(RESEARCH ARTICLE)

Population structure and genetic diversity in sweet cassava accessions from South of Brazil

Filipe Schmidt Schuh, Pedro Soares Vidigal Filho *, Maria Celeste Gonçalves-Vidigal, Mariana Vaz Bisneta, Giselly Figueiredo Lacanallo and Danielle Caroline Manenti

Maringá State University, Departament of Agronomia, Maringá, Brazil.

Abstract

Cassava (*Manihot esculenta Crantz*) is a crop of high socioeconomic importance for several tropical and subtropical regions of the world. It is mainly present in small cultivated areas that unintentionally guard a large part of the species' germplasm. The present work aimed to evaluate the population structure and genetic diversity of 156 traditional sweet cassava accessions from the Western of Paraná and Midwestern regions of Santa Catarina using 29 microsatellite molecular markers. All loci included were considered polymorphic, ranging from 3.00 to 7.00, with an average of 3.93 alleles per locus, and the average value of heterozygosity (Ho) was 0.6185. The polymorphism information content (PIC) presented an amplitude that varied from 0.4887 (GA134) to 0.7041 (GA131), with an average of 0.6130, while the genetic diversity ranged from 0.5688 (GA134) to 0.7424 (GA131), with an average of 0.6751. Analysis of the population structure based on the 29 microsatellite loci demonstrate that the accessions can be separated into two distinct subpopulations - in Santa Catarina and Paraná - with some mixtures observed according to Delta K = 2 groups. The ideal number of groups was found at K = 3, a level in which accessions from Santa Catarina were divided into two subpopulations and accessions from Paraná were grouped into a unique subpopulation. The genetic variability found among the traditional sweet cassava cultivars evaluated was considered wide, and the most dissimilar groups were mostly the accessions from Toledo and Santa Catarina states, constituting a source of genes for the sweet cassava breeding programs and for the development of new sweet cassava cultivars.

Keywords: Genetic diversity; Germplasm characterization; *Manihot esculenta Crantz*; Population structure

1. Introduction

Cassava is a crop of great importance worldwide regarding aspects related to human and animal nutrition [1, 2]. It belongs to the botanical family Euphorbiaceae, and the most widespread species for cultivation and human consumption is *Manihot esculenta* Crantz [1]. Its center of origin is widely discussed, but the most accepted theory is that it is the Southeast Amazon region [3], a site considered the primary center of diversity [4].

The largest global producers of cassava are Nigeria, Thailand, Brazil, Indonesia, and Congo [2]. Brazil produces an average of 23 million tons year $^{-1}$, and the largest producing states are Pará, with 5 million tons year $^{-1}$, followed by Paraná, with an average of 4 million tons year $^{-1}$ [5]. It should be noted that these production values of cassava tuberous roots basically refer to those that are intended for industrial processing, that is, are not intended for '*in natura'* consumption.

* Corresponding author: Pedro Soares Vidigal Filho; Email: psvfilho@uem.br
Maringá State University, Departament of Agronomia, Maringá, Brazil.

In general, cassava used for *fresh* consumption, denominated sweet cassava, is produced by small farmers and rarely reaches large commercial scales. It is highly important for populations with lower availability of economic resources because it is a low-cost energy source that has high simplicity and is easy to cultivate [2, 6]. However, several obstacles that need to be solved are present in its exploitation, the main ones being susceptibility to diseases, such as bacteriosis (*Xanthomonas axonopodis* pv. *manihotis*), productivity, dry mass content, plant architecture [7, 8] and, in some cases, nutritional qualities [9]. Given this, there is a need for the continuous development of genetic improvement programs for the crop, which necessitates sources of genetic variability for the investigation of genes of interest.

In this context, the collection, introduction, characterization, and evaluation of sweet cassava germplasm is of great value for the use and conservation of the species, especially in its use in breeding programs [10]. With these activities, the risk of genetic erosion due to the reduction of the cultivation of subsistence areas, which has been occurring due to the exodus of small farmers to urban centers, is reduced [11].

Some tools are available to assist in the germplasm characterization process and subsequent evaluation of its level of genetic diversity. Among them, we have morpho-agronomic markers, which were routinely used to perform this task over time, initially due to the unavailability of more advanced techniques and later because of the low cost and low need for laboratory processes [12].

With the advancement of molecular biology, new techniques have emerged in the field of germplasm characterization, including molecular markers based on DNA sequences. Such markers have been used with high frequency in the estimation of the genetic variability in the germplasm of cassava [1, 8, 13]. They are not affected by the environment, are applicable to any part of the genome (*introns*, *exons*, and regulatory regions), do not exhibit pleiotropy or epistatic effects, and are able to distinguish polymorphisms that do not produce phenotypic variation and some of them have codominance [14].

Among these molecular markers, microsatellites, or simple sequence repeats (SSRs), stand out as an excellent tool for the evaluation of genetic diversity among accessions [15, 16, 12]. This is due to their high reproducibility, simplicity, speed, small amount of DNA required, low cost of use, great resolution power, wide distribution by the genome, and reproducibility [8, 17, 18]. In this context, the objective of this study was to characterize and to evaluate the genetic diversity of sweet cassava accessions collected in Western of Paraná and in the Midwestern of Santa Catarina and to determine the population structure of these accessions based on microsatellite markers.

2. Material and methods

2.1. Collection and introduction of germplasm

Collections of the traditional accessions of sweet cassava (*M. esculenta* Crantz) cultivated by small farmers and often used for in natura consumption by their families were performed in municipalities in Western of Paraná and the Midwestern regions of Santa Catarina. Of the total of 156 accessions used in the present study, 133 accessions were from the Midwestern region of Santa Catarina (Figure 1) and were collected in the following municipalities: Caçador (26° 46' 31" S; 51° 00' 24" W, altitude: 985 m); Macieira (26° 51' 20" S; 51° 22' 41" W; altitude: 840 m); Videira (27° 00' 30" S; 51° 09' 06" W; altitude: 750 m); Rio das Antas (26° 53' 55" S; 51° 04' 28" W; altitude: 830 m); Arroio Trinta (26° 55' 58" S; 51° 20' 21" W; altitude 840 m); Pinheiro Preto (27° 03' 02" S; 51° 13' 51" W; altitude: 650 m), Iomerê (27° 00' 15" S; 51° 14' 32" W; altitude: 847 m) and Salto Veloso (26° 54' 16" S; 51° 24' 23" W; altitude: 820 m).

