

Article - Agriculture, Agribusiness and Biotechnology

Population Structure and Genetic Diversity of Sweet Cassava Accessions from the Midwestern, Southeastern and Southern Regions of Brazil

Alex Henrique Tiene Ortiz^{1*}

<http://orcid.org/0000-0003-0556-8503>

Pedro Soares Vidigal Filho¹

<https://orcid.org/0000-0003-3196-8631>

Vanesca Priscila Camargo Rocha²

<https://orcid.org/0000-0001-6678-9014>

Rebecca Caroline Ulbricht Ferreira³

<https://orcid.org/0000-0001-7624-3744>

Tiago Maretti Gonçalves⁴

<https://orcid.org/0000-0001-8971-0647>

Maria Celeste Gonçalves-Vidigal¹

<https://orcid.org/0000-0002-5157-6881>

¹State University of Maringá, Genetic and Breeding Graduate Program Maringá, Paraná, Brazil; ²Federal University of Technology, Bioinformatic Graduate Program, Cornélio Procópio, Paraná, Brazil. ³State University of Campinas, Genetic and Molecular Biology Graduate Program, Campinas, São Paulo, Brazil. ⁴Federal University of São Carlos, Evolutive Genetic and Molecular Biology Graduate Program, São Carlos, São Paulo, Brazil.

Received: YYYY.MM.DD; Accepted: YYYY.MM.DD.

* Correspondence: alexhtortiz@gmail.com; Tel.: +55-43-99937-9004

HIGHLIGHTS

- The 303 traditional sweet cassava cultivars from Brazil, were divided based on their genetic diversity into four groups, separating them according to the region of origin. The first group contained cultivars from Paraná (South region), the second from Paraná and Santa Catarina (South), the third from Mato Grosso do Sul (Midwest) and fourth one from Minas Gerais (Southeast).
- Highly divergent combinations were found, between cultivars from different regions, which can be used as parents in breeding programs.

Abstract: The objective of this work was to screen sweet cassava accessions collected in smallholding areas in the Midwestern, Southeastern and Southern regions of Brazil, using 15 SSR molecular markers, to determine population structure and genetic diversity. Polymorphism was detected in every loci analyzed, with mean of 6.33 alleles per locus, and mean polymorphism information content (PIC) of 0.6057, pointing out that the primers were highly informative. The observed heterozygosity ranged from 0.0709 (SSRY 101) to 0.9398 (GA 12), with a mean of 0.6511, and mean genetic diversity of 0.6578, ranging from 0.3592 (GA 136) to 0.8116 (SSRY 21). The most dissimilar combinations observed were BGM526PR-BGM596MS, BGM526PR-BGM622MS and BGM526PR-BGM629MS. The traditional cassava cultivars assessed were divided into four distinct groups: two with cultivars from the South, one from the Southeast and one from the Midwestern region of Brazil. The variances among and within groups determined by the analysis of molecular variance were 44 and 56%, respectively. The PhiPT parameter (analogue to Fst) of 0.44 indicates high differentiation among groups. Broad genetic diversity was found among the traditional sweet cassava cultivars assessed, and the most divergent groups were formed by cultivars from the South and the Midwestern regions of Brazil.

Keywords: *Manihot esculenta*, germplasm characterization, genetic variability, SSR markers.

INTRODUCTION

Cassava (*Manihot esculenta* Crantz) is a staple food for approximately 800 million people worldwide, mainly in Sub-Saharan Africa [1], where the per caput consumption is over 108 kg per year [2]. It is mostly cultivated in developing countries located between latitudes 30°N and 30°S [3], usually in low technology conditions [4]. It is a very important crop in these regions because of its adaptability, productive stability in unfavorable environments [3,5] and drought tolerance [6].

Sweet cassava, which contains less than 100 mg kg⁻¹ in raw pulp of cyanogenic glycosides [7], is of utmost importance for food safety, mainly in developing countries. Several reports of the vast genetic diversity in sweet cassava cultivars from Brazil have been found [8-12], mainly grown by smallholders for their subsistence, in backyard cultivations [7]. These small-cultivated areas have decreased significantly because of population migrations to larger cities in the last 50 years [13]. Moreover, the expansion of urban areas and major crops such as corn, soybean, and cotton [14], over smaller properties, has played an important role in the elimination of small sweet cassava cultivations, leading to the loss of unexplored genetic material [7,15].

Thereby, germplasm collection and characterization to determine population structure and genetic divergence among traditional sweet cassava cultivars have critical importance [16, 8, 9, 17, 18]. Consequently, molecular markers become an essential tool to screen germplasm [19-21], especially microsatellite markers, which are co-dominant [22], have high informative content [20, 23] and are frequent in the sweet cassava genome [21].

Therefore, this study aimed to define the population structure and genetic diversity of 300 traditional sweet cassava cultivars collected in the Midwestern, Southwestern and Southern regions of Brazil, and 3 commercial cultivars using simple sequence repeat markers (SSR).

MATERIAL AND METHODS

Plant Material

The population analyzed consisted of 300 traditional cultivars of sweet cassava and 3 commercial cultivars. Forty-nine of the traditional cultivars were collected in the region of Vale do Ivaí, in the State of Paraná, Southern region of Brazil, in the cities of Apucarana,

Borrazópolis, Cambira, Jandaia do Sul, Kaloré, Marumbí and Novo Itacolomé (cultivar names followed by PR). Other 124 of the traditional cultivars were collected in the cities of Bandeirantes, Cianorte, Maringá, Santa Mariana, and Toledo, in the State of Paraná (PR), and 14 in the State of Santa Catarina, in the city of Urussanga (SC) [7, 9, 24, 25]. Sixty cultivars were collected in the Midwestern region of Brazil, State of Mato Grosso do Sul, near the city of Campo Grande (MS) [26] and 50 in the Southeastern region of Brazil, on the South of the State of Minas Gerais (MG) [27]. Besides, the commercial cultivar IAC 576/70 was used as a standard for comparison with the other traditional cultivars. IAC 576/70 is originated from the cross between the cultivar SRT 797-Ouro do Vale and the clone IAC 14–18 [28], has broad distribution in the Central-Southern region of Brazil, and has great organoleptic and agricultural characteristics [29] related to yield and bacteriosis resistance (*Xanthomonas axonopodis* pv. *manihotis*, [30]). Other cultivars used as standard, such as Amarelinha, Catarina Branca and Pioneira, are traditional cultivars that have broad distribution in Brazil, mainly in the State of Paraná. The commercial cultivar IPR-Upira (derived from Amarelinha) was used because of its excellent characteristics as sweet cassava (presents high carotenoid content, can be easily peeled and has fast cooking time) and IPR-União (derived from Baianinha) was used because of its industrial aptitude [31].

