Genet Resour Crop Evol (2009) 56:797–807
DOI 10.1007/s10722-008-9402-2

RESEARCH ARTICLE

Microsatellite diversity and effective population size in a germplasm bank of *Hymenaea courbaril* var. *stilbocarpa* (Leguminosae), an endangered tropical tree: recommendations for conservation

Juliana Massimino Feres · Marcela Corbo Guidugli · Moacyr Antonio Mestriner · Alexandre Magno Sebbenn · Ana Yamaguishi Ciampi · Ana Lilia Alzate-Marin

Received: 20 August 2008 / Accepted: 15 December 2008 / Published online: 3 February 2009

Abstract Deforestation in southeast Brazil has led to the extinction of *Hymenaea courbaril* var. *stilbocarpa* and *ex situ* conservation has been established. In this study, the levels of genetic diversity and the effective population size of *H. courbaril* in a germplasm bank were investigated using six nuclear microsatellite loci. A total of 79 and 91 alleles were found in 65 seed-trees and their 176 offspring, respectively. Offspring have a higher average number of alleles per locus (A = 15.2) than seed-trees (A = 13.2), but lower observed heterozygosity (offspring: $H_o = 0.566$; seed-trees: $H_o = 0.607$). The estimate of outcrossing rate shows that the study population is perfectly outcrossed ($t_m = 0.978$, $P > 0.05$). Significant deviations from random mating were detected through mating among relatives and correlated matings. The average variance in effective population size for each family was 2.63, with a total effective population size retained in the bank of 170.1. These results confirm that the preserved population of *H. courbaril* retains substantial genetic variability.

Keywords BG/USP/RP · *Ex situ* conservation · Germplasm bank · *Hymenaea courbaril* · Jatobá · Molecular marker · Tropical tree

Moacyr Antonio Mestriner and Ana Lilia Alzate-Marin—CEEFLORUSP Members.

J. M. Feres · M. C. Guidugli · M. A. Mestriner · A. L. Alzate-Marin (✉)
Plant Genetic Laboratory, Department of Genetics, FMRP, University of São Paulo (USP), Av. Bandeirantes, 3900, 14049-900 Ribeirão Preto, SP, Brazil
e-mail: anaalzate@rge.fmrp.usp.br

J. M. Feres
Graduate Program in Genetics, Department of Genetics, FMRP, University of São Paulo (USP), Ribeirão Preto, SP, Brazil

M. C. Guidugli
Graduate Program in Comparative Biology, Department of Biology, FFCLRP, University of São Paulo (USP), Av. Bandeirantes, 3900, 14040-900 Ribeirão Preto, SP, Brazil

A. M. Sebbenn
Experimental Station Tupi, Forest Institute of São Paulo, CP 339, CEP 13400-970 Piracicaba, SP, Brazil

A. Y. Ciampi
Embrapa Cenargen, PqEB, Final Av. W5 Norte, CP 2372, CEP 70770-900 Brasília, DF, Brazil

Introduction

The Atlantic forest of Brazil is one of the world's most important biodiversity hotspots for conservation priority (Myers et al. 2000). In Brazil, Atlantic forest formations have been almost completely destroyed and less than 5% of the original vegetation remains, dispersed among several thousand fragments (Zorzetto et al. 2003). Landscapes have become increasingly fragmented; in the case of semideciduous

tropical forest, populations of forest species are being reduced and ecosystems are being 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, in Northeastern São Paulo State (southeast Brazil), had remained essentially intact. However, following the rapid expansion of cultivation areas, particularly coffee and sugar-cane monocultures, the original vegetation was destroyed and fragmented causing a drastic loss of genetic forest patrimony. In 1997, the University of São Paulo's Forest Project initiated plantation sites at the Ribeirão Preto Campus, with the goal of rescuing and conserving regional vegetation cover. One of the objectives of this project was to establish a germplasm bank (BG/USP/RP) to rescue native tree species in the Ribeirão Preto region. Among approximately 75 ha available for reforestation, 45 have been reserved for a genetic bank containing the offspring of 44 endangered native forest tree species, including *Hymenaea courbaril* var. *stilbocarpa* (Hayne) Lee et Lang (Leguminosae). This species has also been recognized as a priority species for the conservation program of forest tree genetic resources. *H. courbaril* is an economically and ecologically important tropical tree species of the Brazilian Atlantic rainforest.