The climate prevailing in the Midwestern region of Santa Catarina, according to the Köppen scale, is of the Cfb type and consists of a temperate climate, humid mesothermal conditions, and mild summers [19]. The altitude of the Midwestern region of Santa Catarina ranges from 600 to 1,300 m, with an average annual rainfall of 1,280 mm, an average annual temperature of 16.2 °C, a minimum average temperature of 4.6 °C, and a maximum average temperature of 29.3 °C [20].

In turn, from the Western region of Paraná, 23 accessions were collected in 2008 in areas of small farmers in the municipality of Toledo (Figure 2). The municipality of Toledo is located between the geographic coordinates 24° 42′ 50″ S, 53° 44′ 34″ W. The climate of Toledo is humid subtropical and mesothermal, with hot summers, a tendency for rain concentration, and maximum mean temperature higher than 22 °C. In the winter period, frosts are frequent, with a minimum temperature below 18 °C and without the occurrence of a defined dry season. The collections of the accessions were performed in the periods between the end of April and mid-May 2018 for the Midwestern region of Santa Catarina and between July and August 2008 for Western of Paraná when the plant stems were mature and suitable for use as vegetative propagation material.
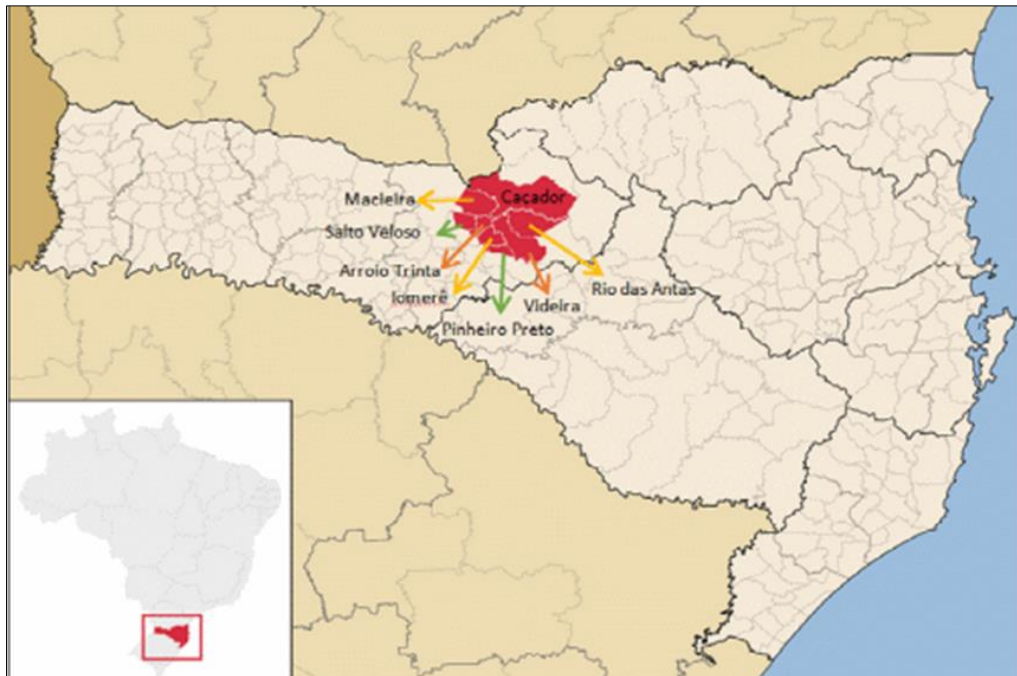
Figure 1 Map of the location of the municipalities of Arroio Trinta, Caçador, Iomerê, Macieira, and Pinheiro Preto, Rio das Antas, Salto Veloso and Videira. Midwestern region of Santa Catarina. Source: Wikipedia (2019)**.**

Figure 2 Location map of the municipality of Toledo. Western region of Paraná. Source: Wikipedia (2019)**.**

After collection, the samples of each accession were transported to the Fazenda Experimental Iguatemi (FEI) of the Universidade Estadual de Maringá, (UEM), located in the northwest region of the state of Paraná. The geographic location of the FEI is described by the coordinates 23° 21' 06” S, 51° 04' 11” W, with an average altitude of 596 m. The planting of the stems of each accession was performed in the area belonging to the Active Germplasm Cassava Bank (BAG-Mandioca) of the FEI-UEM in predefined rows spaced 1.0 m apart, parallel to each other, with 0.60 m between

plants, and each row of each accession consisted of 4 plants. For the planting, the stems of each accession were manually selected, with pieces (stakes) 0.10 to 0.15 m in length.

2.2. Molecular characterization by microsatellites

2.2.1. Extraction and quantification of genetic material

As a prior preparation for the extraction of genetic material (DNA) from each accession, the propagation of plant material from each accession was performed. For this purpose, mature branches were selected, and after sectioning to approximately 0.20 m long trunks, they were planted in plastic bags containing substrate + soil at a ratio of 1:1 by volume and kept in a greenhouse of the Núcleo de Pesquisa Aplicada à Agricultura (Nupagri-UEM) until the beginning of budding. In this stage, the young leaves of each accession were collected in 1.5 mL microtubes and taken to the Nupagri Biotechnology Laboratory.

Total DNA extraction from young leaves was performed using a PureLink® Genomic DNA Minikit DNA extraction kit from Thermo Fischer Scientific. Subsequently, a Qubit® Fluorometer (Qubit Fluorometer Invitrogen) was used to quantify the extracted DNA. Based on the data obtained from the quantification process, the DNA samples from each accession were diluted to a final concentration of 50 ng μL^{-1} DNA [21].

2.2.2. DNA amplification and electrophoresis

Each polymerase chain reaction (PCR) of 25 μL was composed of 50 ng of DNA; 0.25 mM of each of the deoxyribonucleotides (dATP, dCTP, dGTP, and dTTP); 1.5 mM $MgCl_2$; 10 mM 10x PCR Buffer - $MgCl_2$ (Invitrogen); 0.2 μM of each primer (sense and antisense), one unit of Taq DNA Polymerase (Perkin Elmer-Cetus Corp.); and ultrapure water (qsp). The microtubes with all the solutions already prepared were placed in a thermocycler (Techne Endurance TC-512, Analytical) to be subjected to specific amplification cycles for each primer used.

A total of 29 pairs of microsatellite primers (Table 1) were used in the molecular characterization process. Of these, 12 were from the GA series developed by Chavarriaga-Aguirre et al. [22], and 17 were from the SSRY series developed by Mba et al. [23].