DNA extraction and amplification

Stem pieces of about 0.20 m length from the 303 cultivars were grown in plastic bags containing washed sand and kept in a greenhouse. After sprouting, young leaves were collected, placed in aluminum foil packs, which were conditioned in a thermal box filled with liquid nitrogen, and DNA was immediately extracted following the protocol developed by Dellaporta *et al.* [32]. The quantification of the DNA using a fluorometer (Qubit® Fluorometer Invitrogen) was followed by the standardization of the samples at 50 ng μL^{-1} . The PCR reactions consisted of 25 μL containing 50 ng of DNA, 0.25 mM of each dNTP, 1.5 mM MgCl_2 , 10 mM buffer $\times 10$ (Invitrogen), 0.2 μM of each primer (forward and reverse), 1 U of *Taq* polymerase (Invitrogen), and ultrapure water (q.s.p) [23, 33].

Fifteen microsatellite markers from the groups GA (GA 12, GA 21, GA 131, GA 134, and GA 136) [23] and SSRY (SSRY 13, SSRY 19, SSRY 21, SSRY 27, SSRY 28, SSRY 45, SSRY 51, SSRY 100, SSRY 101, and SSRY 135) [33] were analyzed. The same set of primers was previously successfully used by Ferreira *et al.* [26] and Gonçalves *et al.* [27]. The PCRs were performed using a thermocycler Techne Endurance TC-512 (Analítica), and the amplifications were executed according to Mba *et al.* [33] and Chavarriaga-Aguirre *et al.* [23].

The amplified fragments were separated by electrophoresis in non-denaturing 10% polyacrylamide gels ran for approximately 3h 30 min at 100V. SYBR® Safe DNA gel stain (Life Technologies™) was used and the images were digitalized using the L-Pix EX photo documentation system (Loccus Biotechnology). For fragment length assignment, a 100bp standard molecular marker (DNA Ladder, Invitrogen) was used.

Statistical analyses

The software Structure 2.3.4 was used to assess the population structure of the plant material. The probabilistic method was conducted with 10,000 replications in burn-in and 100,000 replications in the Markov-Monte Carlo chain (MCMC) [34, 35]. Maintaining the standardized parameters of the program and the admixture model, 15 clustering simulations were made with the K factor (number of groups) ranging from 1 to 15, and the probabilities P (K) of the individuals to belong to the k-eth group in the bar plot were provided on the output file generated by the program [36]. The Structure Harvester application [37] was used to assess the results of the probabilistic method and determine the ΔK value, and consequently, the optimum K value which represents the number of groups which divide the population according to the genetic diversity of the cassava population.

The genetic divergence in the population was assessed building the C.S. Chord distance matrix [38] using the program Powermarker 3.25 [39]. Based on this distance matrix the Neighbor-joining tree was created using the program MEGA 7 [40]. The polymorphism information content (PIC), observed heterozygosity (H_o) and genetic diversity were also provided by Powermarker, whereas the analysis of molecular variance (AMOVA), and a Principal Coordinate Analysis were performed using the GenAlEx 6.5 [41,42]. In addition, the molecular genetic distance generated by SSR markers and the geographic distance were compared by the Mantel test using Genes software [43].

RESULTS AND DISCUSSION

Genetic diversity

The 15 loci analyzed were polymorphic, with a total of 95 alleles identified. The number of alleles per locus ranged from 4 (GA 12, GA 21 and SSRY 101) to 10 (GA 131) with a mean of 6.33. The mean of the most frequent alleles was 0.4531, and their frequency values ranged from 0.2424 (SSRY 21) to 0.7893 (GA 136) (Table 1).

A total of 31 rare alleles, which have frequencies lower than 0.05 [18], were found. Due to their low frequency, the rare alleles are highly susceptible to elimination under evolutionary forces, mainly genetic drift. For instance, the locus SSRY 101, which presents the lowest observed heterozygosity (H_o) (0.0709) and the 224 bp allele, which is the least frequent in the population (0,0017), is the most susceptible to loss of diversity due to this phenomenon. The elimination of these rare SSR alleles is likely to occasion the loss of other linked alleles, sometimes not explored or known, that occur clustered in the chromosome [44]. This fact indicates the importance of the collection and characterization of germplasm to preserve the genetic diversity. Furthermore, the rare alleles found can be useful when they are linked to genes of agronomic interest. In these cases, the molecular marker-assisted selection is possible [45]; consequently, the rare alleles observed represent opportunities for future studies [12].

The mean H_o was 0.6511. It surpassed the results obtained by Fregene *et al.* [46], Elias *et al.* [47], Rocha *et al.* [48], Costa *et al.* [9] and Ortiz *et al.* [12], who obtained 0.5136, 0.506, 0.4241, 0.5580 and 0.4762, respectively. Some studies identified higher means for H_o , such as that by Turyagyenda *et al.* [49], who found H_o of 0.726 in accessions from Uganda. The high heterozygosity found in the present study indicates the presence of wide heterosis among the sweet cassava plants cultivated by smallholders, which is a typical characteristic of plants that have clonal propagation and a crossed mating system (allogamy) such as *M. esculenta* [46].

The genetic diversity of the loci SSRY21, SSRY 45 and GA 131 presented the highest values (0.8116, 0. 7827, and 0.7774, respectively). In turn, the loci GA136 and GA 21 presented the lowest values (0.3592 and 0.4677, respectively). For GA 21, the low diversity observed was caused mainly by the reduced number of alleles of the locus, whereas for GA 136, it can be attributed to the unbalance observed among the allelic frequency of the locus, in which the 148 bp allele preponderates, and the others appear as rare and low frequency alleles. In general, the loci analyzed presented high genetic diversity (0.6578). The mean genetic diversity obtained corroborates with most results observed in literature for cassava [10, 18, 26, 49, 50].

The primers were highly informative [51], as the mean PIC was 0.6057. According to Xia *et al.* [52], the PIC for SSR in sweet cassava ranges from 0.5 to 0.7; thus, the mean value found was considered adequate. The mean PIC found corroborates with most studies observed, although great variation can be found. Results like Raghu *et al.* [53] show higher values of PIC (from 0.842 to 0.987) whereas Ferreira *et al.* [26] obtained lower values (0.525).

