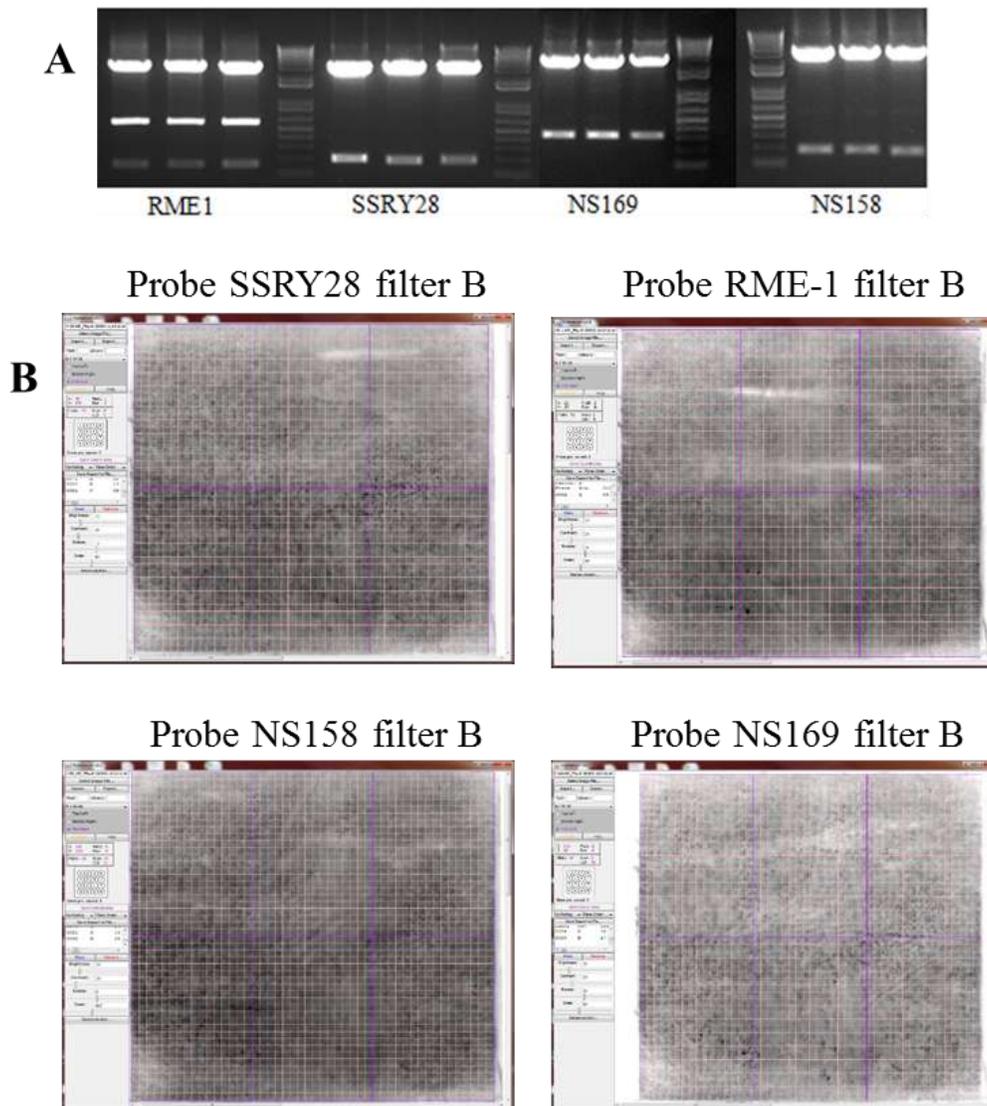


Figure 2. TME3 BAC pool libraries. (A) Restriction digestion of CMD2 locus flanking markers cloned into TOPO vector using EcoRI enzyme. Three independent clones per marker were evaluated. (B) Hybridization of BAC pools libraries on high density filter using CMD2 marker probes. Filters were deconvoluted and coordinates of positive clones assigned using Hybdecon software.

(Supplementary Table 2). The discrepancy was higher in SSR markers when compared with SCAR marker (RME1). The discrepancy between the two assays may be attributable to contamination of clones during transfer or perhaps non-specific hybridization.

Eleven BAC clones that constructed around BAC33a and SBAC33b as reported by CIAT (2007) were procured from CUGI and further screened with CMD2 mapping markers. PCR analysis of RME1 marker identified a fragment corresponding to 740 bp from five clones (29E13, 70C06, 145G15, 34L16 and 17N21) indicating the presence of RME1 marker (Figure 3E).

Thirteen clones derived from contigs constructed on the

bases of mapping marker RME1 and BAC33b (CIAT, 2007) were screened for the presence of CMD2 locus flanking markers by PCR (Supplementary Figure 1). The RME1 marker was detected in BAC23 and BAC32, respectively, while marker NS198 was detected in BAC12 and BAC22 (Supplementary Figure 1 PCR) indicating they were derived upstream of RME1. However, none of the thirteen BACs contained markers NS158, NS169, while SSRY28 mapped downstream of the CMD2 locus (Supplementary Figure 1 PCR). As a result, information pertaining to the position of marker BAC33b or the marker specific primer sequences are not clear and the contigs may therefore have been constructed upstream

Table 2. BAC pool coordinates on high density filters indicating the position of BAC clones that positively hybridized with each of the four markers flanking CMD2 locus.

Markers Flanking CMD2 Locus						
	RME1			NS169	NS158	SSRY28
011E08	085P18	110C04	132K11	003H20	003H20	003H20
017N21	086H08	110N24	134H12	012C24	025N10	011E04
020O14	092L05	110O12	136J08	021C10	026L16	021C10
021E05	096M15	114L14	137O22	026N16	026N16	024E16
021O07	096P20	114O03	138A05	033J15	053C09	026L16
022C04	097K13	116A10	144A14	036H04	065H06	026N16
024E14	098B13	116D18	146B02	039D12	070I19	033J15
025E15	098C13	118J05	146G08	042G04	082G13	035I20
025H17	098J09	118P02	146I10	084D15	092N24	079M09
026D16	098J19	119C14	148K05	094B17	099D14	082G13
027O12	099D23	120N03	151C02	106M01	101C17	091J09
027O15	099K04	121G10	154P16	110N21	109O10	100E16
029L10	099M21	122A19	158M04	116N24	110N21	104D06
033P23	099P02	122C19	159C12	126O21	115L19	108L20
049F21	100G16	123L02	162F05	127L07	116N24	109O10
050O10	102B03	125A09	166I15	128N20	120D06	110N21
050P20	102J10	128J08	169G18	146N21	121M18	112N16
051B06	102O06	129B17	171K03	180M18	122F10	116N24
051J24	104E13	130D14	174L05			118F13
064A11	105B12	130D18	184P08			126O21
064F16	105P05	131G02				139N09
069I12	106D17	131O02				159A06
085P12	109B01	132I02				188A02

of the NS158, NS169 and SSRY28 markers.

Mapping of TME3 BAC sequences to the cassava genome

Ten clones identified using hybridization and confirmed using PCR, were fully sequenced (Table 3). Sequences derived from seven BACs met the quality control threshold, allowing further analysis to be performed. The longest contig assembled was 93474 bp in size derived from BAC 52H02, whereas the majority of contigs from other BACs ranged from 5200 to 42000 bp (Table 3). At the chromosomal scale, five of the seven BAC clones were mapped to chromosome XII with two BACs mapping on chromosome II and XVII of cassava genome version 6.1 respectively (Table 3). All five BAC clones identified on chromosome XII were all positive for the presence of the RME1 marker, and based on chromosomal coordinates they were all constructed from the same genomic region (Table 3). However, none of the BACs sequences were homologous to any of the SSR maker sequences (Table 3). Therefore, it was not

possible to construct a contiguous sequences spanning from the RME1 marker to SSR markers SSRY28, NS158, and NS169.

Sequencing of BAC clones constructed from the region around CMD2 locus, from markers RME1 marker and BAC33b

Among the thirteen BAC clones previously constructed as reported by CIAT (2007), only one BAC, referred to as BAC29, contained opening reading frames (ORFs) for R genes (Figure 4). This entire BAC was assembled to 100 kb and annotated (Figure 4). Analysis of the BAC29 sequences revealed three tandem repeats of putative R-genes (CMD2a, CMD2b and CMD2c) and two partial R-genes (CMD2d and CMD2e) ORFs in opposite direction (Figure 4). The three NBS-LRR encoding genes identified from TME3 BAC29 were 3.5 kb long with 93-98% nucleotide identity to each other. The truncated R genes were missing two motifs, whereby CMD2d covered only 636 amino acid sequences. While CMD2e was interrupted by a repetitive sequence insertion and



Figure 3. Screening of TME3 BAC clones for the presence of CMD2 flanking marker fragments using PCR. The BACs were screen for the presence of RME1 (A), SSRY28 (B), NS158 (C) and NS169 (D). Numbers denotes each BAC clone as follows: 1: 08J12, 2: 92L05, 3: 98J19, 4: 102O06, 5: 110N21, 6: 109O10, 7: 126O21, 8: 116N24, 9: 17N21, 10: 3H20, 11: 146I10, 12: 159C12, 13: 136J08, 14: 137O22, 15: 22C04, 16: 130D18, 17: 26N16, 18: 21C10, 19: 27O15, 20: 26L16, 21: 33J15, 22: Cassava DNA. (E) BAC clones constructed around RME1 and sBAC33b markers. The clones are identified as follows: 1: 55H06, 2: 53H13, 3: 55H04, 4: 29E13, 5: 70C06, 6: 145G15, 7: 52H02, 8: 55H12, 9: 52H20, 10: 94M08, 11: 34L16, 12: 17N21, 13: Cassava DNA.

contained 30 amino acids deletion at the 3' end (Figure 4).

