

## Molecular-assisted selection for resistance to cassava mosaic disease in *Manihot esculenta* Crantz

Cátia Dias do Carmo<sup>1</sup>, Maiane Suzarte da Silva<sup>1</sup>, Gilmar Alvarenga Fachardo Oliveira<sup>1</sup>, Eder Jorge de Oliveira<sup>2\*</sup>

<sup>1</sup>Federal University of Bahia Reconcavo, R. Rui Barbosa, 710, Centro – 44380-000 – Cruz das Almas, BA – Brazil.

<sup>2</sup>Embrapa Cassava and Tropical Fruits, R. da Embrapa, s/n, C.P. 007 – 44380-000 – Cruz das Almas, BA – Brazil.

\*Corresponding author <eder.oliveira@embrapa.br>

Edited by: Martin Boer

Received October 02, 2014

Accepted June 12, 2015

### Introduction

Among the major global cassava (*Manihot esculenta* Crantz) diseases, particularly in Africa, there are the viruses known as Cassava Mosaic Disease - CMD (Legg and Thresh, 2000; Legg and Fauquet, 2004). CMD is transmitted by *Bemisia tabaci* (Genn) biotype B at different stages, as well as by the use of infected propagation material and cutting tools that are commonly used in cassava cultivation (De Bruyn et al., 2012). The damage that is caused by this disease can reach 100 % of the crop (Thresh et al., 1994), with estimated production losses of approximately \$ 440 million annually in Africa (Thresh et al., 1997).

Among the main measures for controlling CMD, the use of resistant varieties is the most effective solution because it reduces both production losses that are caused by the disease and the inoculum source in crops, especially in varieties that suppress virus accumulation (Rabbi et al., 2014). The CMG (Cassava Mosaic Geminiviruses) complex that is related to CMD has not been reported in the Americas, although the insect vector is present in this region (Herrera Campo et al., 2011). Therefore, while not a current problem in Brazil, the intense exchange of propagative materials between countries and favorable conditions for the establishment of the virus in Brazil justify the targeting of research activities that are related to preventive control. However, the quarantine ban in Brazil prevents artificial inoculations, whether under controlled conditions or in the field to

ABSTRACT: The geminivirus complex known as cassava mosaic disease (CMD) is one of the most devastating viruses for cassava (*Manihot esculenta* Crantz). The aim of this study was to use molecular-assisted selection (MAS) to identify CMD-resistant accessions and ascertain promising crosses with elite Brazilian varieties. One thousand two hundred twenty-four accessions were genotyped using five molecular markers (NS169, NS158, SSRY028, SSRY040 and RME1) that were associated with resistance to CMD, along with 402 SNPs (single-nucleotide polymorphism). The promising crosses were identified using a discriminant analysis of main component (DAPC), and the matrix of genomic relationship was estimated with SNP markers. The *CMD1* gene, previously described in *M. glaziovii*, was not found in *M. esculenta*. In contrast, the *CMD2* gene was found in 5, 4 and 5 % of cassava accessions, with flanking markers NS169+RME1, NS158+RME1 and SSRY28+RME1, respectively. Only seven accessions presented all markers linked to the CMD resistance. The DAPC of the seven accessions along with 17 elite cassava varieties led to the formation of three divergent clusters. Potential sources of resistance to CMD were divided into two groups, while the elite varieties were distributed into three groups. The low estimates of the genomic relationship (ranging from -0.167 to 0.681 with an average of 0.076) contributed to the success in identifying contrasting genotypes. The use of MAS in countries where CMD is a quarantine disease constitutes a successful strategy not only for identifying the resistant accessions but also for determining the promising crosses.

Keywords: *Manihot esculenta*, breeding, disease, virus

identify CMD resistance sources. In addition, because the disease is controlled by just a few dominant effect genes, it is possible to implement marker-assisted selection (MAS) aimed at the indirect selection of resistant genotypes (Okogbenin et al., 2007).

The first markers to be associated with CMD were reported by Fregene and Puonti-Kaerlas (2002), who described one microsatellite linked to the *CMD1* gene. Subsequently, other markers were identified flanking the *CMD2* gene (Akano et al., 2002; Lokko et al., 2005; Fregene et al., 2006; Okogbenin et al., 2007). Faced with the prospect of this disease arriving and taking hold in Brazil, there is a pressing need to direct efforts towards the preventive identification of resistance sources, similar to what is being done in Africa where CMD is an endemic disease that causes significant economic loss. The aim of this study was to identify CMD resistance sources in the cassava germplasm from Latin America using molecular markers that are associated with the disease, in addition to suggesting contrasting crosses with elite varieties to generate segregating populations.

### Materials and Methods

#### Plant material

The identification of CMD-resistance sources was performed using 1,274 germplasm accessions belonging to the Cassava Germplasm Bank (CGB) (Cruz das Almas, Brazil) derived from several ecosystems in Brazil, as well as Colombia, Venezuela and Nigeria. This bank was

constructed with local and improved varieties resulting from conventional breeding procedures, such as crossing and selection, as well as the selection of local varieties with high yield potential as identified by farmers or research institutions.

The TME-14 variety from Nigeria (passport data available at <http://genbank.iita.org/>), previously reported as highly resistant to CMD, was used as a control for the presence of fragments.