The amplification cycles for the SSRY primers consisted of the following steps: an initial denaturation step at 94 °C for 5 minutes, followed by 30 cycles consisting of a denaturation step at 95 °C for 1 minute, a primer pairing step with DNA with variable temperature dependent on the primer and duration of 2 minutes, and an elongation step of the strand at 72 °C for 2 minutes. At the end of the 30 cycles, a final extension step was performed at 72 °C for 5 minutes [22, 23].

Table 1 Microsatellite primers used for molecular characterization of cassava accessions collected in Western Paraná and Midwestern Santa Catarina, with their respective sense and antisense sequences, fragment size generated, and annealing temperature

Microsatellite primer	Left primer sequence	Right primer sequence	Size*	AT (°C)
GA 012	GATTCCTCTAGCAGTTAAGC	CGATGATGCTCTTCGGAGGG	131-157	45
GA 013	TTCCCTCGCTAGAACTTGTC	CTATTTGACCGTCTTCGCCG	137-139	45
GA 016	GTACATCACCACCAACGGGC	AGAGCGGTGGGGCGAAGAGC	89-129	45
GA 021	GGCTTCATCATGGAAAAACC	CAATGCTTTACGGAAGAGCC	104-126	58
GA 057	AGCAGAGCATTTACAGCAAGG	TGTGGAGTTAAAGGTGTGAATG	153–183	59
GA 126	AGTGGAAATAAGCCATGTGATG	CCCATAATTGATGCCAGGTT	178–214	58
GA 127	CTCTAGCTATGGATTAGATCT	GTAGCTTCGAGTCGTGGGAGA	203–239	57
GA 131	TTCCAGAAAGACTTCCGTTCA	CTCAACTACTGCACTGCACTC	75-119	45
GA 134	ACAATGTCCCAATTGGAGGA	ACCATGGATAGAGCTCACCG	309–337	59
GA 136	CGTTGATAAAGTGGAAAGAGCA	ACTCCACTCCCGATGCTCGC	145–161	55
GA 140	TTCAAAGGAAGCCTTCAGCTC	GAGCCACATCTACTGCACACC	154–164	55

GA 161	TGTTCTTGATCTTCTGCTGCA	TGATTGTGGACGTGGGTAGA	64-140	45
SSRY 006	TTTGTTGCGTTTAGAAAGGTGA	ACAAATCATTACGATCCATTTG	298	45
SSRY 013	GCAAGAATTCCACCAGGAAG	CAATGATGGTAAGATGGTGCAG	234	55
SSRY 019	TGTAAGGCATTCCAAGAATTATCA	TCTCCTGTGAAAAGTGCATGA	214	55
SSRY 021	CCTGCCACAATATTGAAATGG	CAACAATTGGACTAAGCAGCA	192	55
SSRY 027	CCATGATTGTTTAAGTGGCG	CCATTGGAGAACTTGGCAAC	277	55
SSRY 028	TTGACATGAGTGATATTTTCTTGAG	GCTGCGTGCAAAACTAAAAT	180	55
SSRY 035	GCAGTAAAACCATTCCTCCAA	CTGATCAGCAGGATGCATGT	282	55
SSRY 045	TGAAACTGTTTGCAAATTACGA	TCCAGTTCACATGTAGTTGGCT	228	55
SSRY 047	GGAGCACCTTTTGCTGAGTT	TTGGAACAAAGCAGCATCAC	244	55
SSRY 050	CCGCTTAACTCCTTGCTGTC	CAAGTGGATGAGCTACGCAA	271	55
SSRY 051	AGGTTGGATGCTTGAAGGAA	GGATGCAGGAGTGCTCAACT	298	55
SSRY 061	GGCTGCTTTACCTTCTACTCAGA	CAAGAACGCCAATATGCTGA	233	55
SSRY 065	CATCGCCAAATCGTCAAGTA	TGATGCCATGCATTTCACTT	299	55
SSRY 085	AAGGTGGCAGCACTTTTCTG	AAGAATACTATACGGACTACATGCCA	292	55
SSRY 100	ATCCTTGCCTGACATTTTGC	TTCGCAGAGTCCAATTGTTG	210	55
SSRY 101	GGAGAATACCACCGACAGGA	ACAGCAGCAATCACCATTTC	213	55
SSRY 135	CCAGAAACTGAAATGCATCG	AACATGTGCGACAGTGATTG	253	45

Primers with SSRY nomenclature [22], primers with GA nomenclature [23]; AT: Annealing temperature. *Fragment size in base pairs.

The electrophoresis process consisted of the preparation of solutions containing the products resulting from PCR amplification with the addition of 8.0 µL of 6x loading buffer in each tube, totaling 10 µL of solution. These samples were applied to a polyacrylamide gel at a concentration of 10% in the arrangement of one sample per well formed after gel polymerization.

The gel run time was 120 minutes at a voltage of 80 V using a buffer electric current conductor of 1X tris borate EDTA (TBE). Subsequently, the development of the bands in the gels was performed on an L-Pix EX (Loccus Biotechnology) photodocumenter after staining the gel using SYBR® Safe DNA gel stain (Life TechnologiesTM).

2.2.3. Statistical analysis

The images obtained from the gels with the DNA fragments separated by electrophoresis were subjected to fragment size verification to obtain data on the alleles. For this purpose, a comparison between the bands and the 50-bp ladder (bp) pattern was performed using the photo documentation program LABIMAGE 1D, Revision 1.10 (Loccus Biotecnologia).

The Structure 2.3.4 program [24] was used to analyze the population structure of the 156 sweet cassava accessions from Western Paraná (23) and Midwestern Santa Catarina (133). The analysis was performed with Markov Chain Monte Carlo (MCMC) with 200,000 repetitions for burn-in and 1,000,000 normal repetitions [24, 25]. With the standard parameters of the program and maintaining the admixture model, 14 clustering simulations were performed, with the K factor (number of groups or populations) ranging from 2 to 14. From the output file, the probabilities P (K) of the individuals belong to the k-th group were evaluated, both numerically and in graphical form (bar plot), with the computer program Structure Harvester [26]. In the determination of allelic frequencies, the program GenAlEx 6.5 [27, 28] and the other genetic diversity parameters were calculated using the PowerMarker 3.25 computational program [29]. Additionally, Principal Coordinate Analysis (PCoA), the determination of allele frequencies per evaluated *locus*, and the analysis of molecular variance were performed using the GenAlEx 6.5 Program [27, 28].