Sequence analysis of R genes isolated from different cassava genotypes

The full length and truncated R genes were identified in

cassava genotypes; TME3, Oko-iyawo, TME7, TME204, TMS30001, TMS30572, TMS97/2205, TMS98/0505, 60444 and Ebwanatereka using PCR and sequencing (Figure 5A). Alignment of contig sequences revealed that the R genes were highly conserved in different cassava genotypes and were clustered into five gene families corresponding to the BAC29 R genes (Figure 5B). However, there were divergent sequences in CMD2A that

Table 3. TME3 BAC clone sequences mapped to AM560-5 genome version 6.1.

BAC	Contig size (bp)	Cassava genomic region	Chromosome Coordinates
130D18	5346 - 34974	Chromosome12	5,604,726 - 5,639,235
136J08	5372 - 35989	Chromosome12	5,697,035 - 5,805,960
145G15	5362 - 40052	Chromosome12	5,609,999 - 5,713,990
52H02	93474	Chromosome2	6,681,446- 6,773,105
52H13	5608 - 39099	Chromosome17	1,014,513-1,079,127
70C06	5362 - 42845	Chromosome12	5,604,775 - 5,663,227
17N21	5346 - 29687	Chromosome12	5,604,726 - 5,699,674

showed 100% nucleotide identities with CMD2B and CMD2C (Figure 5B).

Northern blot analysis of putative CMD2 genes in cassava

Expression of CMD2A, CMD2B and CMD2C in cassava genotypes TME3, Oko iyawo, TME7, TME204, TMS30001, TMS30572, TMS97/2205, TMS98/0505, 60444 and Ebwanatereka was determined from 10 µg of cassava total RNA extracted from leaves of greenhouse grown plants. Based on the intensity of the signal on Northern blot, the expression of CMD2A, CMD2B and CMD2C was uniform across all cassava cultivars regardless of their CMD resistant status (Figure 6). This result indicates CMD2A, CMD2B and CMD2C genes are constitutively upregulated in all cassava genotypes investigated in this study.

Transient expression of GUS driven by putative CMD2 promoter

The ability of the promoter of CMD2A, CMD2B and CMD2C to drive expression of the GUS visual marker gene in *N. benthamiana* was studied. GUS staining was observed tobacco leaves under control of by the putative native CMD2 promoter and 35S promoters (Figure 7) and no difference was observed in the intensity of GUS staining between these two chimeric expression cassettes. GUS was also transiently expressed by putative CMD2 promoters derived from cassava genotypes TME3 and 60444. Based on this result and Northern blot analysis (Figure 6), it can be concluded that the isolated R genes are actively transcribed in different cassava genotypes regardless of their CMD resistant status.

Tobamovirus multiplication protein 1 homolog in cassava

In addition to the disease resistant gene cluster identified

(Supplementary Figure 2), a tobamovirus multiplication protein 1 (TOM1) homolog was identified in the same genomic region (chromosome 16:1,005,841- 1,011,074), indicating a locus rich in defense genes. The TOM1 homolog in cassava is encoded by Manes.16G009700 that share 87.5% nucleotide identity with Manes.17G037400 located on chromosome 17:16,718,544 – 16,723,634. Expression of a cassava homolog of TOM1 was identified in genotypes TME3, TME204, 60444 and Ebwanatereka through RT-PCR analysis (Figure 8). The TOM1 family comprises vacuolar membrane proteins that act to inhibit replication of tobamoviruses in tomato when present as mutated alleles (Ishibashi et al., 2010; Yamanaka et al., 2002). Presence of such mutations in CMD resistant or CMD susceptible cassava genotype may be could be implicated in antiviral defense. However, sequence analysis revealed 100% nucleotide identities of TOM1 gene homolog in cassava genotypes TME3, TME204 (both CMD resistant), 60444 and Ebwanatereka (both CMD susceptible) (Figure 8). Based on high sequence similarity of TOM1 homologs in CMD susceptible and resistant cassava genotypes, their involvement in antiviral defense would seem to be unlikely.

DISCUSSION

In this study, molecular markers associated with CMD resistance as reported by Okogbenin et al. (2012) and Rabbi et al. (2014) were identified in all genotypes irrespective of their response to CMD infection. These results agree with Asare et al. (2014) who detected CMD2 flanking markers in both CMD susceptible and resistant genotypes, and contradicts Bi et al. (2010) who reported absence of these markers in the susceptible genotype 60444. Studies by Parkes et al. (2015), reported absence of all marker alleles associated with CMD2 resistance in 18% of F1 progenies derived from a cross between TME11 (CMD2 type) and Dabodabo (local cultivar), despite being resistant to CMD, indicating a different source of CMD resistance. Alternatively, recombination may have occurred in subsequent crossing delinking the marker from the gene, indicating

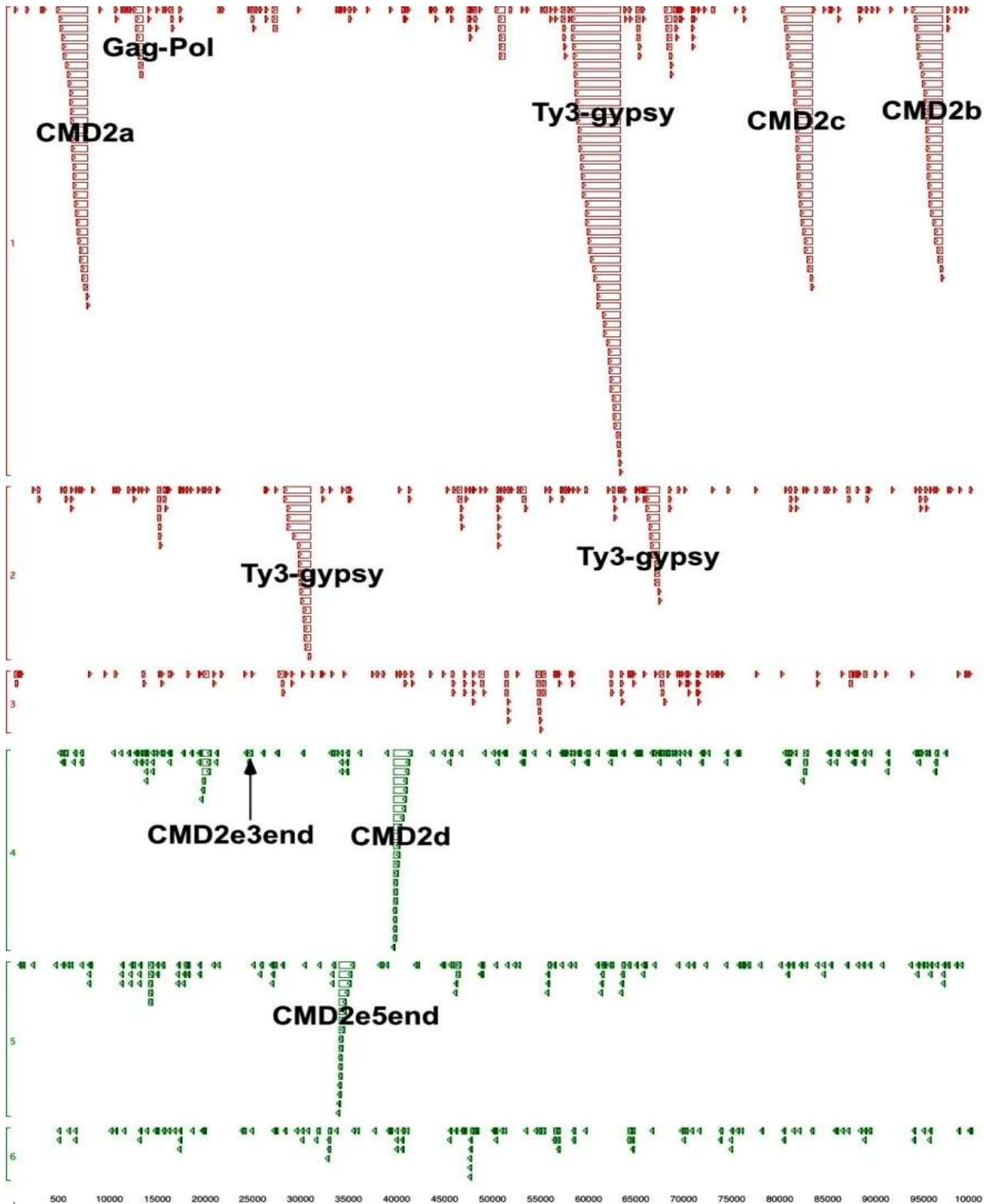


Figure 4. Assembly and annotation of BAC29 derived from TME3 cassava genotype. Several features were identified as follows: CMD2a to e encode putative R genes whereas Gag-pol and Ty3-gypsy represents transposons.