### Genomic DNA

Genomic DNA was extracted according to the CTAB protocol (cetyltrimethylammonium bromide) following the methodology of Doyle and Doyle (1987). Aiming to evaluate DNA integrity and standardize its concentration, 1 % (w/v) agarose gels were stained with ethidium bromide ( $1.0 \text{ mg L}^{-1}$ ) by visual comparison with a series of DNA phage Lambda concentrations.

### Assisted molecular marker

The molecular markers that were used in MAS to identify potential resistance sources included an RME1 marker (Fregene et al., 2006), SCAR (Sequence-Characterized Amplified Region) and the four microsatellites NS158 (Fregene et al., 2006), NS169 (Okogbenin et al., 2012), SSRY28 (Akano et al., 2002) and SSRY40 (Mba et al., 2001) (Table 1), while the linkage map for the markers that were associated with *CMD2* is presented in Figure 1.

Polymerase Chain Reaction (PCR) was done in a final volume of 15  $\mu\text{L}$ , containing 10 ng DNA, Taq polymerase buffer (50 mM KCl, 10 mM Tris-HCl [pH 8.8], and 1 % Nonidet P40), 1.5 to 2.5 mM  $\text{MgCl}_2$ , 0.2 mM of each dNTP, 0.2  $\mu\text{M}$  of each primer, and 1.0 U of Taq DNA Polymerase. The amplification program consisted of one cycle at 94 °C for 5 min; 27 to 30 cycles at 94 °C for 40 s, 55, 56 or 58 °C for 40 s, 72 °C for 60 s; and a final extension at 72 °C for 5 min in a 96-well thermocycler. The amplified products were subjected to electrophoresis in 2 % w/v agarose for the RME1 markers and in 3 % w/v agarose 1000 gels for the SSRY28, SSRY40,

NS169 and NS158 markers. Electrophoresis was performed in  $0.5 \times$  TBE buffer (45 mM Tris-borate, 1.0 mM EDTA), and fragment visualization was performed using ethidium bromide ( $1.0 \text{ mg L}^{-1}$ ) and a photodocumentor. The size of the fragments was determined visually with 50-bp and 100-bp molecular weight markers.

The cassava accessions were classified as having alleles that were associated with CMD resistance (considered resistant - R) or holding other amplified alleles (considered susceptible - S) according to the presence or absence, respectively, of the disease-associated allele.

### Identification of contrasts to generate segregating populations

The accessions of germplasm-containing alleles that were associated with CMD resistance, as identified when using MAS, along with two elite hybrids (9624-09 and 98150-06), nine improved varieties (BRS Aipim Brazil, BRS Caipira, BRS Formosa, BRS Guaíra, BRS Kiriris, BRS Mulatinha, BRS Tapioqueira, BRS Verdinha and Equador 72) and six local varieties (Baianinha, Cidade Rica, Iará, Lagoa, Mani Branca and Sergipe), were analyzed with data from a previous genotyping of 402 SNP-like markers (Single Nucleotide polymorphism) that were obtained by Oliveira et al. (2014).

The SNP marker information was used for grouping the accessions that were potentially resistant to

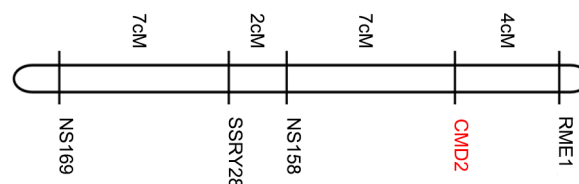


Figure 1 – Linkage map of the *CMD2*-associated markers, adapted from Akano et al. (2002), Fregene et al. (2006) and Mohan et al. (2013); The numbers above refer to the genetic distances in cM = centiMorgan.

Table 1 – List and sequence of the primers that were used in assisted selection for resistance to Cassava Mosaic Disease (CMD), together with their location on the genome and expected allele size (EAS).

Gene	Primer	Location (Scaffold)	Primer sequence	EAS (bp)
CMD1	SSRY40	06352*	TGCATCATGGTCCACTCACT CATTCTTTTTTCGGCATTCCAT	293
	NS158	06906**	GTGCGAAATGGAATCAATG TGAAATAGTGATACATGCAAAGGA	166
CMD2	NS169	06906**	GTGCGAAATGGAATCAATG GCCTTCTCAGCATATGGAGC	319
	SSRY28	05214**	TTGACATGAGTGATATTTCTTGAG GCTGCGTGCAAACTAAAAT	180
	RME1	Not located	ATGTTAATGTAATGAAAGAGC AGAAGAGGGTAGGAGTTATGT	700

\*According to Whankaew et al. (2011); \*\*According to Rabbi et al. (2014).

CMD, along with elite varieties with high yield potential, starch quality, appropriate plant size for the mechanized cultivation system, and resistance to foliar diseases, such as anthracnose and bacteriosis. This grouping aims to identify the most contrasting parental combinations that will generate segregating populations for the development of domestic varieties with the above agronomic characteristics.

The discriminant analysis of main components (DAPC) that was available in the *adegenet* package for R, version 3.0.1 (R Development Core Team 2013) was used to define the clusters of cassava hybrids and elite varieties because this technique does not require the prior definition of genetic groups (Jombart et al., 2010). Successive clustering with the K-means and the Bayesian Information Criterion (BIC) method was used to set the number of groups, where the K with the lowest BIC value is the most likely element of the data set under analysis. K values from 1 to 10 were tested, with 10 runs for each K. After defining the number of groups, the main component analysis axes that explained more than 80 % of the total variance were retained.

