

High levels of genetic divergence and inbreeding in populations of cupuassu (*Theobroma grandiflorum*)

Rafael M. Alves · Alexandre M. Sebbenn ·
Angela S. Artero · Charles Clement · Antonio Figueira

Received: 22 June 2006 / Revised: 6 September 2006 / Accepted: 20 September 2006 / Published online: 22 November 2006
© Springer-Verlag 2006

Abstract *Theobroma grandiflorum* (cupuassu) is an important fruit tree native to the Brazilian Amazon. Establishing the genetic diversity and structure of populations is critical to define long-term strategies for cupuassu conservation presently threatened by rapid deforestation. Three natural populations collected at the putative center of diversity, three groups of accessions established at a germplasm collection, and one derived from commercial plantings were analyzed. The genetic diversity was assessed using 21 polymorphic microsatellite loci originally developed for *Theobroma cacao*, disclosing a total of 113 alleles. The estimated genetic diversity parameters averaged over cupuassu populations ($A=3.53$ alleles per locus; $H_e=0.426$; $H_o=0.346$) were lower than the values reported for other Neotropical tree species. The three natural populations presented a positive and significant fixation index (f),

ranging from 0.133 to 0.234. Cupuassu apparently adhered to a general pattern of genetic diversity structure of some Neotropical tree species occurring at low densities, with a low intrapopulation genetic diversity and important levels of endogamy, possibly due to biparental inbreeding derived from the presence of spatial genetic structure in the populations. A high level of genetic divergence was detected among the natural populations ($\theta_p=0.301$), a strong differentiation caused by limited gene flow, and suggesting that human interference in spreading and/or stimulating plantings might have had a smaller effect than expected. The approximate location of the *T. grandiflorum* center of diversity could not be confirmed by analyzing natural populations from the putative region.

Keywords Cacao · Cocoa · Genetic diversity · Malvaceae *sensu lato* · Simple sequence repeat

Communicated by D. Grattapaglia

R. M. Alves
Embrapa Amazônia Oriental,
Tv. Dr. Enéas Pinheiro s/n, CP 48, Belém, PA 66095-100, Brazil

A. M. Sebbenn
Instituto Florestal de São Paulo,
CP 1322, São Paulo,
SP 01059-970, Brazil

A. S. Artero · A. Figueira (✉)
Centro de Energia Nuclear na Agricultura,
Universidade de São Paulo,
Av. Centenário 303, CP 96,
Piracicaba, SP 13400-970, Brazil
e-mail: figueira@cena.usp.br

C. Clement
Instituto de Pesquisa da Amazônia,
Manaus, CP 478,
Manaus, AM 66011-970, Brazil

Introduction

Theobroma grandiflorum (Willd ex Spreng) Schum. (cupuassu) is a fruit tree native to the Brazilian Amazon, with its putative center of diversity located at the south/southeastern region of Pará state, Brazil (Fig. 1; Cuatrecasas 1964). Cupuassu is a diploid species ($2n=20$), member of the Sterculiaceae, and related to *Theobroma cacao* L. (cacao), the species with the greatest economic importance of the genus. The strong tropical flavor from cupuassu seed-surrounding pulp is highly appreciated in juices, ice creams, jams, candies, desserts, and liquors (Velho et al. 1990). The seed pulp represents up to 45% of the fruit fresh weight, while a product similar to cocoa powder (called “cupulate” in Brazil) can be obtained from fermented seeds. Cupuassu seed fat is distinct from

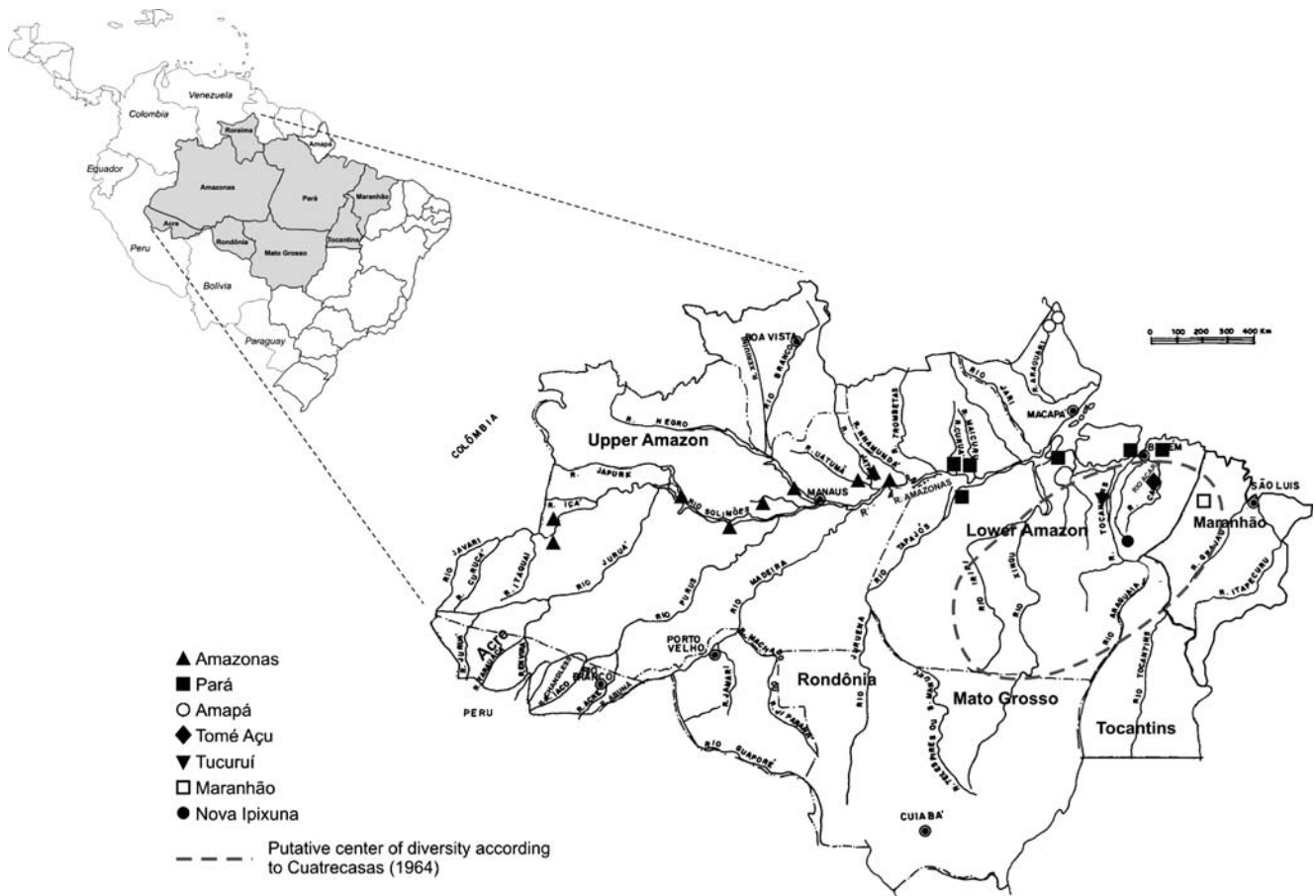


Fig. 1 Approximate location of collection in the Brazilian Amazon of the seven analyzed *Theobroma grandiflorum* groups of accessions, including three from the germplasm collection located at “Embrapa Amazônia Oriental”, Belém, Pará state (1°28’S; 48°27’W), denomi-

nated as “Amazonas”, “Pará”, and “Amapá”; one derived from commercial areas from Tomé-açu, Pará state (2°25’S; 48°09’W); and the three natural populations “Nova Ipixuna”, “Tucuruí”, and “Maranhão”

cocoa butter (Gilabert-Escrivá et al. 2002), but it has the potential to be applied in various cosmetic and food products, including ice creams and cakes.