Conservation and utilization strategies require fundamental knowledge about levels of genetic diversity, population genetic structure and effective population size, because these are the key elements that determine a species' ability to respond to selection, either natural or artificial procedures (Reis and Grattapaglia 2004). *Ex situ* conservation procedures which prioritize the preservation of germplasm, making it available for research and the recovery of degraded areas are rare (FAO 1984). The present studies are, therefore, not only important to describe the existing level and distribution of genetic variation for conservation, but are also essential to optimize the allocation of resources for germplasm conservation and enrichment projects (Reis and Grattapaglia 2004).

At present, a range of powerful molecular genetic techniques, based on the polymerase chain reaction (PCR), are available for studying genetic diversity and structure, mating systems and gene flow (White and Powell 1997). Molecular genetic techniques, primarily using genetic markers, can also aid in the management of *ex situ* populations, by confirming the identity of accessions and monitoring genetic changes in collections. In particular, microsatellite markers (SSRs–Simple Sequence Repeats) are among the best molecular markers for the study of a population's genetic diversity, due to the higher level of information they produce (Chase et al. 1996). SSRs are highly polymorphic, inherited in a co-dominant Mendelian manner, neutral and multiallelic.

In the present study, the level of genetic diversity and the effective population size of a germplasm bank of *Hymenaea courbaril* are investigated and quantified using six nuclear microsatellite loci.

Material and methods

Species studied

Hymenaea courbaril var. *stilbocarpa* (jatobá or guarnipol) is a wide-ranging forest species occurring only in Brazil. It is mostly found in gallery forests in the Cerrado (Savannah) region, but radiates into the northeastern Caatinga and southern rainforests (Leite 2007). Although the species has such a wide distribution, it generally occurs in a low-density population (<1 tree/ha). Individual size varies markedly and is directly related to the habitat occupied: populations that grow in tropical forests can reach heights of 65 m, while those in the Cerrado and semi-arid forests reach a height of just a few meters (Lacerda et al. 2008a). *H. courbaril* is a diploid ($n = 12$) species with alternated leaves, bifoliolate and shortly petiolate; it is hermaphroditic with inflorescences composed of a few racemes; the fruits are indehiscent pods 5–12 cm long and 3–5 cm wide with a thick, floury and sweet endocarp (Lee and Langenheim 1974, 1975; Cowan and Polhill 1981; Leite 2007). Pollination is performed by bats, bees, hummingbirds and flies, with bat pollination occurring at night in the sparsely distributed trees, and other pollinators visiting diurnally (Crestana and Mariano 1985). Bawa (1974) and Jansen (1983) report this species to be self-incompatible. Studies by Dunphy et al. (2004) and Lacerda et al. (2008a) use genetic markers to demonstrate that *H. courbaril* is perfectly outcrossed, confirming the results of Bawa (1974) and Jansen (1983). According to Gibbs et al.

(1999), this apparent incompatibility may result from post-pollination events. Hydrochory is the main mechanism for seed dispersal of the buoyant fruits, but zoochory is also involved, with rodents and other mammals aiding dispersing by transporting fruits or eating seeds (Lee and Langenheim 1975). In addition, the trees produce a high density, extremely decay resistant wood, intensively used and valued in carpentry and construction ranging from civil projects to naval building and furniture (Lee and Langenheim 1975). The fruits and resin are also commercially important products, with the latter exploited in Brazil for use in varnish, incense and folk medicine (Lee and Langenheim 1975). The dimensions reached by the typical forest species (height and diameter), combined with the good physical features of the timber, make this species one of the most valuable and most intensively exploited wood species in the Brazilian Amazon (Lacerda et al. 2008a).

Sampling sizes

The study was conducted in an *ex situ* germplasm bank established by the University of São Paulo, at the Campus of Ribeirão Preto (northeastern São Paulo State, Brazil 21°10′S 47°48′W) (BG/USP/RP), and on its respectively seed-trees. Seeds used in the germplasm bank were collected for three years (1999–2001) in different rainforest remnants across the Ribeirão Preto region, between Pardo and Mogi-Guaçu basins (Fig. 1), mostly in farms. Aiming to minimize genetically related samples, open-pollinated seeds were collected from 75 seed-trees located at least 500 m apart each other. On average, ten offspring were generated from each seed-tree used to establish the germplasm bank (750 offspring). However, many offspring died and ten families were completely lost, leaving the remaining 65 families and about 360 offspring to form the bank. Aiming to quantify the genetic diversity in seed-trees and the germplasm bank, leaves were collected from 65 of the 75 seed-trees and from 176 offspring, with 1–8 individuals sampled from each family (an average of three offspring per family). Approximately five leaves were collected from each tree and placed into plastic sealable bags with a small amount of silica gel to assist in rapid drying and preservation of DNA.