Table 1. Values of allelic frequency, polymorphism information content, observed heterozygosity and genetic diversity for the 15 microsatellite loci analyzed in the 303 traditional sweet cassava cultivars collected in the Southern, Southeastern and Midwestern regions of Brazil

Locus	Number of alleles	Alleles(bp)	Frequency	PIC ¹	H _o ²	Genetic diversity
SSRY13	9	192	0.4073	0.5747	0.6457	0.6444
		201	0.4123			
		205	0.0033*			
		208	0.0033*			
		218	0.1391			
		220	0.0017*			
		230	0.0149*			
		240	0.0066*			
		252	0.0116*			
SSRY19	7	182	0.0135*	0.7025	0.6734	0.7420
		190	0.0606			
		200	0.3468			
		205	0.3333			
		215	0.1179			
		225	0.0825			
		235	0.0455*			
SSRY21	7	163	0.0068*	0.7845	0.6475	0.8116
		171	0.0780			
		182	0.2000			
		186	0.2424			
		191	0.2305			
		195	0.0983			
		205	0.1441			
SSRY45	9	172	0.0102*	0.7495	0.7891	0.7827
		185	0.1105			
		189	0.0017*			
		194	0.2551			
		207	0.2959			
		216	0.2109			
		220	0.0017*			
		223	0.0306*			
		228	0.0833			
SSRY101	4	198	0.5878	0.3982	0.0709	0.5030
		217	0.3885			
		218	0.0220*			
		224	0.0017*			
SSRY135	5	255	0.2075	0.6473	0.6735	0.6985
		269	0.2500			
		272	0.0034*			
		276	0.4286			
		284	0.1105			
GA21	4	108	0.0133*	0.3796	0.6060	0.4677
		110	0.6556			
		112	0.0116*			
		130	0.3195			

		307	0.4043			
		318	0.4025			
GA134	6	328	0.0869	0.5958	0.4716	0.6590
		340	0.0088*			
		351	0.0887			
		360	0.0088*			
		148	0.7893			
GA136	5	150	0.0234*	0.3369	0.3846	0.3592
		157	0.0920			
		163	0.0937			
		170	0.0017*			
		258	0.2247			
SSRY27	5	274	0.3885	0.6686	0.6284	0.7183
		281	0.2534			
		287	0.0068*			
		292	0.1267			
		154	0.0033*			
SSRY28	7	157	0.0569	0.7169	0.7425	0.7512
		162	0.2057			
		170	0.3963			
		180	0.1254			
		187	0.1689			
		190	0.0435*			
		253	0.1672			
SSRY51	7	260	0.1639	0.6204	0.8649	0.6606
		270	0.5203			
		287	0.0253*			
		304	0.1149			
		322	0.0017*			
		336	0.0068*			
		182	0.4503			
SSRY100	6	187	0.0033*	0.6095	0.7450	0.6692
		201	0.2864			
		215	0.2103			
		224	0.0414*			
		243	0.0083*			
		133	0.2759			
GA12	4	143	0.5050	0.5514	0.9398	0.6216
		169	0.2174			
		180	0.0017*			
		91	0.1694			
GA131	10	99	0.0017*	0.7493	0.8837	0.7774
		104	0.0083*			
		108	0.0033*			
		110	0.0017*			
		113	0.3738			
		125	0.1728			
		133	0.0897			
		136	0.0964			
		156	0.0831			
Mean	6.3333		0.4531 ³	0.6057	0.6511	0.6578

¹Polymorphism information content; ²Observed heterozygosity per locus; ³Mean of high allelic frequencies; *rare alleles: frequency < 0.05 [18].

Despite the high PIC observed, the loci SSRY 101, GA 21 and GA 136 showed lower values (0.3982, 0.3796 and 0.3369, respectively). However, all loci analyzed were considered adequate for genetic diversity studies, since even these three were classified as reasonably informative. Therefore, the set of primers studied presented high potential to determine the divergence among the traditional sweet cassava cultivars analyzed.

The combinations between the cultivars BGM526PR-BGM596MS, BGM526PR-BGM622MS and BGM526PR-BGM629MS were the most divergent, with a genetic distance of 0.8487 (Table 2). It can be highlighted that all the combinations shown, with distances from 0.8036 to 0.8487, are highly divergent and contain the most genetically distant individuals among the possible 45753 combinations obtained (Table 2).

From the 18 most divergent combinations of cultivars shown, 61.11% were from Paraná and Mato Grosso do Sul, 5.56% between Santa Catarina and Mato Grosso do Sul, 22.22% between Paraná and Minas Gerais, and 11.11% exclusively from Paraná. This followed the tendency among the 303 cultivars of greater divergence between the groups from Mato Grosso do Sul and Paraná. The most divergent traditional cultivars discovered can be useful in breeding programs as parents for production of heterotic clones [54]. For this use, it is previously necessary to determine their agronomic characteristics (storage root dry matter content, yield, organoleptic characteristics, and the presence of flowering) to determine the most promising combinations in controlled crosses [12].

Table 2. Divergent and similar combinations among the 303 traditional sweet cassava cultivars collected in the Southern, Southeastern and Midwestern regions of Brazil, according to C.S. Chord distances

Most divergente combinations	Most similar combinations
BGM526PR* - BGM596MS = 0,8487	BGM572MS - BGM582MS = 0,0000
BGM526PR - BGM622MS = 0,8487	BGM579MS - BGM581MS = 0,0000
BGM526PR - BGM629MS = 0,8487	BGM583MS - BGM585MS = 0,0000
BGM348PR - BGM619MS = 0,8452	BGM583MS - BGM586MS = 0,0000
BGM355PR - BGM619MS = 0,8277	BGM583MS - BGM587MS = 0,0000
BGM358PR - BGM619MS = 0,8277	BGM583MS - BGM590MS = 0,0000
BGM373PR - BGM642MG = 0,8277	BGM585MS - BGM586MS = 0,0000
IPR União - BGM601MS = 0,8200	BGM585MS - BGM587MS = 0,0000
BGM51PR - BGM601MS = 0,8200	BGM585MS - BGM590MS = 0,0000
BGM526PR - BGM570MS = 0,8157	BGM586MS - BGM587MS = 0,0000
BGM339PR - BGM5PR = 0,8143	BGM586MS - BGM590MS = 0,0000
BGM55PR - BGM658MG = 0,8077	BGM587MS - BGM590MS = 0,0000
BGM234PR - BGM619MS = 0,8065	BGM334PR - BGM555PR = 0,0000
BGM234PR - BGM614MS = 0,8036	BGM639MG - BGM644MG = 0,0000
BGM485SC - BGM629MS = 0,8036	BGM668MG - BGM669MG = 0,0000
BGM526PR - BGM641MG = 0,8036	BGM683MG - BGM690MG = 0,0000
BGM373PR - BGM645MG = 0,8036	BGM88PR - BGM144PR = 0,0000
BGM236PR - BGM544PR = 0,8036	BGM166PR - BGM287PR = 0,0000

*MS=Mato Grosso do Sul, PR=Paraná, MG=Minas Gerais and SC=Santa Catarina.

Regarding the most similar cultivars, high occurrence of duplicates can be observed (Table 2). In all cases, the plants considered duplicates were collected in locations close to each other. This fact indicates that inside the regions, the exchange of plant material among

the farmers is a common practice. The transportation of traditional sweet cassava cultivar propagules can lead to the exchange of alleles between materials from different regions, which can create an increase in local diversity [16, 55].

Population structure

Based on the ΔK obtained using the data from the program Structure 2.3.4 [34] and the application Structure Harvester [37] (Fig. 1), the optimum number of groups that suits the distribution of the similar cultivars inside the population was four. All individuals were attributed to their K group with a Q-value (probability based on the inferred ancestry of the individuals) superior to 0.91. The population structure determined through the probabilistic method and the divergence among the most and least similar cultivars (Table 2) were coincident in the four groups observed (Fig. 2). Therefore, the most divergent traditional cultivars, such as BGM526PR and BGM596MS, were separated in different groups (Group 1 and Group 2). Similarly, the least divergent cultivars, such as BGM572MS and BGM582MS were grouped together (Group 3).