3. Results and discussion

3.1. Evaluation of genetic diversity and population structure

Based on the estimates calculated for the indices related to genetic diversity, all 29 SSR *loci* analyzed were polymorphic, ranging from 3.00 to 7.00, with an average of 3.93 alleles per locus (Table 2). These results were similar to those obtained in other studies on the genetic diversity of cassava. Similarly, Siqueira *et al.* [30] used 9 microsatellite markers to evaluate the genetic diversity in 42 cassava accessions from several regions in Brazil and obtained an average of 5 alleles per marker. The genetic diversity and population structure in 51 farmer-preferred cassava landraces and 15 elite accessions grown in Uganda were investigated by using 26 SSR markers. In this case, a total of 154 alleles with an average of 6.77 alleles per marker were reported. Sixty cassava accessions from Mato Grosso do Sul State, Brazil were evaluated with 19 SSR markers and a mean of 4.63 alleles per marker were observed [8]. A total of 36 SSR markers were used to analyze the genetic diversity of 163 accessions of cultivated cassava (*M. esculenta* Crantz), with 94 from Cuba and 69 from different countries [16]. The authors noted that the 94 accessions in Cuba showed the highest average allele number per locus, with 5.8. Twenty-two accessions of cassava were evaluated in the Tapajós, Pará, using 11 markers and 67 alleles were found and the average of alleles per locus was 6.09 [32]. All loci were polymorphic. There was a variation of three (GA5 and GA136) to 11 alleles (RY93) per locus. Studies conducted by Gonçalves *et al.* [33] evaluated 51 traditional cassava accessions collected in southern Minas Gerais state, Brazil, using 20 microsatellites and found an average of 3.4 alleles per marker. Three hundred traditional cultivars of cassava and 3 commercial cultivars from Brazil were evaluated using 15 markers, and a mean of 6.33 alleles per locus were obtained [18]. More recently, 144 cassava accessions collected in seven places in Southern Brazil were evaluated using 25 SSR markers [34]. The authors noticed that all the loci analyzed were polymorphic and showed several alleles per locus, with a mean of 3.36 alleles [34]. Eighty-nine accessions from Ghana were evaluated using 35 SSRs resulting in 2 to 10 alleles per locus, with a mean of 4.77 alleles per locus [31].

The values of the most frequent alleles ranged from 0.3237 (SSRY 019) and 0.3269 (SSRY045) to 0.5513 (GA 134), with a mean value of 0.4135 (Table 2). In turn, the polymorphism information content (PIC) ranged from 0.4887 to 0.7041, with a mean of 0.6130, indicating that the primers used were mostly highly informative (96.55%). The highest value of PIC was 0.7041 (GA 131), and the lowest was 0.4887 (GA 134). According to Xia *et al.* [21], the PIC for SSR in sweet cassava ranges from 0.5 to 0.7; thus, the mean value found was considered adequate. In general, all *loci* used to characterize the population (Table 2) can be considered highly informative, values greater than 0.5 characterize the locus as highly informative, while values between 0.5 and 0.25 are moderately informative, and those lower than 0.25 are minimally informative.

Additionally, it was observed that the SSRs were highly heterotic, enabling the detection of high levels of genetic diversity among the accessions evaluated (Table 2). The heterozygosity (H_o) per locus varied between 0.0000 (GA 013), 0.2179 (GA134), and 0.2436 (GA140) to 0.0833 (SSRY019), 0.8333 (GA 057), and 0.8718 (SSRY 050), with a mean of 0.6185, while the observed genetic divergence ranged from 0.5688 (GA 134), 0.5887 (SSRY 101), and 0.6145 (SSRY065) to 0.7417 (SSRY019) and 0.7424 (GA 131), with a mean of 0.6751 (Table 2). Similar results were found in other studies on genetic divergence of cassava plants using microsatellite *loci*, where the majority of alleles had frequencies below 0.95 [17, 11, 15, 8, 32, 12, 33, 18, 34, 31].

As shown in Table 2, the values of intralocus genetic diversity (Ho) were high, considering that the lowest diversity value presented was 0.5688 (GA134) and the highest value was 0.7424 (GA 131), with a mean value of 0.6751. In similar studies, the mean values of intralocus genetic diversity found by Costa *et al.* [11] and Agre *et al.* [12] were approximately 0.51 [11, 12], while Ortiz *et al.* [13], Gonçalves *et al.* [33], and Rocha *et al.* [34] found values of 0.658, 0.6487, and 0.644, respectively.

In the present study we observed 85 alleles in the 29 *loci* evaluated, of which only six rare alleles 89 (GA 016), 108 (GA 021), 152 (GA 136), 234 (SSRY 13), 20 (SSRY 019), and 220 (SSRY 019) corresponding to 0.061% of the total in the population. Similar values were found in previous studies of genetic divergence conducted in cassava crop [11, 35]. However, studies conducted by Ferreira *et al.* [8], Ortiz *et al.* [18], Rocha *et al.* [34], and Gonçalves *et al.* [33] found 17, 15, 31, and 19 rare alleles, respectively.

The mean value of heterozygosity observed was higher than those found by Rocha *et al.* [36], who found H_o values of 0.424 for indigenous accessions and 0.4757 for commercial cultivars. Similar results were observed by other authors, who found mean values of 0.506 [37], 0.5581 [11], and 0.56 [15]. In contrast, Montero-Rojas *et al.* [38], Gonçalves *et al.* [33], and Agre *et al.* [12] observed mean H_o values higher than those of this study, of 0.7357, 0.6487 and 0.65, respectively. These differences observed in the number, frequency of alleles, and observed values of heterozygosity

compared with those of studies of different populations may be related to the evolutionary factors that act on these populations. Such factors, such as mutation, selection, and migration, can cause changes in allele frequencies and may fix, reduce (presence of rare alleles), or even extinguish the alleles in a given population [39]. The emergence of new alleles may be the result of genetic recombination, mutation, or introgression of genes from wild species within the genus *Manihot* [37].

