

MOLECULAR CHARACTERIZATION AND CORE COLLECTION EVALUATION OF *Manihot esculenta* Crantz

CARACTERIZAÇÃO MOLECULAR E AVALIAÇÃO DE COLEÇÃO NUCLEAR DE *Manihot esculenta* Crantz

Nathalia Almeida COSTA¹; Hellen Sandra Freires da Silva AZÊVEDO²;
Luciêlio Manoel da SILVA³; Elisa Ferreira Moura Cunha⁴; Amauri SIVIERO³;
Tatiana de CAMPOS³

1. Postgraduate Program in Science, Innovation and Technology for the Amazon, Federal University of Acre, Rio Branco, AC, Brazil; 2. Postgraduate Program in Biodiversity and Biotechnology, Oswaldo Cruz Foundation - Bionorte, Porto Velho, RO, Brazil; 3. Brazilian Agricultural Research Corporation, Embrapa Acre, Rio Branco, AC, Brazil. tatiana.campos@embrapa.br; 4. Brazilian Agricultural Research Corporation, Embrapa Amazônia Oriental, Belém, PA, Brazil.

ABSTRACT: Cassava is one of the most important subsistence crops in tropical regions. It is necessary to preserve and to know the genetic diversity existent for the adequate use of genetic resources. The evaluation of genetic diversity among genotypes results in information about potential parents in breeding programs, allows duplicates identification, and facilitates germplasm exchange between research institutions. The objective of this study was to characterize the genetic diversity of cassava accessions of North Brazil region. A total of 106 accessions were analyzed using ten microsatellite markers. The genetic parameters estimated were: expected heterozygosity (H_E), observed heterozygosity (H_O) and polymorphic information content (PIC). Clustering was performed using the UPGMA and Neighbor-Joining (NJ) method. Bayesian analysis, analysis of principal coordinates and identification of a core collection were also used. The ten loci amplified 8,40 alleles on average. The average heterozygosity estimates were: $H_E = 0.71$, $H_O = 0.58$ and $PIC = 0.72$. Genetic distances ranged from 0.158 to 0.908. Six (5,66%) accesses were redundant. Clustering and dispersion analysis didn't differentiate bitter from sweet cassava, and there wasn't correlation between groups and collect origin. The core collection consisted of 22 individuals that represented 94% of total allelic diversity and 20,75% of the base collection. The results indicate high dissimilarity between the accessions and allowed the detection of redundant genotypes, showing the use of genetic markers as informative tools for the management of collections.

KEYWORDS: Cassava. Core collection. Genetic resources. *Manihot esculenta*. Microsatellite markers. Variability.

INTRODUCTION

In Latin America, Africa and Asia cassava (*Manihot esculenta* Crantz) is one of the main food sources for about 800 million people due mainly to the starch content in the roots (ALBUQUERQUE, 2017). Associated with high adaptation to unfavorable environments, such as low fertility soils and drought tolerance (ORTIZ et al., 2016), cassava has now become the fifth most important crop in the world (SOUSA et al., 2017). Brazil is the fourth largest producer in the world (FAO, 2017) and the North region accounts for most of this production (IBGE, 2018). Acre State presented the third largest cassava production in the North region of Brazil with 641,967 t/ha.

Is also known for its wide diversity of varieties, estimated at approximately 7000 varieties scattered around the world (HERSHEY, 1994). They are classified into two major groups based on

the levels of hydrocyanic acid (HCN) in the roots: "bitter" cassava (above 100 ppm fresh weight) and "sweet" cassava (below 100 ppm fresh weight) (DUFOUR, 1988). Bitter varieties are toxic if consumed without the appropriate pre-cooking treatment for HCN removal (BRADBURY et al., 2013).

Most of the cassava production is conducted by traditional small farmers who use the slash and burn system. Through vegetative propagation, they fix genotypes which are available to hybridize with different clones in field and probably generating new varieties through interspecific hybridization (SAMBATTI et al., 2001, FREGENE et al., 2003). The cultivation system has demonstrated a great preservation challenge due to selection, incorporation and exchange of varieties between farmers (ELIAS et al., 2001; EMPERAIRE; PERONI, 2007; COSTA et al., 2013; MOURA; FARIAS NETO, 2017; SOUSA et al., 2017).

It is essential to characterize the available diversity to adequately conserve the genetic resources and prevent genetic erosion, which would lead to loss of genes or favorable gene combinations (VIEIRA et al., 2010). Considering that Brazil is the center of origin and diversity (OLSEN, 2004) and that Amazon basin is the possible place of domestication of *Manihot esculenta* (ALLEM, 1994; OLSEN; SCHAAL, 2001), collections are essentials to maintain a germplasm of the species, under field conditions, for immediate and future use.

The use of molecular markers is an efficient tool to differentiate accessions, to identify redundancies and to select genotypes with characters of interest (COSTA et al., 2013; MOURA et al., 2013; GONÇALVES et al., 2017; SOUSA et al., 2017).

The objective of this work was to identify the genetic divergence of accessions of *Manihot esculenta* from the collection present in Embrapa Acre, to analyze the genetic structure of the accessions.

MATERIAL AND METHODS

Genetic material

The cassava collection located at Embrapa Acre, Rio Branco, Brazil, is composed of 106 accessions from states of Acre (74), Amazonas (24), Roraima (1), São Paulo (3) and accessions of unknown origin (4). The local names were maintained for each traditional variety and the materials were classified into two types of cassava, “bitter” and “sweet” (Table 1). 21 accessions were collected at Vale do Juruá, Cruzeiro do Sul, Acre and are used to produce the regional flour with geographic indication (GI).

DNA extraction

The leaves were lyophilized, ground in a mill and transferred to 5 mL microtubes. The material was stored at -20 °C. Total genomic DNA was extracted using the protocol described by Hoisington et al. (1994) modified. DNA was quantified on agarose gel (1%), using standard lambda phage DNA at different quantities (50, 100 and 200 ng).

SSR genotyping

Ten microsatellite markers developed by Chavariaga-Aguirre et al. (1998) and Mba et al. (2001) were synthesized (Table 2). The

amplification reactions contained 5 ng of DNA, 0.25 mM of each dNTP, 0.2 µM of each primer, 2 mM of MgCl₂, 1 U *Taq* polymerase and 1X PCR buffer in a total volume of 15 µL. The amplifications were conducted using a thermal cycler (Analytikjena®) with the following program: an initial cycle of 94 °C for 3 min, followed by 40 cycles of 94 °C for 30 s, 30 s at appropriate annealing temperature for each primer, 72 °C for 30 s and 72 °C for 30 min for a final extension. The amplification products were visualized on 3% agarose gel, using a 1 kb plus standard (Invitrogen®). Each forward primer was labeled with 6-carboxy-fluorescein fluorescence (FAM) for analysis in an AB 3500xL automated sequencer (Applied Biosystems Hitachi®), using GeneScan 600 LIZ as an internal marker to infer the size of the alleles. Data collection and analysis were carried out using GeneMarker v.2.7.4 software.