Cupuassu is considered an outbreeding species because of its floral morphology, adapted to pollination by insects, and the occurrence of a complex self-incompatibility system (Alves et al. 1997), probably similar to *T. cacao* (Cope 1976). Cupuassu seeds are recalcitrant (Velho et al. 1990), and seed dispersion is thought to be performed by large rodents (such as agouti) and primates, including humans. Fruit drop from tall trees (often >30 m) may facilitate self-dispersal, but usually, seedlings are not found around mature trees. Similar to other Amazonian crops, cupuassu might have been stimulated by native Amazonian people, resulting in a partial domestication (Clement 1999). Cupuassu is still one of the most common fruits found in agriculturist tribes of the eastern Amazon.

The Amazon region has been suffering from intense colonization pressure, with constant deforestation and timber exploration, particularly at the putative center of diversity of *T. grandiflorum* (Homma et al. 1996; Cuatrecasas 1964),

leading to irreversible losses of genetic resources. Furthermore, few cupuassu accessions have been collected and maintained in ex situ germplasm collections in the Amazon region and elsewhere. There is no information about the genetic structure of *T. grandiflorum* populations, while a few studies have been conducted with the sympatric congener *T. cacao* (Ronning and Schnell 1994; N’Goran et al. 2000; Motamayor et al. 2002; Sereno et al. 2006). Natural cacao populations appeared to have a strong intrapopulation structure with small differentiation among populations (Sereno et al. 2006). Tropical tree species present unique population characteristics and functioning related to the low plant densities and pollen dispersion processes (Ward et al. 2005). Intrapopulation diversity of tropical trees would depend on gene flow, mainly by pollen dispersal, and tolerance to biparental inbreeding and/or selfing due to spatial genetic structure (Latouche-Hallé et al. 2004). Most of the population genetic studies of tropical tree species have been conducted with forest species, while little is known about tree species bearing fruits valued by humans, and putatively proto-domesticated, such as cupuassu.

Establishing the genetic diversity and structure of cupuassu populations is critical to define strategies for long-term in situ and ex situ conservation of genetic resources to minimize further losses of this promising new crop (Velho et al. 1990). There is an urgent need to define areas with important genetic diversity to establish conservation units and/or to accelerate germplasm collection for conservation in ex situ repositories. On the other hand, the few established collections need to be evaluated to help optimize diversity conservation because cupuassu requires in vivo maintenance due to seed recalcitrance. Because of the lack of improved selected cupuassu cultivars, commercial plantings have been established mainly using seeds collected from natural populations, resulting in a great heterogeneity but preserving an important sample of the natural diversity.

The use of microsatellites to characterize the genetic structure and diversity of tropical plant populations has increased in recent years, but a major drawback is the high cost of primer development. Cross-species amplification of microsatellite loci represent an appealing alternative, with transferability rates between congeners typically above 50%, decreasing proportionally to phylogenetic divergence (Varshney et al. 2005), with reported successful application in population genetics studies (Brondani et al. 2005; Varshney et al. 2005; Zucchi et al. 2003). Microsatellite primers developed for *T. cacao* were tested in *T. grandiflorum* with success (Lanaud et al. 1999; Alves et al. 2006), and they have been used to determine the mating system of a natural population of cupuassu, detecting at least one maternal allele in all individuals and without indications of non-Mendelian inheritance (Alves et al. 2003).

The objective of this work was to characterize, using microsatellite markers, the genetic diversity and structure of three types of *T. grandiflorum* populations, including three natural populations collected at the putative center of diversity, three groups of accessions established at a germplasm collection, and one derived from commercial plantings. We expected that *T. grandiflorum* contained a comparable level of within-population genetic diversity as *T. cacao*, a congener with analogous biology and life history traits (Cuatrecasas 1964). In comparison to other species, we hypothesized that *T. grandiflorum* contained a large within-population diversity, considering it as a perfect outbreeder (Alves et al. 2003), and a low genetic differentiation among populations, derived from extensive gene flow (Ward et al. 2005). The natural populations, collected in the wild near the putative center of diversity, were expected to present private alleles and higher levels of genetic diversity than all the other groups. For the germplasm collection, it was considered that, because *T. grandiflorum* did not occur naturally along the Amazon river, the “Amazonas” group should derive from material

originated from Pará by human dispersion and cultivation, with more restricted diversity than the natural populations. The sample derived from commercial plantings (“Tomé-açu”) was expected to contain a lower level of diversity in comparison to the natural populations because it was derived from cultivated stands.

Materials and methods

Populations and groups of accession sampling

Seven samples of *T. grandiflorum* accessions were analyzed (Fig. 1), including three natural populations (“Nova Ipixuna”, “Tucuruí”, and “Maranhão”); three groups of accessions from a cupuassu germplasm collection located at “Embrapa Amazônia Oriental”, Belém, Pará state (1°28'S; 48°27'W), denominated as “Amazonas”, “Pará”, and “Amapá”; and one sample derived from commercial areas from Tomé-açu, Pará state (2°25'S; 48°09'W). The natural populations “Nova Ipixuna” and “Tucuruí” were collected at the putative center of diversity of *T. grandiflorum* in Pará state (Cuatrecasas 1964). The population from Nova Ipixuna (4°53'S; 49°22'W) was represented by 40 individuals (minimum average distance between trees of ca. 500 m), sampled by us on December 2001 at a single site over an approximately 5-km radius. The “Tucuruí” population contained 40 individuals originally collected by the “Instituto de Pesquisa da Amazônia” (INPA) in Tucuruí (4°3'S; 49°5'W) within a few kilometers from the area later flooded by the dam of the Tucuruí hydroelectric power plant. The “Tucuruí” accessions have been maintained at the germplasm collection of INPA, Manaus, Amazonas state. The other natural population (“Maranhão”) was represented by 40 individuals, originally collected by INPA along the Pindaré river, Maranhão state (2°0'S; 46°4'W), within an approximately 5-km radius. These accessions are also maintained by INPA.

The second set analyzed contained three samples of accessions from the cupuassu germplasm collection. The original accessions had been collected from backyard gardens, small plantings, and spontaneous areas in the Brazilian Amazon (Lima and Costa 1991). The “Amazonas” group contained 29 accessions, originally collected in various sites along the Amazon river (Fig. 1). Ten accessions were obtained from various areas of “Pará” state, and seven accessions were from the state of “Amapá”. The third group of accessions (“Tomé-açu”) was derived from a sample of trees from commercial plantings. From approximately 30,000 trees from 12 commercial areas, a sample of 50 accessions were selected, cloned, and established in another field at Embrapa in Tomé-açu and used in the present analysis.