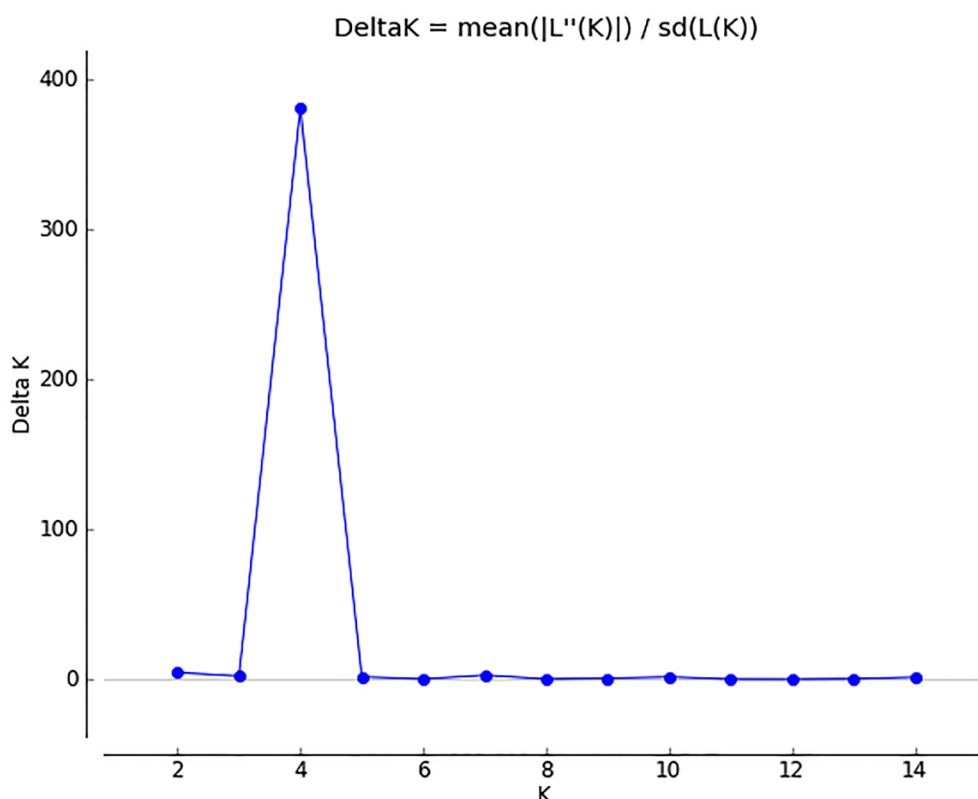


Figure 1. Inference of the K number of groups from 303 traditional sweet cassava cultivars from the Southern, Southeastern and Midwestern regions of Brazil, obtained by the program Structure Harvester [37].

The traditional cultivars were divided into an amount of groups superior to the number of regions analyzed. However, the cultivars from the regions studied (Southern, Midwestern and Southeastern) were separated in distinct groups. The ones from the Midwest and from the Southeast got in two individual groups, whereas the South got divided into two different groups. It indicates a direct relation between the geographic and genetic distance of the cultivars. The correlation between genetic and geographic distances was confirmed by the Mantel test, which showed a correlation of 0.5948, significant at an 1% probability level.

It can be highlighted that the commercial cultivar IAC 576/70, which was used as a standard, was placed in Group 1. So, this cultivar was considered genetically close to the others belonging to this group, which come mainly from the State of Paraná, South of Brazil.

Most of the commercial cultivar also belonged to this group. The cassava cultivar IAC 576/70 was originated from mating between the clone IAC 14–18 and the cultivar SRT 797-Ouro do Vale. According to Valle¹, Ouro do Vale was commonly found in the South of Brazil, State of Paraná. So, its alleles may have been shared among the cultivars collected in the State, leading to greater genetic similarity to the traditional cultivars collected in this region.

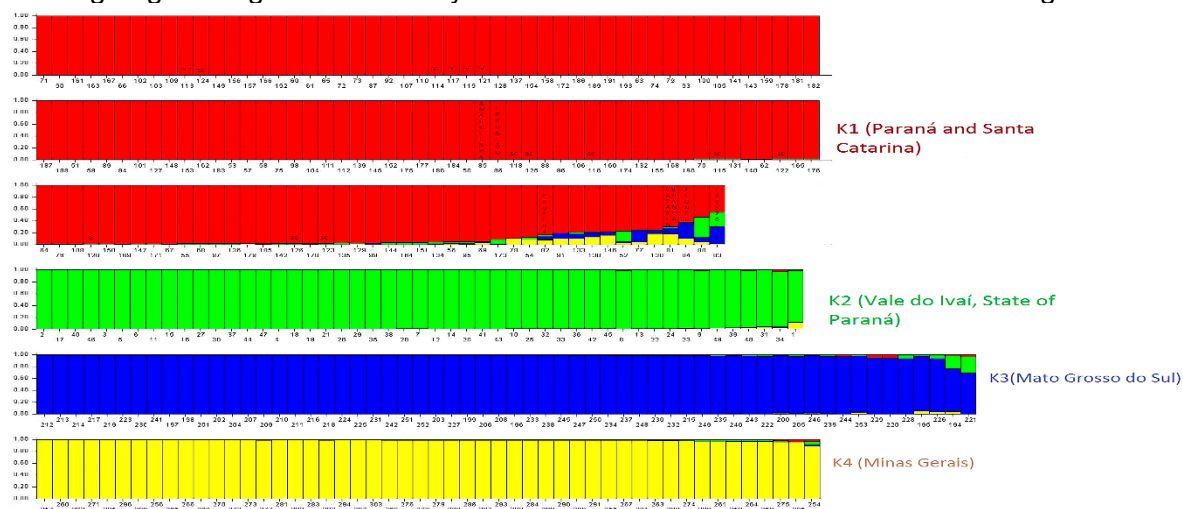


Figure 2. Population structure analysis of 303 traditional sweet cassava cultivars from the Southern, Southeastern and Midwestern regions of Brazil, for $K = 4$ groups given by the software Structure 2.3.4. [34]. The ordinate axis shows the probabilities of the individuals belonging to the K -eth group, while the abscissa axis shows every traditional cultivar on its respective bar. In the group K1, cultivars from Santa Catarina are marked with the initials SC, and the cultivars Amarelinha, Catarina Branca, Pioneira, IPR Upira, IPR União and IAC 576/70 are marked with their names inside its correspondent bars.

