GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Genetic diversity in natural populations of *Theobroma subincanum* Mart. in the Brazilian Amazon

L.H. Rivas[1], L.D. Giustina[1], L.N. Luz[2], I.V. Karsburg[1], T.N.S. Pereira[2] and A.A.B. Rossi[1]

[1]Laboratório de Genética Vegetal e Biologia Molecular,
Centro de Tecnologia da Amazônia Meridional,
Faculdade de Ciências Biológicas e Agrárias,
Campus Universitário de Alta Floresta,
Universidade do Estado de Mato Grosso, Alta Floresta, MT, Brasil
[2]Laboratório de Melhoramento Genético Vegetal,
Centro de Ciências e Tecnologias Agropecuárias,
Universidade Estadual do Norte Fluminense Darcy Ribeiro,
Campos dos Goytacazes, RJ, Brasil

Corresponding author: A.A.B. Rossi
E-mail: anabanrossi@gmail.com

Genet. Mol. Res. 12 (4): 4998-5006 (2013)
Received March 14, 2013
Accepted August 17, 2013
Published October 24, 2013
DOI http://dx.doi.org/10.4238/2013.October.24.12

ABSTRACT. The genus *Theobroma*, recently reclassified in the family Malvaceae, comprises some species with high economic potential, including the cupuí, *Theobroma subincanum* Mart., which has not yet been domesticated, and whose genetics and population structure are mostly unknown. This study aimed to assess the population structure and genetic diversity in natural populations of *T. subincanum* Mart., using inter-simple sequence repeat (ISSR) markers. A total of 59 individuals were sampled in three geographically separate populations, CFA, CMN, and CPT. Nei's genetic distance was estimated to characterize populations with the use of 13 polymorphic primers. The analysis of

molecular variance revealed that the variability between populations (51.71%) was higher than that within populations (48.29%). Among the three populations, CPT showed the highest diversity index and percentage of polymorphism. The ISSR molecular markers were efficient and presented sufficient polymorphism to estimate genetic diversity in populations of *T. subincanum* Mart.

Key words: Cupuí; ISSR; Genetic diversity

INTRODUCTION

Cupuí (*Theobroma subincanum* Mart.) is one of the most important species of the family Malvaceae (Alverson et al., 1999; Whitlock et al., 2000), formerly Sterculiaceae. It is distributed in the tropical rainforests of the southern hemisphere, from Mexico to the Southern Amazon rainforest. The genus *Theobroma* is comprised of 22 species (Cuatrecasas, 1964), but only four have been described as economically relevant, including cacao, *T. cacao* L. and cupuaçu, *T. grandiflorum* Willd. ex Spreng. K. Schum, in addition to *T. bicolor* Humb. & Bonpl. and *T. speciosum* Willd. ex Spreng, which are not as well known, but are believed to have economic potential.

Several other species of the genus *Theobroma* may have some practical use, which is the case of cupuí, *T. subincanum* Mart. This species can be found in the Amazon basin and spreads into the State of Mato Grosso. It is widely used in traditional communities because of its edible fruit, whose flesh can be consumed fresh or as juice. Additionally, its seeds are commonly used in the preparation of homemade chocolate (Souza et al., 1996). Wild species, such as cupuí, which provide income to traditional communities, are very useful for regional development, favor the rational use of genetic resources, and stimulate traditional populations to preserve plant genetic resources.

Furthermore, knowledge of the wild relatives of cultivated species can also potentially contribute to the breeding of cultivated species. In this respect, knowledge about the genetic and population relationships in wild species is particularly important because they are a great reservoir of genes for cultivated species (Silva et al., 2005, 2007). Despite the significant potential of cupuí, there is almost no information about its population dynamics or genetic behavior. Such information is important for the conservation of this species and others that are currently endangered by deforestation.

Effective strategies for the preservation of wild species and their genetic heritage should seek to quantify genetic diversity between native populations to support the rescue of threatened populations and to identify populations with greater genetic variability for use and conservation. Knowledge about genetic resources both within and between populations provides support for the conservation of genetic resources and increases understanding of the evolutionary dynamics of species (Hamrick, 1983). The understanding of variability within a population is essential for the study of evolutionary diversification and knowledge of different conservation and management actions.