Table 1. Accessions of cassava (*Manihot esculenta* Crantz) belonging to the collection located in Embrapa Acre.

Nº	Varieties	Origin	Use	Nº	Varieties	Origin	Use	Nº	Varieties	Origin	Use	Nº	Varieties	Origin	Use
1	Muxuanga	AM	B	28	João Grande	AM	B	55	Goela de Jacu	AC	S	82	Branquinha 1* ¹	AM	B
2	Gavião	AM	B	29	IM194	AM	B	56	Primavera	AC	S	83	Branquinha 2* ¹	AC*	B
3	06 Meses	AM	B	30	IM196	AM	B	57	Fécua Branca	SP	S	84	Paxiúba 1* ²	AC	B
4	João Gonçalves	AM	B	31	IM198	AM	B	58	Corcunda	SP	B	85	Paxiúba 2* ²	AC	S
5	Bonitinha	AM	B	32	IM319	AM	B	59	IAC Caipora	SP	S	86	Milagrosa 1* ³	AC	B
6	Acari	AC	B	33	IM322	AM	B	60	Capora	AC	S	87	Milagrosa 2* ³	AC*	B
7	IM201	AM	B	34	IM486	AM	B	61	N2 Cascuda	AC	B	88	Chico Anjo 1* ⁴	AC*	B
8	IM204	AM	S	35	Peruana	AM	B	62	FB2	AC	S	89	Chico Anjo 2* ⁴	AC*	B
9	IM205	AM	B	36	Noé	AC	B	63	BRS Kiriris	NI	B	90	Caboquinha 1* ⁵	AC	S
10	IM214	AM	B	37	Curumim Doida	AC	B	64	BRS rosada	NI	S	91	Caboquinha 2* ⁵	AC*	B
11	IM215	AM	B	38	Fortaleza	AC	S	65	BRS Gema de Ovo	NI	S	92	Pirarucu 1* ⁶	AC	B
12	IM220	AM	B	39	Curumim Mansa	AC	S	66	BRS Jari	NI	S	93	Pirarucu 2* ⁶	AC*	B
13	IM221	AM	B	40	Amarelinha	AC	B	67	Roxinha	AC*	B	94	Varejão 1* ⁷	AC	B
14	Verdinha	AC	B	41	Baianinha	AC	B	68	Curumiem Doida	AC*	B	95	Varejão 2* ⁷	AC	B
15	Cariri	AC	B	42	Rosada	AC	B	69	Curumiem Branca	AC*	B	96	Amarela 1* ⁸	AC	S
16	Chica de Coca	AC	B	43	Emílio Borges	AC	S	70	Pinhão Manso	AC*	B	97	Amarela 2* ⁸	AC	S
17	Cabocla	AC	B	44	Paranacre	AC	S	71	Juriti	AC*	B	98	Cumarú 1* ⁹	AC	S
18	Ferreirão	AC	B	45	Manteguinha	AC	S	72	Boa Fé	AC*	B	99	Cumarú 2* ⁹	AC	B
19	Baiana Roxa	AC	S	46	Caipora	AC	S	73	Curumiem Roxa	AC*	B	100	Rasgadinha 1* ¹⁰	AC	B
20	Pão	AC	B	47	MX-26	RR	S	74	Mulatinha	AC*	B	101	Rasgadinha 2* ¹⁰	AC	S
21	Aruari	AC	S	48	MX_2	AM	S	75	Santa Maria	AC*	B	102	Metro 1* ¹¹	AC	S
22	Camparia	AC	S	49	MD_35	AM	B	76	Curumiem Torcida	AC*	B	103	Metro 2* ¹¹	AC	S
23	Baixinha	AC	B	50	Cruzeiro do Sul	AC	S	77	Mansa Brava	AC*	B	104	Agromazon 1* ¹²	AC	S
24	Xerém	AC	S	51	Hélia	AC	S	78	Mansi Brava	AC*	B	105	Agromazon 2* ¹²	AC	S
25	Casquinha	AC	S	52	Novo Ideal	AC	S	79	Paxiubão	AC*	B	106	Agromazon 3* ¹²	AC	S
26	Pretinha	AC	S	53	Sutinga	AC	S	80	BRS Ribeirinha	AC*	B				
27	Grilo Roxo	AM	S	54	Pãozinho	AC	S	81	BRS Tapioqueira	AC*	B				

AC = State of Acre, Brazil. AC* = Vale do Juruá, Cruzeiro do Sul, Acre, Brazil. AM = State of Amazonas, Brazil. RR = State of Roraima, Brazil. SP = State of São Paulo, Brazil. NI = No information. B = Bitter cassava. S = Sweet cassava. *ⁿ = Genotypes with the same nomenclature collected at different places.

Table 2. Sequence of the ten forward (F) and reverse (R) primers used with their respective annealing temperatures (Ta °C) and allelic size range in base pairs (bp).

Microsatellite Loci	Primer sequences (5' – 3')	Ta (°C)	Allelic size range (bp)
GA12*	F: GATTCCTCTAGCAGTTAAGC R: CGATGATGCTCTTCGGAGGG	55	131 – 157
GA126*	F: AGTGGAAATAAGCCATGTGATG R: CCCATAATTGATGCCAGGTT	55	178 – 214
GA131*	F: TTCCAGAAAGACTTCCGTTCA R: CTCAACTACTGCACTGCACTC	55	75 – 119
GA134*	F: ACAATGTCCCAATTGGAGGA R: ACCATGGATAGAGCTCACCG	55	309 – 337
GA136*	F: CGTTGATAAAGTGGAAAGAGCA R: ACTCCACTCCCGATGCTCGC	55	145 – 161
GA140*	F: TTCAAGGAAGCCTTCAGCTC R: GAGCCACATCTACTCGACACC	55	154 – 164
GAGG5*	F: TAATGTCATCGTCGGCTTCG R: GCTGATAGCACAGAACACAG	55	109 – 127
SSRY13**	F: GCAAGAATTCCACCAGGAAG R: CAATGATGGTAAGATGGTGCAG	55	193 – 245
SSRY89**	F: AGTTGAGAAAACCTTGCATGAG R: GGCTGTTTCGTGATCCTTATTAAC	55	102 – 117
SSRY164**	F: TCAAACAAGAATTAGCAGAACTGG R: TGAGATTTTCGTAATATTCATTTCACTT	45	134 – 182

*CHAVARRIAGA-AGUIRRE et al. (1998). **MBA et al. (2001).

