

Genetic diversity and mating system of *Copaifera langsdorffii* (Leguminosae/ Caesalpinioideae)

A. Gonela¹, A.M. Sebbenn², H.H. Soriani^{1,3}, M.A. Mestriner¹,
C.A. Martinez³ and A.L. Alzate-Marin^{1,4}

¹Laboratório de Genética Vegetal, Departamento de Genética,
Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo,
Ribeirão Preto, SP, Brasil

²Estação Experimental de Tupi, Instituto Florestal de São Paulo,
Piracicaba, SP, Brasil

³Departamento de Biologia,
Faculdade de Filosofia Ciências e Letras de Ribeirão Preto,
Universidade de São Paulo, Ribeirão Preto, SP, Brasil

⁴Programa de Pós-Graduação em Genética,
Departamento de Genética, Universidade de São Paulo,
Ribeirão Preto, SP, Brasil

Corresponding author: A.L. Alzate-Marin
E-mail: anaalzate@fmrp.usp.br

Genet. Mol. Res. 12 (1): 569-580 (2013)

Received June 11, 2012

Accepted October 13, 2012

Published February 27, 2013

DOI <http://dx.doi.org/10.4238/2013.February.27.6>

ABSTRACT. *Copaifera langsdorffii*, locally known as copaíba, is a valuable tropical tree with medicinal properties of its oil. We studied the genetic variation, genetic structure, and the mating system of trees in stands of *C. langsdorffii* (Leguminosae/Caesalpinioideae) located in an extensive area between the Pardo and Mogi-Guaçu basins in São Paulo State, Brazil, and their offspring, conserved in an *ex situ* germplasm bank at the University of São Paulo in Ribeirão Preto, SP, Brazil, using six microsatellite loci. Leaves were collected from 80 seed trees and from 259 offspring and their DNA extracted. A total of

140 and 175 alleles were found in the seed trees and their offspring, respectively. Low genetic differentiation was observed between stands, indicating intense gene flow due to efficient pollen dispersion vectors. An estimation of the outcrossing rate showed that these stands are outcrossed ($t_m = 0.98$, $P > 0.05$). The mean variance of the effective population size of each family in two of the stands was 3.69 and 3.43, while the total effective population size retained in the germplasm bank was between 81 and 96. The paternity correlation was low, ranging from 0.052 to 0.148, demonstrating that the families implanted in this germplasm bank are composed predominantly of half-sibs.

Key words: Copaiba; Tropical tree species; Conservation; Gene bank

INTRODUCTION

The Atlantic forest is a hotspot for conservation priorities (Myers et al., 2000). In Brazil, the formations of Atlantic Forest have been almost completely destroyed, and less than 5% of its original vegetation remains, dispersed into several thousand fragments (Zorzetto et al., 2003). As the Atlantic Forest landscapes, including semi-deciduous forests, become increasingly fragmented, populations of forest species are reduced and ecosystems are altered, resulting in a progressive erosion of biological diversity (Tilman et al., 1994).

Until the mid-nineteenth century, the natural vegetation of the Ribeirão Preto region located in northeastern São Paulo, in southwest Brazil, was essentially intact. However, the rapid expansion of farming, particularly the monocultures of coffee and sugar cane, led to the devastation and fragmentation of the original vegetation (Kotchetkoff-Henriques, 2003; Zorzetto et al., 2003), causing a drastic loss of forest genetic resources.

With the objective to rescue and conserve these regional genetic resources, the University of São Paulo (USP) Forest Project at the Ribeirão Preto campus was established in 1997. One of the purposes of this project was to establish a gene bank (BG-USP/RP) to rescue native tree species in the Ribeirão Preto region. Among approximately 75 ha available for reforestation, 45 of them have been reserved as a gene bank shelter containing the offspring of 44 endangered native forest tree species, including *Copaifera langsdorffii*, known as copaiba. The conservation of this species, whose oil has high economic and medicinal value, is important to the income of many of Brazil's native populations and is also considered a priority by Food and Agriculture Organization of the United Nations (FAO's report, 2001).

Forest genetic resources are invaluable to humanity, not only as suppliers of products and services or as an aid to economic development, but also for their untapped potential in areas such as medical research. Although there are a large number of species that are used, only approximately 500 species have been systematically studied for their current utility and potential (FAO, 2011). Conservation and utilization strategies require fundamental knowledge about the levels of genetic diversity, population genetic structure and effective population size, because these are the key elements that determine the ability to respond to selection, either by natural or artificial procedures (Reis and Grattapaglia, 2004).

Microsatellites, also called simple sequence repeats (SSRs), are among the molecular markers most suitable for studying the genetic diversity of populations due to the high level

of information produced (Ashley, 2010). SSRs are highly polymorphic and are inherited in a codominant Mendelian manner, both in neutral and multiallelic forms. Their successful use to resolve the genetic relatedness of tree species has already been extensively documented (Ferreira-Ramos et al., 2008; Feres et al., 2009; Guidugli et al., 2012).