DNA extraction

Leaf samples from all the individuals were oven-dried (40°C). Samples of ca. 150 mg of leaf tissue were ground in liquid nitrogen, and DNA was extracted using the procedure described by Sereno et al. (2006). DNA was quantified by fluorimetry in a DyNA Quant 2000 fluorometer (Amersham Biosciences, Buckinghamshire, UK), and a 5-ng μl^{-1} stock was prepared.

Microsatellite analyses

The total amplification reaction volume was 13 μl , containing 15 ng of genomic DNA, 1.5 mM MgCl_2 , 100 μM of each dNTP, 0.2 μM of each primer, and 1.2 U *Taq* polymerase in an appropriate buffer (Invitrogen do Brasil, São Paulo, Brazil). A total of 21 primer pairs developed for *T. cacao* (Lanaud et al. 1999) were synthesized by Invitrogen and used in amplifications conducted using a touchdown program as described by Alves et al. (2006). From the 21 loci used here, ten were previously confirmed in cupuassu to be homologous to *T. cacao* microsatellite loci by hybridization (Lanaud et al. 1999), while five other loci (*mTcCIR06*, *mTcCIR31*, *mTcCIR33*, *mTcCIR43*, and *mTcCIR61*) had been analyzed for segregation in eight families to determine the mating system of a natural population of *T. grandiflorum*, without disclosing non-Mendelian inheritance or any bias in allele transmission (Alves et al. 2003). Amplification products were separated in denaturing sequencing gels (6% polyacrylamide; 7 M urea) run in TBE at 55 W for 2 h, and products were visualized by silver staining according to the procedure described by Creste et al. (2001).

Statistical analyses

The intrapopulation genetic diversity was characterized by average number of alleles per locus (A), observed heterozygosity (H_o), expected heterozygosity in Hardy–Weinberg equilibrium (H_e), and fixation index (f) estimated using GDA-NT (Degen 2006). Single and multilocus departures from Hardy–Weinberg expected ratios were tested in natural populations based on fixation index. The statistical significance of individual and multilocus F values was calculated based on Monte-Carlo method (Manly 1997). A total of 1,000 permutations of homologous alleles among individuals were run to test the significance of the F -value for each locus using GDA-NT (Degen 2006). The distribution of the genetic variability between and within natural populations was analyzed using F -statistics based on Weir and Cockerham (1984). Genetic divergence between populations (θ_p) was estimated, and the 95 and 99% confidence intervals were calculated by bootstrap based on 10,000 replications using GDA (Lewis and Zaykin 1999). Historical gene flow (Nm)

among natural populations was estimated according to Crow and Aoki (1984) as $Nm = (1/4\alpha)[(1/F_{ST}) - 1]$, where F_{ST} was the measured genetic differentiation among populations, and $\alpha (\alpha = [a/(a-1)])^2$ was the correction for finite number of populations (a). We substituted F_{ST} for θ_p , following Cockerham and Weir (1993). The distances between pairs of populations were measured by the unbiased minimum genetic distance (D_j) of Nei (1978) and clustered based on UPGMA using TFPGA (Miller 1997). The clustering stability was tested by bootstrap with 1,000 resamplings.

Results

The analyses of 216 cupuassu accessions using the 21 microsatellite loci revealed a total of 113 alleles. The number of alleles per polymorphic locus ranged from 2 to 11, with an average of 5.38 alleles per locus (Table 1). The mean expected heterozygosity (H_e) estimated for each locus was higher than the mean observed heterozygosity (H_o) for most cases, except for loci *mTcCIR43* and *mTcCIR58* (Table 1). The overall mean expected heterozygosity (H_e) estimated for each locus for the species was 0.614, ranging from 0.032

Table 1 Genetic diversity of 21 microsatellite loci of *T. grandiflorum*, with the total number of accessions analyzed for each locus (n), total number of alleles for each locus (n_A), mean expected heterozygosity (H_e), and mean observed heterozygosity (H_o) averaged over all accessions

Locus	n	n_A	H_e	H_o
<i>mTcCIR04</i>	213	3	0.521	0.169
<i>mTcCIR06</i>	209	6	0.743	0.349
<i>mTcCIR09</i>	216	2	0.336	0.018
<i>mTcCIR11</i>	215	4	0.658	0.144
<i>mTcCIR12</i>	213	3	0.501	0.141
<i>mTcCIR17</i>	215	8	0.838	0.558
<i>mTcCIR18</i>	216	8	0.722	0.361
<i>mTcCIR19</i>	212	5	0.757	0.410
<i>mTcCIR22</i>	213	7	0.670	0.263
<i>mTcCIR25</i>	216	7	0.816	0.592
<i>mTcCIR26</i>	214	10	0.804	0.369
<i>mTcCIR30</i>	216	4	0.032	0.014
<i>mTcCIR31</i>	215	7	0.726	0.340
<i>mTcCIR33</i>	216	5	0.740	0.324
<i>mTcCIR43</i>	215	6	0.633	0.693
<i>mTcCIR45</i>	216	3	0.106	0.069
<i>mTcCIR48</i>	214	3	0.558	0.280
<i>mTcCIR54</i>	213	2	0.490	0.221
<i>mTcCIR57</i>	212	5	0.716	0.259
<i>mTcCIR58</i>	216	4	0.665	0.829
<i>mTcCIR61</i>	208	11	0.873	0.808
Grand mean	214	5.38	0.614	0.343
Total		113		

Table 2 Estimates of genetic parameters in seven *T. grandiflorum* samples based on 21 microsatellite loci

Groups	<i>n</i>	<i>n_A</i>	<i>n_{pa}</i>	<i>A</i>	<i>H_e</i>	<i>H_o</i>	<i>f</i>
Natural populations							
Nova Ipixuna	40	67	5	3.19	0.388	0.353	0.133*
Tucuruí	40	68	4	3.24	0.422	0.339	0.234*
Maranhão	40	77	10	3.67	0.447	0.374	0.209*
Mean	40	70.7	6.3	3.37	0.419	0.355	0.192
Germplasm groups							
Amazonas	29	106	12 ^a	5.05	0.489	0.355	n.e.
Pará	10	68	1 ^a	3.24	0.405	0.376	n.e.
Amapá	7	58	0 ^a	2.76	0.402	0.319	n.e.
Growers							
Tomé-açu	50	75	0 ^a	3.57	0.353	0.305	n.e.

n is the number of accessions analyzed for each sample, *n_A* is the total number of alleles per population, *n_{pa}* is the number of private alleles, *A* is the mean number of alleles per locus, *H_e* is the mean expected heterozygosity, *H_o* is the mean observed heterozygosity, and *f* is the mean coefficient of inbreeding within population

**P*<0.01

^aIn relation to total sample

n.e. Not estimated

(*mTcCIR30*) to 0.873 (*mTcCIR61*) among loci, while the observed heterozygosity (*H_o*) was 0.343, ranging from 0.014 (*mTcCIR30*) to 0.829 (*mTcCIR58*) among loci (Table 1).