Statistical analysis

For analysis of genetic diversity, the following parameters were estimated: number of alleles (N) per locus, observed (H_o) and expected heterozygosity (H_e). Analysis were carried out using the Genetic Data Analysis – GDA v.1.1 software (LEWIS; ZAYKIN, 2001). The mean number of alleles per locus was obtained by dividing the total number of alleles by the total number of loci.

The polymorphic information content (PIC) was calculated with TFPGA v.1.3 (MILLER, 1997) using the following equation by Botstein et al. (1980):

$$PIC = 1 - \sum_{i=1}^n f_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2f_i^2 f_j^2$$

Where f_i is the frequency of the i -th allele, f_j is the frequency of the j -th allele and the sum extends over n alleles.

Genetic distances were calculated using the modified Rogers' genetic distance matrix (WRIGHT, 1978) obtained by the TFPGA v.1.3 software (MILLER, 1997), with UPGMA (Unweighted Pair Group Method with Arithmetic Mean) hierarchical method and used in the

dendrogram construction in the MEGA X software (KUMAR et al., 2018).

Using the Bayesian method in the STRUCTURE software v.2.3 (PRITCHARD et al., 2000) the number of different gene pools (K) was verified, assuming Hardy-Weinberg equilibrium and absence of linkage disequilibrium between analyzed loci. Thus, groups of individuals sharing the same gene pool are grouped, with no need for prior information on their origin. The non-admixture model was used. Following a burn-in period of 100,000, 20 independent runs were carried out for each value of K (from 1 to 10), with 1000 simulations of MCMC (Markov Chain Monte Carlo). The determination of K (ΔK) was according to Evanno, Regnaut & Goudet (2005) by the Structure Harvester v.0.6.9 software (EARL; VON HOLDT, 2012).

The same genetic distance matrix of modified Rogers (WRIGHT, 1978), obtained by the TFPGA software, were used to verify the relationship between the accessions through Principal Coordinates Analysis (PCoA) calculated in Darwin software v.6.0.15 (PERRIER; JACQUEMOUND-COLLET, 2006). Also, in this software, the clustering analysis was performed

using the Neighbor-Joining (NJ) method.

To determine the core collection, the CoreFinder v.1.1 software was used based on the M strategy (POLICRITI; SGARRO, 2018). This strategy uses Monte Carlo simulations to select a group of genotypes with higher allelic richness (GOUESNARD et al., 2001).

RESULTS AND DISCUSSION

All loci were polymorphic. The ten microsatellites amplified 84 alleles with an average of 8.4 per locus. The GA12 locus had the lowest number of alleles, which identified only four alleles.

And the locus with the highest number of alleles was SSRY13 with 15 alleles (Table 3).

Similar values were found by Sousa et al. (2017) with 470 accessions of the germplasm bank located at Embrapa Amazônia Ocidental. They detected 113 alleles, with an average of 11.3 alleles per locus. Mühlen et al. (2013) identified 9.9 alleles per locus in 494 accessions belonging to cassava collections in the Amazon, Central-West and Southeast of Brazil. Even with larger samples size of these studies, the average of alleles per locus was similar to that presently found, demonstrating that the accessions are consistently representative of genetic diversity of the species.

Table 3. Characterization of ten loci in number of alleles per locus (N); Expected heterozygosity (H_E); Observed heterozygosity (H_O) e polymorphism content (PIC) for 106 accessions of *Manihot esculenta* Crantz.

Loci	N	H_E	H_O	PIC
GA12	4	0,54	0,54	0,54
GA126	10	0,83	0,80	0,84
GA131	10	0,80	0,70	0,82
GA134	5	0,49	0,44	0,51
GA136	11	0,81	0,65	0,81
GA140	7	0,72	0,62	0,73
GAGG5	5	0,58	0,51	0,64
SSRY13	15	0,89	0,74	0,89
SSRY89	6	0,64	0,24	0,65
SSRY164	11	0,75	0,56	0,75
Total	84			
Mean	8,4	0,71	0,58	0,72

Plants with crossed mating system (allogamy) and vegetative propagation such as cassava, usually present high values of heterozygosity (ORTIZ et al., 2016). The average expected heterozygosity between the loci was 0.71, ranging from 0.49 (GA134) to 0.89 (SSRY13). The mean observed heterozygosity was 0.58 and ranged from 0.24 (SSRY89) to 0.80 (GA126) (Table 3).

The high levels of heterozygosity may indicate that gene flow and recombination play a significant role in maintaining genetic diversity of *M. esculenta* (MEZETTE et al., 2013). However, it is noted that the values of H_O were lower when compared to those of H_E in most of the loci, indicating a slight deficit of heterozygotes. Other studies from Brazilian Amazon also detected a deficit in the heterozygosity, being a presumable consequence of asexual propagation and immediate

fixation of cassava genotypes (ALVES-PEREIRA et al., 2011; MÜHLEN et al., 2013).

These values of heterozygosity, as average of alleles, were similar to those found in studies with cassava accessions from Pará, Amapá and Amazonas states, that found values from 0.55 to 0.71 (MOURA et al., 2013; MOURA-CUNHA et al., 2016; SILVA et al., 2016; SOUSA et al., 2017). However, the heterozygosity levels were higher than those found in studies carried out with germplasm banks from the Southeast and Southern regions of Brazil, which detected values from 0.26 to 0.47 (SIQUEIRA et al., 2009; SIQUEIRA et al., 2010; ORTIZ et al., 2016). These results demonstrate the wide genetic diversity in Amazon, which corroborates with studies that indicate it as center of origin and domestication.

There was variation in the content of

polymorphism between the loci. The average among the ten microsatellites was 0.72. The locus SSRY13 presented the highest value (0.89) and GA134 the lowest polymorphism content (0.51) (Table 3). All loci had PIC values greater than 0.50. According to Botstein et al. (1980), loci with PIC values above 0.50 are highly informative.

Based on the Rogers's modified genetic distance (WRIGHT, 1978) the accessions were

grouped by the UPGMA method (Figure 1). The greatest distance (0.908) was observed between the four redundant accessions "Caboquinha2", "Paxiubão", "BRS Ribeirinha" and "Pirarucu2", and the variety "IM194". There are differences between these accessions in relation to the color of the new sprouting, petiole color, lobe morphology, branching habits, stem color and productivity (SILVA, 2010; BERGO; SIVIERO; LESSA, 2015).