The purpose of our research was to characterize the genetic diversity, structure and mating system of three stands of *C. langsdorffii* climatic trees and their offspring, which are implanted in an *ex situ* germplasm bank in Ribeirão Preto in northwestern São Paulo State, located in southwestern Brazil, using six microsatellite loci to generate useful information for conservation strategies. We specifically addressed the following questions: a) What are the levels of genetic diversity of seed trees and their offspring? b) What are the levels of correlated paternity, effective number of pollen donors, coancestry coefficient and effective population size within an array of open-pollinated progeny from the stands? and c) How efficient was the seed collection that was performed to establish the gene bank?

MATERIAL AND METHODS

Species studied

C. langsdorffii Desf. (family Leguminosae, subfamily Caesalpinioideae), a tree that is amply distributed in Brazil, is an important species in the Atlantic rain forest's geomorphologic domain, which includes the semi-deciduous "planalto" forests (Leite and Lleras, 1993). *C. langsdorffii* grows naturally in the region extending from the Amazon to the State of Rio Grande do Sul. This species also grows in northeastern Argentina, southern Bolivia and northern Paraguay, and is able to persist in both riparian and savanna forests (Leite and Lleras, 1993; Almeida et al., 1998; Freitas and Oliveira, 2002; Pinto et al., 2004). The species has hermaphroditic greenish-white flowers, which last only one day and are congregated in paniculate inflorescences without petals (Freitas and Oliveira, 2002). The nectariferous flowers of *C. langsdorffii* are intensively visited by diverse insects, mainly of the families *Apinae* and *Vespidae*, with *Trigona* sp and *Apis mellifera* being the most frequent visitors. Social wasps, Diptera, Lepidoptera, and Coleoptera often constitute the secondary pollinators of this species (Freitas and Oliveira, 2002). *C. langsdorffii* forms fruits through preferential crosses, and has been reported as self-incompatible and an apomictic species (Freitas and Oliveira, 2002). However, Oliveira et al. (2002) and Sebbenn et al. (2011) observed significant selfing. Upon fructification, the trees are visited by birds, which are the major carriers responsible for their seed dispersal (Motta Junior and Lombardi, 1990; Pedroni et al., 2002). Copaíba oil, a transparent liquid that varies in color from yellow to light brown that is the exuded material from these trees, can be tapped.

Plant materials

The study was conducted in an *ex situ* gene bank established at the University of São Paulo, at the Ribeirão Preto campus in northeastern São Paulo State, Brazil (21°10'S, 47°48'W), and included its respective seed trees (Figure 1). Seeds used for establishing the gene bank (BG-USP/RP) were collected mostly from farms for three years (1999/stand A, 2000/stand B, 2001/stand C) in different areas of the rainforest that remain greatly impacted

by the agricultural practices of the sugar cane industry in the Ribeirão Preto region between the Pardo and Mogi-Guaçu basins in São Paulo State in the southeastern region of Brazil (Figure 1). Aiming to minimize genetically related samples, naturally pollinated seeds were collected from 80 seed trees (an average of 27 trees/year) located at least 500 m apart from each other. On average, 10 offspring of each seed tree were produced and used to establish the gene bank, totaling 800 offspring. However, many offspring died as well, especially from stand A. Aiming to quantify the genetic diversity in seed trees and the germplasm bank, leaves were collected from 80 seed trees (mapped using GPS - GARMIN eTrex[®]) of stands A, B and C and from 259 of the offspring of BG-USP/RP, and these samples were placed into plastic bags. After collection, the leaves were frozen at -20°C. Samples of adults and an average of 6 and 3 offspring per family from stands B and C, respectively, were used for crossing rate studies.