The mean number of alleles per locus for the natural populations was 3.37 alleles, ranging from 3.19 (“Nova

Ipixuna”) to 3.67 (“Maranhão”) (Table 2). The total number of alleles ranged from 67 for the “Nova Ipixuna” to 77 for the “Maranhão” populations. Considering only the natural populations, “Nova Ipixuna” displayed five private alleles, while “Tucuruí” contained four and “Maranhão” ten private alleles. The average fixation index was positive and significantly different from zero for all natural populations, ranging from 0.133 to 0.234, indicating departure from Hardy–Weinberg proportions caused by excess of homozygotes (Table 2). All loci were tested for Hardy–Weinberg proportions based on Monte-Carlo method, and many loci were not in equilibrium for all the natural populations (Table 3). For the “Nova Ipixuna” population, 13 of the 21 loci (62%) exhibited a departure from Hardy–Weinberg proportions (Table 3), and the same trend was observed for the “Tucuruí” and “Maranhão” populations, with deviation for 12 (57%) and 15 loci (71%), respectively. The genetic divergence among the natural populations was high and statistically significantly different from zero ($\theta_p = 0.301$; *P* < 0.01), suggesting a limited historic gene flow among these populations (*Nm*=0.26).

Considering the three groups of germplasm accessions and the one from growers (“Tomé-açu”), the total number of alleles ranged from 58 (“Amapá”) to 106 for the “Amazonas” group of accessions (Table 2). Considering the total sample of accessions, only the “Amazonas” group displayed 12 private alleles, while the “Pará” displayed one

Table 3 Fixation index (*f*) and *P*-values to Hardy–Weinberg equilibrium for each locus in the three natural populations

Locus	Nova Ipixuna		Tucuruí		Maranhão	
	<i>f</i>	<i>P</i> -value	<i>f</i>	<i>P</i> -value	<i>f</i>	<i>P</i> -value
<i>mTcCIR04</i>	1.000	1.000	0.284	0.972	0.103	0.771
<i>mTcCIR06</i>	0.271	0.998	0.239	0.946	0.201	0.923
<i>mTcCIR09</i>	0.000	–	0.000	–	–0.026	0.874
<i>mTcCIR11</i>	0.212	0.872	0.457	0.997	–0.030	0.799
<i>mTcCIR12</i>	–0.187	0.828	0.375	0.986	0.422	0.989
<i>mTcCIR17</i>	0.185	0.964	0.191	0.968	0.330	1.000
<i>mTcCIR18</i>	0.292	0.957	0.061	0.732	–0.050	0.694
<i>mTcCIR19</i>	0.138	0.840	0.523	1.000	0.309	1.000
<i>mTcCIR22</i>	0.197	0.932	0.575	1.000	0.799	1.000
<i>mTcCIR25</i>	0.360	0.968	–0.370	1.000	–0.590	1.000
<i>mTcCIR26</i>	0.051	0.627	0.672	1.000	0.328	0.998
<i>mTcCIR30</i>	0.000	–	0.000	–	1.000	1.000
<i>mTcCIR31</i>	0.252	0.973	0.257	0.988	0.524	1.000
<i>mTcCIR33</i>	–0.088	0.735	0.549	1.000	0.413	0.999
<i>mTcCIR43</i>	–0.770	1.000	–0.200	0.920	–0.499	1.000
<i>mTcCIR45</i>	0.314	0.971	1.000	1.000	0.000	–
<i>mTcCIR48</i>	0.232	0.928	0.291	0.938	0.292	0.957
<i>mTcCIR54</i>	0.000	–	–0.182	0.870	1.000	1.000
<i>mTcCIR57</i>	0.775	1.000	0.208	0.867	0.566	1.000
<i>mTcCIR58</i>	–0.375	1.000	–0.582	1.000	–0.756	1.000
<i>mTcCIR61</i>	–0.470	1.000	0.094	0.852	–0.152	0.992

– Not tested

private allele. No private allele was found in the “Amapá” and “Tomé-açu” samples. The mean number of alleles per locus for these group of accessions ranged from 2.76 (“Amapá”) to 5.05 (“Amazonas”).

The estimated unbiased Nei’s minimum genetic distance (Nei 1978) for each pair of sample ranged between 0.012 and 0.385 (Fig. 2). Among the natural populations, the “Tucuruí” and “Nova Ipixuna” populations were genetically distant (0.222). The three groups of accessions from the germplasm collection formed a highly homogeneous cluster, with low genetic distance (between 0.012 and 0.033; Fig. 2). The sample from commercial plantings (“Tomé-açu”) exhibited a large genetic distance to the germplasm group of accessions, ranging from 0.350 to 0.376, while being more similar to the natural populations (0.198 to 0.234).

Discussion

The genetic diversity of seven groups of accession, including three natural populations of *T. grandiflorum*, was evaluated using 21 polymorphic microsatellite loci, originally developed for *T. cacao* (Lanaud et al. 1999), disclosing a total of 113 alleles. The number of loci utilized in this study ($n=21$) greatly exceeded the average number commonly employed in similar diversity analyses of tropical tree species (maximum of 16 loci reported for *T. cacao* by Motamayor et al. 2002), as well as in other diversity studies of wild plants ($n = 8.4 \pm 6.7$; Nybom 2004), representing a superior genome coverage.

Changes in microsatellite allele structure and/or decrease of average length may reduce the detection of variability between congeners, a phenomenon known as ascertainment bias (van Treuren et al. 1997; Shepherd et al. 2002). However, the estimated diversity parameters averaged over natural cupuassu populations ($A=3.37$ alleles per locus;

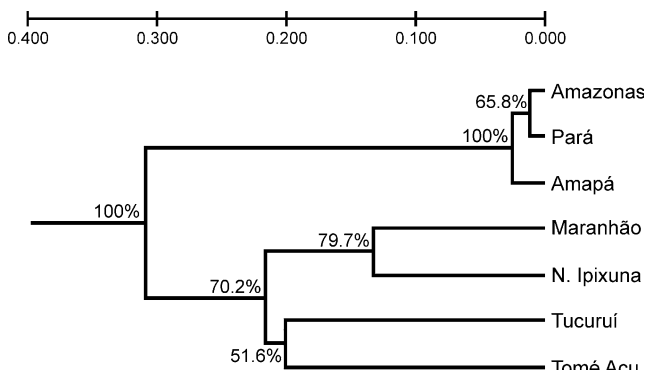


Fig. 2 Graphic representation of the clustering analysis using UPGMA based on nonbiased minimum genetic distance of Nei (1978), containing the *P*-value of bootstrap analyses for seven *T. grandiflorum* samples based on microsatellite markers

$H_e=0.419$; $H_o=0.355$; Table 2) were equivalent to the diversity levels reported based on 11 to 16 microsatellite loci for populations of *T. cacao* ($A = 3.9 - 4.0$; $H_e = 0.540 - 0.566$; $H_o = 0.347 - 0.413$; Motamayor et al. 2002; Sereno et al. 2006), a sympatric congener with analogous biology and life history traits. Both species were analyzed with similar set of primers, and these results confirmed that any ascertainment bias was only mild in cupuassu using cacao microsatellite primers. Thus, *T. grandiflorum* contained a comparable level of within-population genetic diversity to *T. cacao* as hypothesized.