Based on the constitution of Group 1, it is possible to make an inference about the narrow genetic base among IAC 576/70 and the other commercial cultivars analyzed. The convergent selection towards similar desired characteristics caused the cultivars Amarelinha, Catarina Branca and Pioneira, which were introduced traditional sweet cassava cultivars, and IPR Upira, IPR União and IAC 576/70, which are commercially bred cultivars, to be grouped together. Besides, other traditional cultivars, with unknown agronomic potential belong to this same group. This fact raises the need for new studies aiming to produce information about these cultivars, and the possibility for the more promising ones to be used in breeding programs.

The constitution of each group was the following: in the South, Paraná obtained 27% (49 traditional cultivars) in Group 2 and 73% (132) in Group 1 and Santa Catarina 100% (14) in Group 1; in the Midwest, Mato Grosso do Sul had 100% (60) in Group 3 and in the Southeast, Minas Gerais had 100% (50) in Group 4 (Fig. 2). There is a clear separation among the cultivars directly connected to their regions of origin. The relations of proximity among the groups can be better understood analyzing the divergence among them (Table 3).

¹ Personal communication: Valle, Teresa Losada. Researcher at Instituto Agrônomo de Campinas, Brazil.

Table 3. Estimated divergences between groups by program Structure 2.3.4 (Pritchard *et al.* 2000)

Group	1 (PR and SC)	2(PR)	3 (MS)	4 (MG)
1 (PR and SC)	-	-	-	-
2 (PR)	0,1330	-	-	-
3 (MS)	0,2478	0,1754	-	-
4 (MG)	0,1918	0,2249	0,2071	-

The Groups 1 and 2 were the ones with the lowest divergence. This reinforces the connection between genetic and geographic distance, since the cultivars from both groups belonged to the South region of the country. Similarly, the most divergent groups (0.2478) were the ones constituted by cultivars from the most distant regions (Group 1 from the Southern and Group 3 from the Midwestern regions). However, the divergence results between groups are not capable of explaining all the relations observed. Group 2, for instance, contains the cultivars collected in the Vale do Ivaí, State of Paraná, South of Brazil. This group presented a high divergence against Group 4 (0.2249), which contains cultivars from Minas Gerais, and a smaller divergence (0.1754) considering Group 3. These peculiar characteristics can be explained by the complex migratory fluxes among those regions.

The State of Paraná suffered a considerable population growth, mainly in rural areas, from the years of 1940 to 1965, adding 2,700,000 migrants to its territory, mainly coming from the State of São Paulo [56]. This migration was stimulated by the expansion of the coffee production. In the same period migrants from São Paulo also colonized areas from Minas Gerais and Mato Grosso do Sul. This process probably led to the input of cassava cultivars from São Paulo to other regions of Brazil, generating genic flux among cultivars from the Southeast to the South and the Midwest.

Besides, a migratory wave from Paraná, mainly from the region of Vale do Ivaí, which was an important coffee producer, happened in the 1970's. This population efflux, promoted by the crisis in the coffee industry [57], justifies the proximity of cultivars from this region to the ones from Mato Grosso do Sul, which was an important agronomic frontier at the time. The State of Mato Grosso do Sul presented as an important alternative to the producers affected by the crisis, and consequently the state had a 9% growth on its cultivated area until the middle of the 1970's decade [58]. The genetic proximity relations among the regions can also be observed in the Neighbor-joining tree based on the C.S. Chord distance [38] (Fig. 3).

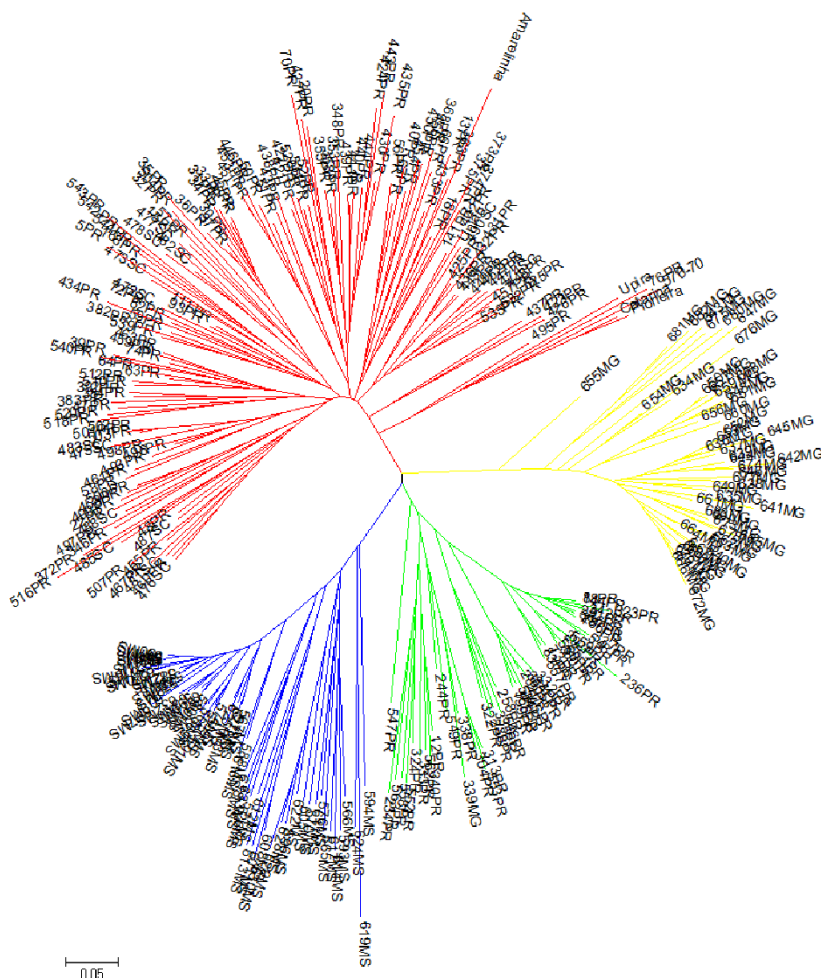


Figure 3. Neighbour-joining tree based on microsatellite data of 303 sweet cassava cultivars collected in the Southern, Southeastern and Midwestern regions of Brazil. The color of the branches is based on the groups formed by the program Structure Harvester [37].

The result observed in the Neighbor-joining is similar to the probabilistic method, where the cultivars from the Midwestern region were grouped in one cluster (blue), and the ones from the Southeastern region got in a separated cluster (yellow). The cultivars from the South were divided into two clusters, one containing cultivars from the States of Paraná and Santa Catarina (red), and other one with cultivars from the region of the Vale do Ivaí, State of Paraná (green). The most distant clusters were the red and the blue one. This result explains the fact that the most divergent combinations, such as BGM526PR-BGM596MS (Table 2), belong to the States of Paraná and Mato Grosso do Sul. The same fact can be observed with the least divergent combinations, which are all next to each other in the Neighbor-joining tree.