DNA extraction

Genomic DNA was extracted from approximately 150 mg of frozen leaves, following by the standard CTAB procedure (Doyle and Doyle 1990). Purity and concentration of genomic DNA were determined using a spectrophotometer (Spectronic Genesys, model 5).

SSR amplification

The SSR markers Hc12, Hc14, Hc17, Hc33, Hc40 and Hc42 (Ciampi et al. 2008) were used for this study. Microsatellite loci were amplified via PCR in a final volume of 10 µl containing 0.3 µM of each

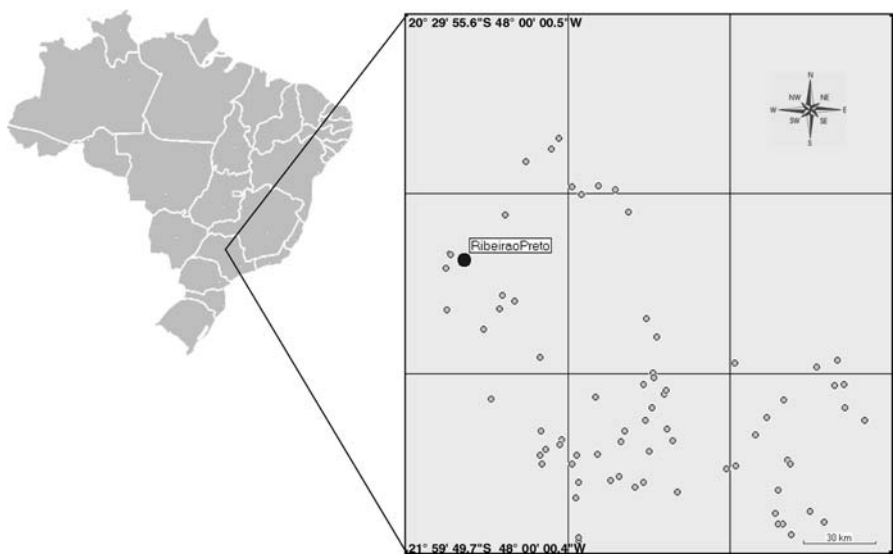

Fig. 1 Geographic localization of the *H. courbaril* seed-trees

primer, 1 U *Taq* DNA polymerase, 0.25 mM of each dNTP, 1× $MgCl_2$-free reaction buffer [75 mM Tris–HCl pH 9.0, 50 mM KCl and 20 mM $(NH_4)_2SO_4$], 1.5 mM $MgCl_2$ and 2.5 ng of template DNA. The PCR profile used to amplify the microsatellites was 96°C for 5 min; 30 cycles of denaturation at 94°C for 30 s, annealing at 58°C for 1 min, 72°C for 1 min, and a final elongation step at 72°C for 7 min. Amplifications were performed in a MasterCycler Eppendorf. PCR products were denatured and separated on 10% denaturing polyacrylamide gels stained with silver nitrate (Sanguinetti et al. 1994). Allele sizes were estimated by comparison to a 10 bp DNA ladder standard (Invitrogen).

Genetic diversity and fixation index analysis

Genetic diversity for each locus and averaged loci were calculated for sampled seed-trees and offspring. The indices examined were the total number of alleles over loci (k), observed heterozygosity (H_o) and expected heterozygosity under the Hardy-Weinberg equilibrium (H_e). All cited indices and paternity exclusion of the first [Pr(Ex_1)] and second parent [Pr(Ex_2)] were estimated using the CERVUS 3.0 program (Kalinowski et al. 2007). The effective number of alleles per locus was estimated by the expression $\hat{A}_e = 1/(1 - \hat{H}_e)$. The fixation index ($F$) was calculated for offspring and reproductive adults as $\hat{F} = 1 - (\hat{H}_o/\hat{H}_e)$. The statistical significance of F values was assessed from 10,000 bootstrap replicates across loci, using GDA (Lewis and Zaykin 2000). Genetic differentiation between seed-trees and offspring was estimated using the F_{st} statistic (method of Weir and Cockerham 1984) and the program GDA.