The matrix of genomic relatedness between the cassava genotypes was obtained according to VanRaden (2008), wherein

$$U = \frac{ZZ'}{2 \sum p_i (1 - p_i)},$$

where:  $Z = M - P$ , with  $M$  being the matrix of the markers and  $P$  the matrix for the allele frequencies as expressed by  $2(p_i - 0.5)$ . Estimates of genomic relatedness were obtained using the *kin* function from the *synbreed* package (Wimmer et al., 2012) in the R platform, version 3.0.1 (R Development Core Team, 2013). These data were used as an auxiliary strategy to indicate the most contrasting crosses.

## Results and Discussion

### MAS Implementation for CMD resistance

No *CMD1*-related allele or SSRY40-associated marker was found in the Brazilian cassava accessions. This result indicates no homology of the *CMD1* gene in the genome of *M. esculenta*, considering that this gene was originally found in *M. glaziovii* (Okogbenin et al., 2012). According to these authors, the genetic resistance that was derived from *M. glaziovii* is polygenic, with recessive inheritance and a heritability of 0.60. Hybrids resulting from a cross between *M. esculenta* and *M. glaziovii*, known as Tropical Manihot Species (TMS), have been used as a CMD resistance source in Africa (Legg and Fauquet, 2004; Thresh and Cooter, 2005), although the results of this introgression in cultivated species have not shown great success in recent years (Rabbi et al., 2014).

Due to the absence of genotypes bearing the *CMD1* gene, efforts in this work aimed at identifying accessions

bearing the *CMD2* gene. SSRY28 was the first codominant inheritance marker to be mapped next to the *CMD2* allele (9 cM) using a mapping population that originated from the cross between the local variety TMe-3 (resistant) and the hybrid TMS30555 (susceptible) (Akano et al., 2002). The TMe-3 variety is the original source of resistance to CMD, but this study used the TMe-14 variety, which is considered a duplicate of the TMe-3 source (Rabbi et al., 2014), whose genotyping results confirmed the size of the alleles as related to the *CMD2* gene for all of the markers.

An individual analysis of the presence of the alleles that were associated with *CMD2* for the NS169, NS158, SSRY28 and RME1 markers showed that 19, 16, 23, and 27 % of the cassava accessions had alleles associated with CMD resistance (Table 2), respectively. In addition, the genotypic coincidence of cassava accessions for CMD resistance was variable according to the markers used (Table 3).

The genotypic coincidence, in relation to the alleged behavior of cassava accessions, was higher (0.74) for the NS158 × NS169 and NS158 × SSRY28 markers. The efficiency of MAS depends on the proximity of the tag to the gene of interest, where distances are estimated by genetic recombination analysis. In fact, the distance between the NS158 × SSRY28 and NS158 × NS169 markers is relatively small, i.e., 2 cM (Mohan et al., 2013) and 9 cM (Fregene et al., 2006), respectively (Figure 1), which certainly contributes to providing a lower recombination rate between these markers. In addition, the NS158 and NS169 markers are anchored in

Table 2 – Number of accessions from the Active Cassava Germplasm Bank that present marks that are linked to *CMD2* gene that attributes resistance to Cassava Mosaic Disease (CMD).

Markers	Genotypes simultaneously containing resistance related alleles	
	Total	%
NS169	206	19
NS158	173	16
SSRY28	236	23
RME1	267	27
NS169 + RME1	61	5
NS158 + RME1	44	4
SSRY28 + RME1	62	5
NS169 + NS158 + RME1	9	1
NS158 + SSRY28 + RME1	16	1
NS169 + NS158 + SSRY28 + RME1	7	1

Table 3 – Coincidence of genotypes for the presence of markers that are linked to Cassava Mosaic Disease (CMD) resistance in accessions in the Active Cassava Germplasm Bank.

Markers	NS169	NS158	SSRY28
NS158	0.74		
SSRY28	0.70	0.74	
RME1	0.72	0.71	0.68

the same *scaffold* (Rabbi et al., 2014), which reinforces the physical connection between these markers in the *M. esculenta* genome.

The RME1 × SSRY28 combinations presented the lowest genotypic coincidence (0.68), while the NS169 × SSRY28, NS169 × RME1, and NS158 × RME1 combinations showed intermediate coincidences (ranging from 0.70 to 0.72) (Table 3). In addition, the greatest genetic distance between the NS169 × RME1 markers (20 cM - Figure 1) tended to result in less genotypic coincidence between these two accessions due to the possibility of historical occurrence of crossing over between these markers, but the results showed a median genotypic coincidence (Table 3).

Still considering the analysis of individual markers, Ribeiro et al. (2012) used the strategies of hybridization, marker-assisted selection and severity assessment for CMD under field conditions in the cassava genetic breeding program in Ghana. These authors demonstrated that selection based on the NS158 marker resulted in the identification of a greater number of resistant progeny compared to selection based on SSRY28 and RME1.

Although Okogbenin et al. (2007) indicated that the cassava germplasm from Latin America is quite susceptible to CMD, the results of this study indicate that the percentage of accessions with possible resistance to CMD can reach more than 20 %, depending on the marker. This result is certainly in line with the high genetic variability of the national cassava germplasm, considering that Brazil is its center of origin (Olsen and Schaal, 2001).