The principal coordinate analysis (PCoA, Fig. 4) presented the same tendency observed, on which the cultivars have separated according to their place of origin, and the most divergent ones are graphically distant, whereas the most similar ones are close. Besides, the dispersion of the points in the Cartesian plane is an indicator of the great diversity among the traditional cultivars studied. The two main coordinates explained 32,59% of the total variation (Coordinate1 explained 19.74% and Coordinate 2 12.85%).

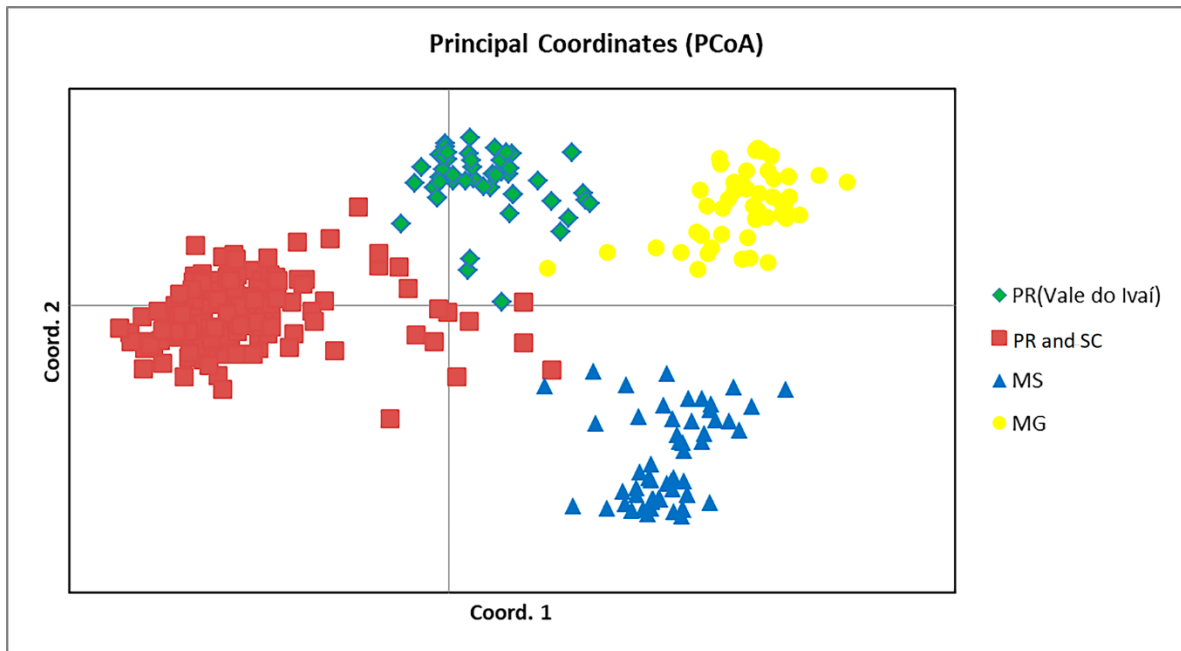


Figure 4. Principal Coordinate Analysis of 303 traditional sweet cassava cultivars from Brazil based on microsatellite data.

The AMOVA showed similar levels of among and within group variability considering the four groups generated by the structure analysis (Table 4). Most of the variation (56%) was allocated in the differentiation of individuals inside the groups, showing that even among individuals clustered by genetic proximity (which were from the same region), there is still high variability. The remaining 44% of the variability found was allocated among groups, which shows the regions (South, Southeast and Midwest) were heterogeneous when compared to each other.

Table 4. Analysis of the molecular variance (AMOVA) of the 303 traditional sweet cassava cultivars from the Southern, Southeastern and Midwestern regions from Brazil, considering the four K groups generated by the structure analysis

Variation source	DF ¹	SS ²	MS ³	EV ⁴	(%)	PhiPT	MPS ⁵
Among groups	3	1147,79	382,59	5,45	44	0,44	0,001
Within groups	299	2073,84	6,93	6,93	56		
Total	302	3221,637		12,39	100		

¹DF, Degrees of freedom; ²SS, square sums; ³MS, Mean square; ⁴EV, estimated variance; ⁵MPS, minimum probabilistic of significance; **significant at 1% of probability.

The differentiation among the groups can be compared by the PhiPT index (analogue to Wright [59] Fst). The 0.44 value obtained (44%) indicates a high differentiation among the groups. This explains the high divergence observed, and the concise division of the population into four groups in all the methods applied.

CONCLUSION

A wide genetic diversity was found in the population analyzed. The traditional sweet cassava cultivars were divided into four groups, connected to the geographic regions

studied. The cultivars from the Southern region were divided into two different groups, the Midwestern in one and the Southeastern in another group. The two most distant groups were formed by traditional sweet cassava cultivars from the State of Mato Grosso do Sul, Midwestern region, and from Paraná and Santa Catarina, Southern region of Brazil. Some of these cultivars formed highly divergent combinations, such as BGM526PR-BGM596MS, BGM526PR-BGM622MS and BGM526PR-BGM629MS, which must have their agronomic traits determined to enable their use as promising parents in breeding programs.