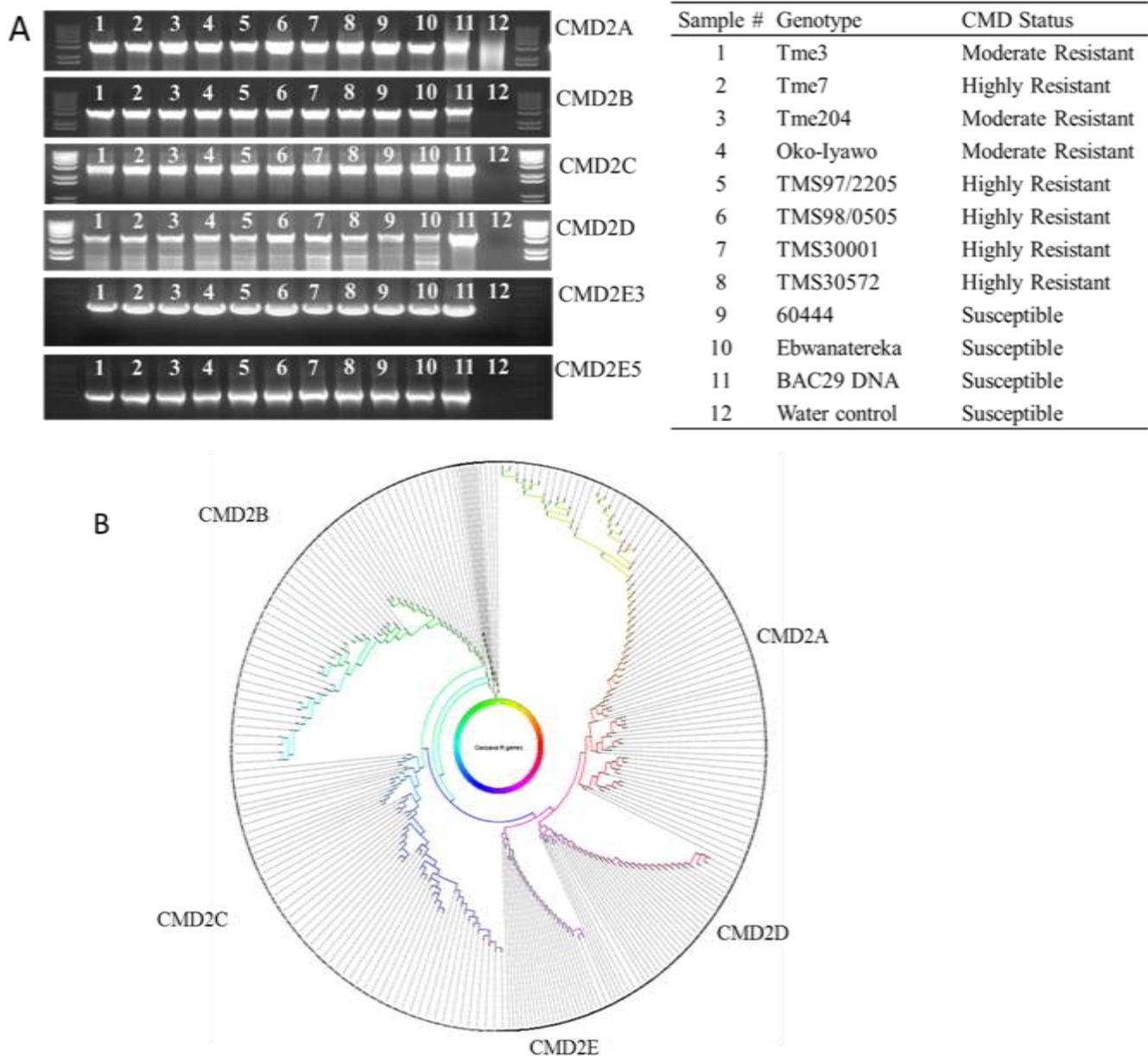


Figure 5. Analysis of R genes from diverse cassava genotypes based on BAC29 sequences. (A) PCR amplification of full length R genes 1: TME3, 2: TME7, 3: TME204, 4: Oko-Iyawo, 5: TMS97/2205, 6: TMS98/0505, 7: TMS30001, 8: TMS30572, 9: 60444, 10: Ebwanatereka, 11: BAC29, 12: Water control. (B) Phylogenetic analysis of putative CMD2 genes from different cassava genotypes. Full length nucleotide sequences were analyzed using aligned MegAlign Pro of DNASTAR Lasergene 12.2.

poor linkage between the markers and the CMD2 resistance mechanism. In cassava genotype TMS97/2205, it was shown here that the RME1 marker was not detected (Figure 1A) as previously reported by Okogbenin et al. (2012). TMS97/2205 has a pedigree of TME6 (CMD2) and TMS30572 (CMD1) (Okogbenin et al., 2012; Dixon et al., 2010) and shows very high resistant to all CMGs under field (Okogbenin et al., 2012) and greenhouse conditions (unpublished data). It is therefore important to stress that data generated in the present

study strongly suggests that the CMD2 markers are not close enough to the gene to be reliably used as a fully diagnostic tools for the presence of functional CMD2 resistance. Based on this result, the presence or absence of the previously described marker alleles may not give conclusive evidence of the presence of gene(s) conferring functional resistance to CMD and, importantly, therefore not completely reliable for screening materials during early stages of breeding programs.

The SCAR marker (RME1) and SSR markers (NS158

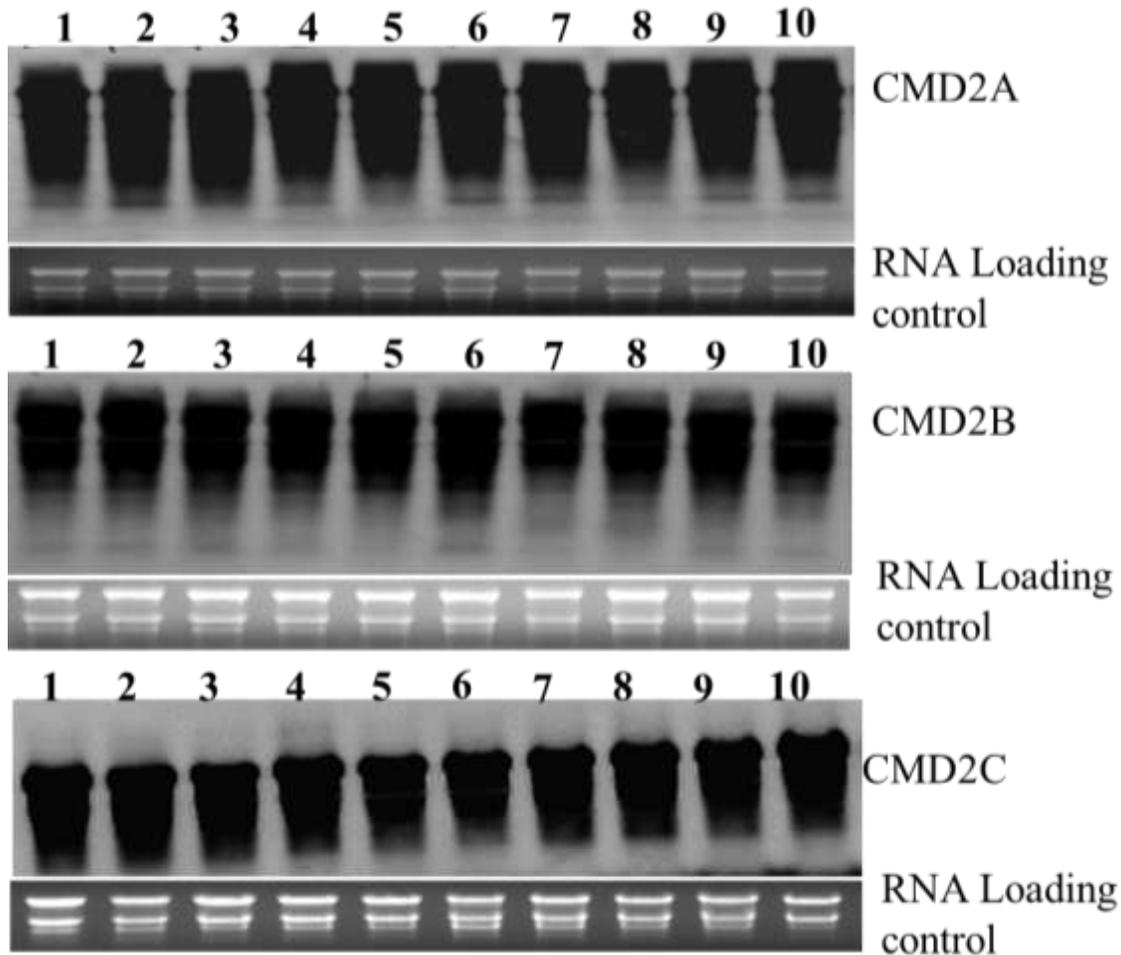


Figure 6. Northern blot analysis of putative CMD resistant gene messenger RNA accumulation in cassava. The following genotypes were analyzed; 1: TME3, 2: TME7, 3: TME204, 4: Oko-Iyawo, 5: TMS97/2205, 6: TMS98/0505, 7: TMS30001, 8: TMS30572, 9: 60444, 10: Ebwanatereka.

and NS169) are located 4.5 CentiMorgans (cM) and 7.1 cM from CMD2 locus (Parkes et al., 2015; Lokko et al., 2005; Akano et al., 2002). However, phenotypic data for these markers in F1 mapping population has not been reported (Okogbenin et al., 2013). Sixty eight percent of the total phenotypic variation of CMD2 is explained by SSRY28 at a distant of 13 cM from the CMD2 locus (Lokko et al., 2005; Akano et al., 2002), whereas NS198, that is reportedly linked with the CMD3 locus, explains 11% of variation in CMD resistance in F1 segregating populations (Okogbenin et al., 2012). One would expect that CMD resistant genotypes would contain 100% marker-trait association. However, CMD resistant loci have been described using eight molecular markers (Rabbi et al., 2014; Okogbenin et al., 2012; Lokko et al., 2005; Akano et al., 2002) that are sparsely distributed, making it impracticable to precisely associate flanking markers with CMD resistant status. High quality genome assembly and dense genetic maps are essential tools for

mapping to allow identification of loci associated with traits (ICGMC, 2015; Gaur et al., 2015; Zhang et al., 2015). In soybean, for example, a high density genetic map saturated with 2500 molecular markers were developed to precisely map QTLs conferring resistance to *Fusarium graminearum* on 300 kb region of chromosome 8 (Acharya et al., 2015). Likewise, Zhang et al. (2015) developed linkage map containing 8007 markers to identify weeping trait in an ornamental woody plant *Prunus mume*. Such quality density maps are not be available for cassava.