Additionally, greater accuracy in the use of the MAS can be obtained using markers flanking the gene of interest, considering the lower chance of recombination between the markers and the gene of interest (Collard et al., 2005; Collard and MacKill, 2008). Thus, considering the genetic distances (Figure 1) and the simultaneous selection of accessions strategy based on flanking markers, i.e., RME1 and other markers, the selection accuracy is 0.987, 0.993 and 0.994 for RME1 × NS169, RME1 × SSRY28 and RME1 × NS158, respectively (Collard and MacKill, 2008).

Using this strategy, a reduction of more than three times the number of cassava accessions containing the markers that were associated with *CMD2*, i.e., 5, 4 and 5 % of accessions that were considered resistant to the NS169 + RME1, NS158 + RME1 and SSRY28 + RME1 combinations, respectively, was observed. In addition, the percentage of accessions containing three marks was even lower, i.e., 0.71 to 1 % for NS169 + NS158 + RME1 and NS158 + SSRY28 + RME1, respectively. Finally, the presence of all four markers was observed in seven cassava accessions only (Table 2). The genotypes containing markers that were related to *CMD2* were resistant when evaluated under natural conditions in fields that were infested with the pathogen in Africa (Bi et al., 2010; Okogbenin et al., 2012; Ribeiro et al., 2012); therefore, the use of accessions with more than one marker linked to

the gene of interest can boost the selection of resistant genotypes in countries such as Brazil, where CMD is a quarantine disease.

When assessing the BC<sub>1</sub> progenies that derived from crosses with CMD resistance sources, Ribeiro et al. (2012) identified 82 % of individuals with at least one allele for RME1, SSRY28 and NS158 markers, but only 62 subjects (28 %) had three allele markers that were associated with the *CMD2* gene. Clearly, the high percentage of CMD-resistant individuals in this BC<sub>1</sub> population occurs because these populations are improved.

The identification of resistance in 2,490 genotypes using RME1 and NS158 markers and the further evaluation under field conditions in an area that is highly affected by the disease in Nigeria revealed that the accuracy of selection based on these two markers ranged from 70 to 80 %, depending on the progeny analyzed (Blair et al., 2007). In another study, Okogbenin et al. (2007) made a first introduction of the germplasm from Latin America in Africa with materials from crosses between elite parental from breeding programs and a second introduction of material that was selected based on MAS (two microsatellites and six SCARs) and a subsequent evaluation of genotypes under natural conditions in a field that was infested with CMD in Nigeria. In the first set of entries, there was a devastating effect of the disease, infecting 100 % of the plants one month after planting. However, the results of the second introduction indicated that an introgression of the *CMD2* gene in the cassava germplasm from Latin America via MAS provided high resistance under field conditions, with 67 % of the genotypes showing grades 1 to 2 (grading score in which 1 = resistant plants and 5 = highly susceptible plants). Therefore, the MAS that is used can present a high level of efficiency in the indication of cassava genotypes with proven resistance under field conditions, which is clearly due to the *CMD2* gene's dominant nature and its efficacy against a broad range of viral strains.

In fact, the introduction of the *CMD2* gene into various germplasm seems to be the most effective strategy in CMD control, as the incorporation of the *CMD1* gene into exotic germplasm and subsequent field testing in Africa indicated that the incorporation of *CMD1* has not been effective because this is a gene of minor quantitative effect (Fregene et al., 2006; Okogbenin et al., 2007). Moreover, the use of the dominant *CMD2* gene has been successful in transferring gene resistance genotypes originally from Latin America (Okogbenin et al., 2007).

#### Analysis of parental diversity available for breeding

The seven cassava accessions (BGM0268, BGM0270, BGM0281, BGM0284, BGM0358, BGM0359 and BGM1370) containing alleles related to *CMD2* for four markers are local varieties with adaptations restricted to their place of origin; therefore, populations segregated from crosses of these CMD resistance sources to national elite varieties must be generated to develop

varieties that are adapted to the major growing regions that bear traits of agronomic interest. To maximize the segregation of progeny, the selection of parents should combine contrasting characteristics that are often inferred by genetic divergence. Therefore, analyzing the divergence of 24 genotypes based on a discriminant analysis of principal components (DAPC) and in the genomic relatedness matrices may indicate the most contrasting crosses.

The Bayesian information criterion (BIC), as implemented in the DAPC using seven CMD resistance sources and 17 Brazilian elite varieties, indicated that three groups of diversity is the optimal number to describe the divergence of this set of materials (Figure 2). Thus, this number of groups was used in the subsequent analyses. The DAPC identifies complex population structure with the probabilistic assignment of individuals into specific groups without requiring assumptions about the Hardy-Weinberg equilibrium and linkage disequilibrium between markers, thus minimizing the cluster variation (Jombart et al., 2010). The DAPC has been used for large data volumes, particularly generated by SNP markers. Recently, the DAPC was used to structure the germplasm of *Carica papaya* L. with microsatellite markers (Matos et al., 2013) and cassava with SNPs markers (Oliveira et al., 2014).

The three diversity groups that were formed by DAPC are disparate, with a clear separation of groups

according to the representation of the first two linear functions (Figure 2). CMD resistance sources were allocated in Groups 1 (BGM0284, BGM0268 and BGM1370) and 2 (BGM0268, BGM0270, BGM0358 and BGM0359). Although Rabbi et al. (2014) suggested that all sources of monogenic resistance in cassava are derived from a single common ancestor, some authors suggest the existence of other genes for CMD resistance (Bi et al., 2010; Okogbenin et al., 2012). The seven resistance sources that were found in this study were clearly allocated to different divergence groups and, therefore, indicate the existence of genetic variability in these genotypes is not associated with resistance loci, which may contribute to the generation of transgressive individuals from crosses with the elite varieties. However, it is not possible to determine whether these resistance sources are different from those that are grown in Africa.