The use of DNA markers in the quantification of genetic diversity is a very important tool for the analysis of the genetic structure of natural populations, since they can detect genetic differences and polymorphisms at the DNA level. Among these, inter-simple sequence

repeat (ISSR) markers have been widely used in studies of the genetic diversity and variability of wild populations. These markers stand out because they do not need prior information of the DNA sequence, have low development costs, present laboratory procedures with good rates of transferability for other plant species (Bhart et al., 2002), and generate a large number of polymorphic fragments in spite of the low-cost of the technique (Brandão et al., 2011). ISSR markers have proven to be efficient in the many studies on genetic variability of populations that have used them (Cidade et al., 2009; Rossi et al., 2009; Brandão et al., 2011).

This study aimed to characterize genetic diversity in three natural populations of *T. subincanum* through the use of ISSR markers, so as to contribute knowledge about the population dynamics of the species and to aid the establishment of programs for its conservation and breeding.

MATERIAL AND METHODS

Collection of genetic material and DNA extraction

Three natural populations of cupuí (*T. subincanum*) were assessed; two in the city of Alta Floresta, namely, Amazon Forest (CFA) and Mundo Novo (CMN), and one in the municipality of Paranaíta, Paranaíta population (CPT). Both cities are located in the far northern Mato Grosso (Table 1 and Figure 1). Leaves at the intermediate stage of maturation were collected for DNA extraction in all 59 individuals selected from the three study populations (Table 1).

Table 1. Populations analyzed and their respective code; number of individuals (N) and geographic location.

Populations	Code	N	Latitude (S)	Longitude (W)
Amazon florest	CFA	19	9°52'45.70"	56°6'5.92"
Mundo Novo	CMN	22	9°51'3.36"	55°53'5.75"
Paranaíta	CPT	18	9°43'30.81"	56°26'41.12"
Total		59		

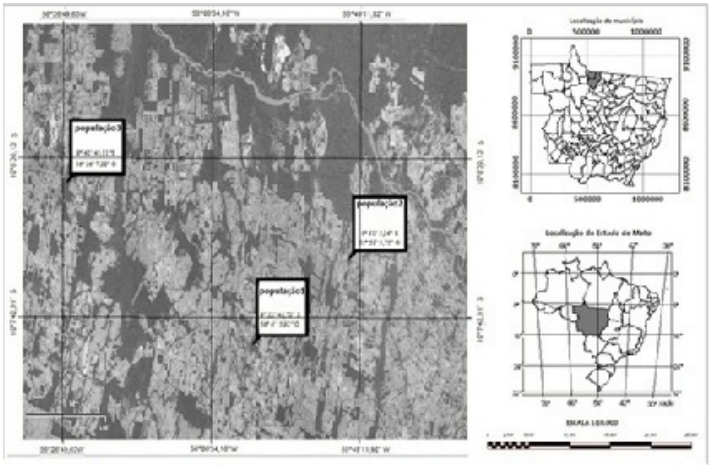

Figure 1. Geographic location of the three populations of *Theobroma subincanum* analyzed (population 1 - CFA; population 2 - CMN; population 3 - CPT). Image provided by Google Earth on May 18, 2012. Georeferenced by the MicroStation and ArcGIS software systems. For abbreviations, see Table 1.

Total DNA extraction was performed following the CTAB protocol developed by Doyle and Doyle (1987), with adaptations suggested by Faleiro et al. (2002) for DNA extraction in *T. cacao*. After extraction, DNA amount and quality were assessed by comparative analysis of the samples on 0.8% agarose gel stained with ethidium bromide. The samples were diluted in ultrapure water and standardized in 10 ng/μL volumes.

Polymerase chain reactions (PCRs)

Thirty-five ISSR primers were tested for initial amplification via PCR. Thirteen of these primers (Table 2) provided reproducible and polymorphic bands and were selected for the final analysis in all individuals.

PCRs were performed in a total volume of 20 μL in an MJ 96 thermocycler (Biocycler) in the Laboratory of Genetics - UNEMAT, Alta Floresta. The amplification program for the species under study complied with descriptions of Charters and Wilkinson (2000): one initial cycle of denaturation at 94°C for 5 min, followed by 30 cycles of 94°C for 1 min, 47°C (depending on the primer used) for 2 min, and 72°C for 30 s, with one final extension cycle at 72°C for 5 min.

The amplification products were separated by electrophoresis on 1.5% agarose gel in 1X TBE running buffer (89.15 mM Tris base, 88.95 mM boric acid, and 2.23 mM EDTA) at a constant voltage (100 V) for 4 h. The gel was stained with 0.6 ng/mL ethidium bromide. The sizes of the amplified fragments were estimated by comparison to the 100-bp DNA Ladder (Invitrogen™) molecular marker.