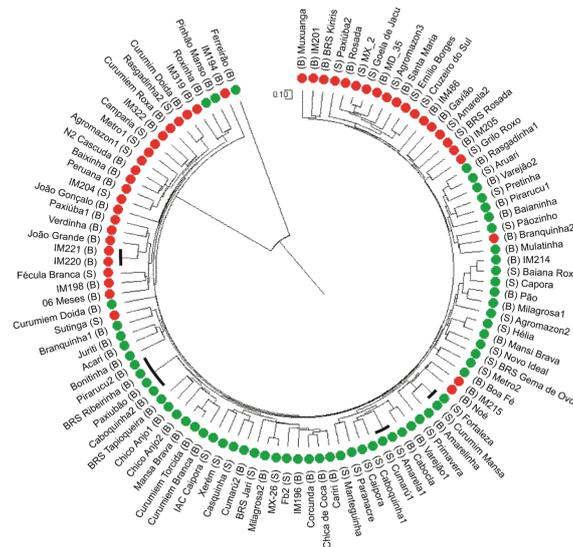


Figure 1. Dendrogram with 106 accessions of *Manihot esculenta* from Embrapa Acre collection obtained by the UPGMA method, using the modified Rogers distance (Wright, 1978) with ten microsatellites. The green and red circles refer to the groupings obtained by the STRUCTURE software for $K = 2$. The highlighted terminals refer to the redundant genotypes.

The accession "Ferreirão" was highly divergent and was positioned as an external group in the dendrogram (Figure 1). Siviero et al. (2009) also detected "Ferreirão" accession highlighted from the others because it presented higher starch content, dry matter and resistance to root rot when compared to the analyzed varieties, associated with a good productivity per t/ha. Thus, it is an interesting divergent genotype and the genetic distances obtained by molecular data also corroborates with the morphological results.

Through the dendrogram it was possible to detect redundancy with individuals with different names. Six accessions (5.66%) presented identical genetic profiles, in three distinct groups: accessions "IM220" and "IM221", bitter genotypes from Amazonas state; "Curumim Mansa" and "Fortaleza", "Amarela 1" and "Cumarú 1" that are sweet varieties from Acre state (Figure 1). A fourth redundant group with "Caboquinha2", "Paxiubão", "BRS Ribeirinha" and "Pirarucu2" presented the same genetic profile in all loci. These four accesses were collected in Vale do Juruá and belongs to the same flour producer. Bergo, Siviero & Lessa (2015)

identified the same productivity averages among these accessions and Silva (2010) found the same morphological characters between them corroborating with the molecular data. These genotypes were named differently by the farmer, but they are clones.

Moura et al. (2013) identified that approximately half of the collection of Embrapa Amazônia Oriental was composed by duplicate accesses.

The identification of redundancies in collections of vegetatively propagated species is common (IRISH et al., 2010; RIBEIRO et al., 2011; GROSS et al., 2012), probably due to material exchanges between farmers (MOURA et al., 2016) where the genotype can be given another name in the new location. The presence of duplicates accessions represents substantial expenditure in the preservation of cassava germplasm under field conditions. For this reason, it is fundamental to use tools to identify redundancies, when incorporating new accessions to a germplasm, especially in cassava collections.

All individuals with the same name, highlighted (*ⁿ) in Table 1, had distances greater than 0.547 with each other (Figure 1). The variety denomination is probably carried out independently by farmers from different locations and represent a distinct genotype belonging to the same variety (MOURA; FARIAS NETO, 2017). Sambatti et al. (2001) analyzed the local names of varieties of *M. esculenta* and perceived the tendency of the use of perceptual characters to designate them, as the petiole color, color of the apical bud, color of the root pulp; and characteristics of cooking and taste.

It can be also considered the introduction of varieties, originated from intervarietal crosses that escape human control and are unconsciously propagated and assigned to an existing variety based on morphological similarity (ELIAS et al., 2001).

Such low accuracy from farmers in varieties differentiation can lead to the incorrect designation. However, the discard of accessions based on the nomenclature should be avoided (MOURA; FARIAS NETO, 2017), requiring the prior characterization of the materials by means of morphological, agronomic and preferentially by molecular markers.

Through the Bayesian analysis implemented by the STRUCTURE software and the statistics of Evanno, Regnaut & Goudet (2005) was calculated the highest delta K value for K=2, highlighting the genetic differentiation of the 106 accessions of *M. esculenta* in two different groups (Figure 2).

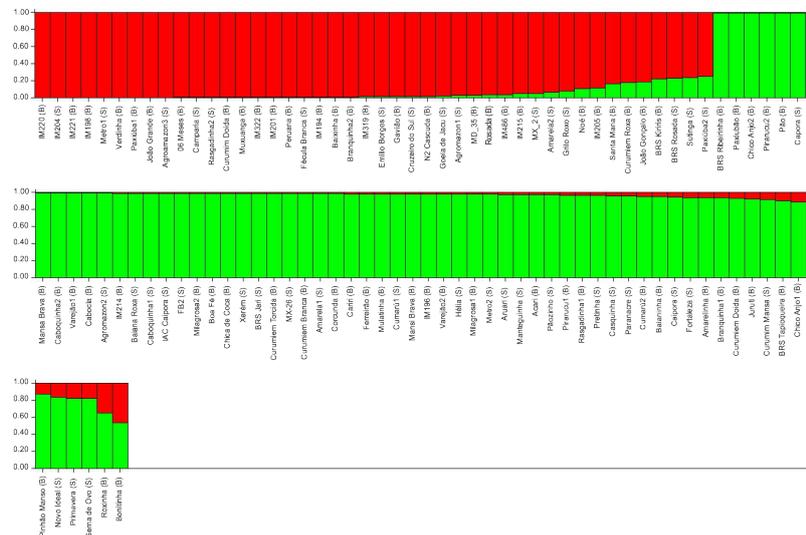


Figure 2. Representation of the 106 accessions of *Manihot esculenta* according to the Bayesian analysis of the STRUCTURE software. Each individual is represented by a column, sorted according to the probability value Q. The evaluated accessions were divided into 2 groups (K = 2): group I represented by red color; group II represented by green color.

The gene pool I represented by the red color (Figure 2) grouped 44 accessions (41.5%), being 28 accessions of “bitter” cassava and 16 of “sweet” cassava. The gene pool presented in green color grouped 62 accessions (58.5%), where 38 are “bitter” and 24 “sweet” cassava. The two gene pools had weak genetic structure corresponding to information on HCN content, as there was practically the same proportion of “bitter” and “sweet” varieties in both groups.

The distinction between “bitter” and “sweet” cassava was detected in some studies (PERONI; KAGEYAMA; BEGOSSI, 2007; MÜHLEN et al., 2013). Sousa et al. (2017) and Bradbury et al. (2013) also didn’t find correlation of the cyanogenic content with the genetic structure.