SSR markers flanking the CMD2 and CMD3 loci, respectively were identified on chromosome 12 of cassava reference genome (AM560-2) (<http://phytozome.jgi.doe.gov>). SSR markers NS169 and NS158 sequences are both derived from the intron of the cassava gene designated Manes.12G074900.1 annotated to code for SWI/itch/Sucrose Non-Fermentable (SWI/SNF). This is part of an ATP-dependent gene family

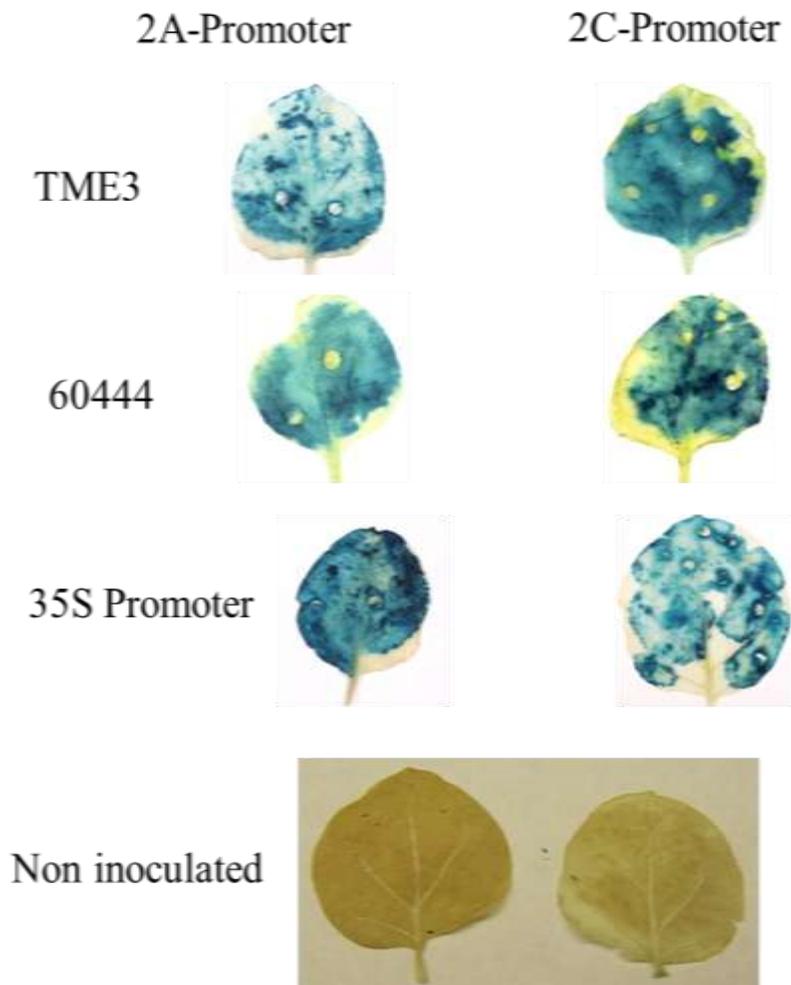


Figure 7. Transient expression of GUS reporter gene driven by endogenous promoter for native R genes in cassava when compared with 35S promoter.

found in both eukaryotes and prokaryotes that plays a central role in DNA packaging (Hargreaves and Crabtree, 2011). Conversely, marker SSRY28 is derived from cassava transcript Manes.12G074000.1 that codes for PREPHENATE DEHYDRATASE (P PROTEIN) involved in phenylalanine and tyrosine biosynthesis (Yoo et al., 2013; Maeda et al., 2011). CMD3 flanking marker NS198 is derived from cassava transcript Manes.12G016800.1 coding for protein phosphatase 2A (PP2A) involved in cell cycle control (Richard and Elder, 2005). The RME1 marker was shown to code for the partial motif of a disease resistance gene and mapped to several genomic regions. All three genes carrying sequences of SSR markers are key in the regulation of plant growth and development. As a result, the markers developed from such genes may not be sufficiently polymorphic to allow distinction between CMD susceptible and resistant genotypes. Single-strand conformation polymorphism (SSCP) of SSR markers on PAGE gel (Figure 1B)

showed potential nucleotide differences in diverse cassava genotypes. Sequence analysis of SSRY28 revealed four nucleotide deletion at nucleotide position 114 in CMD resistant genotype TME3 but not in CMD susceptible genotype 60444.

In plants, it has been demonstrated that disease resistant genes are mostly located as clusters of homologous or heterologous sequences, and in some cases occur as both (Leister, 2004). Global analysis of the BAC clone sequences for the presence of disease resistance genes revealed NBS domains in each of the five BAC clones containing the RME1 fragment that mapped to Chromosome 12 (Table 3). For BAC29, a tandem repeat of NBS-LRR were mapped on chromosome 16 of cassava reference genome version 6. In the same chromosomal region, a cluster of 14 NBS-LRR genes were identified (Supplementary Figure 2). These NBS domains were homologous to NBS described in cassava genotypes TME3, TME117 and wild relative

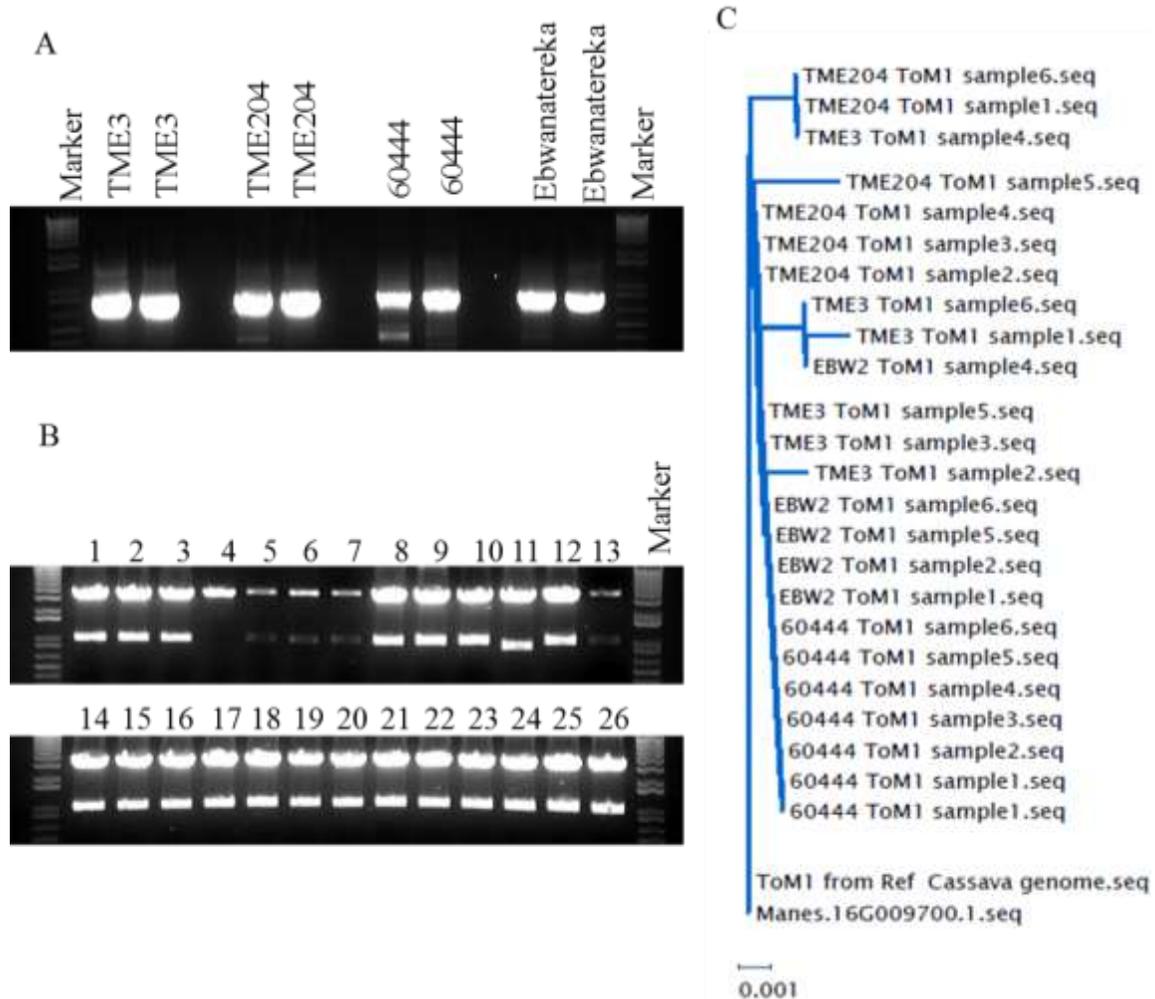


Figure 8. Tobamovirus multiplication protein 1 (TOM1) homolog in cassava. (A) RT-PCR amplification of TOM1 gene from cassava genotypes. (B) *SacI* and *XbaI* restriction digestion of TOM1 like genes from cassava CVs TME3, TME204, 60444 and EBW2 cloned into PCR Blunt II TOPO vector. Lanes: 1-6: TME3; 7-12: TME204; 13-20: 60444; 21-26: Ebwanatereka.

M. euprinosa and *M. brachyandra* (Gedil et al., 2012). The presence of numerous R genes in BAC29 corroborate with previous reports that indicate that R genes are usually present as clusters resulting from evolutionary pressure to counter changes in pathogen virulence proteins (Ashfield et al., 2012). In the same locus, the TOM1 homolog was identified, indicating a genomic region with high density of defense-related genes. Clusters of genes effective against bacteria, fungus, viruses and oomycetes have been mapped within the same vicinity on soybean chromosome 13 (Ashfield et al., 2012; Innes et al., 2008; Sandhu et al., 2005) and in chromosome 1 and chromosome 2 in lettuce (Christopoulou et al., 2015; McHale et al., 2009).