Data analysis

The ISSR fragments (amplification products) were evaluated and coded as binary characters, presence (1) or absence (0) of the bands amplified, for the construction of a matrix. The data of the binary presence-absence matrix were analyzed by means of the POPGENE 1.31 software system (Yeh et al., 1999). Nei's genetic diversity (*H*, 1978), the Shannon index (*I*), and the percentage of polymorphism (*P*) were all estimated. The coefficient of genetic differentiation (G_{ST}) was estimated to assess the population genetic structure among populations. The genetic identity and genetic distance between populations were also assessed using the model presented in Nei (1978).

Analysis of molecular variance (AMOVA) was used to make inferences about the genetic structure of populations by means of the total decomposition of components between and within populations. AMOVA was performed according to Excoffier et al. (1992) in the ARLEQUIN 3.01 software system (Excoffier et al., 2006). The dendrogram representing the genetic distances between populations was generated from the Nei's distance matrix (1978) by the unweighted pair-group method using arithmetic averages (UPGMA), with the use of the MEGA 3.1 software system (Kumar et al., 2004).

RESULTS AND DISCUSSION

ISSR analyses and genetic diversity within populations

The 13 selected ISSR primers produced a total of 61 bands in *T. subincanum*, with an average of 4.69 bands per primer. The minimum number of bands per primer was 3 and the

maximum was 7 (Table 2). Forty-one of the 61 loci (67.2%) were polymorphic at the species level. Primer 827 provided most information for the three populations because it showed the highest number of polymorphic bands, while primer 807 revealed no polymorphism among the amplified fragments in all individuals analyzed. The CPT population presented 60 bands, 29 of which were polymorphic. The CFA and CMN populations presented a total of 59 bands; 21 and 17 of which were polymorphic, respectively.

Table 2. Primers used, their respective sequences and the number of bands produced at the level of species and populations.

Primers	Sequences (5'→3')	Number of bands			
		Species	CFA	CMN	CPT
807	AGAGAGAGAGAGAGT	3 (0)	3 (0)	3 (0)	3 (0)
808	AGAGAGAGAGAGAGAGC	6 (4)	5 (0)	5 (0)	6 (4)
809	AGACAGAGAGAGAGG	4 (4)	4 (2)	4 (1)	4 (4)
811	AGAGAGAGAGAGAGAC	5 (4)	5 (2)	5 (1)	5 (4)
826	ACACACACACACACAC C	4 (4)	4 (2)	4 (2)	4 (3)
827	CACACACACACACACG	5 (5)	5 (4)	5 (3)	5 (4)
834	AGAGAGAGAGAGAGYT	4 (3)	4 (1)	4 (2)	4 (0)
835	AGAGAGAGAGAGAGYC	6 (2)	6 (1)	6 (1)	6 (2)
840	GAGAGAGAGAGAGAGAYT	7 (6)	6 (2)	6 (2)	6 (3)
844	CTCTCTCTCTCTCTCTRC	4 (1)	4 (1)	4 (0)	4 (0)
855	ACACACACACACACYT	4 (2)	4 (2)	4 (2)	4 (0)
873	GACAGACAGACAGACA	5 (4)	5 (2)	5 (2)	5 (4)
889	BDACACACACACACAC	4 (2)	4 (2)	4 (1)	4 (1)
Total		61 (41)	59 (21)	59 (17)	60 (29)

The number of polymorphic bands is shown in parentheses. For abbreviations, see Table 1.

ISSR molecular markers were able to detect polymorphism in the populations analyzed and proved to be reproducible for the species. Therefore, this method is efficient to detect genetic variability within and between populations of *T. subincanum*. The usefulness of this type of marker has been previously described for both cultivated and wild species (Rossi et al., 2009; Almeida et al., 2009; Brandão et al., 2011). This marker also presents high reproducibility and potential to be used in species without detailed DNA sequence information (Rossi et al., 2009; Almeida et al., 2009; Brandão et al., 2011).

H obtained from ISSR marker information ranged from 0.1126 for the CMN population to 0.1759 for the CPT population, while *I* ranged from 0.1653 for CMN to 0.2550 for CPT, presenting an average of $H = 0.2612$ and $I = 0.3889$ at the species level. The CMN population showed the lowest genetic diversity and the lowest percentage of polymorphism, while the CPT population presented the highest genetic diversity and the highest percentage of polymorphism (Table 3). Nonetheless, all three populations studied showed polymorphism and genetic diversity between individuals within each population.