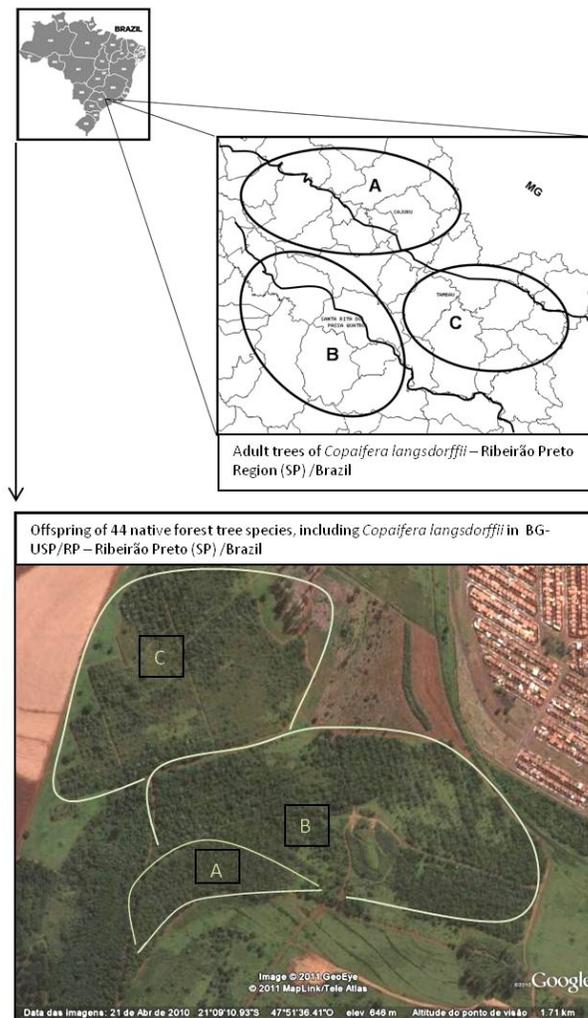


Figure 1. Geographic location of the study areas.

DNA extraction and SSR analyses

Total genomic DNA was extracted from the leaves of each sampled tree and its offspring using the method described by Alzate-Marin et al. (2009). Six microsatellite (SSR) primer pairs (CL01, CL06, CL20, CL27, CL32, and CL34) previously developed by Ciampi et al. (2000) were used for this study. Microsatellite loci were amplified by polymerase chain reaction (PCR) in a final volume of 10 μ L containing 10 ng DNA, 0.2 mM of each dNTP, 1.5 mM $MgCl_2$, 10 mM Tris-HCl, pH 8.9, 50 mM KCl, 0.3 μ L each of primer forward and reverse (Bio-Synthesis, USA), and 0.12 U *Taq* DNA polymerase (Biotools). Amplifications were performed with an Eppendorf MasterCycler using the following protocol: 96°C for 5 min, 28 cycles of denaturation at 94°C for 1 min, annealing at 46°-52°C for 1 min, and extension at 72°C for 1 min, and a final elongation step at 72°C for 7 min. PCR products were denatured and separated on 12% denaturing polyacrylamide gels stained with silver nitrate. Alleles were sized by comparison to a standard 10-bp DNA ladder (Invitrogen, Carlsbad, CA, USA).

Data analysis

Genetic diversity of the adult generation and its families was characterized by the following indices: average number of alleles per locus (A), observed heterozygosity (H_o), expected heterozygosity (H_e) under the assumptions of Hardy-Weinberg equilibrium (HWE), and the effective number of alleles per locus [$\hat{A}_e = 1/(1 - \hat{H}_e)$]. The fixation index (F) was also estimated according to Weir (1996):

$$\hat{F} = \frac{(\hat{H}_e - \hat{H}_o) + \frac{1}{2n} \hat{H}_o}{\hat{H}_e - \frac{1}{2n} \hat{H}_o}, \text{ where } n \text{ is the sample size.}$$

The statistical significance of F values was assessed from 10,000 bootstraps across loci. Hierarchical population genetic structure was assessed (Weir and Cockerham, 1984) by the coefficient of coancestry among plants within populations or genetic divergence between populations (θ_p). To verify whether the average estimation of θ_p was statistically different from zero, the 95 and 99% confidence intervals were estimated using the bootstrap method (10,000 repetitions over loci). All of these analyses were calculated using the GDA program (Lewis and Zaykin, 2002).

The fixation index of the families (F_o) was split into its components resulting from self-fertilization (F_s) and mating among relatives ($F_{t_m-t_s}$), $F_o = F_s + F_{t_m-t_s}$. Inbreeding originating from self-fertilization (s) was estimated by $F_s = 0.5\hat{s}(1 + \hat{F}_p)$, where F_p is the fixation index in seed trees. Inbreeding due to mating among relatives ($1 - s$) was estimated by $\hat{F}_{t_m-t_s} = \hat{\theta}_p$. Thus, $\hat{F}_o = 0.5\hat{s}(1 + \hat{F}_p) + (1 - \hat{s})\hat{\theta}_p$ (Sebbenn, 2006). Therefore, with F_o and F_s , $F_{t_m-t_s}$ can be used for $\hat{F}_{t_m-t_s} = \hat{F}_o - \hat{F}_s$ without presupposing inbreeding equilibrium.

Mating system parameters [population multilocus outcrossing rate (t_m), single-locus outcrossing rate (t_s), outcrossing rate among relatives ($t_m - t_s$), multilocus paternity correlation ($r_{p(m)}$), and single-locus paternity correlation ($r_{p(s)}$)] were estimated based on the