The number of disease resistant genes present in plant genomes varies greatly across species. For example, 143 R genes have been characterized in coconut (Puch-

Hau et al., 2015) whereas cassava proteome has been reported to encode 28 Toll/interleukin-1 receptor Nucleotide-binding domain shared by apoptotic protease activating factor 1 (TIR-NB-ARC-LRR), 177 non-TIR-NB-ARC-LRR putative proteins, two with TIR-LRR domains (Soto et al., 2015). Other studies have identified 228 NBS-LRR genes present in the cassava reference genome, grouped into two subfamilies; 181 CC-domain-containing (CNL) and 47 TIR-domain-containing (TNL) (Lozano et al., 2015). A diverse cluster of 99 genes highly homologous to full-sized NBS-LRR genes have been identified in the reference genome encoding none, or only a small part, of the NBS domain (Lozano et al., 2015). The full length RME1 marker sequences are homologous to cassava genes encoding for NB-ARC (Figure 1). It can therefore be hypothesized that the RME1 marker is homologous to many endogenous disease resistant

genes in the cassava genome while the second RME1 allele (shorter 500 nt) would fit the truncated version of NB-ARC gene.

TME3 BAC library hybridization with RME1 fragment revealed 89 BAC clones positive for RME1 compared with 41 BAC clones for the three SSR markers, confirming a high number of R genes in cassava (Table 2). The RME1 full length sequence hits several places in cassava genome indicating that it is a highly repetitive sequence. This concurs with the fact that NBS-LRR are encoded by one of the largest gene families in plants carrying highly conserved NBS motifs (McHale et al., 2006). NBS-LRR motifs are in general conserved across many genes with pathogen resistance function (Hammond-Kosack and Parker, 2003; Glowacki et al., 2011).

Three full length and two truncated putative R genes occurring as tandem repeats were found in BAC29 sequences (Figure 4). This is in agreement with the recent report by Soto et al. (2015) who described the presence of clusters of full length and shorter versions of genes encoding NB-ARC-LRR in the cassava reference genome. Annotation of BAC29 and five other BAC clones mapping to the CMD2 locus genomic region and carrying RME1 marker sequences, revealed that putative R genes were localized in genomic region rich in transposable elements indicating a region undergoing fast evolution. This result fits with the model of an “arms-race” between pathogens and disease resistant genes that drives selective pressure to evolve R genes with specificity to new pathogen virulence proteins (Christopoulou et al., 2015). The truncated R genes identified in BAC29 showed a large deletion at three prime end (CMD2D) and a large insertion (7 kb) on CMD2e at nucleotide position 2237, indicating potential pseudogenes. Large deletions and insertions leading to frameshifts have been associated with condensed mRNAs and proteins. In most cases, disease resistant genes coding for partial NBS domain are regarded as pseudogenes (Luo et al., 2012). Truncated NBS-LRR with non-coding capacity have been located in genomic regions as clusters adjacent full length functional NBS-LRR in *Lotus japonicus*, *Medicago truncatula* and potato (Li et al., 2010; Lozano et al., 2012; Ameline-Torregrosa et al., 2008).

Cassava disease resistance genes have not been widely reported. Sequence analysis of R genes from different cassava genotypes revealed a high percentage of nucleotide identity indicating high levels of conservation (Figure 5B). Importantly, however, this was without association with presence or absence of functional CMD resistance loci. Resistance to CMD is governed by: single dominant gene(s) found within the CMD2 locus (Rabbi et al., 2014; Akano et al., 2002), multigenic recessive resistance harbored by the CMD1 loci (Fregene and Puonti-Kaerlas, 2002) or synergism between CMD2 and an additional locus referred to as CMD3 (Okogbenin et al., 2012). Therefore, the identified

NBS-LRR may not be involved in CMD resistance. Biolistic inoculation of cassava genotypes with cassava-infecting geminiviruses indicates that CMD resistance lacks the typical R gene mediated hypersensitive response (Patil and Fauquet, 2015). Studies by Louis and Rey (2015) identified resistance gene analogs (RGAs) encoding for resistance protein analogs (RPAs) that were uniquely overexpressed in TME3 recovering from South African cassava mosaic virus (SACMV) infection. However, none of the identified RGAs in TME3 localized with the CMD2 locus.

The cassava reference genome version 6.1 contains NBS-LRR clusters on chromosome 16 that are homologous to BAC29 and cassava isolated R genes. Northern blot analysis performed here indicated constitutive expression of BAC29 derived R gene in all diverse cassava genotypes (Figure 6). This may indicate their involvement in plant developmental functions while their role of antiviral defense remains to be demonstrated.

Through BAC end sequencing, 13 BAC clone libraries were constructed spanning regions around the CMD2 locus using the RME1 and BAC33b markers (CIAT, 2007). Full length BAC clone sequencing and assembly enabled construction of contigs of 100 kb in two independent clones while the shortest BAC contig assembled was 4.0 kb. However, these results indicated that some of the clones, for example BAC29, 52H02 and 52H13 were mapping to different genomic regions that have not been associated with CMD2 resistance. Approximately, 60% of reads derived from BACs such as 003H20, BAC33, BAC31 and BAC38 mapped to the *E. coli* genome, with the remaining 40% either not homologous to any cassava genomic region, or assembled poorly into a few short 1 – 2 kb contigs scattered throughout the genome. Other BACs contained sequence reads that did not map to cloning vector PIndigoBac536 BAC and had no hits within the cassava genome. Additional constraints encountered in using the TME3 BAC clone library were that detailed description of this library, such as position on physical map of cassava, BAC end sequences, and maps to physical clones in the freezer collection seems to be missing. These discrepancies observed in TME3 BACs studies here may have resulted from a mix-up of clones during library construction or processing. Another possibility that cannot be ruled out is low quality of the TME3 BAC libraries. Hybridization of markers to entire TME3 BAC libraries only provided clones originating from the marker region and did not form contiguous sequences that traversed through the CMD2 genomic region. Presence of *E. coli* sequences would also indicate contamination and poor quality of TME3 BAC clones library. This study therefore raises serious questions regarding the reliability of this TME3 BACs collection and the technical challenges of using it for identification of sequences such as those coding for CMD2 gene(s).

Conclusion

In this study, a BAC library was used in an attempt to study and identify genes residing within the CMD2 locus. CMD2 locus was mapped with four markers that did not show polymorphism in different cassava genotypes. Although, BAC libraries have been successfully utilized for map based cloning in other species, the present study was not able to identify clones spanning across the CMD2 locus, despite utilizing contig maps and markers previously reported. Sequences from some of the libraries reported in CMD2 region were found to be located in different chromosomal regions, certainly TME3 BAC library appears to be of insufficient quality at the CMD2 locus to be useful for identifying genes involved in this CMD resistance mechanism. It is likely that it would also not be a functional useful tool for discovery of other genes of importance in cassava. With regards to CMD, the situation is further compounded by the fact that the Cassava genome V6.1 contains gaps in CMD2 region making it difficult to design probes for screening potential BACs for chromosomal walking. Efforts should therefore focus on developing polymorphic molecular markers closely linked with the CMD2 locus that can easily be utilized to screen for CMD resistance. In addition, there is need for whole genome sequencing of CMD2 type cassava to identify sequences coding for genes harbored in CMD2 region. Transcriptome data should also be generated from CMD2 cassava genotypes to unravel the mode of action of CMD2 gene(s) and functional validation was performed through gene knock down or overexpression studies. New BAC libraries should be created from the same DNA source and be highly valuable is allowing accurate cloning of the gene(s) involved in CMD2 resistance.

Conflict of interests

The authors have not declared any conflict of interests.

ACKNOWLEDGEMENTS

The authors are grateful for financial support from Bill and Melinda Gates Foundation. They would like to thank Dr Luis Augusto and Dr Michael Atkins for providing BAC clones and Dr Steve Rounsley for sequencing.

REFERENCES

Acharya B, Lee S, Mian MR, Jun TH, McHale LK, Michel AP, Dorrance AE (2015). Identification and mapping of quantitative trait loci (QTL) conferring resistance to *Fusarium graminearum* from soybean PI 567301B. *Theor. Appl. Genet.* 128(5):827-838.

Akano A, Dixon A, Mba C, Barrera E, Fregene M (2002). Genetic mapping of a dominant gene conferring resistance to cassava mosaic disease. *Theor. Appl. Genet.* 105(4):521-525.

Akinbo O, Labuschagne M, Fregene M (2012). Increased storage

protein from interspecific F1 hybrids between cassava (*Manihot esculenta* Crantz) and its wild progenitor (*M. esculenta* ssp. *flabellifolia*). *Euphytica* 185(2):303-311.

Ameline-Torregrosa C, Wang BB, O'Brien MS, Deshpande S, Zhu H, Roe B, Young ND, Cannon SB (2008). Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant *Medicago truncatula*. *Plant Physiol.* 146(1):5-21.

Asare PA, Galyuon IKA, Asare-Bediako E, Sarfo JK, Tetteh JP (2014). Phenotypic and molecular screening of cassava (*Manihot esculenta* Crantz) genotypes for resistance to cassava mosaic disease. *J. Gen. Mol. Virol.* 6(2), 6-18.

Ashfield T, Egan AN, Pfeil BE, Chen NW, Podicheti R, Ratnaparkhe MB, Ameline-Torregrosa C, Denny R, Cannon S, Doyle JJ, Geffroy V, Roe BA, Saghai Maroof MA, Young ND, Innes RW (2012). Evolution of a complex disease resistance gene cluster in diploid Phaseolus and tetraploid Glycine. *Plant Physiol.* 159(1):336-354.