REFERENCES

1. Anyanwu, C.N.; Ibeto C.N.; Ezeoha, S.L.; Ogbuagu, N.J. Sustainability of cassava (*Manihot esculenta* Crantz) as industrial feedstock, energy and food crop in Nigeria. *Renew Energ* 2015, 81, 745-752.
2. FAO - Food and Agriculture Organization of the United Nations. Food Outlook: Biannual report on global food markets. (2016) Available on-line: <http://www.fao.org/3/a-i6198e.pdf> (accessed in 2017 Aug. 23).
3. Nassar, N.M.A.; Ortiz, R. Cassava improvement: challenges and impacts. *J Agric Sci* 2007, 145, 163–171.
4. El-Sharkawy, M.A. Cassava biology and physiology. *Plant Mol Biol* 2004, 56, 481–501.
5. Hurtado, P.; Olsen, K.M.; Buitrago, C.; Ospina, C.; Marin, J.; Duque, M.; de Vicente, C.; Wongtiem, P.; Wenzel, P.; Killian, A.; Adeleke, A.; Fregene, M. Comparison of simple sequence repeat (SSR) and diversity array technology (DArT) markers for assessing genetic diversity in cassava (*Manihot esculenta* Crantz). *Plant Genet Resour-C*. 2008, 6, 208–2014.
6. Chen, X.; Fu, Y.; Xia, Z.; Jie, L.; Wang, H.; Lu, C.; Wang, W. Analysis of QTL for yield-related traits in cassava using an F1 population from non-inbred parents. *Euphytica* 2012, 187, 227–234.
7. Zuin, G.C.; Vidigal-Filho, P.S.; Kvitschal, M.V.; Vidigal, M.C.G.; Coimbra, G.K. Divergência genética entre acessos de mandioca-de-mesa coletados no município de Cianorte, região Noroeste do Estado do Paraná. *Semin-Cienc Agrar* 2009, 30, 21–30.
8. Rimoldi, F.; Vidigal Filho, O.S.; Kvitschal, M.V.; Gonçalves-Vidigal, M.C.; Prioli, A.J.; Prioli, M.A.P.; Costa, T.R. Genetic divergence in sweet cassava cultivars using morphological agronomics traits and RAPD molecular markers. *Braz Arch Biol Technol* 2010, 53, 1477–1486.
9. Costa, T.R.; Vidigal Filho, P.S.; Gonçalves-Vidigal, M.C.; Galván, M.Z.; Lacanallo, G.F.; Silva, L.I.; Kvitschal, M.V. Genetic diversity and population structure of sweet cassava using simple sequence repeat (SSR) molecular markers. *Afr J Biotechnol* 2013, 12, 1040–1048.
10. Mezette, T.F.; Blumer, C.G.; Veasey, E.A. Morphological and molecular diversity among cassava genotypes. *Pesq Agrop Brasileira* 2013, 48, 510–518.
11. Oliveira, E.J.; Ferreira, C.F.; Santos, V.S.; Jesus, O.N.; Oliveira, G.A.; Silva, M.S. Potential of SNP markers for the characterization of Brazilian cassava germplasm. *Theor Appl Genet* 2014, 127, 1423–1440.
12. Ortiz, A.H.T.; Rocha, V.P.C.; Moiana, L.D.; Gonçalves-Vidigal, M.C.; Galván, M.Z.; Vidigal Filho, P.S. Population structure and genetic diversity in sweet cassava cultivars from Paraná, Brazil. *Plant Mol Biol Rep* 2016, 34, 1153-1166.
13. IBGE – Censo Demográfico 2010 (2017) Available on-line: http://www.ibge.gov.br/home/estatistica/populacao/censo2010/tabelas_pdf/Brasil_tab_1_9.pdf (accessed in 2017 Aug. 23).
14. CONAB - Séries históricas. (2014) Available on-line: http://www.conab.gov.br/conteudos.php?a=1252&t=2&Pagina_objcmsconteudos=1#A_objcmsconteudos (accessed in 2017 Aug. 23).
15. Willemen, L.; Scheldeman, X.; Cabellos, V.; Salazar, S.; Guarino, L. Spatial patterns of diversity and genetic erosion of traditional cassava (*Manihot esculenta* Crantz) in the Peruvian Amazon: an evaluation of socio-economic and environmental indicators. *Genet Resour Crop Ev* 2007, 54, 1599–1612.

16. Kizito, E.B.; Chiwona-Karltun, L.; Egwang, T.; Fregene, M.; Westerbergh, A. Genetic diversity and variety composition of cassava on small-scale farms in Uganda: an interdisciplinary study using genetic markers and farmer interviews. *Genetics* 2007, 130, 301–318.
17. Kawuki, S.R.; Morag, F.; Labuschagne, M.; Herselman, L.; Kim, D.J. Identification, characterization and application of single nucleotide polymorphisms for diversity assessment in cassava (*Manihot esculenta* Crantz). *Mol Breed* 2009, 23, 669–684.
18. Siqueira, M.V.B.M.; Queiroz-Silva, J.R.; Bressan, E.A.; Borges, A.; Pereira, K.J.C.; Pinto, J.G.; Veasey, E.A. Genetic characterization of cassava (*Manihot esculenta*) landraces in Brazil assessed with simple sequence repeats. *Genet Mol Biol* 2009, 32, 104–110.
19. Collard, B.C.Y.; Jahufer, M.Z.Z.; Brouwer, J.B.; Pang, E.C.K. An introduction to markers, quantitative trait loci (QTL) mapping and marker assisted selection for crop improvement: the basic concepts. *Euphytica* 2005, 142, 169–196.
20. Agarwal, M.; Shrivastava, N.; Padh, H. Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Rep* 2008, 27, 617–631.
21. Raji, A.A.; Anderson, J.V.; Kolade, O.; Ugwu, C.D.; Dixon, A.G.O.; Ingelbrecht, I.L. Gene-based microsatellites for cassava (*Manihot esculenta* Crantz): prevalence, polymorphisms and cross-taxa utility. *BMC Plant Biol* 2009, 9, 118–129.
22. Kalia, R.K.; Rai, M.K.; Kalia, S.; Singh, R.; Dhawan, A.K. Microsatellite markers: an overview of the recent progress in plants. *Euphytica* 2011, 177, 309–334.
23. Chavarriaga-Aguirre, P.; Maya, M.M.; Bonierbale, M.W.; Kresovich, S.; Fregene, M.A.; Tohme, J.; Kochert, G. Microsatellites in cassava (*Manihot esculenta* Crantz): discovery, inheritance and variability. *Theor Appl Genet* 1998, 97, 493–501.
24. Rocha, V.C.P. Caracterização morfoagronômica e molecular de acessos tradicionais de mandioca de mesa oriundos do Paraná e de Santa Catarina. Doctorate Thesis, Universidade Estadual de Maringá, Maringá, Paraná, Brazil, 2014.
25. Silva, L.I.; Vidigal Filho, P.S.; Costa, T.R.; Moiana, L.D.; Gonçalves-Vidigal, M.C. Molecular characterization of traditional sweet cassava accessions from the periurban region Toledo, Western Paraná., Southern Brazil. *J Glob Biosci* 2015, 4, 1268–1278.
26. Ferreira, R.C.U.; Vidigal Filho, P.S.; Gonçalves-Vidigal, M.C.; Moiana, L.D.; Kvitschal, M.V. Genetic and population structure of sweet cassava (*Manihot esculenta* Crantz) germplasm collected from Campo Grande, Mato Grosso do Sul, Brazil. *Aust J Crop Sci* 2015, 9, 458–467.
27. Gonçalves, T.M.; Vidigal Filho, P.S.; Gonçalves-Vidigal, M.C.; Ferreira, R.C.U.; Rocha, V.P.C.; Ortiz, A.H.T.; Moiana, L.D.; Kvitschal, M.V. Genetic diversity and population structure of traditional sweet cassava accessions from Southern of Minas Gerais State, Brazil, using microsatellite markers. *Afr J Biotechnol* 2017, 16, 346–358.
28. Villela, O.V.; Pereira, A.S.; Lorenzi, J.O.; Valle, T.L.; Monteiro, D.A.; Ramos, M.T.B.; Schmidt, N.C. Competição de clones de mandioca selecionadas para mesa e indústria. *Bragantia* 1985, 44, 559–568.
29. Mezette, T.F.; Carvalho, C.R.L.; Morgano, M.A.; Silva, M.G.; Parra, E.S.B.; Galera, J.M.S.V.; Valle, T.L. Seleção de clones-elite de mandioca de mesa visando a características agronômicas, tecnológicas e químicas. *Bragantia* 2009, 68, 601–609.
30. Aguiar, E.B.; Valle, T.; Lorenzi, J.; Kanthack, R.; Miranda, H.; Granja, M.P. Efeito da densidade populacional e época de colheita na produção de raízes de mandioca de mesa. *Bragantia* 2011, 70, 561–569.
31. Takahashi, M. Jornal Dia de Campo - IPR Upira: boa para o produtor e o mercado (2012) Available on-line: <http://www.diadecampo.com.br/zpublisher/materias/Materia.asp?id=27159&secao=Pacotes%20Tecnol%F3gicos> (accessed in Aug. 26 2017).
32. Dellaporta, S.L.; Wood, J.; Hicks, J.B. A plant DNA mini preparation: version II. *Plant Mol Biol Report* 1983, 1, 19–21.
33. Mba, R.E.C.; Stephenson, P.; Edwards, K.; Melzer, S.; Nkumbira, J.; Gullberg, U.; Apel, K.; Gale, M.; Tohme, J.; Fregene, M. Simple sequence repeat (SSR) markers survey of the cassava (*Manihot esculenta* Crantz) genome: towards an SSR-based molecular genetics map of