Table 2 Estimated indices of genetic diversity by SSR *locus* evaluated

Locus	N° of alleles	Frequency	PIC	Ho^2	Genetic diversity
GA012	4	0.3750	0.6755	0.7500	0.7247
GA013	3	0.4231	0.5732	0.0000	0.6479
GA016	5	0.4615	0.6421	0.2756	0.6893
GA021	4	0.4519	0.5750	0.5769	0.6468
GA057	4	0.4551	0.5994	0.8333	0.6599
GA126	3	0.4135	0.5802	0.6346	0.6543
GA127	4	0.3526	0.6674	0.7051	0.7198
GA131	5	0.3974	0.7041	0.7244	0.7424
GA134	3	0.5513	0.4887	0.2179	0.5688
GA136	4	0.4071	0.5799	0.7115	0.6537
GA140	3	0.4167	0.5798	0.2436	0.6538
GA161	4	0.3526	0.6584	0.6923	0.7121
SSRY06	4	0.4231	0.5940	0.2692	0.6617
SSRY13	5	0.3718	0.6626	0.7308	0.7145
SSRY019	7	0.3237	0.6983	0.8333	0.7417
SSRY021	4	0.3301	0.6790	0.7821	0.7296
SSRY027	4	0.4071	0.6210	0.5192	0.6800
SSRY028	4	0.4423	0.6347	0.8205	0.6879
SSRY035	4	0.4455	0.6246	0.5897	0.6795
SSRY045	4	0.3269	0.6918	0.7756	0.7398
SSRY047	3	0.4712	0.5378	0.7244	0.6165
SSRY050	4	0.4327	0.5875	0.8718	0.6529
SSRY051	3	0.4615	0.5390	0.7308	0.6183
SSRY061	4	0.3397	0.6834	0.6410	0.7326
SSRY065	3	0.4455	0.5325	0.5321	0.6141
SSRY085	4	0.4359	0.5895	0.7115	0.6554
SSRY100	4	0.3718	0.6216	0.7308	0.6853
SSRY101	3	0.4808	0.5014	0.5641	0.5887
SSRY135	4	0.4231	0.6546	0.7436	0.7046
Mean	3.9310	0.4135	0.6130	0.6185	0.6751

No heterozygosity was detected for locus GA 013 (Table 2) (H_o = 0.000), similar to the results obtained by Costa *et al.* [11] and Rocha *et al.* [34] who observed a value of H_o = 0.05 for this same locus. In addition, apparently, the GA134 locus has a low number of alleles in general, as observed by Olsen *et al.* [40], who found only one allele for this locus. In several of the populations evaluated by Gonçalves *et al.* [33], and Rocha *et al.* [34], they found only three alleles for this *locus* in their studies.

In addition, the values obtained from the evaluation of the accessions collected in Western Paraná and Midwestern Santa Catarina were quantified within the overall mean of the related studies, demonstrating a good level of heterozygosity for the *loci* evaluated (Table 2). This is directly related to the reproductive system of cassava, which is an allogamous plant, and its wide dissemination by producers, who exchange materials, which provides high rates of hybridization and consequently high levels of heterozygosity [17].

Among the 29 *loci* evaluated, the *loci* GA 12, GA 127, GA 131, GA 161, SSRY 13, SSRY 19, SSRY 21, SSRY 45, SSRY 61, and SSRY 135 had genetic diversity values higher than 0.70 (Table 2). These high values of genetic diversity are certainly related to the greater number of alleles present in these *loci* and the better distribution of the frequencies presented by each of these alleles. In the PCoA (Figure 3), the coloration of the points representing the sweet cassava traditional accessions from the two evaluated regions highlights the migration of accessions of both populations. A comparison of the two groups formed in the probabilistic method shows that the points exhibited similar agglomeration trends in both methods, resulting in a cluster of 156 cassava accessions according to their place of origin, and the accessions that were more divergent from each other were more graphically distant.
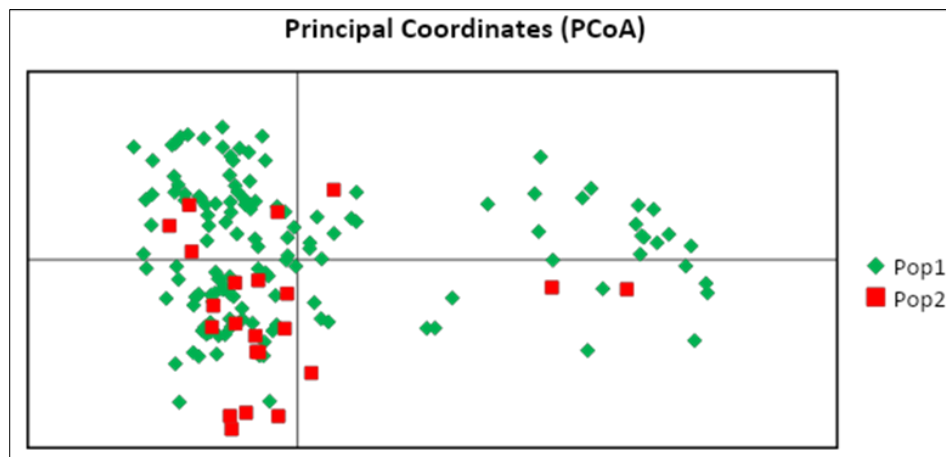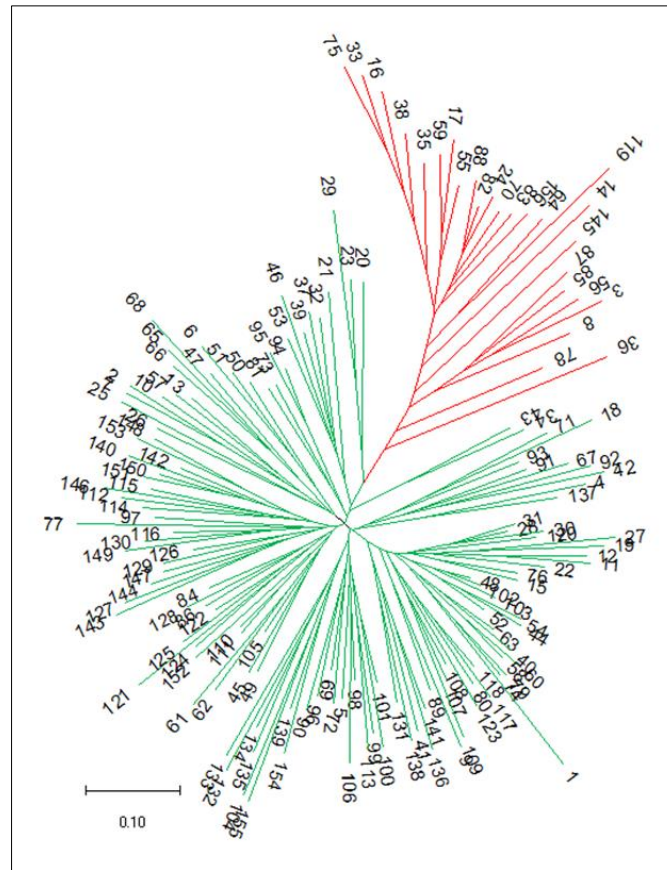

Figure 3 PCoA (PCo1 vs. PCo2), Pop1 refers to the traditional sweet cassava accessions collected in the Midwestern region of Santa Catarina, and the Pop2 accessions collected in the Western of Parana**.**

The first principal coordinate explained > 17% of the total variance, while the second coordinate accounted for 5%. There was also an approximation of the points in the case of traditional accessions that were considered similar or even duplicates (Figure 4) based on the CS chord distance analysis [41].