Those varieties can hybridize between them in the cropping systems, associated with incorporation, exchange and diffusion of hybrid seedlings between farmers, and leading to low genetic distinction between “sweet” and “bitter” varieties, as a consequence. In addition, the differentiation between “bitter” and “sweet” cassava by farmers can vary regionally and even among communities (MOURA; FARIAS NETO, 2017). Therefore, mixtures due to intraspecific hybridization between varieties with different cyanogenic contents at the farming fields; error in the distinction of varieties by the farmers, which are based, mostly, on the taste of the uncooked roots (BRADBURY et al., 2013), thus allowing subjectivity in the classification of the cultivars; and as a consequence, incorrect passport

data at the time of introduction of the accessions in the collection may have contributed to the low consistency of the genetic structure found in this study.

The two subpopulations resulting from the Bayesian analysis ($K=2$) did not correlate with the origin of the accessions. The gene pool I grouped most accessions from the state of Amazonas, while the second gene pool gathered most of the varieties from the state of Acre, which was expected since these accessions were collected in the same region (MÜHLEN et al., 2013; MOURA et al., 2016; RABBI et al., 2015). However, some accessions were grouped in a way that had no correspondence to the collection sites. This weak correlation is possibly due to the lack of geographical barriers promoted by the intense exchange of genotypes among the farmers of the Amazon region (MEZETTE et al., 2013; SOUSA et al., 2017), in such a way that the collection site no longer reflects the origin of the genotype. Besides, the inflorescence architecture and seed dispersal mechanism of *M. esculenta* facilitates cross-fertilization between neighboring plants (KAWUKI

et al., 2013) favoring allele exchange, which is associated with vegetative propagation and trading between farmers, enables the dissemination of genotypes (MEZETTE et al., 2013).

The association between the accessions was presented in a multidimensional plane through the PCoA (Figure 3). The first two principal coordinates represented 20.54% of the total variation. This means that in two dimensions it was possible to observe most of the variation found. The accumulated variation in the first two principal coordinates was low, as in the study by Sousa et al. (2017), that analyzed cassava varieties and identified only 17.84% of the variation in two coordinates and, as in the present study, were not able to detect genetic structure corresponding to the cyanogenic content of the varieties, attributing this to misidentification of “sweet” or “bitter” varieties and possible mixture of accessions during clonal propagation (on farm and *ex situ*). However, Costa et al. (2013) and Gonçalves et al. (2017) when evaluating “sweet” cassava managed to explain much of the variation in two coordinates (53.25% and 48.21%, respectively).

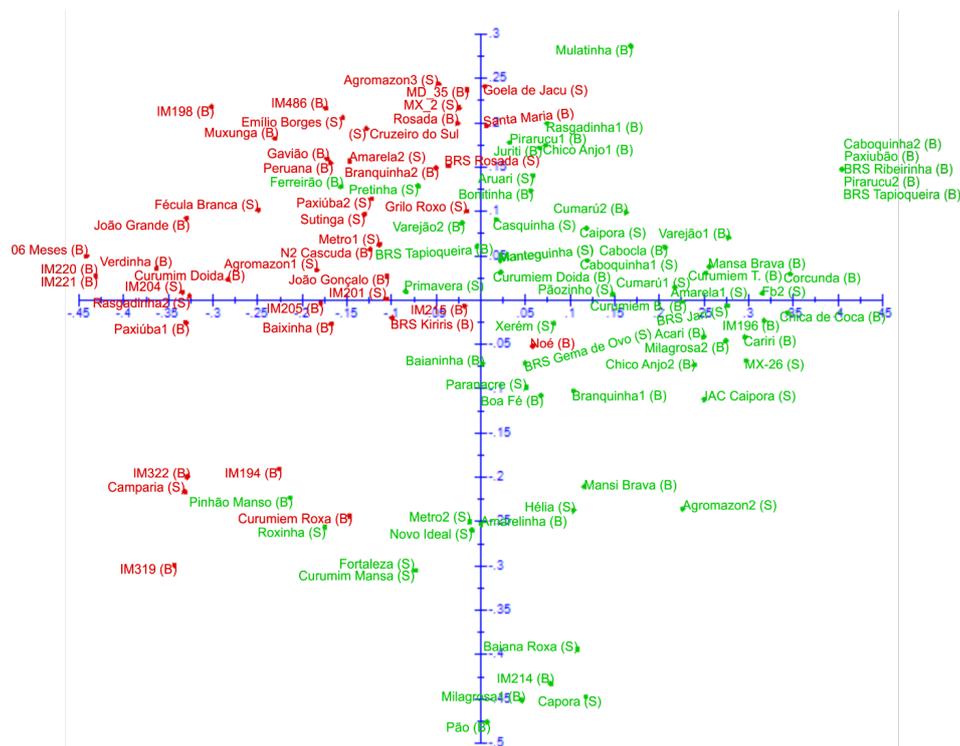


Figure 3. Graphical representation of the 106 accessions of *M. esculenta* based on the principal coordinates analysis. The colors are in agreement with the formation of the groups generated by the STRUCTURE software, considering $K = 2$.

Comparing the Bayesian analysis data with PCoA, the observed dispersion supported the establishment of the two groups identified by the Bayesian analysis, emphasizing the grouping of the

duplicate accessions and also showing no correspondence either with cyanogenic content of the accessions and with the genotypes collection site.

The 84 identified alleles can be found in 30 individuals representing total allelic diversity, with 28,30% of the accessions that constitute the collection. However, for the establishment of a core collection, this percentage is considered high. This subset should represent between 70 and 80% of the diversity of the base collection with only 5 to 20% of individuals (BROWN; SPILLANE, 1999).

Therefore, to identify the core collection containing 20,75% of the accessions of the base

collection, 22 individuals representing 94% of the allelic diversity were required (Table 4). Of these, 10 accessions were grouped in the gene pool I, and 12 accessions were allocated in the gene pool II. As for the cyanogenic content of the individuals that composed the core collection, there is 16 “bitter” varieties and 6 “sweet” varieties. Figure 4 shows the number of accessions of the core collection and the allelic diversity represented in this subset.

Table 4. Identification of the 22 accessions of *Manihot esculenta* for the formation of the core collection of the cassava collection of Embrapa Acre.

Nº	Varieties	Use	Nº	Varieties	Use
1	Bonitinha	B	17	Pão	B
2	Curumim Doida	B	18	BRS Jari	S
3	Chico Anjo2	B	19	Metro 2	S
4	Curumim Doida	B	20	BRS Rosada	S
5	IM201	B	21	IM198	B
6	FB2	S	22	Ferreirão	B
7	Amarela2	S	23	Baixinha	B
8	IM205	B	24	Pinhão Manso	B
9	Caboquinha2	B	25	IM319	B
10	Paxiúba2	S	26	MX-26	S
11	IM215	B	27	Curumim Torcida	B

B = “Bitter” cassava; S = “Sweet” cassava.