Beyene G, Chauhan RD, Wagaba H, Moll T, Alicai T, Miano D, Carrington J, Taylor NJ (2015). Loss of CMD2-mediated resistance to cassava mosaic disease in plants regenerated through somatic embryogenesis. *Mol. Plant Pathol.*

Bi H, Aileni M, Zhang P (2010). Evaluation of cassava varieties for cassava mosaic disease resistance jointly by agro-inoculation screening and molecular markers. *Afr. J. Plant Sci.* 4(9):330-338.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7(5):335-336.

Chauhan RD, Beyene G, Kalyaeva M, Fauquet CM, Taylor N (2015). Improvements in Agrobacterium-mediated transformation of cassava (*Manihot esculenta* Crantz) for large-scale production of transgenic plants. *Plant Cell Tiss. Organ Cult.* 121(3):591-603.

Christopoulou M, McHale L, Kozik A, Reyes-Chin Wo S, Wroblewski T, Michelmore RW (2015). Dissection of Two Complex Clusters of Resistance Genes in Lettuce (*Lactuca sativa*). *Mol. Plant Microbe Interact.* 28(7):751-765

CIAT (2006). Annual Report IP3. Improved cassava for the developing world, CIAT, Cali, Colombia.

CIAT (2007). Annual Report IP3. Improved cassava for the developing world, CIAT, Cali, Colombia.

Fregene M, Puonti-Kaerlas J (2002). Cassava biotechnology. *Cassava: Biol. Prod. Util.* 10:179-207.

Gaur R, Jeena Gaur G, Shah N, Gupta S, Pradhan S, Tyagi AK, Jain M, Chattopadhyay D, Bhatia S (2015). High density linkage mapping of genomic and transcriptomic SNPs for synteny analysis and anchoring the genome sequence of chickpea. *Sci. Rep.* P 5.

Gedil M, Kumar M, Igwe D (2012). Isolation and characterization of resistant gene analogs in cassava, wild Manihot species, and castor bean (*Ricinus communis*). *Afr. J. Biotechnol.* 11(85):15111-15123.

Glowacki S, Macioszek V, Kononowicz A (2011). R proteins as fundamentals of plant innate immunity. *Cell. Mol. Biol. Lett.* 16(1):1-24.

Hammond-Kosack KE, Parker JE (2003). Deciphering plant-pathogen communication: fresh perspectives for molecular resistance breeding. *Curr. Opin. Biotechnol.* 14(2):177-193.

Hargreaves DC, Crabtree GR (2011). ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res.* 21(3):396-420.

Harimalala M, Chiroleu F, Giraud-Carrier C, Hoareau M, Zinga I, Randriamampianina JA, Velombola S, Ranomenjanahary S, Andrianjaka A, Reynaud B, Lefeuvre P, Lett JM (2015). Molecular epidemiology of cassava mosaic disease in Madagascar. *Plant Pathol.* 64(3):501-507.

Innes RW, Ameline-Torregrosa C, Ashfield T, Cannon E, Cannon SB, Chacko B, Chen NW, Couloux A, Dalwani A, Denny R, Deshpande S, Egan AN, Glover N, Hans CS, Howell S, Ilut D, Jackson S, Lai H, Mammadov J, Del Campo SM, Metcalf M, Nguyen A, O'Brien MS, Pfeil BE, Podicheti R, Ratnaparkhe MB, Samain S, Sanders I, Séguens B, Sévignac M, Sherman-Broyles S, Thareau V, Tucker DM, Walling J, Wawrzynski A, Yi J, Doyle JJ, Geffroy V, Roe BA, Maroof MA, Young ND (2008). Differential accumulation of retroelements and diversification of NB-LRR disease resistance genes in duplicated regions following polyploidy in the ancestor of

- soybean. *Plant Physiol.* 148(4):1740-1759.
- International Barley Genome Sequencing Consortium (2012). A physical, genetic and functional sequence assembly of the barley genome. *Nature* 491(7426):711-716.
- International Cassava Genetic Map Consortium (2015). High-Resolution Linkage Map and Chromosome-Scale Genome Assembly for Cassava (*Manihot esculenta* Crantz) from 10 Populations. *G3 (Bethesda)* 5(1):133-144.
- Ishibashi K, Nishikiori M, Ishikawa M (2010). Interactions between tobamovirus replication proteins and cellular factors: Their impacts on virus multiplication. *Mol. Plant Microbe Interact.* 23(11):1413-1419.
- Jørgensen K, Bak S, Busk PK, Sørensen C, Olsen CE, Puonti-Kaerlas J, Møller BL (2005). Cassava plants with a depleted cyanogenic glucoside content in leaves and tubers. Distribution of cyanogenic glucosides, their site of synthesis and transport, and blockage of the biosynthesis by RNA interference technology. *Plant Physiol.* 139(1): 363-374.
- Kaweesi T, Kawuki R, Kyaligonza V, Baguma Y, Tusiime G, Ferguson ME (2014). Field evaluation of selected cassava genotypes for cassava brown streak disease based on symptom expression and virus load. *Virology* 11(1):216.
- Legg JP (1999). Emergence, spread and strategies for controlling the pandemic of cassava mosaic virus disease in east and central Africa. *Crop Prot.* 18(10):627-637.
- Legg JP, Jeremiah SC, Obiero HM, Maruthi MN, Ndyetabula I, Okao-Okuja G, Bouwmeester H, Bigirimana S, Tata-Hangy W, Gashaka G, Mkamilo G, Alicai T, Kumar PL (2011). Comparing the regional epidemiology of the cassava mosaic and cassava brown streak virus pandemics in Africa. *Virus Res.* 159(2):161-170.
- Legg JP, Kumar PL, Makeshkumar T, Tripathi L, Ferguson M, Kanju E, Ntawuruhunga P, Cuellar W (2015). Cassava virus diseases: Biology, epidemiology, and management. *Adv. Virus Res.* 91:85-142.
- Leister D (2004). Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. *Trends Genet.* 20(3):116-122.
- Li X, Cheng Y, Ma W, Zhao Y, Jiang H, Zhang M (2010). Identification and characterization of NBS-encoding disease resistance genes in *Lotus japonicus*. *Plant Syst. Evol.* 289(1-2):101-110.
- Lokko Y, Danquah EY, Offei SK, Dixon AG, Gedil MA (2005). Molecular markers associated with a new source of resistance to the cassava mosaic disease. *Afr. J. Biotechnol.* 4(9):873-881.
- Louis B, Rey C (2015). Resistance gene analogs involved in tolerant cassava-geminivirus interaction that shows a recovery phenotype. *Virus Genes* 51(3):393-407.
- Lozano R, Hamblin MT, Prochnik S, Jannink JL (2015). Identification and distribution of the NBS-LRR gene family in the Cassava genome. *BMC Genomics* 16(1):360.
- Lozano R, Ponce O, Ramirez M, Mostajo N, Orjeda G (2012). Genome-wide identification and mapping of NBS-encoding resistance genes in *Solanum tuberosum* group phureja. *PLoS One* 7(4):e34775.
- Luo S, Zhang Y, Hu Q, Chen J, Li K, Lu C, Liu H, Wang W, Kuang H (2012). Dynamic nucleotide-binding site and leucine-rich repeat-encoding genes in the grass family. *Plant Physiol.* 159(1):197-210.
- Maeda H, Yoo H, Dudareva N (2011). Prephenate aminotransferase directs plant phenylalanine biosynthesis via arogenate. *Nat. Chem.* 7(1):19-21.
- Martin M (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* 17(1):10.
- McHale L, Tan X, Koehl P, Michelmore RW (2006). Plant NBS-LRR proteins: Adaptable guards. *Genome Biol.* 7(4):212.
- McHale LK, Truco MJ, Kozik A, Wroblewski T, Ochoa OE, Lahre KA, Knapp SJ, Michelmore RW (2009). The genomic architecture of disease resistance in lettuce. *Theor. Appl. Genet.* 118(3):565-580.
- Okogbenin E, Egesi CN, Olanami B, Ogundapo O, Kahya S, Hurtado P, Marin J, Akinbo O, Mba C, Gomez H, de Vicente C, Baiyeri S, Uguru M, Ewa F, Fregene M (2012). Molecular marker analysis and validation of resistance to cassava mosaic disease in elite cassava genotypes in Nigeria. *Crop Sci.* 52(6):2576-2586.
- Okogbenin E, Moreno I, Tomkins J, Fauquet, CM, Mkamilo G, Fregene M (2013). Marker-Assisted Breeding for Cassava Mosaic Disease Resistance. *Translat. Genom. Crop Breed.: Biotic Stress* 1:291-325.
- Okogbenin E, Porto MCM, Egesi C, Mba C, Espinosa E, Santos LG, Ospina C, Marín J, Barrera E, Gutiérrez J, Ekanayake I, Iglesia C, Fregene MA. (2007). Marker-assisted introgression of resistance to cassava mosaic disease into Latin American germplasm for the genetic improvement of cassava in Africa. *Crop Sci.* 47(5):1895-1904.
- Parkes E, Fregene M, Dixon A, Okogbenin E, Boakye-Peprah B, Labuschagne M (2015). Developing Cassava Mosaic Disease resistant cassava varieties in Ghana using a marker assisted selection approach. *Euphytica* 203(3):549-556.
- Patil BL, Fauquet CM (2009). Cassava mosaic geminiviruses: Actual knowledge and perspectives. *Mol. Plant Pathol.* 10(5):685-701.
- Patil BL, Fauquet CM (2015). Studies on differential behavior of cassava mosaic geminivirus DNA components, symptom recovery patterns, and their siRNA profiles. *Virus Genes* 50(3):474-486.
- Prochnik S, Marri PR, Desany B, Rabinowicz PD, Kodira C, Mohiuddin M, Rodriguez F, Fauquet C, Tohme J, Harkins T, Rokhsa DS, Rounsley S (2012). The cassava genome: current progress, future directions. *Trop. Plant Biol.* 5(1):88-94.
- Puch-Hau C, Oropeza-Salín C, Peraza-Echeverría S, Gongora-Paredes M, Córdova-Lara I, Narvaez-Cab M, Zizumbo-Villareal D, Sáenz-Carbonell L (2015). Molecular cloning and characterization of disease-resistance gene candidates of the nucleotide binding site (NBS) type from *Cocos nucifera* L. *Physiol. Mol. Plant Pathol.* 89:87-96.
- Rabbi IY, Hamblin MT, Kumar PL, Gedil MA, Ikpan AS, Jannink JL, Kulakow PA (2014). 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.
- Ragupathy R, Rathinavelu R, Cloutier S (2011). Physical mapping and BAC-end sequence analysis provide initial insights into the flax (*Linum usitatissimum* L.) genome. *BMC Genom.* 12(1):217.
- Richard YZ, T Elder R (2005). Viral infections and cell cycle G2/M regulation. *Cell Res.* 15(3):143-149.
- Sandhu D, Schallack KG, Rivera-Velez N, Lundeen P, Cianzio S, Bhattacharyya MK (2005). Soybean Phytophthora resistance gene Rps8 maps closely to the Rps3 region. *J. Hered.* 96(5):536-541.
- Shi X, Zeng H, Xue Y, Luo M (2011). A pair of new BAC and BIBAC vectors that facilitate BAC/BIBAC library construction and intact large genomic DNA insert exchange. *Plant Methods* 7(1):33.
- Siritunga D, Arias-Garzon D, White W, Sayre RT (2004). Over-expression of hydroxynitrile lyase in transgenic cassava roots accelerates cyanogenesis and food detoxification. *Plant Biotechnol. J.* 2(1):37-43.
- Soto JC, Ortiz JF, Perlaza-Jiménez L, Vásquez AX, Becerra LA, Mathew B, León J, Bernal AJ, Ballvora A, López CE (2015). A genetic map of cassava (*Manihot esculenta* Crantz) with integrated physical mapping of immunity-related genes. *BMC Genom.* 16(1):190.
- Taylor N, Gaitan-Solis E, Moll T, Trauterman B, Jones T, Pranjal A, Trembley C, Abernathy V, Corbin D, Fauquet CM (2012). A high-throughput platform for the production and analysis of transgenic cassava (*Manihot esculenta*) plants. *Trop. Plant Biol.* 5(1):127-139.
- Tomkins J, Fregene M, Main D, Kim H, Wing R, Tohme J (2004). Bacterial artificial chromosome (BAC) library resource for positional cloning of pest and disease resistance genes in cassava (*Manihot esculenta* Crantz). *Plant Mol. Biol.* 56(4):555-561.
- Vanderschuren H, Moreno I, Anjanappa RB, Zainuddin IM, Gruijssem W (2012). Exploiting the combination of natural and genetically engineered resistance to cassava mosaic and cassava brown streak viruses impacting cassava production in Africa. *PLoS One* 7(9):e45277.
- Vezi F, Del Fabbro C, Tomescu AI, Policriti A (2012). rNA: a fast and accurate short reads numerical aligner. *Bioinformatics* 28(1):123-124.
- Wang W, Feng B, Xiao J, Xia Z, Zhou X, Li P, Zhang W, Wang Y, Møller BL, Zhang P, Luo MC, Xiao G, Liu J, Yang J, Chen S, Rabinowicz PD, Chen X, Zhang HB, Ceballos H, Lou Q, Zou M, Carvalho LJ, Zeng C, Xia J, Sun S, Fu Y, Wang H, Lu C, Ruan M, Zhou S, Wu Z, Liu H, Kannangara RM, Jørgensen K, Neale RL, Bonde M, Heinz N, Zhu W, Wang S, Zhang Y, Pan K, Wen M, Ma PA, Li Z, Hu M, Liao W, Hu W, Zhang S, Pei J, Guo A, Guo J, Zhang J, Zhang Z, Ye J, Ou W, Ma Y, Liu X, Tallon LJ, Galens K, Ott S, Huang J, Xue J, An F, Yao Q, Lu X, Fregene M, López-Lavalle LA,