Based on the results of the population structure analysis, using the software Structure [24], for which a variation of *K* groups from 2 to 14 was simulated, the formation of two populations was observed. Statistical analyses using the program Structure [42] were used to obtain the Delta K value and, consequently, to determine the optimal value of K, which represents the number of groups that best represents the distribution of the most similar traditional accessions (Figure 5). Figure 5A best represents the population, with the formation of two groups, Group 1 (red) and Group 2 (green), in which there was a lower relative number of individuals with ancestry belonging to a single population, less than 90% (*Q* <0.9) [43, 44]. The accessions that demonstrate the presence of their respective area are represented with more than one color, assuming, arbitrarily, a cutoff region of 10% (Figure 5). Considering this aspect, these two groups showed approximately 13 and 18% accessions with this characteristic. Conversely, in Group 3 (blue), the presence of accessions with ancestry lower than 90% was approximately 22% and 19.5%, respectively.

Figure 4 Distribution of cassava accessions from the Midwestern part of Santa Catarina and from the Western of Parana, according to the Neighbor-joining Tree.
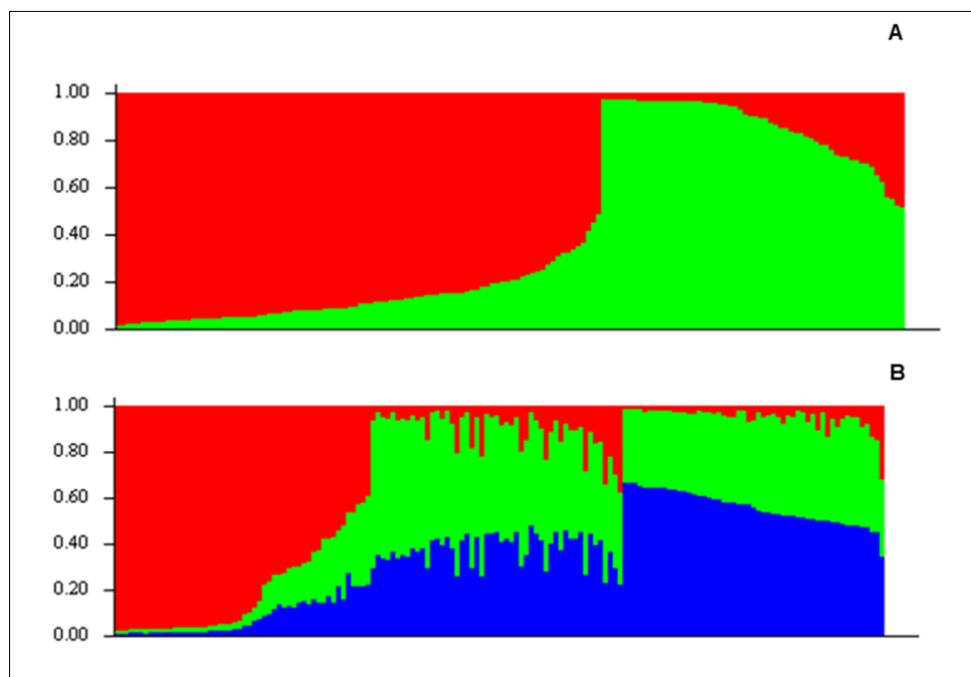

Figure 5 Population structure and allocation probabilities of sweet cassava accessions collected in the Midwestern of Santa Catarina and in the Western of Parana. A Group identified for *K = 2.* B Groups identified for *K = 3*.

According to Carmo *et al.* [44], the participation of accessions in different populations indicates that these individuals share some genomic regions among themselves, which makes it difficult to allocate them to a well-defined diversity group, even with robust population structuring models such as that implemented by Structure.

However, as a criterion for classifying the accessions in the groups, it was considered that most of the ancestry of the individual belonged to the group in which it was allocated. Regarding the size of the groups (Figure 5), the largest group was Group 1 (red), with 134 accessions, followed by Group 2 (green), with 22 accessions. With regard to the composition of the groups formed, most of them consisted of wide variability in relation to the collection municipalities.

To determine the population structure that best represented the organization of the collected accessions, the computer program Structure Harvester was used [42], which calculates a *Delta K* value corresponding to the optimal number of groups within an initial population. For the present study, Figure 6 shows that the peak value of *Delta K* was congruent with $K = 3$ and that this number of groups could best represent the population structure according to this methodology.

Figure 6 shows the graph of the population structure obtained by Structure 2.3.4, where the ordinate axis represents the probabilities of the individuals belonging to the K-th group while the abscissa axis represents the numerical codes for each accession collected. Based on this, the formation of two groups (subpopulations) can be observed, delimited by the orange lines superimposed on the graph and corresponding to the *Delta K* obtained by the Structure Harvester.
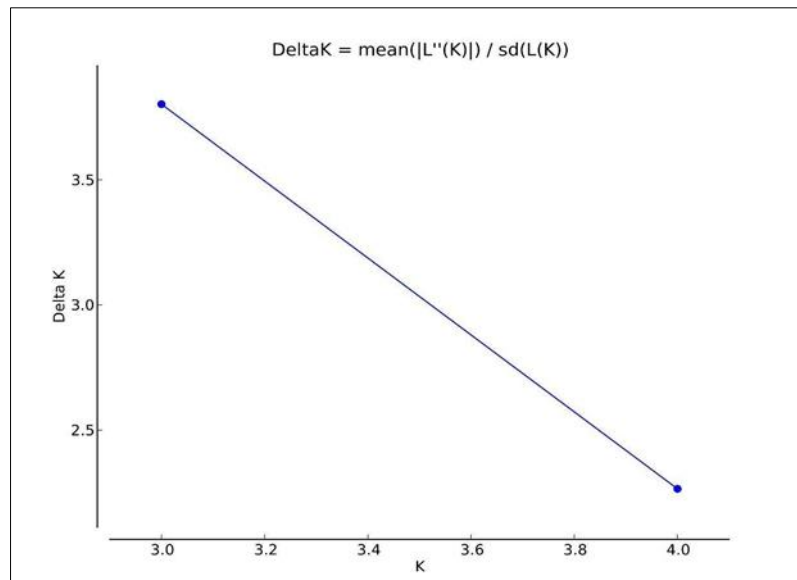
Figure 6 Inference of the number of K groups of the study population, obtained using the software Structure Harvester [42]

Given the above, the high genetic diversity found in the present study may be useful for the development of sweet cassava breeding programs, since it is essential that accessions with high genetic divergence be used as parents in the programs.