Regarding the elite cassava varieties, there was no tendency to cluster according to their degree of improvement, as both the hybrids and improved varieties and landraces were divided into different groups. The BRS Aipim Brazil, BRS Kiriris and Ecuador 72 varieties were grouped into Group 1, while the hybrid 9624-09 and BRS Caipira, BRS Mulatinha and BRS Tapioqueira varieties were allocated to Group 2. The other hybrid (98150-06) and the BRS Verdinha, BRS Formosa, BRS Guaira, Baianinha, Mani Branca, Cidade Rica, Irará, Lagoa and Sergipe varieties were placed in Group 3 (Figure 3, Table 4).

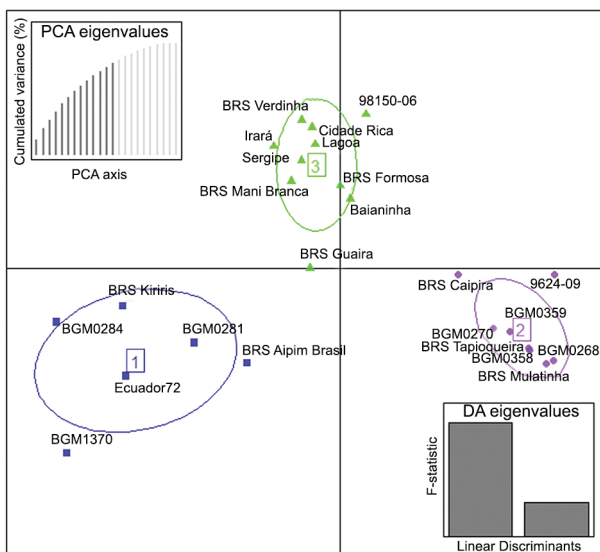


Figure 2 – Scattergraph of the first and second main components of the discriminant analysis of principal components (DAPC) based on the analysis of 24 cassava genotypes with 402 Single-Nucleotide Polymorphism (SNP)-type markers; Clusters are represented by numbers, and the accessions analyzed by different colors represent the analyzed accessions; the eigenvalues of the principal component analysis (PCA) and discriminant analysis (DA) are also presented.

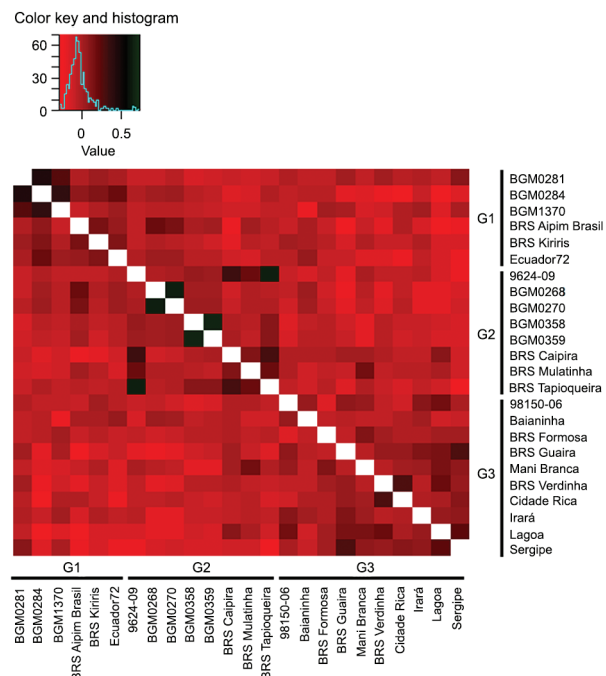


Figure 3 – Genomic relatedness heat map for 24 cassava genotypes based on an analysis of the 402 Single-Nucleotide Polymorphism (SNP)-type markers; the three groups that are formed by the discriminant analysis of main components are represented in the figure as G1, G2 and G3.

Table 4 – Estimates of genomic relatedness among the seven sources of Cassava Mosaic Disease (CMD) resistance and 17 elite cassava varieties. The selected crosses for each resistance source are shown in bold and underlined.

Group	Elite variety	CMD resistance source						
		Group 1			Group 2			
		BGM 0281	BGM 0284	BGM 1370	BGM 0268	BGM 0270	BGM 0358	BGM 0359
1	BRS Aipim Brazil	-0.018	0.087	-0.102	0.192	0.163	-0.001	0.013
	BRS Kiriris	0.036	0.125	-0.032	-0.034	-0.040	-0.059	-0.109
	Ecuador72	0.000	0.210	0.037	0.034	0.045	-0.047	-0.048
2	9624-09	-0.092	-0.018	-0.063	-0.018	-0.007	0.069	0.052
	BRS Caipira	-0.089	-0.168	-0.117	-0.103	-0.084	-0.032	0.007
	BRS Mulatinha	<b>-0.109</b>	-0.131	-0.112	-0.034	-0.032	-0.035	-0.077
	BRS Tapioqueira	<b>-0.094</b>	-0.012	-0.065	-0.036	-0.025	0.123	0.106
3	98150-06	-0.022	-0.069	-0.074	-0.076	-0.106	<b>-0.166</b>	-0.143
	Baianinha	-0.062	-0.053	<b>-0.202</b>	0.036	0.023	-0.069	-0.038
	BRS Formosa	-0.074	-0.049	0.027	-0.104	-0.094	-0.049	-0.043
	BRS Guaira	0.028	-0.179	-0.007	<b>-0.211</b>	<b>-0.176</b>	-0.115	-0.092
	Mani Branca	-0.057	-0.160	-0.125	-0.039	-0.077	<b>-0.177</b>	<b>-0.178</b>
	BRS Verdinha	0.018	-0.205	-0.114	-0.181	-0.162	-0.045	-0.055
	Cidade Rica	-0.052	<b>-0.267</b>	-0.023	-0.058	-0.047	-0.083	-0.109
	Irará	-0.013	-0.100	0.016	-0.179	<b>-0.176</b>	-0.100	-0.093
	Lagoa	-0.060	<b>-0.283</b>	<b>-0.168</b>	<b>-0.186</b>	-0.175	-0.115	-0.125
Sergipe	0.111	-0.168	0.004	-0.135	-0.109	-0.128	-0.146	