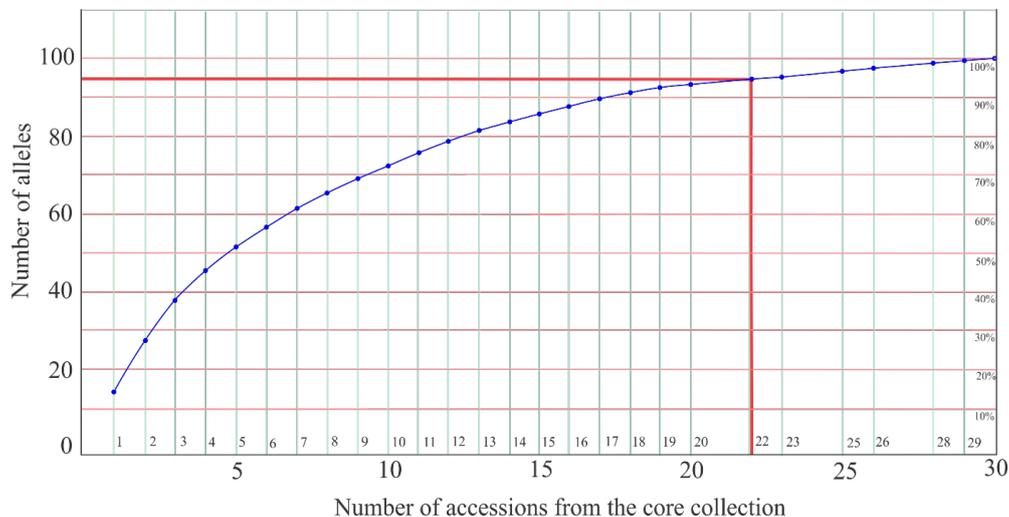


Figure 4. Total number of accessions and percentage of alleles included in the cassava core collection.

The core collection may guide the preservation of the genetic variability detected in the collection, as well as to facilitate access to genetic information contained in these genotypes by cassava breeding programs. Since cassava is a vegetatively propagated crop and its conservation in the Embrapa Acre collection is maintained under field conditions, the identification of redundancies may impact the maintenance costs of the germplasm and the use of

these accessions in breeding programs, avoiding crosses between duplicates.

ACKNOWLEDGEMENTS

The authors gratefully thank Embrapa, CNPq for the financial support through the project 407991/2013-1 FERRAMENTAS MOLECULARES PARA ESTUDO DA

VARIABILIDADE GENÉTICA DA MANDIOCA NA AMAZÔNIA, and Capes for granting scholarships.

RESUMO: A mandioca é uma das mais importantes culturas de subsistência em países tropicais. É preciso conservar e conhecer a diversidade genética para o uso adequado dos recursos genéticos. A avaliação da diversidade genética entre os genótipos resulta em informações sobre potenciais genitores em programas de melhoramento, possibilita a identificação de duplicatas, além disso, facilita o intercâmbio de germoplasma entre instituições de pesquisa. O objetivo deste trabalho foi caracterizar a diversidade genética dos acessos de mandioca da Região Norte do Brasil. Foram analisados 106 acessos por meio de dez marcadores microsatélites. Os parâmetros de diversidade genética estimados foram: heterozigosidade esperada (H_E), heterozigosidade observada (H_O) e conteúdo de informação polimórfica (PIC). Agrupamentos foram realizados pelo método UPGMA e *Neighbor-Joining* (NJ). Utilizou-se também análises bayesianas, dispersão por coordenadas principais e a identificação de uma coleção nuclear. Os dez locos amplificaram 8,40 alelos em média. A média das estimativas de diversidade foram altas: $H_E = 0,71$, $H_O = 0,58$ e $PIC = 0,72$. As distâncias genéticas variaram de 0,158 a 0,908. Seis (5,66%) acessos estão redundantes. Os agrupamentos e análises de dispersão não evidenciaram distinção entre variedades bravas e mansas e não foi identificada estrutura genética correspondente a origem dos acessos. A coleção nuclear foi formada por 22 indivíduos, que representaram 94% da diversidade alélica total e 20,75% da coleção base. Os resultados indicam alta dissimilaridade entre os acessos e permitiram a detecção de genótipos redundantes, mostrando o uso de marcadores genéticos como ferramentas informativas para o manejo de coleções.

PALAVRAS-CHAVE: Coleção nuclear. Mandioca. *Manihot esculenta*. Marcadores microsatélites. Recursos genéticos. Variabilidade.

REFERENCES

- ALBUQUERQUE, H. Y. G. **Diversidade genética e identificação de duplicatas de *Manihot esculenta* Crantz com base em marcadores Single-Nucleotide Polymorphism (SNP)**. 2017. 113 f. Dissertação (Mestrado em Recursos Genéticos Vegetais) – Programa de Pós-Graduação em Recursos Genéticos Vegetais, Universidade Federal do Recôncavo da Bahia, Cruz das Almas, 2017.
- ALLEM, A. C. The origin of *Manihot esculenta* Crantz (Euphorbiaceae). **Genetic Resources and Crop Evolution**, Bari, v. 41, n. 3, p. 133-150, 1994. <https://doi.org/10.1007/BF00051630>
- ALVES-PEREIRA, A., et al. Genetic structure of traditional varieties of bitter manioc in three soils in Central Amazonia. **Genetica**, Bowling Green, v. 139, n. 10, p. 1259-1271, 2011. <https://doi.org/10.1007/s10709-011-9627-4>
- BERGO, C. L.; SIVIERO, A.; LESSA, L. S. Produtividade e teor de amido de sete genótipos de mandioca em diferentes épocas em Rio branco, Acre. In: CONGRESSO BRASILEIRO DE MANDIOCA E CONGRESSO LATINO-AMERICANO E CARIBENHO DE MANDIOCA, 16., Foz do Iguaçu. **Anais...** Foz do Iguaçu: SBM, 2015.
- BOTSTEIN, D., et al. Construction of a genetic map in man using restriction fragment length polymorphism. **American Journal Human Genetics**, Birmingham, v. 32, n. 3, p. 314-331, 1980.
- BRADBURY, E. J., et al. Geographic differences in patterns of genetic differentiation among bitter and sweet manioc (*Manihot esculenta* subsp. *esculenta*; Euphorbiaceae). **American Journal of Botany**, St. Louis, v. 100, n. 5, p. 857-866, 2013. <https://doi.org/10.3732/ajb.1200482>
- BROWN, A. H. D.; SPILLANE, C. Implementing core collections principles procedures, progress, problems and promise. In: JOHNSON, R. C.; HODGKIN, T. (Ed.) **Core collections for today and tomorrow**. Roma: IPGRI, 1999. p. 1-9.