However, the allelic diversity parameters disclosed for *T. grandiflorum* were lower than values commonly reported for wild plants analyzed using microsatellite markers. Nybom (2004) reviewed 106 intraspecific genetic diversity studies in wild plants with a wide range of biology and life history traits based on microsatellites, reporting an average of 83.2 alleles from 8.4 loci (average 9.9 alleles per locus), with a mean expected heterozygosity of 0.61 ± 0.21 and observed heterozygosity of 0.58 ± 0.22 . Furthermore, when the diversity parameters of cupuassu were compared to those from other Neotropical tree species with high outcrossing rate analyzed by microsatellites, such as *Pithecellobium elegans* ($A=6.5$; $H_e=0.645$; $H_o=0.710$ Chase et al. 1996), *Carapa guianensis* ($A=7.6$; $H_e=0.666$; $H_o=0.683$; Dayanandan et al. 1999), *Swietenia humilis* ($A=7.1$; $H_e=0.524$; $H_o=0.482$; White et al. 1999), *Caryocar brasiliensis* ($A=10.6$; $H_e=0.856$; $H_o=0.765$; Collevatti et al. 2001a), *Dinizia excelsa* ($A=18.2$; $H_e=0.611$; $H_o=0.651$; Dick et al. 2003), *Swietenia macrophylla* ($A=9.5$; $H_e=0.781$; $H_o=0.751$; Lemes et al. 2003), *Dicorynia guianensis* ($A=7.3$; $H_e=0.629$; $H_o=0.621$; Latouche-Hallé; et al. 2003), *Symphonia globulifera* ($A=8$; $H_e=0.870$; $H_o=0.730$; Degen et al. 2004), and *Manilkara huberi* ($A=6.4$; $H_e=0.813$; $H_o=0.688$; Azevedo et al. 2005), the same trend could be established, with *T. grandiflorum* displaying a lower allelic diversity and reduced expected and observed heterozygosities. Most tropical forest tree species typically presented a larger number of alleles per locus and a consequent higher potential expected heterozygosity. On the other hand, lower diversity had been detected in *T. cacao* ($A = 3.9 - 4.0$; $H_e = 0.54 - 0.566$; $H_o = 0.347 - 0.413$; Motamayor et al. 2002; Sereno et al. 2006), *Eugenia dysenterica* ($A=3.1$; $H_e=0.442$; $H_o=0.458$; Zucchi et al. 2003), and *Vitellaria paradoxa* ($A=2.3$; $H_e=0.512$; $H_o=0.306$; Sanou et al. 2005). Therefore, it can be concluded that *T. grandiflorum* did not exhibit the large genetic diversity as was originally hypothesized.

The microsatellite allelic richness displayed by a particular locus can be influenced by the method of locus identification, which can be either from genomic libraries (random or microsatellite-enriched) or expressed genic sequences (Varshney et al. 2005). However, all loci are commonly

preselected for the highest diversity, maximizing length and number of alleles per locus (van Treuren et al. 1997). Considering that all loci analyzed in the studies of tropical tree species (listed above) were obtained from enriched genomic libraries by similar approaches, the difference in the number of alleles per locus should not be biased by the microsatellite development method. One possible explanation for the low number of alleles detected in cupuassu might derive from a supposed partial domestication by native Amazonian peoples (Clement 1999), reducing allelic diversity probably because of a founder effect. A similar phenomenon could have also occurred with *T. cacao* (Serenó et al. 2006), *E. dysenterica* (Zucchi et al. 2003), and *V. paradoxa* (Sanou et al. 2005), all fruit-bearing species valued by humans, who may act as a seed disperser, in opposition to the other commonly analyzed tropical tree species mostly logged for timber. Proto-domestication should not affect the number of alleles, assuming that microsatellites are considered to be neutral markers (Sanou et al. 2005). However, microsatellite loci are sensitive to sample size due to the high level of allelic polymorphism, and cultivation/domestication tends to be based on limited number of individuals. Hollingsworth et al. (2005) detected lower allelic richness in five planted stands directly compared to geographically matching natural stands of *Inga edulis* in Amazônia. Therefore, cultivation/domestication of *T. grandiflorum* might have affected the allelic diversity of cupuassu populations by genetic drift.

Many Neotropical tree species occur at a considerably lower density than temperate equivalents, with agglomeration of individuals, which tends to reduce intrapopulation genetic diversity due to small effective population size, while favoring population differentiation (Lemes et al. 2003; Novick et al. 2003; Dutech et al. 2004). The lower observed heterozygosity of some of these tropical trees might, therefore, derive from the mating system, with increased biparental inbreeding and/or selfing due to limited seed and pollen dispersal and spatial genetic structure (Collevatti et al. 2001b; Dutech et al. 2004). An important consequence of the spatial genetic structure detected in various tropical tree species studied to date (e.g., Degen et al. 2004; Hardy et al. 2006) has been the common observation of outcrossing between relatives, resulting in a significant excess of homozygotes compared to expected Hardy–Weinberg proportions. Excess of homozygotes has been reported for *Melaleuca alternifolia* in Australia (Rossetto et al. 1999), *S. humilis* in Honduras (White et al. 1999), *C. brasiliensis* in Brazil (Collevatti et al. 2001a), *S. macrophylla* in Meso-America (Novick et al. 2003) and Brazil (Lemes et al. 2003), and *D. guianensis* in French Guyana (Latouche-Hallé et al. 2003). The deficit in heterozygotes has also been a current observation in *T. cacao* populations, estimated by isozymes (Ronning and

Schnell 1994), RFLP (N’Goran et al. 2000), and microsatellites (Serenó et al. 2006).