Estimates of genomic kinship based on the SNP markers between all pairs of cassava genotypes showed variation of -0.167 to 0.681, averaging 0.076 (Figure 3). The negative values indicate genomic kinship pairs of individuals that share few alleles compared to those expected based on the allelic frequencies, whereas positive values represent the opposite. Therefore, these estimates indicate a low relatedness between individuals, which certainly contributes to the maximization of the expected gains with selection in the resulted segregated populations.

According to Figure 3, the highest predicted genomic kinship was observed among CMD resistance sources: BGM0281 × BGM0284 (0.457) and BGM0284 × BGM1370 (0.373); as well as among 9624-09 × BRS Caipira (0.334), 9624-09 × BRS Tapioqueira (0.666) and BRS Caipira × BRS Tapioqueira (0.324). This result is consistent with DAPC, as all of these genotype combinations belong to the same group. Based on these results, crosses involving these genotypes would not be recommended.

The first separation of cassava accessions into three groups based on the DAPC of SNP markers indicated a hierarchical level for use in targeting possible contrasting intersections based on molecular information. Therefore, considering the groups that are formed by DAPC, the best crosses among the CMD resistance sources and the cassava elite varieties that are most frequently planted in Brazil were selected considering the two most contrasting combinations regarding genomic relatedness (Table 4).

For resistance sources BGM0284 and BGM1370 in Group 1, combinations of crosses with a lower kinship

degree were found among varieties in Group 3 (Table 4). For the BGM0281 accession, the two most contrasting genotype combinations were those with the BRS Mulatinha (-0.109) and BRS Tapioqueira (-0.094) varieties, both belonging to Group 2. However, for the source BGM0284, the lowest kinship estimates were obtained with the Cidade Rica (-0.267) and Lagoa (-0.283) varieties, and for the source BGM1370, the best combinations were with the Baianinha (-0.202) and Lagoa (-0.168) varieties. Even with a median kinship degree, the resistance sources BGM0281 and BGM0284 showed a lower relatedness degree with different varieties, which can certainly help to increase the genetic variability in the populations carrying this highly divergent genetic background.

For the CMD resistance sources from Group 2 (BGM0268, BGM0270, BGM0358 and BGM0359), there was more complementarity (lower degree of genomic relatedness) with the varieties belonging to Group 3. The best combinations of crosses for source BGM0268 were observed between BRS Guaira (-0.211) and Lagoa (-0.186), while for BGM0270, the best combinations were between BRS Guaira and Irará, both with the same degree of kinship (-0.176). For the source BGM0358, the best combinations were between the hybrid 98150-06 (-0.166) and the Mani Branca variety (-0.177). Furthermore, the combinations with a lower relatedness degree with the source BGM0359 were Mani Branca (-0.178) and Sergipe (-0.146).

The use of resistant genotypes is a good medium- and long-term strategy to address CMD. However, the development of new cassava varieties should not focus exclusively on the introduction of the *CMD2* gene be-

cause there is a strong preference of farmers for local or improved varieties that meet specific quality characteristics and root flavor (when intended for *in natura* consumption), as well as those with high starch yields, which are intended for the starch industry.

In fact, there are classic examples in Africa, in which CMD-resistant varieties were rejected by farmers in Uganda because they produced lower-quality roots compared to those of locally adapted varieties (Kawuki et al., 2011). The introgression of CMD resistance genes in varieties with desirable characteristics becomes crucial for the adoption of new varieties, as performed by other authors (Okogbenin et al., 2007; Kawuki et al., 2011; Ribeiro et al., 2012).

Being a heterozygous species, there will surely be high segregation in  $F_1$  cassava progenies derived from the crosses suggested in this work, which will result in the generation of different parental genotypes. However, the recombination of genes that are present in the Brazilian elite varieties that have already been planted by farmers, represents an important advance in breeding new varieties with different genetic backgrounds but with desirable characteristics that have already been accepted by both farmers and consumers.

### Prospects for cassava breeding aiming to improve CMD resistance

Currently, CMD does not occur in the Americas, but the main insect vector (whitefly) has been reported in this region, while its alternative form, biotype B (*B. argentifolia*) is widespread, and according to Polston and Anderson (1997), its host range includes cassava. Therefore, the prospect for the arrival of the disease in Brazil and the possible adaptation of biotype B for feeding on cassava require breeding programs to develop prevention measures. The incorporation of CMD resistance in Brazilian varieties is one of the goals of the cassava breeding program in recent years.