---

4. Conclusion

All *loci* analyzed were considered polymorphic, ranging from 3.00 to 7.00, with a mean of 3.93 alleles per locus and a mean value of heterozygosity (H_o) of 0.6185. The polymorphism information content (PIC) ranged from 0.4887 (GA134) to 0.7041 (GA131), with a mean of 0.6130, while the genetic diversity ranged from 0.5688 (GA134) to 0.7424 (GA131), with a mean of 0.6751. The analysis of the population structure based on the 29 microsatellite loci showed that the sweet cassava traditional accessions can be separated into two distinct subpopulations - Santa Catarina and Paraná - with some mixtures observed according to groups Delta K = 2. The ideal number of groups was found at K = 3, a level at which the accessions of Santa Catarina were divided into two subpopulations, and the accessions of Paraná were clustered in one group. The genetic variability found among the sweet cassava accessions evaluated was considered wide, and the most dissimilar groups were those that included accessions from Western of Paraná and the Midwestern of Santa Catarina, constituting a source of genes for the sweet cassava breeding programs and for the development of new sweet cassava cultivars.

Supporting information

Supplemental material is available online for this article.

S1 Table Traditional sweet cassava accessions and sampling sites in the Western region of Paraná and in the Midwestern region of Santa Catarina.

Compliance with ethical standards

Disclosure of conflict of interest

The authors declare that No conflict of interest.

References

[1]   Vieira EA, Fialho JDF, Faleiro FG, Bellon G, Fonseca KGD, Castelo Branco Carvalho LJ, Silva MS. Caracterização molecular e variabilidade genética de acessos elite de mandioca para fins industriais. Ciência Rural. 2010; 40: 2467-2471.

[2]   FAO – Food and Agriculture Organization of the United Nations. Faostat [Internet]. Food and Agriculture Organization of the United Nations; 2020 [cited 2020 Jan 20]. Available from: http://www.fao.org/faostat/en/#country.

[3]   Léotard G, Duputié A, Kjellberg F, Douzery EJ, Debain C, De Granville JJ, Mckey D. Phylogeography and the origin of cassava: new insights from the northern rim of the Amazonian basin. Mol Phylogenet Evol. 2009; 53, 329-334.

[4]   Nassar N. Wild cassava, *Manihot* spp.: biology and potentialities for genetic improvement. Genet Mol Biol. 2000; 23: 201-212.

[5]   Groxko M. Prognóstico Mandioca [Internet]**.** Estado do Paraná: Secretária de Estado da Agricultura e do Abastecimento – Departamento de Economia Rural; 2019. [cited 2020 Feb 20]. Available from http://www.agricultura.pr.gov.br/sites/default/arquivos_restritos/files/documento/2019-12/Mandioca%202020.pdf

[6]   Ortiz AHT, Rocha VPC, Moiana LD, Gonçalves-Vidigal MC, Galván MZ, Vidigal Filho PS. Population structure and genetic diversity in sweet cassava cultivars from Paraná, Brazil. Plant Mol Biol Rep. 2016; 6: 1153-1166.

[7]   Rimoldi F, Vidigal Filho PS, Kvitschal MV, Gonçalves-Vidigal MC, Prioli AJ, Prioli MAP, Costa TR. Genetic divergence in sweet cassava cultivars using morphological agronomic traits and RAPD molecular markers. Braz Arch Biol Technol. 2010; 53: 1477-1486.

[8]   Ferreira RCU, Vidigal Filho PS, Goncalves-Vidigal MC, Moiana LD, Kvitschal MV. Genetic and population structure of sweet cassava ('*Manihot esculenta*'Crantz) germplasm collected from Campo Grande, Mato Grosso do Sul, Brazil. Aust J of Crop Sci. 2015; 9: 458-467.

[9]   Esuma W, Pariyo A, Kawuki R, Wanjala B, Nzuki I, Harvey JJW, Baguma Y. Genetic diversity of provitamin A cassava in Uganda. J of Plant Stud. 2012; 1: 60-71.

[10]   Mohammadi SA, Prasanna BM. Analysis of genetic diversity in crop plants—salient statistical tools and considerations. Crop Sci. 2003; 43: 1235-1248.

[11]   Costa TR, Vidigal Filho PS, Gonçalves-Vidigal MC, Galván MZ, Lacanallo GF, Silva LI, Kvitschal MV. Genetic diversity and population structure of sweet cassava using simple sequence repeat (SSR) molecular markers. Afr J Biotechnol. 2013; 12: 1040-1048.

[12]   Agre AP, Bhattacharjee R, Rabbi IY, Alaba OA, Unachukwu NN, Ayenan MAT, Loko YL, Bauchet GJ, Dansi A. Classification of elite cassava varieties (*Manihot esculenta* Crantz) cultivated in Benin Republic using farmers'

knowledge, morphological traits and simple sequence repeat (SSR) markers. Genet Resour Crop Evol. 2017; 65, 513-525.

[13] Ortiz, AHT, Rocha VPC, Moiana LD, Gonçalves-Vidigal MC, Galván MZ, Vidigal Filho PS. Population structure and genetic diversity in sweet cassava cultivars from Paraná, Brazil. Plant Mol Biol Rep. 2017; 6: 1153-1166.

[14] Mondini L, Noorani A, Pagnotta MA. Assessing plant genetic diversity by molecular tools. Diversity. 2009; 1: 19-35.

[15] Emmy C, Miriam K, Oliver K, Arunga EE, Kimno S. Genetic diversity of cassava mutants, hybrids and landraces using simple sequence repeat markers. Am J Exp Agric. 2015; 5: 287-294.