mixed-mating model and the correlated-mating model (Ritland and Jain, 1981). These parameters were calculated using the MLTR program (Ritland, 2002) and the maximum likelihood method (Algorithm EM, Expectation-Maximization). The standard deviation (SD) of the estimates of the parameters was calculated by means of 500 bootstrap replicates and was used to calculate a 95% confidence interval for the error [$IC_{95\%} = 1.96(DP/\sqrt{m})$, where m represents the number of families]. For the analyses, the parameter self-fertilization correlation (r_s) was fixed as having a value of zero because the rate of self-fertilization in the populations sampled had been low ($t_m > 0.9$), and practically no information on the self-fertilization correlation is extracted in such circumstances (Sampson, 1998). Mating system parameters were used for estimating other demographic and genetic parameters. The effective number of pollen donors was calculated from the paternity correlation using $\hat{N}_{ep} = 1/\hat{r}_p$. The frequency of half-sibs [$\hat{P}_{hs} = \hat{t}_m^2(1 - \hat{r}_p)$] and full-sibs ($\hat{P}_{fs} = \hat{t}_m^2\hat{r}_p$) within families was estimated. The average coefficient of coancestry (Θ) within families was estimated from the coefficient of relatedness, $\hat{r} = 0.25(1 + \hat{F}_p)[4\hat{s} + (\hat{t} + \hat{t}\hat{s}_s)(1 + \hat{r}_p)]$, considering that in diploid species without inbreeding, the coancestry coefficient is half of the coefficient of relatedness ($\Theta = r/2$). The correlation coefficient (r_s) was assumed to be zero r_s , and the coancestry was calculated using $\hat{\Theta} = 0.125(1 + \hat{F}_p)[4\hat{s} + \hat{t}(1 + \hat{r}_p)]$. The genetic representation of the families was estimated by the variance effective size:

$$\hat{N}_{e(v)} = \frac{0.5}{\hat{\Theta}\left(\frac{n-1}{n}\right) + \frac{1 + \hat{F}_o}{2n}} \quad (\text{Cockerham, 1969}).$$

The number of mother trees for seed collection was calculated assuming that infinite seeds would be collected from each tree ($n = \infty$) and that the objective would be retained the total effective sample size of 100, $\hat{m} = N_{e(\text{reference})} / \hat{N}_{e(v)}$ (Sebbenn, 2006).

The total effective population size in the germplasm bank was calculated multiplying the number of families sampled (m) of each stand (22 for B and 29 for C) by the averaged variance effective size: $\hat{N}_{e(v)t} = m\hat{N}_{e(v)}$.

RESULTS

All 6 microsatellite loci analyzed were polymorphic (Table 1). A total of 140 and 175 alleles were found in 80 seed trees and their offspring, respectively. In adults, the number of alleles among loci ranged from 22 to 30 and in offspring from 24 to 34, with a mean of 23.33 (adults) and 29.20 (offspring) alleles per locus. The H_o ranged from 0.637 to 0.780, with a mean of 0.715, and H_e ranged from 0.817 to 0.928, with a mean of 0.882. All loci showed a significant excess of homozygosity in comparison to the expected ratio for the HWE model (Table 1).

Table 1. Characteristics of the 6 microsatellite loci used in *Copaifera langsdorffii* studies.

Locus	N	Ka	Kp	\hat{H}_E	\hat{H}_O	\hat{F}
CL01	339	20	24	0.854	0.780	0.087**
CL06	339	25	31	0.898	0.753	0.162**
CL20	339	30	34	0.928	0.643	0.307**
CL27	339	22	26	0.882	0.637	0.277**
CL32	339	22	32	0.913	0.767	0.160**
CL34	339	21	28	0.817	0.713	0.128**
Mean		23.3	29.2	0.882	0.715	0.189**
Total		140	175			

**P < 0.01. N = sample size; Ka and Kp = allele number detected in adult and progeny, respectively; \hat{H}_O = observed heterozygosity; \hat{H}_E = expected heterozygosity; \hat{F} = fixation index.

A larger average number of alleles per locus was observed in the offspring ($\hat{A} = 20.11$) in comparison to the seed trees ($\hat{A} = 16.06$), suggesting that the reproductive area of the populations can be bigger than the area sampled (Table 2). The effective number of alleles per locus was smaller than the average number of alleles per locus, and this was higher in the adult trees ($\hat{A}_E = 8.37$) than the offspring ($\hat{A}_E = 7.19$), showing that a large number of the alleles have low frequency. The H_E and H_O were statistically similar between seed trees and offspring (Table 2). The genetic differentiation between stands was low for both adults ($\hat{\theta}_p = 0.007$) and offspring ($\hat{\theta}_p = 0.026$), although offspring differentiation significantly differed from zero (Table 3).

Table 2. Estimates of genetic diversity indexes for *Copaifera langsdorffii* adult trees and their offspring.

Stand	N	\hat{A}	\hat{A}_E	\hat{H}_E	\hat{H}_O
Adults					
A	30	15.83	8.39	0.881	0.743
B	22	15.00	8.48	0.882	0.908
C	28	17.33	8.26	0.879	0.706
Mean		16.06	8.37	0.881 ± 0.00153^{ns}	0.786 ± 0.10755^{ns}
Offspring					
A	40	16.00	6.05	0.835	0.649
B	127	23.17	7.43	0.865	0.727
C	92	21.17	8.54	0.883	0.676
Mean		20.11	7.19	0.861 ± 0.02425^{ns}	0.684 ± 0.03961^{ns}

N = sample size; \hat{A} = average number of alleles for locus; \hat{A}_E = effective number of alleles for locus; \hat{H}_O = observed heterozygosity; \hat{H}_E = expected heterozygosity. ns = not statistically significant (paired sample *t*-test).