Because CMD is a quarantine disease in Brazil, MAS is one of the few ways to identify resistance sources in the absence of the pathogen. This need is especially true for long-cycled crops (> 12 months), such as cassava, which require 8 to 12 years for the development of a variety using conventional procedures (Okogbenin et al., 2007; Oliveira et al., 2014). In the specific case of breeding cassava for CMD resistance, MAS was used for the initial identification of resistance sources, while SNP markers were used to estimate the relatedness of individuals seeking to guide the crossings. After obtaining  $F_1$  segregating populations, MAS will again be used for identifying segregating alleles with CMD resistance to reduce the number of plants to be evaluated at the *seedling* stage and subsequent stages of the selection process. Additionally, using MAS shortens the detection period of CMD-resistant genotypes to only two years (from crossing to obtaining the  $F_1$  plants) compared to eight years based on conventional breeding (from crossing to the completion of the advanced resistance and production

tests). Thus, as cassava is cultured by vegetative propagation, the use of MAS allows breeders to focus their efforts just on the genotypes that are identified as resistant during the early stages of the breeding program, whose genotypes are fixed and can thus reduce the number of evaluation cycles. This strategy will reduce the selection cycles and, thus, the costs involved in the genotype evaluation under field conditions, similar to what has been accomplished in other programs (Blair et al., 2007; Okogbenin et al., 2007; 2012; Ribeiro et al., 2012).

The genetic improvement of any crop relies on the extent of genetic variability that is present in the germplasm, although cassava breeding programs have historically used few parental crosses, mainly due to the lack of familiarity with the agronomic potential and available genetic variation. In this sense, the results that were obtained in this study regarding the identification of possible CMD resistance sources, together with the ascertainment of varieties with desirable traits, such as resistance to some foliar (BRS Caipira, BRS Mulatinha, BRS Tapioqueira and BRS Formosa) and root diseases (BRS Kiriris, BRS Aipim Brazil and Mani Branca), resistance to mites and bedbugs (Ecuador72), high root and starch yield (9624-09, 98150-06, Baianinha, Cidade Rica, Irará, Lagoa and Sergipe) and tolerance to water deficit (BRS Guaíra, BRS Verdinha), constitute a major breakthrough in targeting specific crosses. Therefore, it is expected that the preventive improvement strategies that have been implemented in breeding programs via MAS can help reduce the number of selection cycles that are required for the transfer of CMD-resistant alleles and the generation of new cassava varieties with characteristics that are desirable for industry and for fresh consumption.

### Acknowledgments

The authors thank the Bahia State Foundation for Research Support (FAPESB), Coordination for the Improvement of Higher Level Personnel (CAPES) and Brazilian National Council for Scientific and Technological Development (CNPq) for financial assistance and scholarship support.

### References

- Akano, A.O.; Dixon, A.; Mba, C.; Barrera, E.; Fregene, M. 2002. Genetic mapping of a dominant gene conferring resistance to cassava mosaic disease. *Theoretical and Applied Genetics* 105: 521-525.
- Bi, H.; Aileni, M.; Zhang, P. 2010. Evaluation of cassava varieties for cassava mosaic disease resistance jointly by agro-inoculation screening and molecular markers. *African Journal of Plant Science* 4: 330-338.
- Blair, M.W.; Fregene, M.A.; Beebe, S.E.; Ceballos, H. 2007. Marker-assisted selection in common beans and cassava. p. 81-115. In: Guimarães, E.; Ruane, J.; Scherf, B.; Sonnino, A.; Dargie, J., eds. *Marker-assisted selection: current status and future perspectives in crops, livestock, forestry and fish*. FAO, Rome, Italy.