The three natural populations exhibited positive and significant fixation index (Table 2). The excess of homozygotes in *T. grandiflorum* populations might derive from mating system by biparental inbreeding or by Wahlund effect or as a result of null alleles. Long-distance pollen dispersal and tolerance to selfing are two strategies to overcome demographic restrictions to allow efficient mating (Latouche-Hallé et al. 2004). Cupuassu appears to be pollinated by generalist insects (Falcão and Lleras 1983), which tend to restrict pollen flow in comparison to foraging habits of specialized pollinators. Natural seed dispersal of cupuassu is conducted either by gravity or mainly by animals, such as primates and large rodents (*Dasyprocta* sp. *Agouti paca*), responsible for scatter hoarding. *T. grandiflorum* appears to be a typical outbreeding species because of its floral syndrome and the occurrence of an efficient self-incompatibility system (Alves et al. 1997, 2003), possibly similar to the one described in *T. cacao* (Cope 1976). The unique gametophytic self-incompatibility system of *T. cacao* is not absolute, but quantitative, depending on the ratio of fused to nonfused ovules (Knight and Rogers 1955; Cope 1962), and can be overcome by employing a mixture of compatible and incompatible pollen with successful self-fertilization, or even naturally at a very low rate, affected by environment and maternal genotype (Glendinning 1960). A similar situation might occur in cupuassu under natural conditions, with the mixture of compatible and incompatible pollen allowing self-pollination and/or crossing between relatives. In fact, Alves et al. (2003) evaluated the mating system of cupuassu of the natural population “Nova Ipixuna” using microsatellites, concluding that cupuassu appeared as a perfect outbreeder, but showing some level of biparental inbreeding. Inbreeding in an offspring derived from crosses between relatives is equal to the half of the coefficient of relatedness between parents or to the mean coancestry between parents. The estimated fixation indices for the *T. grandiflorum* natural populations were high, ranging from 0.133 to 0.209. Assuming that selfing was absent because of self-incompatibility, the observed coefficient of inbreeding could have derived from mating between half-sib (0.125) and/or full-sib families (0.25). Alves et al. (2003) detected that 93% of the individuals from open-pollinated progenies appeared to be full sibs. The outcrossing between half and full sibs could explain the high levels of inbreeding in the populations. Biparental inbreeding might be a main reason for the high fixation index in some *T. grandiflorum* populations, but the occurrence of Wahlund effect cannot be entirely discarded due to possible existence of temporal breeding subunits inside the sampled regions. The alternative hypotheses for

the deficit of heterozygotes include the occurrence of null alleles and/or allele dominance (Nyblom 2004) or biased transmission of alleles. Presence of null alleles (which fail to amplify because of mutations at the flanking primer sites) or allele dominance is difficult to demonstrate, requiring specific experiments, as described by Morand et al. (2002). However, the low frequency of homozygous individuals for null alleles represented by total lack of amplification (as demonstrated by the total number of individuals scored for each locus on Table 1) indicated that it is highly unlikely that null alleles were present in a frequency that affected the level of observed heterozygosity. Additionally, Alves et al. (2003) detected at least one maternal microsatellite allele in all individuals from eight families for eight loci, when investigating the mating system of cupuassu, which corroborated a low frequency of null alleles. In terms of biased transmission of alleles, there was no evidence indicating non-Mendelian inheritance of alleles for most of the loci used here, since some of the loci (*mTcCIR06*, *mTcCIR17*, *mTcCIR19*, *mTcCIR22*, *mTcCIR31*, *mTcCIR33*, *mTcCIR43*, and *mTcCIR61*) were previously analyzed in families, without disclosing any bias in allele transmission (Alves et al. 2003). Considering that some of the cupuassu loci were shown to be homologous in cacao (Lanaud et al. 1999), where they have been characterized and mapped without any evident, highly abnormal behavior, it seemed unlikely that the excess of homozygotes derived exclusively from biased transmission of alleles.

The three germplasm groups of accessions were genetically similar (Table 2) despite the differences in observed heterozygosity, the number of alleles, and the occurrence of private alleles. The “Amazonas” set displayed the highest expected heterozygosity (Table 2) and most of the private alleles ($n_{pa}=12$). This sample was composed by accessions collected from a wide geographic range (Fig. 1), likely containing alleles from diverse origins, which might justify the large genetic diversity and rare alleles detected. Because *T. grandiflorum* is considered a species with pre-Colombian use and not native to the state of Amazonas (Ducke 1946; Cuatrecasas 1964), specimens may be regarded as subspontaneous or cultivated, probably introduced from other provenances by migrating people. The genetic distance between “Amazonas”, “Pará”, and “Amapá” was low (Fig. 2) even though they were originally sampled at distant locations (Fig. 1). There is no information about the exact origin of the genetic materials collected, but because of their low divergence, it can be suggested that they shared a common origin, probably predominantly from Pará state.

On the other hand, the “Pará” set, which supposedly originated from the species center of diversity (Cuatrecasas 1964), displayed a single private allele (Table 2). But the “Pará” group was poorly represented at the germplasm collection ($n=10$) because the collecting expeditions covered

a limited number of sites away from the region of natural occurrence (Lima and Costa 1991; Fig. 1). Curiously, the two natural populations (“Nova Ipixuna” and “Tucuruí”), sampled closer to the *T. grandiflorum* putative center of diversity at the south/southeastern region of Pará state (Fig. 1), did not contain any private alleles when the total sample was considered. The approximate location of the *T. grandiflorum* center of diversity remains to be confirmed. The natural populations (“Nova Ipixuna”, “Tucuruí”, and “Maranhão”) displayed more genetic diversity than the growers’ sample (“Tomé-açu”) in terms of mean expected and observed heterozygosities (Table 2), but not in mean number of alleles per locus (except for “Maranhão”), suggesting that in fact selection/cultivation might have reduced the level of diversity, as initially expected.

T. grandiflorum apparently adhered to a general pattern of genetic diversity structure of some tropical trees occurring at low densities, with a low intrapopulation genetic diversity, with important levels of endogamy, likely due to biparental inbreeding, probably derived from spatial genetic structure and a partial tolerance to selfing (Rossetto et al. 1999; Collevatti et al. 2001a,b; Lemes et al. 2003; Novick et al. 2003; Latouche-Hallé et al. 2004; Sanou et al. 2005; Sereno et al. 2006). The isolation of the populations by limited pollen and seed dispersal, together with the demography of individuals at low density and occurring in clumps, may favor population differentiation. The genetic divergence among natural populations was high and significant ($\theta_p=0.301$), indicating a very limited gene flow among populations ($Nm=0.26$). Moderate, but significant, genetic differentiation among populations has been described for various Neotropical trees, usually in studies on a large geographical scale, analyzing populations at average distances over 500 km, including *S. macrophylla* in Brazil ($\theta_p=0.097$; Lemes et al. 2003) and Meso-America ($\theta_p=0.109$; Novick et al. 2003) and *C. brasiliense* in Brazil ($\theta_p=0.07$; Collevatti et al. 2001a), while in *S. humilis* (White et al. 1999) and *C. guianensis* (Dayanandan et al. 1999), much lower nonsignificant genetic differentiation was detected among populations from a limited geographical range. The high level of genetic differentiation observed between two natural populations of *T. grandiflorum* (“Nova Ipixuna” and “Tucuruí”) distant by ca. 100 km was similar to the differentiation between populations of *S. macrophylla* less than 20 km apart described by Lemes et al. (2003). On the other hand, Dutech et al. (2004) described a nonsignificant low differentiation ($F_{ST}=0.08$) between *Vouacapoua americana* populations distant by 300 km, while Sanou et al. (2005) reported a low population differentiation ($F_{ST}=0.047$) for *V. paradoxa* populations across up to 900 km in Mali, possibly, in the latter case, because of human impact on gene dispersal.