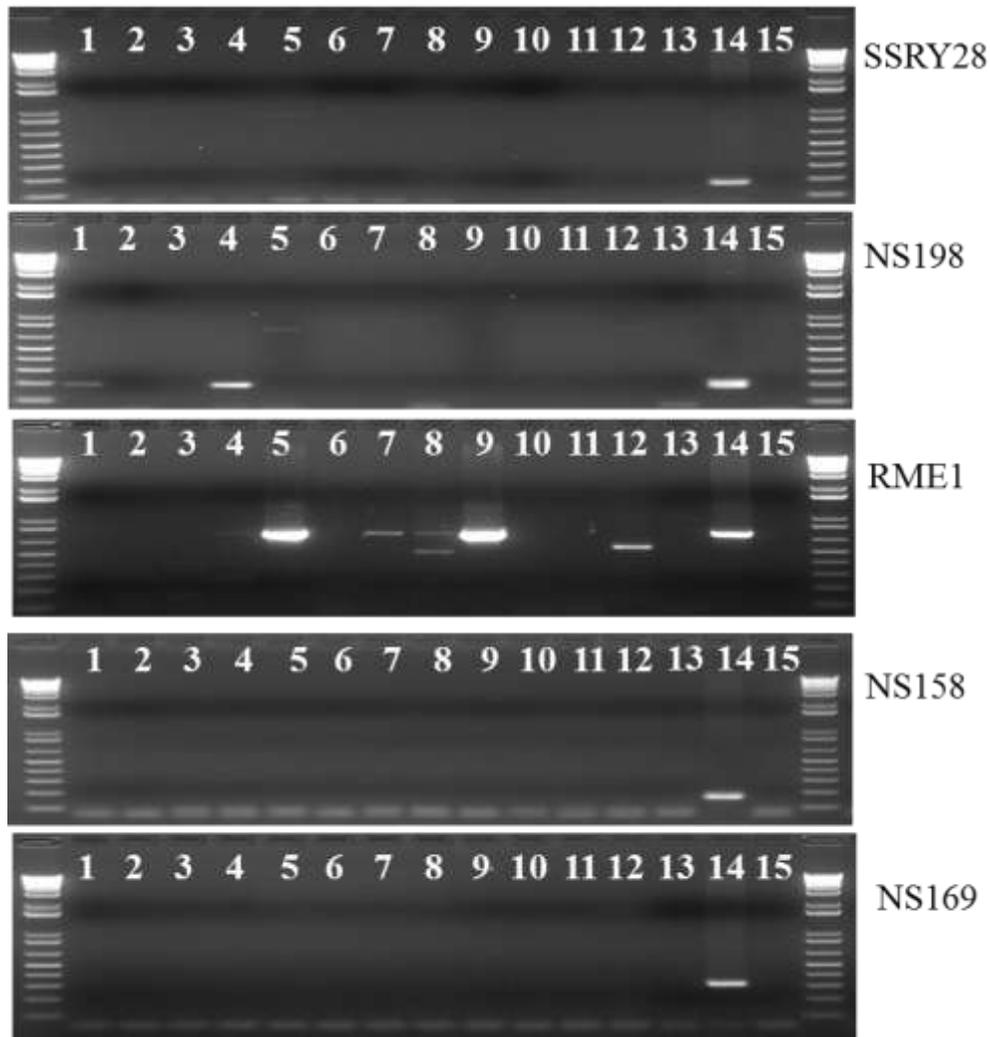
- Wu J, You FM, Chen M, Hu S, Wu G, Zhong S, Ling P, Chen Y, Wang Q, Liu G, Liu B, Li K, Peng M (2014). Cassava genome from a wild ancestor to cultivated varieties. *Nat. Commun.* 5:5110.
- Welsch R, Arango J, Bär C, Salazar B, Al-Babili S, Beltrán J, Chavarriaga P, Ceballos H, Tohme J, Beyer P (2010). Provitamin A accumulation in cassava (*Manihot esculenta*) roots driven by a single nucleotide polymorphism in a phytoene synthase gene. *Plant Cell* 22(10):3348-3356.
- Wroblewski T, Tomczak A, Michelmore R (2005). Optimization of Agrobacterium-mediated transient assays of gene expression in lettuce, tomato and Arabidopsis. *Plant Biotechnol. J.* 3(2):259-273.
- Yoo H, Widhalm JR, Qian Y, Maeda H, Cooper BR, Jannasch AS, Gonda I, Lewinsohn E, Rhodes D, Dudareva N (2013). An alternative pathway contributes to phenylalanine biosynthesis in plants via a cytosolic tyrosine: Phenylpyruvate aminotransferase. *Nat. Commun.* P 4.
- Zhang HB, Scheuring CF, Zhang M, Zhang Y, Wu CC, Dong JJ, Li Y (2012). Construction of BIBAC and BAC libraries from a variety of organisms for advanced genomics research. *Nat. Protoc.* 7(3):479-499.
- Zhang J, Zhang Q, Cheng T, Yang W, Pan H, Zhong J, Huang L, Liu E (2015). High-density genetic map construction and identification of a locus controlling weeping trait in an ornamental woody plant (*Prunus mume* Sieb. et Zucc). *DNA Res.* 22(3):183-191.

Supplementary Table 1. List of primers used in this study.