- Collard, B.C.Y.; Jahufer, M.Z.Z.; Brouwer, J.B.; Pang, E.C.K. 2005. An introduction to markers, Quantitative Trait Loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica* 142: 169-196.
- Collard, B.C.Y.; MacKill, D.J. 2008. *Marker-assisted selection: an approach for precision plant breeding in the twenty-first century*. *Philosophical Transactions of the Royal Society B* 363: 557-572.
- De Bruyn, A.; Villemot, J.; Lefeuvre, P.; Villar, E.; Hoareau, M.; Harimalala, M.; Abdoul-Karime, A.L.; Abdou-Chakour, C.; Reynaud, B.; Harkins, G.W.; Varsani, A.; Martin, D.P.; Lett, J. 2012. East African cassava mosaic-like viruses from Africa to Indian ocean islands: molecular diversity, evolutionary history and geographical dissemination of a bipartite begomovirus. *BMC Evolutionary Biology* 12: 228.
- Doyle, J.J.; Doyle, J.L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11-15.
- Fregene, M.; Morante, N.; Sánchez, T.; Marin, J.; Ospina, C.; Barrera, E.; Gutierrez, J.; Guerrero, J.; Bellotti, A.; Santos, L.; Alzate, A.; Moreno, S.; Ceballos, H. 2006. Molecular markers for introgression of useful traits from wild *Manihot* relatives of cassava, marker-assisted selection (MAS) of disease and root quality traits. *Journal of Root Crops* 32: 1-31.
- Fregene, M.; Puonti-Kaerlas, J. 2002. Cassava biotechnology. p. 179-207. In: Hillocks, R.J.; Thresh, J.M.; Bellotti, A.C., eds. *Cassava: biology, production and utilization*. CABI, Wallingford, UK.
- Herrera Campo, B.V.; Hyman, G.; Bellotti, A. 2011. Threats to cassava production: known and potential geographic distribution of four key biotic constraints. *Food Security* 3: 329-345.
- Jombart, T.; Devillard, S.; Balloux, F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structure populations. *BMC Genetics* 11: 94.
- Kawuki, R.S.; Pariyo, A.; Amuge, T.; Nuwamanya, E.; Ssemakula, G.; Tumwesigye, S.; Bua, A.; Baguma, Y.; Omongo, C.; Alicai, T.; Orone, J. 2011. A breeding scheme for local adoption of cassava (*Manihot esculenta* Crantz). *Journal of Plant Breeding and Crop Science* 3: 120-130.
- Legg, J.P.; Fauquet, C.M. 2004. Cassava mosaic geminiviruses in Africa. *Plant Molecular Biology* 56: 585-599.
- Legg, J.P.; Thresh, J.M. 2000. Cassava mosaic virus disease in East Africa: a dynamic disease in a changing environment. *Virus Research* 71: 135-149.
- Lokko, Y.; Danquah, E.Y.; Offei, S.K.; Dixon, A.G.O.; Gedil, M.A. 2005. Molecular markers associated with a new source of resistance to the cassava mosaic disease. *African Journal of Biotechnology* 4: 873-881.
- Matos, E.L.S.; Oliveira, E.J.; Jesus, O.N.; Dantas, J.L.L. 2013. Microsatellite markers of genetic diversity and population structure of *Carica papaya*. *Annals of Applied Biology* 163: 298-310.
- Mba, R.E.C.; Stephenson, P.; Edwards, K.; Melzer, S.; Nkumbira, J.; Gullberg, U.; Apel, K.; Gale, M.; Tohme, J.; Fregene, M. 2001. Simple Sequence Repeat (SSR) markers survey of the cassava (*Manihot esculenta* Crantz) genome: towards an SSR-based molecular genetic map of cassava. *Theoretical and Applied Genetics* 102: 21-31.
- Mohan, C.; Shanmugasundaram, P.; Maheswaran, M.; Senthil, N.; Raghu, D.; Unnikrishnan, M. 2013. Mapping new genetic markers associated with CMD resistance in cassava (*Manihot esculenta* Crantz) using simple sequence repeat markers. *Journal of Agricultural Science* 5: 57-65.
- Okogbenin, E.; Egesi, C.N.; Olasanmi, 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 Science* 52: 2576-2586.
- Okogbenin, E.; Porto, M.C.M.; Egesi, C.; Mba, C.; Ospinosa, E.; Santos, L.G.; Ospina, C.; Marin, J.; Barera, E.; Gutierrez, J.; Ekanayake, I.; Iglesias, C.; Fregene, M. 2007. Marker aided introgression of CMD resistance in Latin American germplasm for genetic improvement of cassava in Africa. *Crop Science* 47: 1895-1904.
- Oliveira, E.J.; Ferreira, C.F.; Santos, V.S.; Jesus, O.N.; Oliveira, G.A.F.; Silva, M.S. 2014. Potential of SNP markers for the characterization of Brazilian cassava germplasm. *Theoretical and Applied Genetics* 127: 1423-1440.
- Olsen, K.M.; Schaal, B.A. 2001. Microsatellite variation in cassava (*Manihot esculenta*, Euphorbiaceae) and its wild relatives: further evidence for a southern Amazonian origin of domestication. *American Journal of Botany* 88: 131-142.
- Polston, J.P.; Anderson, P.K. 1997. The emergence of whitefly-transmitted Gemini viruses in tomato in the Western hemisphere. *Plant Disease* 81: 1358-1369.
- Rabbi, I.Y.; Hamblin M.T.; Kumar, P.L.; Gedil, M.A.; Ikpan, A.S.; Jannink, J.L.; Kulakow, P.A. 2014. High-resolution mapping of resistance to cassava mosaic geminiviruses in cassava using genotyping-by-sequencing and its implications for breeding. *Virus Research* 186: 87-96.
- Ribeiro, P.F.; Akromah, R.; Manu-Aduening, J. 2012. Using marker assisted selection to hasten screening of cassava cultivars developed through introgression of Cassava Mosaic Disease (CMD) resistance into cassava landraces in Ghana. *Journal of Agricultural Science and Technology B* 2: 74-80.
- Thresh, J.M.; Cooter, R.J. 2005. Strategies for controlling cassava mosaic virus disease in Africa. *Plant Pathology* 54: 587-614.
- Thresh, J.M.; Fargette, D.; Otirn-Nape, G.W. 1994. Effects of African cassava mosaic geminivirus on the yield of cassava. *Tropical Science* 34: 26-42.
- Thresh, J.M.; Otim-Nape, G.W.; Legg, J.P.; Fargette, D. 1997. African cassava mosaic virus disease: The magnitude of the problem. *African Journal of Root and Tuber Crops* 2: 13-18.
- VanRaden, P. 2008. Efficient methods to compute genomic predictions. *Journal of Dairy Science* 91: 4414-4423.
- Wimmer, V.; Albrecht, T.; Auinger, H.J.; Schön, C.C. 2012. Synbreed: a framework for the analysis of genomic prediction data using R. *Bioinformatics* 28: 2086-2087.