Table 3. Comparison of populations of *Theobroma subincanum* for measurements of Nei's genetic diversity (*H*); Shannon genetic diversity index (*I*) and percentage of polymorphism (% *P*).

Populations	N	% *P*	*H*	*I*
CFA	20	37.70	0.1648	0.2366
CMN	22	29.51	0.1126	0.1653
CPT	18	47.54	0.1759	0.2550
Average		38.25	0.1511	0.2190
Level of species	59	72.13	0.2612	0.3889

N = sample size. For abbreviations, see Table 1.

Results similar to those of the present study for *T. subincanum* were found by Rossi (2007) for *Mauritia flexuosa* (H = 0.206 and I = 0.308), and by Müller (2010) in studying *Bertholletia excelsa*, who reported genetic diversity ranging from 0.1984 to 0.2475. Although the species used in the studies cited above are very far from the one reported in the present research, these studies were carried out in natural populations with some level of fragmentation, which is a condition similar to that of this study. Thus, they can help in the interpretation of the data obtained for *T. subincanum*.

Similar studies on the genus *Theobroma* are scarce. However, while studying *Theobroma speciosum*, Giustina (2011) found values close to those reported here (H = 0.1672 and I = 0.2549). Araújo (2011) analyzed the diversity of *Theobroma grandiflorum* and reported values of H = 0.2437 and I = 0.3715. These values are compatible with those found in the present study. These data suggest that genetic diversity ranges from moderate to low within populations in *Theobroma*. This is an expected result, since gene flow tends to be contained within populations; however, there is currently no information about the movement of genes between populations in these species. Genetic information is critical for the management of these natural populations.

Genetic differentiation between populations

AMOVA showed that 51.71% of the total variance is between populations and 48.29% is within populations. This indicated that genetic differentiation was greater in the inter-population component compared to the intrapopulation component (Table 4). There was significant genetic differentiation ($P < 0.000$) between populations. This variability distribution pattern was different from that found in studies on other species of tropical plants (Rossi, 2007; Rossi et al., 2009; Bertoni et al., 2010; Brandão et al., 2011). However, the same distribution pattern between populations was found in a study carried out by Silva et al. (2007), who found high inter-population genetic differentiation between populations of *Oryza glumaepatula* Steud. (Poaceae).

Table 4. Analysis of molecular variance of the three populations of *Theobroma subincanum* studied based on 3 ISSR markers.

Source of variation	d.f.	SS	CV	VT (%)
Between populations	2	153.068	4.16799	51.71
Within populations	57	198.288	3.89246	48.29
Total	59	351.356	8.06045	

d.f. = degrees of freedom; SS = sum of squares; CV = coefficient of variation; VT = total variance.

Hamrick (1983) suggested that the reproductive system is important for the formation of the genetic structure of a population. In general, autogamous species have low genetic diversity within populations and high genetic differentiation between populations when compared with allogamous species (Hamrick and Godt, 1996). However, the results of the present study showed high genetic differentiation between populations, which can be explained by the low gene flow (N_m = 0.6668) found between populations (Table 5), since, according to Figueira and Cascardo (2001), the genus *Theobroma* is composed of allogamous species. The gene flow estimate found in this study for the cupuí populations was close to that found by Zimback et al. (2004) in their study on *Trichilia pallida*, in which gene flow was estimated at N_m = 0.78.

Table 5. Genetic parameters of the *Theobroma subincanum* population.

	H_T	H_S	G_{ST}	N_m	F_{ST}
Average	0.2615	0.1494	0.4285	0.6668	0.51709
Standard deviation	0.0374	0.0197			

H_T = total heterozygosity; H_S = average genetic diversity within populations; G_{ST} = genetic divergence between populations; N_m = gene flow; F_{ST} = genetic differentiation between populations.

Forest fragmentation in the study region may be related to the data obtained by AMOVA, since the isolation of tree species populations reduces the amount of reproductive individuals and the population density. It may also affect genetic processes, including genetic drift, gene flow, selection, and the mating system (Yang and Meerow, 1996).

These data corroborate with those obtained by Rossi (2007), who found a relationship between genetic and geographic structures for the species *M. flexuosa* and *Acrocomia aculeata* (Jacq.), respectively.

The dendrogram showed that CFA and CMN were the most genetically similar populations, and that CPT was the most genetically distant population (Figure 2). Populations with higher genetic similarity were the closest geographically (Table 6). This suggests that genetic structure in *T. subincanum* is determined by geographic distance.