Table 3. *F*-statistics for 2 generations of *Copaifera langsdorffii* populations.

Generation	\hat{f}	\hat{F}	$\hat{\theta}_p$
Adults	0.114**	0.120**	0.007
Progenies	0.197*	0.217**	0.026**

*P < 0.05; **P < 0.01.

The multilocus outcrossing rate was not significantly different from unity in both stands B and C (ranging from 0.984 ± 0.054 to 0.989 ± 0.046 ; Table 4). The single-locus outcrossing rate was significantly lower than the multilocus outcrossing rate (ranging from

0.718 ± 0.012 to 0.759 ± 0.010). The difference between the multilocus and single-locus rate ($t_m - t_s$) was significantly different from zero in the two stands (ranging from 0.225 to 0.271), suggesting mating among relatives in the populations. The multilocus paternity correlations ($r_{p(m)}$) were also low (varying from 0.052 to 0.148), but significantly different from zero in the stands studied (Table 4). The effective number of pollen donors ranged from 7 to 19.

Table 4. Parameters of mating system for two stands of *Copaifera langsdorffii*.

Parameters	Stand B	Stand C
Reproduction system		
Multilocus outcrossing rate: t_m	0.984 (0.054)	0.989 (0.046)
Single-locus outcrossing rate: t_s	0.759 (0.010)	0.718 (0.012)
Mating rate among relatives: $t_m - t_s$	0.225 (0.053)	0.271 (0.042)
Multilocus paternity correlation: $r_{p(m)}$	0.052 (0.012)	0.148 (0.025)
Single-locus paternity correlation: $r_{p(s)}$	0.083 (0.010)	0.162 (0.023)
Crossings involving pollinator tree relatives: $r_{p(s)} - r_{p(m)}$	0.030 (0.012)	0.014 (0.017)
Average number of pollen donors: $1/r_{p(m)}$	19.2	6.8
Inbreeding and genetic structure		
Fixation index in adult population: F_a	-0.030	0.200**
Fixation index in the offspring: F_o	0.160**	0.235**
Inbreeding by self-fertilization: $F_s = 0.5s(1 + F_{mat})$	0.008	0.006
Biparental inbreeding: $F_{in-sibs} = F_o - F_s$	0.152	0.229
Average proportions of half-sibs: $\hat{p}_{hs} = \hat{t}_m^2(1 - \hat{r}_{p(m)})$	0.902	0.829
Average proportions of full-sibs: $\hat{p}_{fs} = \hat{t}_m^2 \hat{r}_{p(m)}$	0.050	0.144
Average coefficient of coancestry within families: $\hat{\Theta}_{xy}$	0.135	0.146
Variance effective size: $N_{e(v)}$	3.69	3.43
Total variance effective size: $\hat{N}_{e(v)r} = m\hat{N}_{e(v)}$	81.18	96.04
Number of seed trees to retain the effective size of 100	27.1	29.2

**P < 0.01. Confidence intervals for 95% error probability are reported in parentheses.

The fixation index was high and statistically significant in the seed trees of stand C, and in offspring of both stands B and C, suggesting inbreeding in the populations (Table 4). The partition of the inbreeding of the families into its components of self-fertilization and biparental inbreeding indicates that mating among relatives was the main cause of inbreeding. The progeny were predominantly half-sibs (83 to 90%, $\hat{\Theta}$ ranging from 0.135 to 0.146).

The variance effective population size was high (ranging from 3.43 to 3.69). This information can also be used to determine the effective size retained in the bank. Seeds were originally collected from 22 (stand B) and 28 (stand C) seed trees, and as each family represents the variance effective size range from 3.69 to 3.43, it is possible to conclude that the effective size retained is between 81 and 96 for stands B and C, respectively. The number of seed trees necessary for collecting seeds for the purpose of retaining the effective size of 100 was estimated to be 28 (Table 4).

DISCUSSION

Our results indicate that more than 97% of the genetic diversity was distributed within stands and suggest that these stands are within the same genetic zone (Graudal et al., 1997), since the genetic divergence among populations was low. This pattern of genetic structure is

expected for tree species with efficient mechanisms of gene flow, such as tropical tree species that are pollinated by animals and have seed dispersal. Previous studies with *C. langsdorffii* observed a similar pattern with low genetic differentiation among populations (Carvalho and Oliveira, 2004; Martins et al., 2008). The observed low genetic differentiation between the populations indicates the lack of isolation and intense gene flow between the populations, which homogenizes its genetic diversity. *C. langsdorffii* is pollinated by bees, which can travel kilometers in search of nectar and pollen, and the dispersal of seeds occurs by zoochory, in this case birds, which have the capacity to cover great distances. Thus, the observed model of genetic structure can be explained by the efficient vectors of pollen and seed dispersal in this species.