[16] Beovides Y, Fregene M.; Gutiérrez, J.P.; Milián, M.D.; Coto, O.; Buitrago, C.; Cruz, J.A.; Ruiz, E.; Basail, M.; Rayas, A.; Rodríguez, D.; Santos, A.; López, J.; Medero, V. Molecular diversity of Cuban cassava (*Manihot esculenta* Crantz) cultivars assessed by simple sequences repeats (SSR). Biotechnol Agron Soc Environ. 2015; 19: 364–377.

[17] Fregene MA, Suarez M, Mkumbira J, Kulembeka H, Ndedya E, Kulaya A, Mitchel S, Gullberg U, Rosling H, Dixon AGO, Dean R, Kresovich S. Simple sequence repeat marker diversity in cassava *landraces*: genetic diversity and differentiation in an asexually propagated crop. Theor and Appl Genet. 2003; 107: 1083-1093.

[18] Ortiz AHT, Vidigal Filho PS, Rocha VPC, Ferreira RCU, Gonçalves TM, Gonçalves-Vidigal MC. Population structure and genetic diversity of sweet cassava accessions from the Midwestern, Southeastern and Southern Regions of Brazil. Brazilian Archives of Biology and Technology. 2019; 62: 1-16.

[19] Pandolfo C, Braga HJ, Silva Júnior VD, Massignan AM, Pereira ES, Thomé VMR, Valci FV. Atlas climatológico do estado de Santa Catarina. Florianópolis: Epagri. 2002; 1: 13.

[20] Mattos PP, Santos AT, Rivera H, Oliveira YMM, Rosot MAD, Garrastazu MC. Crescimento de *Araucaria angustifolia* na Reserva Florestal Embrapa/Epagri, Caçador, SC. Pesqui FlorestBras. 2007; 55: 107.

[21] Xia L, Peng K, Yang S, Wenzl P, De Vicente MC, Fregene M, Kilian A. DArT for high throughput genotyping of cassava (*Manihot esculenta*) and its wild relatives. Theor and Appl Genet. 2005; 110: 1092-1098.

[22] Chavarriaga-Aguirre P, Maya MM, Bonierbale MW, Kresovich S, Fregene MA, Tohme J, Kochert G. Microsatellites in cassava (*Manihot esculenta* Crantz): discovery, inheritance and variability. Theor and Appl Genet. 1998; 97: 493-501.

[23] Mba REC, 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 and Appl Genet. 2001; 12: 21-31.

[24] Pritchard JK, Stephensand M, Donelly P. Inference of population structure using multilocus genotype data. Genetics. 2000; 155: 945–959.

[25] 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.

[26] Kwak M, Gepts P. Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). Theor and Appl Genet. 2009; 118: 979-992.

[27] Peakall R, Smousse PE. Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes. 2006; 6: 288-295.

[28] Peakall R, Smousse PE. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinform. 2012; 28: 2537-2539.

[29] Liu KJ, Muse SV. Power Marker: an integrated analysis environment for genetic marker analysis. Bioinform. 2005; 21: 2128-2129.

[30] Siqueira MVBM, Queiroz-Silva JR, Bressan EA, Borges A, Pereira KJC, Pinto JG, Veasey EA. Genetic characterization of cassava (*Manihot esculenta*) landraces in Brazil assessed with simple sequence repeats. Genet Mol Biol. 2009; 32: 104-110.

[31] Adjebeng-Danquah J, Manu-Aduening J, Asante IK, Agyare RY, Gracen V, Offei SK. Genetic diversity and population structure analysis of Ghanaian and exotic cassava accessions using simple sequence repeat (SSR) markers. Heliyon. 2020; 6: e03154, 2020.

[32] Silva RS, Moura EF, Farias Neto JT, Souza NR, Moura MF, Sampaio JE. Divergência genética entre acessos de mandioca (*Manihot esculenta* Crantz) coletados na região do Tapajós, Estado do Pará, por meio de caracteres agronômicos e marcadores microssatélites. Semina - Ciênc Agrár. 2016; 37: 2989-3004.

[33] Gonçalves TM, Vidigal Filho PS, Gonçalves-Vidigal MC, Ferreira RCU, Rocha VPC, Ortiz AHT, Moiana LD, Kivitschal MV. 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.

[34] Rocha VPC, Gonçalves-Vidigal MC, Ortiz AHT, Valetini G, Ferreira RCU, Gonçalves TM, Lacanallo GF, Vidigal-Filho PS. Population structure and genetic diversity in sweet cassava accessions in Paraná and Santa Catarina, Brazil. Plant Mol Biol Rep. 2020; 38: 25-38.

[35] 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 Evol. 2007; 54: 1333–1349.

[36] Rocha OJ, Zaldivar ME, 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.

[37] Elias M, Mühlen GS, Mckey D, Roa AC, Tohme J. Genetic diversity of traditional South American landraces of cassava (*Manihot esculenta* Crantz): an analysis using microsatellites. Econ Bot. 2004; 58, 242-256.

[38] Montero-Rojas M, Correa AM, Siritunga D. Molecular differentiation and diversity of cassava *(Manihot esculenta)* taken from 162 locations across Puerto Rico and assessed with microsatellite markers. AoB Plants. 2011; 1-10.

[39] Kesawat MS, Das BK. Molecular markers: it's application in crop improvement. J Crop Sci Biotechnol. 2009; 12: 169-181.

[40] Olsen KM, Schaal BA. Microsatellite variation in cassava (*Manihot esculenta*, Euphorbiaceae) and its wild relatives: further evidence for a southern Amazonian origin of domestication. Am J Bot. 2001; 88: 131-142.

[41] Cavalli-Sforza LL, Edwards AWF. Phylogenetic analysis: models and estimation procedures. Am J Hum Genet.1967; 19: 233-257.

[42] Earl DA, Vonholdt BM. Structure Harvester: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Res. 2015; 4: 359-361.

[43] Emanuelli F, Lorenzi S, Grzeskowiak L, Catalano V, Stefanini M, Troggio M, Myles S, Martinez-Zapater JM, Zyprian E, Moreira FL, Grando MS. Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grapes. BMC Plant Biol. 2013; 13: 1-17.

[44] Carmo CD, Santos DB, Alves LB, Oliveira GAF, Oliveira EJ. Development of TRAP (Target Region Amplification Polymorphism) as new tool for molecular genetic analysis in cassava. Plant Mol Biol Rep. 2015; 33: 1953-1966.