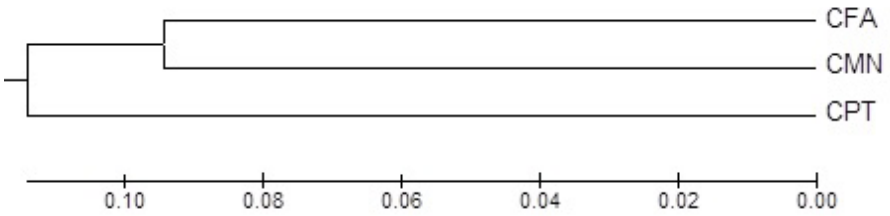

Figure 2. UPGMA dendrogram of three populations of *Theobroma subincanum* based on Nei's divergence. For abbreviations, see Table 1.

Table 6. Geographic distance, genetic distance and Nei's genetic identity (1978) between three populations of *Theobroma subincanum*.

Populations	Genetic distance	Genetic identity	Geographic distance*
CFA with CMN	0.1885	0.8282	24 km
CFA with CPT	0.2080	0.8122	41 km
CMN with CPT	0.2484	0.7801	65 km

*Distance provided by Google Earth, in straight line. For abbreviations, see Table 1.

The results of this study revealed genetic variability in natural populations of *T. subincanum* and that the levels of genetic polymorphism maintained in this species may favor conservation practices. However, constant forest fragmentation may increasingly reduce the levels of diversity and variability in the population. *In situ* conservation strategies are among the most acceptable practices to maintain the structure and genetic integrity of forest species populations. *T. subincanum* is an important reservoir of genes for breeding and for species of the genus *Theobroma*, especially cocoa, which is a major representative of the genus, with high potential for commercialization.

ISSR molecular markers were efficient in detecting polymorphism within populations of *T. subincanum* and at the species level. The CPT population presented the highest rate of genetic diversity and the highest percentage of polymorphism compared with the CFA and CMN populations; however, all populations showed some level of diversity and polymorphism.

The gene flow estimate showed that these populations may be genetically isolated, a factor that can be related to forest fragmentation. There is greater genetic diversity between populations than within populations. The levels of diversity found in *T. subincanum* are compatible with conservation and breeding programs within the genus *Theobroma*.

ACKNOWLEDGMENTS

We are grateful to Fundação de Amparo à Pesquisa do Estado de Mato Grosso (FAPEMAT) (Foundation for the Support of Research of the State of Mato Grosso), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (National Council for Scientific and Technological Development), and Universidade do Estado de Mato Grosso (UNEMAT) for the execution of the research and the granting of an Undergraduate Research scholarship.

REFERENCES

Almeida CMA, Lima GSA, Brito JZ, Donato VMTS, et al. (2009). Caracterização molecular de cultivares de cana-de-açucar utilizando marcadores ISSR. *Ciênc. Agrotecnol.* 33: 1771-1776.

Alverson WS, Whitlock BA, Nyffeler R, Bayer C, et al. (1999). Phylogeny of the core Malvales: evidence from ndhF sequence data. *Am. J. Bot.* 86: 1474-1486.

Araújo VAAC (2011). Diversidade genética em populações de *Theobroma grandiflorum* W. EX S, cultivadas no norte do estado de Mato Grosso. Monograph thesis, Universidade do Estado de Mato Grosso, UNEMAT, Alta Floresta.

Barth S, Melchinger AE and Lubberstedt TL (2002). Genetic diversity in *Arabidopsis thaliana* L. Heynh. investigated by cleaved amplified polymorphic sequence (CAPS) and inter-simple sequence repeat (ISSR) markers. *Mol. Ecol*. 11: 495-505.

Bertoni W, Telles MPC, Malosso MG, Torres SC, et al. (2010). Genetic diversity in natural populations of *Jacaranda decurrens* Cham. determined using RAPD and AFLP markers. *Genet. Mol. Biol.* 33: 532-538.

Brandão MM, Vieira FA and Carvalho D (2011). Estrutura genética em microescala espacial de *Myrcia splendens* (Myrtaceae). *Rev. Árvore* 35: 957-964.

Charters YM and Wilkinson MJ (2000). The use of self-pollinated progenies as 'in-groups' for the genetic characterization of cocoa germplasm. *Theor. Appl. Genet.* 100: 160-166.

Cidade FW, Souza-Chies TT, Batista LAR, Dall'agnol M, et al. (2009). Isolation and characterization of microsatellite loci in *Paspalum notatum* Flüggé (Poaceae). *Conservat. Genet.* 10: 1977-1980.