For a species considered to be an outbreeder, the cupuassu population differentiation was extremely high,

with 30.1% of the genetic diversity allocated among populations and 69.9% within populations. The genetic divergence among populations revealed by neutral loci, such as microsatellites, can derive from high mutation rate for the various alleles of the same locus in distinct populations, from genetic drift, and from population isolation. In *T. grandiflorum*, all three factors might have occurred. Species with wide geographic distribution often develop into locally adapted populations, which can become genetically distinct (Seoane et al. 2001). Colonization or foundation of new populations based on few genotypes may also lead to a strong genetic divergence by genetic drift. It is possible that this might have happened to cupuassu considering its seed dispersal by gravity and pollen flow by generalist insects. Forest fragmentation could have been another reason, causing population reduction and isolation, with the subsequent increase in genetic drift, and limiting or totally voiding gene flow among populations. Therefore, it appears that cupuassu may suffer a limited gene flow, and human interference in spreading and/or stimulating plantings might have had a smaller effect than expected in natural populations.

The high genetic divergence detected among natural populations suggested that the approach for in situ conservation would require the establishment of a large number of areas for medium- to long-term preservation of the genetic diversity. According to Namkoong (1988), the probability to keep or sample an allele at q frequency, occurring in a fraction p of populations (under Hardy–Weinberg equilibrium) when n individuals are maintained from S populations (where S is the number of populations necessary to retain or sample a target allele frequency), can be estimated by $P = \left[\frac{(1-p) + p(1-q)^{2n}}{S} \right]^S$, which can be solved for as $S = \frac{1nP}{1n \left[\frac{(1-p) + p(1-q)^{2n}}{S} \right]}$. Brown and Hardner (2000) suggested that common alleles (frequency >0.05) of localized occurrence in some populations (occurring in less than 25% of the populations; Adams 1981) must be prioritized for species conservation because they may possibly represent adaptations for specific environments, being invaluable against sudden environmental changes. Using these common alleles as reference ($q > 0.05$), occurring in 25% of the populations ($p = 0.25$), and assuming that at least 100 nonrelated individuals per population can be maintained ($n = 100$), the number of populations S , required for in situ conservation at a confidence probability of $P = 0.95$, is estimated to be a minimum of 11 populations. However, because *T. grandiflorum* populations may not be at a Hardy–Weinberg equilibrium (e.g., natural populations), a larger number of populations may be required for conservation. For ex situ conservation, germplasm collection should be planned to include many sites and sample various individuals within each site to cover more genetic diversity of the species.

Based on the same estimation of the number of populations as above ($S > 11$) for localized common alleles, seeds from at least 25 trees per population should be sampled to represent an effective number of 50 (Alves et al. 2003).

Acknowledgments We would like to thank for the financial support from the Brazilian National Research Council (CNPq), International Foundation of Science (IFS), and FAPESP.

References

- Adams WT (1981) Population genetics and gene conservation in Pacific Northwest conifers. In: Scudder GGE, Revel JL (eds) Evolution today. Proceedings of the 2nd International Congress of Systematic and Evolution Biology. Hunt Institute for Botanical Documentation, Carnegie-Mellon University, Pittsburgh, PA, USA, pp 401–415
- Alves RM, Araújo DG, Fernandes JRQ (1997) Compatibilidade entre clones de cupuaçuzeiro (*Theobroma grandiflorum*). Rev Bras Genet 20(3):148 (abstract)
- Alves RM, Artero AS, Sebbenn AM, Figueira A (2003) Mating system in a natural population of *Theobroma grandiflorum* (Willd. Ex Spreng.) Schum., by microsatellite markers. Genet Mol Biol 26:373–379
- Alves RM, Sebbenn AM, Artero AS, Figueira A (2006) Microsatellite loci transferability from *Theobroma cacao* to *Theobroma grandiflorum*. Mol Ecol Notes 6:583–586
- Azevedo VCR, Vinson CC, Ciampi AY (2005) Twelve microsatellite loci in *Manilkara huberi* (Ducke) Standl (Sapotaceae), an Amazonian timber species. Mol Ecol Notes 5:13–15
- Brondani RPV, Zucchi MI, Brondani C, Rangel PHN, Borba TCO, Rangel PN, Magalhães MR, Vencovsky R (2005) Genetic structure of wild rice *Oryza glumaepatula* populations in three Brazilian biomes using microsatellite markers. Genetica 125:115–123
- Brown AHD, Hardner CM (2000) Sampling the gene pools of forest trees for ex situ conservation. In: Young A, Boshier D, Boyle T (eds) Forest conservation genetics: Principles and practice. CSIRO, Australia, pp 183–196
- Chase MR, Moller C, Kesseli R, Bawa KS (1996) Distant gene flow in tropical tree. Nature 383:398–399
- Clement CR (1999) 1492 and the loss of Amazonian crop genetic resources. I. The relation between domestication and human population decline. Econ Bot 53:188–202
- Cockerham CC, Weir BS (1993) Estimation of gene flow from F -statistics. Evolution 47:855–863
- Collevatti RG, Grattapaglia D, Hay JD (2001a) Population genetic structure of the endangered tropical tree species *Caryocar brasiliense*, based on variability at microsatellite loci. Mol Ecol 10:349–356
- Collevatti RG, Grattapaglia D, Hay JD (2001b) High resolution microsatellite based analysis of the mating system allows the detection of significant biparental inbreeding in *Caryocar brasiliense*, and endangered tropical tree species. Heredity 86:60–67
- Cope FW (1962) The mechanism of pollen incompatibility in *Theobroma cacao*. Heredity 17:157–182
- Cope FW (1976) Cacao, *Theobroma cacao*. In: Simmonds NW (ed) Evolution of crop plants. Longman, London, pp 285–289
- Creste S, Tulmann-Neto A, Figueira A (2001) Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. Plant Mol Biol Rep 19:299–306
- Crow JF, Aoki K (1984) Group selection for a polygenic behavioural trait: estimating the degree of population subdivision. Proc Natl Acad Sci USA 81:6073–6077