The high multilocus outcrossing rate estimates for stands B and C indicate that the *C. langsdorffii* species is outcrossed. Such estimates are slightly higher than what has been reported for a natural *C. langsdorffii* population ($\hat{t}_m = 0.917 \pm 0.041$, Oliveira et al., 2002). The present estimates are probably overestimated, considering that these were calculated using families established in an *ex situ* bank. There is a great probability that selection against self-fertilizing individuals was increased between the reproductive event and the moment that the genotypes were accessed. Self-fertilization seeds could have been aborted before germination, died in the nursery phase or in the *ex situ* germplasm bank. Thus, the differential mortality of self-seeds and offspring can be overestimated by the outcrossing rate. Selection between the reproductive event and seed production has already been documented in *Platypodium elegans* (Hufford and Hamrick, 2003). Alternatively, the outcrossing rate can vary between trees of a population, different reproductive events (Rocha and Aguilar, 2001; Hufford and Hamrick, 2003) and between populations (Ribeiro and Lovato, 2004), and be affected by anthropogenic interferences such as fragmentation (Doligez and Joly, 1997) and logging (Obayashi et al., 2002). These factors could also explain the small differences observed between the studies.

The fixation index estimated in the adult generation was lower than that in the offspring in both stands (B and C). These results suggest the presence of selection against homozygous individuals between the seedling and adult stages. This phenomenon has also been observed in other studies with tropical forest species (Alves et al., 2003).

Although the outcrossing rate was high for the stands, a substantial part of these matings occur among related individuals (minimum 22.5%). In plants with hermaphroditic flowers, the mating among relatives is another source of inbreeding (biparental inbreeding). The partition of the fixation index into its components of self-fertilization and mating among relatives indicates that nearly 100% of the inbreeding occurred due to mating among relatives. The mating among relatives produces inbreeding equal to the coancestry between the crossed parents ($F_o = \theta_{xy}$). Thus, the observed inbreeding can be explained by mating among half-sibs and full-sibs. This fact also suggests the presence of spatial genetic structure within the population.

The correlation of paternity estimated in the stands of *C. langsdorffii* was low (Table 4), suggesting that a few matings were correlated and families were composed mainly of half-sibs. This demonstrates that the families of *C. langsdorffii* implanted in the USP/RP germplasm bank are composed predominantly of half-sibs, suggesting that high genetic diversity was retained *ex situ*. Estimates of the correlation of paternity in other tropical tree species have generally shown higher values. For example, in *Senna multijuga*, the estimated paternity correlation ranged from 0.249 to 0.349 (Ribeiro and Lovato, 2004) and in *Ceiba pentandra* from 0.400 to 0.600 (Lobo et al., 2005). The lower correlation of paternity in *C. langsdorffii*

could be explained by the fact that each fruit contains only one seed, reducing the probability of paternity among seeds from different fruits. This hypothesis is based on studies of paternity correlation estimates between and within fruits, which shows a lower correlation of paternity among seeds of different fruits, compared to within fruits (Ritland, 1989; Lobo et al., 2005).

The effective number of pollen donors (minimum of 7) observed in this study was higher than what has been detected in other tropical tree species such as *Hymenaea courbaril*, *Enterolobium cyclocarpum*, *Theobroma grandiflorum*, and *Myracrodruon urundeuva* (Rocha and Aguilar, 2001; Alves et al., 2003; Moraes et al., 2004; Feres et al., 2009). Also, a high number of pollen donors have been detected in species including *Carapa pocera* ($\hat{N}_{ep} = 12$) and *Tabebuia roseo-alba* (among fruits) ($\hat{N}_{ep} = 8-9$) (Doligez and Joly, 1997; Feres et al., 2012).

The average coefficient of coancestry and the variance in the effective population size within families observed in this study (Table 4) was approximately what was expected for half-sib families (0.125 and 4, respectively), confirming that the offspring of the BG-USP/RP bank possess the high genetic variability of *C. langsdorffii*. Also, effective sizes on the order of 80 (see Table 4) can retain the original genetic variation in the short-term (<10 generations), where in this species it can represent at least 500 years.

In conclusion, the present results confirm that the BG-USP/RP gene bank retains substantial genetic diversity of samples of *C. langsdorffii* populations. Therefore, this study validates the representation within the existing germplasm collection and helps to delineate future strategies for conservation of the species.

ACKNOWLEDGMENTS

Research supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, Grant #03/04199-4 to A.L. Alzate-Marin), Provost for Research of São Paulo University (A.L. Alzate-Marin) and Fundação de Apoio ao Ensino, Pesquisa e Assistência (FAEPA; M.A. Mestriner) grants. This research is part of the post-doctoral fellowship of A. Gonela. H.H. Soriani was supported by an undergraduate fellowship from CNPq. A.M. Sebben and C.A. Martinez are recipients of CNPq research fellowships. A.L. Alzate-Marin was supported by a research assistantship from FAPESP (#2004/01378-8) and CNPq (PD Senior #150277/2009-1, PV #300140/2011-8). We are grateful to Maria C.T. Canas and Ana L. Pimentel for their technical support in the microsatellite analyses and Antonio Justino da Silva and Marco Aurelio Candin for their sampling support.