Cuatrecasas J (1964). Cacao and its allies A taxonomic revision of the genus *Theobroma*. *Contrib. U.S. Nat. Herb.* 35: 379-614.

Doyle JJ and Doyle JL (1987). A rapid DNA isolation procedure for small amounts of fresh leaf tissue. *Phytochem. Bull.* 19: 11-15.

Excoffier L, Smouse PE and Quattro JM (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479-491.

Excoffier L, Laval G and Schneider S (2006). Arlequin ver. 3.01. An Integrated Software Package for Population Genetics Data Analysis. Computational and Molecular Population Genetics Lab (CMPG). Institute of Zoology, University of Bern, Bern.

Faleiro FG, AraújoI S, Bahia RCS, Santos RF, et al. (2002). Otimização de extração e amplificação de DNA de *Theobroma cacao* L. visando obtenção de marcadores RAPD. *Agrotrópica* 14: 31-34.

Figueira AVO and Cascardo JCM (2001). Marcadores moleculares no melhoramento. In: Melhoramento genético do cacaueiro (Dias LAS, ed.). FUNAPE, Viçosa, 385-438.

Giustina LD (2011). Análise da Estrutura e Diversidade Genética em Populações Naturais de *Theobroma speciosum* (L). W. (Malvaceae) por Meio de Marcadores Moleculares ISSR. Mater's thesis, Universidade do Estado de Mato Grosso, UNEMAT, Alta Floresta.

Hamrick JL (1983). The distribution of genetic variation within and among natural plant populations. In: Genetic and conservation (Schonewald-Cox CM, Chambers SM, Macbryde B and Thomas WL, eds.). Benjamin Cummings, Menlo Park, 287-310, 335-348.

Hamrick JL and Godt MJW (1996). Conservation genetics of endemic plant species. In: Conservation genetics, case histories from nature (Avise JC and Hamrick JL, eds.). Chapman and Hall, New York, 281-304.

Kumar S, Tamura K and Nei M (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief. Bioinform.* 5: 150-163.

Müller KE (2010). Diversidade genetica de populações naturais de *Bertholletia excelsa* H.B.K. (Lecythidaceae) por meio de marcadores ISSR. Monograph thesis, Universidade do Estado de Mato Grosso, UNEMAT, Alta Floresta.

Nei M (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89: 583-590.

Rossi FS (2007). Avaliação da diversidade genética entre e dentro populações naturais de *Mauritia flexuosa* (Arecaceae) no município de Alta Floresta-MT, com marcadores de ISSR. Monograph thesis, UNEMAT, Alta Floresta.

Rossi AA, de Oliveira LO, Venturini BA and dos Santos SR (2009). Genetic diversity and geographic differentiation of disjunct Atlantic and Amazonian populations of *Psychotria ipecacuanha* (Rubiaceae). *Genetica* 136: 57-67.

Silva CR, Venturieri GA and Figueira A (2005). Description of Amazonian *Theobroma* L. collections species indentification and characterization of interspecific hybrids. *Acta Bot. Brasílica* 18: 333-341.

Silva CR, Venturieri GA and Figueira A (2007). Description of Amazonian *Theobroma* L. collections species identification and em *Oryza glumaepatula* Steud. (Poaceae) avaliada com microssatélites. *Biota Neotropica* 7: 165-172.

Souza AGC, Silva SEL, Tavares AM and Rodrigues MRL (1996). A Cultura do Cupuaçu (*Theobroma grandiflorum* Willd. ex Spreng. Schum. EMBRAPA, Manaus, 6-7.

Whitlock MC, Ingvarsson PK and Hatfield T (2000). Local drift load and the heterosis of interconnected populations. *Heredity* 84 (Pt 4): 452-457.

Wright S (1931). Evolution in Mendelian populations. *Genetics* 16: 97-159.

Yang SL and Meerow AW (1996). The *Cycas pectinata* (Cycadaceae) complex structure and gene flow. *Int. J. Plant Sci.* 157: 468-483.

Yeh FC, Yang RC and Boyle T (1999). POPGENE. Microsoft Windows-based freeware for population genetic analysis. Release 1.31. University of Alberta, Edmonton.

Zimback L, Mori ES, Kageyama PY, Veiga RFA, et al. (2004). Estrutura genética de populações de *Trichillia pallida* Swartz (Meliaceae) por marcadores RAPD. *Sci. Florest.* 65: 114-119.