CHAVARRIAGA-AGUIRRE, P., et al. Microsatellites in cassava (*Manihot esculenta* Crantz): discovery, inheritance and variability. **Theoretical and Applied Genetics**, Stuttgart, v. 97, n. 3, p. 493-501, 1998. <https://doi.org/10.1007/s001220050922>

COSTA, T. R., et al. Genetic diversity and population structure of sweet cassava using simple sequence repeat (SSR) molecular markers. **African Journal of Biotechnology**, Sapele, v. 12, n. 10, p. 1040-1048, 2013. <http://doi.org/10.5897/AJB12.2727>

DUFOUR, D. L. Cyanide content of cassava (*Manihot esculenta*, Euphorbiaceae) cultivars used by Tukanoan indians in Northwest Amazonia. **Economic Botany**, Fullerton, v. 42, n. 2, p. 255-266, 1988. <https://doi.org/10.1007/BF02858929>

EARL, D. A.; VON HOLDT, B. M. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. **Conservation Genetics Resources**, Cardiff, v. 4, n. 2, p. 359–361. 2012. <https://doi.org/10.1007/s12686-011-9548-7>

ELIAS, M., et al. Unmanaged sexual reproduction and the dynamics of genetic diversity of a vegetatively propagated crop plant, cassava (*Manihot esculenta* Crantz), in a traditional farming system. **Molecular Ecology**, Vancouver, v. 10, n. 8, p. 1895-1907, 2001. <https://doi.org/10.1046/j.0962-1083.2001.01331.x>

EMPERAIRE, L.; PERONI, N. Traditional management of agrobiodiversity in Brazil: a case study of manioc. **Human Ecology**, New York, v. 35, n. 6, p. 761-768, 2007. <https://doi.org/10.1007/s10745-007-9121-x>

EVANNO, G.; REGNAUT, S; GOUDET, J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. **Molecular Ecology**, Vancouver, v. 14, n. 8, p. 2611- 2620, 2005. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>

FAO, 2017. **Food Outlook: Biannual report on global food markets – November 2017**. Available at: <<http://www.fao.org>>. Accessed in: 03. June. 2018.

FREGENE, M. A., et al. Simple sequence repeat marker diversity in cassava landraces: genetic diversity and differentiation in an asexually propagated crop. **Theoretical and Applied Genetics**, Stuttgart, v. 107, n. 6, p. 1083-1093, 2003. <https://doi.org/10.1007/s00122-003-1348-3>

GONÇALVES, T. M., et al. Genetic diversity and population structure of traditional sweet cassava accessions from Southern of Minas Gerais State, Brazil, using microsatellite markers. **African Journal of Biotechnology**, Sapele, v. 16, n. 8, p. 346-358, 2017. <https://doi.org/10.5897/AJB2016.15768>

GOUESNARD, B., et al. MSTRAT: an algorithm for building germplasm core collections by maximizing allelic or phenotypic richness. **Journal of Heredity**, College Station, v. 92, n. 1, p. 93–94. 2001. <https://doi.org/10.1093/jhered/92.1.93>

GROSS, B. L., et al. Identification of “duplicate” accessions within the USDA-ARS National plant germplasm system *Malus* collection. **Journal of the American Society for Horticultural Science**, Watsonville, v. 137, n. 5, p. 333–342, 2012. <https://doi.org/10.21273/JASHS.137.5.333>

HERSHEY, C. H. *Manihot* genetic diversity. In: MEETING OF THE INTERNATIONAL NETWORK FOR CASSAVA GENETIC RESOURCES, 1., Cali. **Anais...** Cali: CIAT, 1994.

HOISINGTON, D.; KHAIRALLHAH, M.; DE LEON, D. G. **Laboratory protocols: CIMMYT applied molecular genetics laboratory**. 2 ed. México: CIMMYT, 1994. 102 p.

IBGE (2018) **Instituto Brasileiro de Geografia e Estatística**. Available at: <<http://ibge.com.br>>. Accessed in: 07. June. 2018.

- IRISH, B. M., et al. Microsatellite fingerprinting of the USDA-ARS tropical agriculture research station cacao (*Theobroma cacao* L.) germplasm collection. **Crop Science**, Madison, v. 50, n. 2, p. 656–667, 2010. <https://doi.org/10.2135/cropsci2009.06.0299>
- KAWUKI, R. S., et al. Genetic diversity of cassava (*Manihot esculenta* Crantz) landraces and cultivars from southern, eastern and central Africa. **Plant Genetic Resources**, Wageningen, v. 11, n. 2, p. 170-181, 2013. <https://doi.org/10.1017/S1479262113000014>
- KUMAR, S., et al. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. **Molecular Biology and Evolution**, Lawrence, v. 35, n. 6, p. 1547-1549, 2018. <https://doi.org/10.1093/molbev/msy096>
- LEWIS, P.O.; ZAYKIN, D. GDA (Genetic Data Analysis): **Computer Program for the Analysis of Allelic Data**. Version 1.1, University of Connecticut, Storrs. 2001. Available at: <<http://en.bio-soft.net/dna/gda.html>>. Accessed in: 27. May. 2018.
- MBA, R. E. C., et al. Simple sequence repeat (SSR) markers survey of the cassava (*Manihot esculenta* Crantz) genome: towards an SSR-based molecular genetic map of cassava. **Theoretical and Applied Genetics**, Stuttgart, v. 102, n. 1, p. 21-31, 2001. <https://doi.org/10.1007/s001220051614>
- MEZETTE, T. F.; BLUMER, C. G.; VEASEY, E. A. Morphological and molecular diversity among cassava genotypes. **Pesquisa Agropecuária Brasileira**, Brasília, v. 48, n. 5, p. 510-518, 2013. <https://doi.org/10.1590/S0100-204X2013000500007>
- MILLER, M. P. Tools for population genetic analyses (TFPGA): **A Windows program for the analysis of allozyme and molecular population genetic data, version 1.3**. Northern Arizona University, Arizona, 1997.
- MOURA, E. F.; FARIAS NETO, J. T. **Diferenciação genética entre variedades de mandioca com a mesma nomenclatura coletadas na Amazônia Brasileira**. Belém: Embrapa Amazônia Oriental, 2017. 20 p. (Embrapa Amazônia Oriental. Boletim de pesquisa e desenvolvimento, 113).
- MOURA, E. F., et al. Identification of duplicates of cassava accessions sampled on the North Region of Brazil using microsatellite markers. **Acta Amazonica**, Manaus, v. 43, n. 4, p. 461-468, 2013. <https://doi.org/10.1590/S0044-59672013000400008>
- MOURA, E. F., et al. Molecular characterization of accessions of a rare genetic resource: sugary cassava (*Manihot esculenta* Crantz) from Brazilian Amazon. **Genetic Resources and Crop Evolution**, Bari, v. 63, n. 4, p. 583-593, 2016. <https://doi.org/10.1007/s10722-016-0378-z>
- MOURA-CUNHA, E. F., et al. Molecular characterization of “sweet” cassavas (*Manihot esculenta*) from a germplasm bank in Brazilian Eastern Amazonia. **Crop Breeding and Applied Biotechnology**, Viçosa, v. 16, n. 1, p. 28-34, 2016. <https://doi.org/10.1590/1984-70332016v16n1a5>
- MÜHLEN, G. S., et al. Genetic diversity and differentiation of Brazilian bitter and sweet manioc varieties (*Manihot esculenta* Crantz, Euphorbiaceae) based on SSR molecular markers. **Tipiti: Journal of the Society for the Anthropology of Lowland South America**, Amherst, v. 11, n. 2, p. 66-73, 2013.
- OLSEN, K. M. SNPs, SSRs and inferences on cassava’s origin. **Plant Molecular Biology**, Yokohama, v. 56, n. 4, p. 517-526, 2004. <https://doi.org/10.1007/s11103-004-5043-9>
- OLSEN, K. M.; SCHAAL, B. A. Microsatellite variation in cassava (*Manihot esculenta*, Euphorbiaceae) and its wild relatives: further evidence for a southern Amazonian origin of domestication. **American Journal of Botany**, St. Louis, v. 88, p. 131–142, 2001. <https://doi.org/10.2307/2657133>