REFERENCES

- Almeida SP, Proença CEB, Sano SM and Ribeiro LF (1998). Cerrado - Espécies Vegetais Úteis. CPAC - Embrapa, Planaltina.
- Alves RM, Artero AS, Sebbenn AM and Figueira A (2003). Mating system in natural population of *Theobroma grandiflorum* (Willd. ex Spreng.) Shumm., by microsatellite markers. *Genet. Mol. Biol.* 26: 373-379.
- Alzate-Marin AL, Guidugli MC, Soriani HH, Martinez CA, et al. (2009). An efficient and rapid DNA miniprep procedure suitable for PCR/SSR and RAPD analyses in tropical forest tree species. *Braz. Arch. Biol. Technol.* 52: 1217-1224.
- Ashley MV (2010). Plant parentage, pollination, and dispersal: How DNA microsatellites have altered the landscape. *Crit. Rev. Plant Sci.* 29: 148-161.
- Carvalho D and Oliveira AF (2004). Genetic structure of *Copaifera langsdorffii* Desf. natural populations. *Cerne* 10: 137-153.
- Ciampi AY, Brondani RPV and Grattapaglia D (2000). Otimização de Sistemas Fluorescentes de Genotipagem Multiloco

- e Desenvolvimento de Marcadores Microsatélites para *Copaifera langsdorffii* Desf. (Copaíba) Leguminosae-Caesalpinioideae. Embrapa Recursos Genéticos e Biotecnologia, Brasília, 40.
- Cockerham CC (1969). Variance of gene frequencies. *Evolution* 23: 72-84.
- Doligez A and Joly H (1997). Mating system of *Carapa procera* (Meliaceae) in the french guiana tropical forest. *Am. J. Bot.* 84: 461.
- FAO's report (2001). Global forest resources assessment, tropical South America. *FAO Forestry Paper* 140: 287-291.
- FAO (2011). Recursos Genéticos Forestales. Available at [<http://www.fao.org/forestry/fgr/es/>]. Accessed September 10, 2011.
- Feres JM, Guidugli MC, Mestriner MA, Sebbenn AM, et al. (2009). Microsatellite diversity and effective population size in a germplasm bank of *Hymenaea courbaril* L. var *stilbocarpa* (Leguminosae), an endangered tropical tree: recommendations for conservation. *Genet. Resour. Crop Evol.* 56: 797-807.
- Feres JM, Sebbenn AM, Guidugli MC, Mestriner MA, et al. (2012). Mating system parameters at hierarchical levels of fruits, individuals and populations in the Brazilian insect-pollinated tropical tree, *Tabebuia roseo-alba* (Bignoniaceae). *Conserv. Genet.* 13: 393-405.
- Ferreira-Ramos R, Laborda PR, Santos MO, Mayor MS, et al. (2008). Genetic analyses of forest species *Eugenia uniflora* L. through of newly developed SSR markers. *Conserv. Genet.* 9: 1281-1285.
- Freitas CV and Oliveira PE (2002). Reproductive biology of *Copaifera langsdorffii* Desf. (Leguminosae, Caesalpinioideae). *Rev. Bras. Bot.* 25: 311-321.
- Graudal L, Kjaer E, Thomsen A and Larsen AB (1997). Planning national programmes for conservation of forest genetic resources. Danida Forest Seed Centre, Humlebaek, Denmark Technical Note No. 48. Available at [<http://curis.ku.dk/portal-life/files/20710949/tn48.pdf>]. Accessed January 30, 2013.
- Guidugli MC, Ferreira-Ramos R, de Sousa AC, Cidade FW, et al. (2012). Genetic diversity of *Metrodorea nigra* (Rutaceae) from a small forest remnant in Brazil assessed with microsatellite markers. *Genet. Mol. Res.* 11: 10-16.
- Hufford KM and Hamrick JL (2003). Viability selection at three early life stages of the tropical tree, *Platydictyon elegans* (Fabaceae, Papilionoideae). *Evolution* 57: 518-526.
- Kotchetkoff-Henriques O (2003). Characterization of the Natural Vegetation in Ribeirão Preto, SP: Bases for Conservation. DSc thesis, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto. Available at [<http://www.bv.fapesp.br/pt/dissertacoes-teses/6724/caracterizacao-vegetacao-natural-ribeirao-preto/>]. Accessed January 30, 2013.
- Leite AMC and Lleras E (1993). Priority areas for genetic resources of native forest species: preliminary phase. *Acta Bot. Bras.* 7: 61-94.
- Lewis PO and Zaykin D (2002). GDA - Genetic Data Analysis: Version 1.1 for Windows 95/NT. Available at [<http://hydrodictyon.eeb.uconn.edu/people/plewis/software.php>]. Accessed January 30, 2013.
- Lobo JA, Quesada M and Stoner KE (2005). Effects of pollination by bats on the mating system of *Ceiba pentandra* (Bombacaceae) populations in two tropical life zones in Costa Rica. *Am. J. Bot.* 92: 370-376.
- Martins K, Santos JD, Santos JD, Moreno MA, et al. (2008). Estrutura genética populacional de *Copaifera langsdorffii* Desf. (Leguminosae - Caesalpinioideae) em fragmentos florestais no Pontal do Paranapanema, SP, Brasil. *Rev. Bras. Bot.* 31: 61-69.
- Moraes MLT, Kageyama PY and Sebbenn AM (2004). Correlated mating in dioecious tropical tree species, *Myracrodruon urundeuva* Fr. All. *Forest Genet.* 11: 53-59.
- Motta Junior JC and Lombardi JA (1990). Aves como agentes dispersores da copaíba (*Copaifera langsdorffii*, Caesalpinioideae) em São Carlos, estado de São Paulo. *Ararajuba* 1: 105-106.
- Myers N, Mittermeier RA, Mittermeier CG, Fonseca GAB, et al. (2000). Biodiversity hotspots for conservation priorities. *Nature* 403: 853-858.
- Obayashi K, Tsumara Y, Ihara-Unino T, Niyama K, et al. (2002). Genetic diversity and outcrossing rate between undisturbed and selective logged forest of *Shorea curtisii* using microsatellite DNA analysis. *Int. J. Plant. Sci.* 163: 151-158.
- Oliveira AF, Carvalho DES and Rosado CS (2002). Outcrossing rate and mating system of a natural population of *Copaifera langsdorffii* Desf. in the region of Lavras, MG, starting from isozyme. *Rev. Bras. Bot.* 25: 331-338.
- Pedroni F, Sanchez M and Santos FAM (2002). Phenology of copaíba (*Copaifera langsdorffii* Desf. Leguminosae, Caesalpinioideae - in a semideciduous forest, southeastern Brazil. *Rev. Bras. Bot.* 25: 183-194.
- Pinto SIC, Souza AM and Carvalho D (2004). Variabilidade genética por isoenzimas em populações de *Copaifera langsdorffii* Desf. em dois fragmentos de mata ciliar. *Sci. Forest.* 65: 40-48.
- Reis AMM and Grattapaglia D (2004). RAPD variation in a germoplasm collection of *Myracrodruon urundeuva* (Anacardiaceae), an endangered tropical tree: Recommendations for conservation. *Genet. Resour. Crop Evol.* 51:

- 529-538.
- Ribeiro RA and Lovato MB (2004). Mating system in a neotropical tree species, *Senna multijuga* (Fabaceae). *Genet. Mol. Biol.* 27: 418-424.
- Ritland K (1989). Correlated matings in the partial selfer *Mimulus guttatus*. *Evolution* 43: 848-859.
- Ritland K (2002). Estimation of gene frequency and heterozygosity from pooled samples. *Mol. Ecol. Notes* 2: 370-372.
- Ritland K and Jain S (1981). A model for the estimation of outcrossing rate and gene frequencies using n independent loci. *Heredity* 47: 35-52.
- Rocha OJ and Aguilar G (2001). Variation in the breeding behavior of the dry forest tree *Enterolobium cyclocarpum* (guanacaste) in Costa Rica. *Am. J. Bot.* 88: 1600-1606.
- Sampson JF (1998). Multiple paternity in *Eucalyptus rameliana* (Myrtaceae). *Heredity* 81: 349-355.
- Sebbenn AM (2006). Sistema de Reprodução em Espécies Arbóreas Tropicais e suas Implicações para a Seleção de Árvores Matrizes para Reflorestamentos Ambientais (Higa AH and Silva L, eds.). Pomar de Sementes de Espécies Florestais Nativas. FUPPEF, Curitiba.
- Sebbenn AM, Carvalho AC, Freitas ML, Moraes SM, et al. (2011). Low levels of realized seed and pollen gene flow and strong spatial genetic structure in a small, isolated and fragmented population of the tropical tree *Copaifera langsdorffii* Desf. *Heredity* 106: 134-145.
- Tilman D, May RM, Lehman CL and Nowak MA (1994). Habitat destruction and the extinction debt. *Nature* 371: 65-66.
- Weir BS (1996). Genetic Data Analysis II: Methods for Discrete Population Genetic Data. Sinauer Associates, Sunderland.
- Weir BS and Cockerham CC (1984). Estimating F-statistics for the analysis of population-structure. *Evolution* 38: 1358-1370.
- Zorzetto R, Fioravanti C and Ferroni MA (2003). The forest is reborn. *Pesq. FAPESP* 91: 48-52.