- Cuatrecasas JA (1964) Cocoa and its allies: a taxonomic revision of the genus *Theobroma*. *Contrib US Natl Herb* 35(6):32–46
- Dayanandan S, Bawa KS, Kesseli R (1999) Population structure delineated with microsatellite markers in fragmented populations of a tropical tree, *Carapa guianensis* (Meliaceae). *Mol Ecol* 8:1585–1592
- Degen B (2006) Genetic data analysis and numerical test. GDA-NT. Beta version 1.0
- Degen B, Bandou E, Caron H (2004) Limited pollen dispersal and biparental inbreeding in *Symphonia globulifera* in French Guiana. *Heredity* 93:585–591
- Dick CW, Etchelecu G, Austerlitz F (2003) Pollen dispersal of tropical trees (*Dinizia excelsa*: Fabaceae) by native insects and African honeybees in pristine and fragmented Amazonian rainforest. *Mol Ecol* 12:753–764
- Ducke A (1946) Plantas de cultura precolombiana na Amazônia Brasileira: notas sobre as espécies ou formas espontâneas que supostamente lhes teriam dado origem. *Boletim Técnico*, 8, 24 p. Belém, IAN, Brazil
- Dutech C, Joly HI, Jarner P (2004) Gene flow, historical population dynamics and genetic diversity within French Guiana populations of a rainforest tree species, *Vouacapoua americana*. *Heredity* 92:69–77
- Falcão MA, Lleras E (1983) Aspectos fenológicos ecológicos e de produtividade do cupuaçu, *Theobroma grandiflorum* (Willd ex Spreng) Schum. *Acta Amazonica* 13(5–6):725–735
- Gilbert-Escrivá MV, Gonçalves LAG, Figueira A, Silva CRS (2002) Fatty acid and triacylglycerol composition and thermal behavior of fats from seeds of Brazilian Amazonian *Theobroma* species. *J Sci Food Agric* 82:1425–1431
- Glendinning DR (1960) Selfing of self-incompatible cocoa. *Nature* 187:170
- Hardy OJ, Maggia L, Bandou E, Breyne P, Caron H, Chevallier MH, Doligez A, Dutech C, Kremer A, Latouche-Hallé C, Troispoux V, Veron V, Degen B (2006) Fine-scale genetic structure and gene dispersal inferences in 10 Neotropical tree species. *Mol Ecol* 15:559–571
- Hollingsworth PM, Dawson IK, Goodall-Copestake WP, Richardson JE, Weber JC, Sotelo-Montes C, Pennington RT (2005) Do farmers reduce genetic diversity when they domesticate tropical trees? A case study from Amazonia. *Mol Ecol* 14:497–501
- Homma AKO, Walker RT, Carvalho RA, Conto AJ, Ferreira CAP (1996) Razões de risco e rentabilidade na destruição de recursos florestais: o caso de castanhais em lotes de colonos no Sul do Pará. *Rev Econômica do Nordeste* 27(3):515–535
- Knight R, Rogers H (1955) Incompatibility in *Theobroma cacao*. *Heredity* 9:67–69
- Lanaud C, Risterucci AM, Pieretti I, Falque M, Bouet A, Lagoda PJJ (1999) Isolation and characterization of microsatellites in *Theobroma cacao* L. *Mol Ecol* 8:2141–2143
- Latouche-Hallé C, Ramboer A, Bandou E, Caron H, Kremer A (2003) Nuclear and chloroplast genetic structure indicate fine-scale spatial dynamics in a neotropical tree population. *Heredity* 91:181–190
- Latouche-Hallé C, Ramboer A, Bandou E, Caron H, Kremer A (2004) Long-distance pollen flow and tolerance to selfing in a neotropical tree species. *Mol Ecol* 13:1055–1064
- Lemes MR, Gribel R, Proctor J, Grattapaglia D (2003) Population genetic structure of mahogany (*Swietenia macrophylla* King, Meliaceae) across the Brazilian Amazon, based on variation at microsatellite loci: implications for conservation. *Mol Ecol* 12: 2875–2883
- Lewis PO, Zaykin D (1999) GDA—Genetic data analysis: version 1.0 (d12) for Windows (software), 39 p. The University of New Mexico, Albuquerque, NM, USA. <http://www.lewis.eeb.uconn.edu/lewishome/software.html>
- Lima RR, Costa JPC (1991) Registro de introduções de plantas de cultura pré-colombiana coletadas na Amazônia Brasileira. *Série Documentos*, 58, 191 p. Belém, Embrapa, Brazil
- Manly BFJ (1997) Randomization, bootstrap and Monte Carlo Methods in biology. Chapman & Hall, London, UK
- Miller MP (1997) Tools for Population Genetic Analyses (TFPGA) version 1.3: a windows program for the analysis of allozyme and molecular population genetic data©. Flagstaff Northern Arizona University, Flagstaff, AZ
- Morand ME, Brachet S, Rossignol P, Dufour J, Frascaria-Lacoste N (2002) A generalized heterozygote deficiency assessed with microsatellites in French common ash populations. *Mol Ecol* 11:377–385
- Motamayor JC, Risterucci AM, Lopez PA, Ortiz CF, Moreno A, Lanaud C (2002) Cacao domestication I: the origin of the cacao cultivated by the Mayas. *Heredity* 89:308–386
- Namkoong G (1988) Sampling for germplasm collections. *Hortscience* 23:79–81
- Nei M (1978) Estimation of average heterozygosity and genetic distance from small number of individuals. *Genetics* 89:583–590
- N’Goran JAK, Laurent V, Risterucci AM, Lanaud C (2000) The genetic structure of cocoa (*Theobroma cacao* L.) populations revealed by RFLP analysis. *Euphytica* 115:83–90
- Novick RR, Dick CW, Lemes MR, Navarro C, Caccone A, Bermingham E (2003) Genetic structure of Mesoamerican populations of big-leaf mahogany (*Swietenia macrophylla*) inferred from microsatellite analysis. *Mol Ecol* 12:2885–2893
- Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Mol Ecol* 13:1143–1155
- Ronning CM, Schnell RJ (1994) Allozyme diversity in a germplasm collection of *Theobroma cacao* L. *J Hered* 85:291–295
- Rossetto M, Slade RW, Baverstock PR, Henry RJ, Lee LS (1999) Microsatellite variation and assessment of genetic structure in tea (*Melaleuca alternifolia*—Myrtaceae). *Mol Ecol* 8:633–643
- Sanou H, Lovett PN, Bouvet JN (2005) Comparison of quantitative and molecular variation in agroforestry populations of the shea tree (*Vitellaria paradoxa* C.F. Geartn) in Mali. *Mol Ecol* 14:2601–2610
- Seoane CES, Sebbenn AM, Kageyama PY (2001) Sistema de reprodução em populações de *Esenbeckia leiocarpa* Engl. *Rev do Instituto Florestal* 13(1):19–26
- Sereno ML, Albuquerque PSB, Vencovsky R, Figueira A (2006) Genetic diversity and natural population structure of cacao (*Theobroma cacao* L.) from the Brazilian Amazon evaluated by microsatellite markers. *Conserv Genet* 6:13–24
- Shepherd M, Cross M, Maguire TL, Dieters MJ, Williams CG, Henry RJ (2002) Transpecific microsatellites for hard pines. *Theor Appl Genet* 104:819–827
- van Treuren K, Karkkainen H, Baena-Gonzalez K, Savolainen E (1997) Evolution of microsatellites in *Arabis petraea* and *Arabis lyrata*, outcrossing relatives of *Arabidopsis thaliana*. *Mol Biol Evol* 14:220–229
- Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. *Trends Biotechnol* 23:48–55
- Velho CC, Whipkey A, Janick J (1990) Cupuaçu: a new beverage crop for Brazil. In: Janick J, Simon JE (eds) *Advances in new crops*. Timber, Portland, OR, USA, pp 372–375
- Ward M, Dick CW, Gribel R, Lowe AJ (2005) To self, or not to self: a review of outcrossing and pollen-mediated gene flow in neotropical trees. *Heredity* 95:246–254
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370
- White GM, Boshier DH, Powell W (1999) Genetic variation within a fragmented population of *Swietenia humilis* Zucc. *Mol Ecol* 8:1899–1909
- Zucchi MI, Brondani RP, Pinheiro JB, Chaves LJ, Coelho AS, Vencovsky R (2003) Genetic structure and gene flow in *Eugenia dysenterica* DC in the Brazilian Cerrado utilizing SSR markers. *Genet Mol Biol* 26:449–457