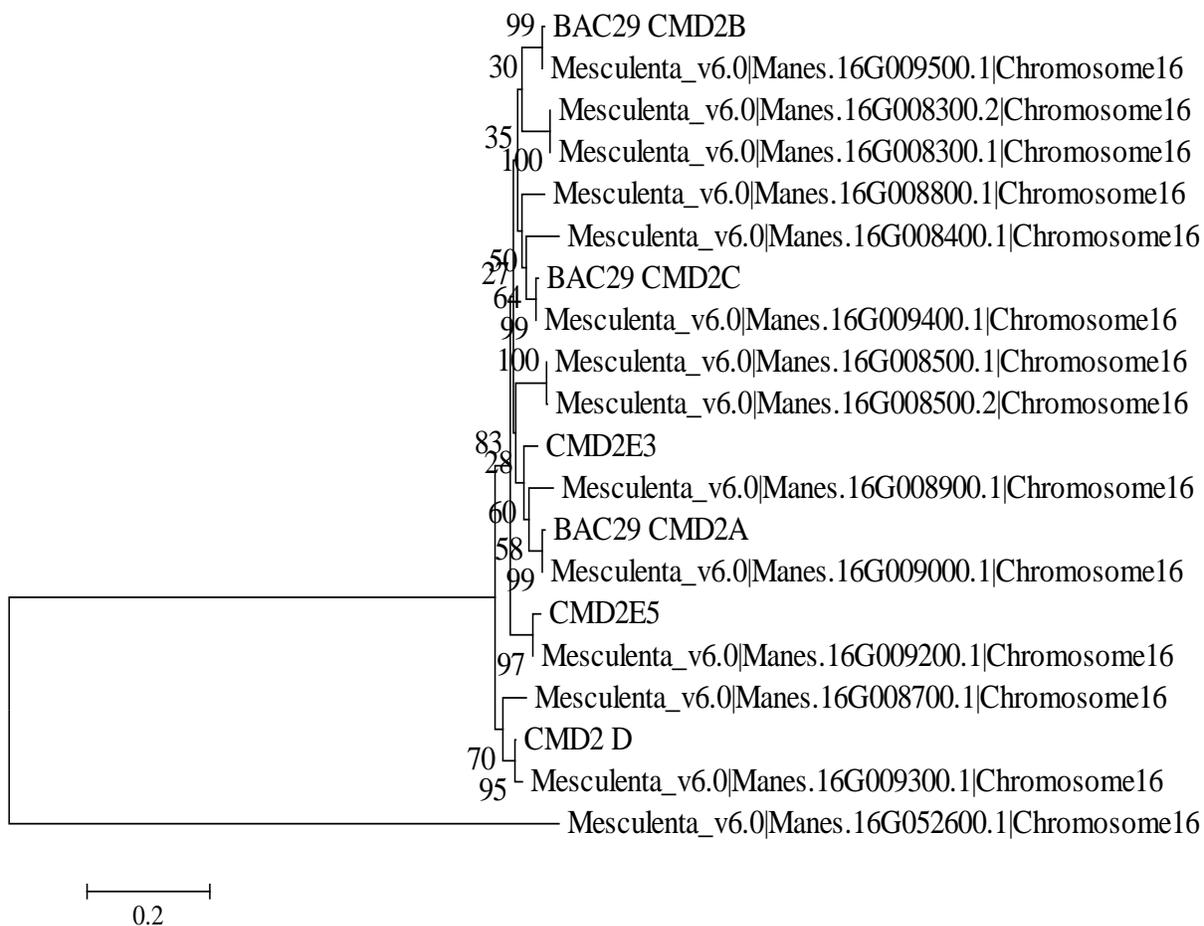
Primer name	Sequence (5' 3')
CMB2A R - KpnI	5' GGTACCTTTAATAGTACTAGATATGGCAGCACGC 3'
CMD2A F - SacI	5' GAGCTCTTCATTTAGTGCAAAATTC AAGTAATC 3'
CMD2B R2-SacI	5' GAGCTCCCATGATGCTATTTGCTGTCTCTTGCTGC 3'
CMD2B F2-XmaI	5' CCCGGGGAGCCAAGCATGATTTTACCTTTTCGTGG 3'
CMD2C F - SacI	5' CGAGCTCAACTCAAAATTGAGTGATTCTAGCCG 3'
CMD2C R - KpnI	5' GGTACCATAAAGTATTCCATGCTGCCCATTTG 3'
CMD2A-prom-Sac II F	5' GACCGCGGAAGCTTCCATGTGCCTAAATTATTTAT 3'
CMD2A-prom- BamMHI R	5' GAGGATCCACCATGTCTGTGGACTTAACTAGTT 3'
CMD2B P SAC II - ASCI F- 1	5' GACCGCGGCGCGCCGAGAGGTACATTCATAATCCAATCCCC 3'
CMD2B P BamMHI R - 2	5' GAGGATCCCCACGAAAGGTAATCATGCTTGGCTCTTCATC 3'
CMD2C P SAC II - ASCI F-1	5' GACCGCGGCGCGCCGTGCCTTGTATAGGAGTTAGCATTG 3'
CMD2C P BamHI R - 2	5' GAGGGATCCCGCTAGAATCACTCAATTTTGAGTTATG 3'
CMD2A ORF F XbaI	5' GATCTAGAATGGATGTTGTGACTTGTATCGC 3'
CMD2A ORF R BamHI	5' CTCGGATCCTTAAAATGATAAAACGGTACGACTTCCTC 3'
CMD2B ORF F XbaI	5' GATCTAGAATGGATGTTGTGACTTGTATCGCTGG 3'
CMD2B ORF R BamHI	5' CTCGGATCCCAATTAATAATGCTAAAACGATATGTACCTCCTC 3'
CMD2C ORF F XbaI	5' GCTCTAGAATGGATGTTGTGACTTGTATCGCTGG 3'
CMD2C ORF R BamHI	5' CTCGGATCCCACAATTAATAATGCTAAAACGATACGTACCTC 3'
CMD2D F	5' CCACATCCAGTAACCTTAGATTAGTCAACTTCCC 3'
CMD2D R	5' GTTAAGTCCACGGACATGGATGTTGTGAGTTG 3'
CMD2.e.5 F	5' CCTGGGTGTGCCTACAGCACAAAGTTGG 3'
CMD2.e.5 R	5' GTCCACAGACATGGATGTTGTGAC 3'
CMD2.e.3 F	5' TGAGAGAAGCAGCAAATGTC 3'
CMD2.e.3 R	5' GGGCTGGGCCTAACACTAGTGATTGTC 3'
CMD2.e.5 F1	5' GCTGATAGCCAGAGACAGTGATTCTCAGCAAG 3'
CMD2.e.5 R1	5' AATAGAGGGATTGACAACCTGAAAACATC 3'
CMD2abc 950R	5' AAGTCAGAGACTCAATCTCAGCGCCT 3'
CMD2abc 943F	5' AGAATGCAGGCGCTGAGATTGAGT 3'
CMD2a 1912R	5' TGGAAGAAGAGTCGCCATCAAGAG 3'
CMD2abc 1801F	5' GGGAGCTGAAGATGCTTGAATTCT 3'
CMD2ab 2753R	5' TGTAGCTTGGCAGATGCAACAAC 3'
CMD2a 2207F	5' AAGCCGGAGACGCCTTAGAGTAAA 3'
CMD2a 3177R	5' CGGTACGTACTTCTCTTTGTTGTG 3'
CMD2c 1958R	5' CGGGACAGAGTAGACATAGCATTG 3'
CMD2c 2870R	5' CCCTTTCTGCAACAATTGCTTCTAAT 3'
CMD2bc 2500F	5' ATGCTTTGTTGCCGGAAGTGAAG 3'
CMD2C Promoter 253F	5' CGTTCATAATCCACTGCCCTTCTCTTATTG 3'
CMD2C Promoter 1666R	5' CATGACATGGTTAACAGGCTGGAGTC 3'

Supplementary Table 2. Description of TME3 BAC clones hybridized with markers and tested with PCR.

Clone Number	Hybridized Marker	Markers associated with CMD2 locus			
		RME1	SSRY28	NS158	NS169
003H20	NS158/NS169/SSRY28	Positive	Negative	Positive	Negative
026L16	NS158/SSRY28	Positive	Negative	Positive	Negative
021C10	NS169/SSRY28	Positive	Negative	Negative	Negative
026N16	NS158/NS169/SSRY28	Positive	Negative	Positive	Negative
033J15	NS169/SSRY28	Positive	Negative	Positive	Negative
082G13	NS158/SSRY28	Negative	Negative	Negative	Negative
109O10	NS158/SSRY28	Negative	Negative	Negative	Negative
110N21	NS158/NS169/SSRY28	Negative	Negative	Positive	Negative
126O21	NS169/SSRY28	Positive	Negative	Negative	Negative
116N24	NS158/NS169/SSRY28	Positive	Negative	Positive	Negative
017N21	RME-1	Positive	Negative	Positive	Negative
022C04	RME-1	Negative	Negative	Negative	Negative
027O15	RME-1	Positive	Negative	Negative	Negative
092L05	RME-1	Positive	Negative	Positive	Negative
098J19	RME-1	Positive	Negative	Negative	Negative
102O06	RME-1	Positive	Negative	Positive	Negative
130D18	RME-1	Positive	Positive	Positive	Negative
136J08	RME-1	Positive	Positive	Positive	Negative
137O22	RME-1	Positive	Negative	Positive	Negative
146I10	RME-1	Positive	Negative	Positive	Negative
159C12	RME-1	Positive	Negative	Positive	Negative



Supplementary Figure 1. PCR analysis of thirteen BAC clones spanning across CMD locus region based on markers RME1 and BAC33b located on either side of the CMD2 locus using CMD2 mapping markers. The clones were represented as follows: 1, BAC12; 2, BAC16; 3, BAC21; 4, BAC22; 5, BAC23; 6, BAC26; 7, BAC29; 8, BAC31; 9, BAC33; 10, BAC38; 11, BAC40; 12, BAC44; 13, BAC45; 14, TME3 DNA; 15, negative control.



Supplementary Figure 2. Molecular phylogenetic analysis of NBS-LRR cluster in cassava chromosome 16 showing high nucleotide identities with BAC29 derived R genes.