ORTIZ, A. H. T., et al. Population structure and genetic diversity in sweet cassava cultivars from Paraná, Brazil. **Plant Molecular Biology Reporter**, Düsseldorf, v. 34, n. 6, p. 1153-1166, 2016.

<https://doi.org/10.1007/s11105-016-0995-x>

PERONI, N.; KAGEYAMA, P. Y.; BEGOSSI, A. Molecular differentiation, diversity, and folk classification of “sweet” and “bitter” cassava (*Manihot esculenta*) in Caiçara and Caboclo management systems (Brazil).

Genetic Resources and Crop Evolution, Bari, v. 54, n. 6, p. 1333-1349, 2007. <https://doi.org/10.1007/s10722-006-9116-2>

PERRIER, X.; JACQUEMOUND-COLLET, J. P. 2006. **DARwin software**. Available at:

<<http://darwin.cirad.fr/product.php>>. Accessed in: 02. June. 2018.

POLICRITI, A; SGARRO, A. **CoreFinder v.1.1**. 2018. Available at:

<<http://services.appliedgenomics.org/software/corecollections/>>. Accessed in: 05. April. 2019.

PRITCHARD, J. K.; STEPHENS, M.; DONNELLY, P. Inference of population structure using multilocus genotype data. **Genetics**, Aurora, v. 155, n. 2, p. 945-959, 2000.

RABBI, I. Y., et al. Tracking crop varieties using genotyping-by-sequencing markers: a case study using cassava (*Manihot esculenta* Crantz). **BioMed Central Genetics**, London, v. 16, n. 1, p. 1-11, 2015.

<https://doi.org/10.1186/s12863-015-0273-1>

RIBEIRO, M. N. O., et al. Genetic variability among cassava accessions based on SSR markers. **Crop**

Breeding and Applied Biotechnology, Viçosa, v. 11, n. 3, p. 263-269, 2011. <https://doi.org/10.1590/S1984-70332011000300009>

SAMBATTI, J. B. M.; MARTINS, P. S.; ANDO, A. Folk taxonomy and evolutionary dynamics of cassava: a case study in Ubatuba, Brazil. **Economic Botany**, Fullerton, v. 55, n. 1, p. 93-105, 2001.

<https://doi.org/10.1007/BF02864549>

SILVA, B. S. **Caracterização botânica e agrônômica da coleção de trabalho de mandioca da Embrapa Acre**. 2010. 75 f. Dissertação (Mestrado em Produção Vegetal) – Programa de Pós-Graduação em Produção Vegetal, Universidade Federal do Acre, Rio Branco, 2010.

SILVA, R. S., et al. Genetic divergence among accessions of cassava (*Manihot esculenta* Crantz) sampled in the Tapajós region, State of Pará, using agronomic characters and microsatellite markers. **Semina: Ciências Agrárias**, Londrina, v. 37, n. 5, p. 2989-3004, 2016. <https://doi.org/10.5433/1679-0359.2016v37n5p2989>

SIQUEIRA, M. V. B. M., et al. Genetic characterization of cassava (*Manihot esculenta*) landraces in Brazil assessed with simple sequence repeats. **Genetics and Molecular Biology**, Ribeirão Preto v. 32, n. 1, p. 101-110, 2009. <https://doi.org/10.1590/S1415-47572009005000010>

SIQUEIRA, M. V. B. M., et al. Microsatellite polymorphisms in cassava landraces from the Cerrado Biome, Mato Grosso do Sul, Brazil. **Biochemical Genetics**, Azores, v. 48, n. 9-10, p. 879-895, 2010.

<https://doi.org/10.1007/s10528-010-9369-5>

SIVIERO, A.; EVANGELISTA, R. C.; SCHOTT, B. Produtividade, teor de amido, matéria seca e resistência a podridão radicular de cultivares de mandioca do Acre. In: CONGRESSO BRASILEIRO DE MANDIOCA, 13., Botucatu. **Anais...** Botucatu: UNESP, 2009.

SOUSA, S. B., et al. Farmer variety exchange along Amazonian rivers influences the genetic structure of manioc maintained in a regional Brazilian GeneBank. **Genetics and Molecular Research**, Ribeirão Preto, v. 16, n. 3, p. 1 - 11, 2017. <https://doi.org/10.4238/gmr16039690>

Molecular characterization...

COSTA, N. A. et al.

VIEIRA, E. A., et al. Caracterização molecular e variabilidade genética de acessos elite de mandioca para fins industriais. **Ciência Rural**, Santa Maria, v. 40, n. 12, p. 2467-2471, 2010. <https://doi.org/10.1590/S0103-84782010001200005>

WRIGHT, S. **Evolution and the genetics of populations, vol 4: variability within and among natural populations**. Chicago: University of Chicago Press, 1978. 590 p.