- cassava. *Theor Appl Genet* 2001, 12, 21–31.
34. Pritchard, J.K.; Stephensand, M.; Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics*. 2000, 155, 945–959.
 35. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol* 2005, 14, 2611–2620.
 36. Kwak, M.; Gepts, P. Structure of genetic diversity in two major gene pools of common bean (*Phaseolus vulgaris*). *Theor Appl Genet* 2009, 118, 979–992.
 37. Earl, D.A.; Vonholdt, B.M. Structure Harvester: a website and program for visualizing structure output and implementing the Evanno method. *Conserv Genet Resour* 2012, 4, 359–361.
 38. Cavalli-Sforza, L.L.; Edwards, A.W.F. Phylogenetic analysis: models and estimation procedures. *Am J Hum Genet* 1967, 19, 233–257.
 39. Liu, K.J.; Muse, S.V. Power marker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 2005, 21, 2128–2129.
 40. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for bigger datasets. *Mol Biol Evol* 2016, 33, 1870–1874.
 41. Peakall, R.; Smousse, P.E. GENALEX 6: genetic analysis in excel. Population genetic software for teaching and research. *Mol Ecol Notes* 2006, 6, 288–295.
 42. Peakall, R.; Smousse, P.E. GenALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*. 2012, 28, 2537–2539.
 43. Cruz, C.D. *Programa Genes: Diversidade Genética*; Editora UFV: Viçosa, MG, Brazil, 2008; 278p.
 44. Reed, D.H.; Frankham, R. Correlation between fitness and genetic diversity. *Conserv Biol* 2003, 17, 230–237.
 45. Weiser, E.L.; Grueber, C.E.; Jamieson, I.G. Simulating retention of rare alleles in small populations to assess management options for species with different life histories. *Conserv Biol* 2013, 27, 335–344.
 46. Fregene, M.A.; Suarez, M.; Mkumbira, J.; Kulembeka, H.; Ndedya, E.; Kulaya, A.; Mitchel, S.; Gullberg, U.; Rosling, H.; Dixon, A.G.; Dean, R.; Kresovich, S. Simple sequence repeat marker diversity in cassava landraces: genetic diversity and differentiation in an asexually propagated crop. *Theor Appl Genet* 2003, 107, 1083–1093.
 47. Elias, M.; Mühlen, G.S.; Mckey, D.; Roa, A.C.; Tohme, J. Genetic diversity of traditional South American landraces of cassava (*Manihot esculenta* Crantz): An analysis using microsatellites. *Econom Bot* 2004, 58, 242–256.
 48. Rocha, O.J.; Zaldivar, M.E.; Castro, L.; Castro, E.; Barrantes, R. Microsatellite variation of cassava (*Manihot esculenta* Crantz) in home gardens of Chibchan Amerindians from Costa Rica. *Conserv Genet* 2008, 9, 107–118.
 49. Turyagyenda, L.F.; Kizito, E.B.; Ferguson, M.E.; Baguma, Y.; Harvey, J.W.; Gibson, P.; Wanjala, B.W.; Osiru, D.S.O. Genetic diversity among farmer-preferred cassava landraces in Uganda. *Afr Crop Sci J* 2012, 20, 15–30.
 50. Peroni, N.; Kageyama, Y.; Begossi, A. Molecular differentiation, diversity and folk classification of “sweet” and “bitter” cassava (*Manihot esculenta*) in Caiçara and Caboclo management systems (Brazil). *Genet Resour Crop Ev* 2007, 54, 1333–1349.
 51. Botstein, D.; White, R.L.; Skalnick, M.H.; Davies, R.W. Construction of a genetic linkage map in man using restriction fragment length polymorphism. *Am J Hum Genet* 1980, 32, 314–331.
 52. Xia, L.; Peng, K.; Yang, S.; Wenzhi, P.; Carmem De Vicente, M.; Fregene, M.; Kilian, A. DArT for high-throughput genotyping of cassava (*Manihot esculenta*) and its wild relatives. *Theor Appl Genet* 2005, 110, 1092–1098.
 53. Raghu, D.; Senthil, N.; Saraswathi, T.; Raveendran, M.; Gnanam, R.; Venkatachalam, R.; Shanmugasundaram, P.; Mohan, C. Morphological and simple sequence repeats (SSR) based finger printing of south Indian cassava germplasm. *Int J Integr Biol* 2007, 1, 141–148.
 54. Gonçalves-Vidigal, M.C.; Vidigal filho, P.S.; Amaral Júnior, A.T.; Braccini, A.L. Divergência genética entre cultivares de mandioca por meio de estatística multivariada. *Bragantia* 1997, 56, 263–271.

55. Kawuki, R.S.; Herselman, L.; Labuschagne, M.T.; Nzuki, I.; Ralimanana, I.; Bidiaka, M.; Kanyange, M.C.; Gashaka, G.; Masumba, E.; Mkamilo, G.; Gethi, J.; Wanjala, B.; Zacarias, A.; Madabula, F.; Ferguson, M.E. Genetic diversity of cassava (*Manihot esculenta* Crantz) landraces and cultivars from Southern, Eastern and Central Africa. *Plant Genet Resour* 2013, 11, 170–181.
56. Nicholls, W.H. A Fronteira Agrícola na História Recente do Brasil: O Estado do Paraná, 1920-65. *Rev Bras Econ* 1970, 24, 33-91.
57. Camarano, A.A.; Abramovay, R. IPEA. Êxodo rural, envelhecimento e masculinização no Brasil: panorama dos últimos 50 anos (1999) Available on-line: http://www.ipea.gov.br/pub/td/1999/td_0621.pdf (accessed in 2017 Aug. 23)
58. Mueller, C.C. Dinâmica, condicionantes e impactos socioambientais da evolução da fronteira agrícola no Brasil. *Rev Admin Publ* 1992, 26, 64-87.
59. Wright, S. The genetic structure of populations. *Ann Eugenics* 1951, 15, 323–354.



© 2018 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY NC) license (<http://creativecommons.org/licenses/by/4.0/>).