Mating system analysis

Mating system analyses were conduced in two steps: (1) all families (65) were used to estimate outcrossing rates, even those families that had only one offspring; and (2) families with three or more offspring (17) were used to estimate paternity correlations, or the proportion of full-sibs within families.

The mixed-mating and correlated mating models were used to characterize the mating system, using the software Multilocus MLTR version 3.1 (Ritland 2002). The calculated parameters were: multilocus outcrossing rate (t_m); single-locus outcrossing rate (t_s); mating among relatives rate ($t_m - t_s$); and multilocus paternity correlation ($r_{p(m)}$). The number of effective pollen donors was calculated from paternity correlations by the expression: $\hat{N}_{ep} = 1/\hat{r}_p$.

This analysis was performed at the population level, using the Newton-Raphson numerical method. The 95% confidence interval of the parameters was calculated from 1,000 bootstraps. For the population analysis, the standard deviation was calculated by resampling individuals within families. The average proportions of self-sibs (P_{ss}), half-sibs (P_{hs}) and full-sibs (P_{fs}) within families were estimated as $\hat{P}_{ss} = s$, $\hat{P}_{hs} = \hat{t}_m(1 - \hat{r}_{p(m)})$ and $\hat{P}_{fs} = \hat{t}_m\hat{r}_{p(m)}$, respectively.

Estimation of effective population size from mating system parameters

The effective population size was calculated using the variance effective population size. For mating system analysis, the average variance effective size ($N_{e(v)}$) was estimated using the Cockerham (1969) estimator, in which an idealized population was taken as reference (infinite size; random matings; no selection, mutation and migration):

$$\hat{N}_{e(v)} = \frac{0.5}{\hat{\Theta}_{xy}},$$

where Θ_{xy} is the coancestry coefficient among plants within families, estimated from mating system parameters and from the averaged relatedness coefficient within families (r_{xy}):

$$\hat{r}_{xy} = 0.25(1 + \hat{F}_p)[4\hat{s} + (\hat{t}_m^2 + \hat{r}_s\hat{t}_m\hat{s})(1 + \hat{r}_p)]$$

(Ritland 1989), where F_p is the inbreeding coefficient in the parental generation, s is the self-fertilization rate ($1 - \hat{t}_m$), and r_s is the selfing correlation, assumed to be zero ($r_s = 0$) as the study species is reportedly self-incompatible (see Results and Discussion); other parameters follow the definitions presented above. Thus, assuming that $\hat{\Theta}_{xy} = \hat{r}_{xy}/2$, the average coancestry coefficient within families was calculated as:

$$\hat{\Theta}_{xy} = 0.125(1 + \hat{F}_p)[4\hat{s} + \hat{t}_m^2(1 + \hat{r}_{p(m)})].$$

The total effective population size in the germplasm bank was calculated multiplying the number of sampled families (m) by the averaged variance effective size:

$$\hat{N}_{e(v)(total)} = m\hat{N}_{e(v)}.$$

Results

Genetic diversity in seed-trees

All six microsatellite loci analyzed in seed-trees were polymorphic (Table 1). A total of 79 alleles were found in 65 seed-trees, where three low frequency alleles were privates (not found in any offspring). The number of alleles per locus ranged from six (Hc33) to 22 (Hc42) with a mean of 13.2. The effective number of alleles per locus ranged from 3.60 to 7.87, averaging 5.82, which suggests that although many alleles were found at the loci, a large proportion were of low frequency. The observed heterozygosity ranged from 0.292 to 0.828, with a mean of 0.607; this is lower than the expected heterozygosity, which ranged from 0.722 to 0.873, averaging 0.813. A significant deficiency in heterozygosity was observed in three of six loci. The other loci fitted to Hardy-Weinberg proportions. The average fixation index was significantly different from zero ($\hat{F} = 0.259$, $P < 0.05$). The probability exclusion of the first and second parent was high, 0.9799 and 0.9982, respectively, indicating that the number of loci used here is sufficient for parentage studies.

Genetic diversity in offspring

A total of 91 alleles were found in 176 offspring (Table 1). Fifteen low frequency alleles were privates and were not found in the seed-trees gene pool. The number of alleles per locus ranged from seven (Hc33) to 27 (Hc42); average of 15.2. The effective number of alleles per locus ranged from 3.82 to 7.69, averaging 5.64. Observed heterozygosity ranged from 0.291 to 0.824, averaging 0.566, and expected heterozygosity ranged from 0.738 to 0.870, averaging 0.813. The fixation index was significantly different from zero ($P < 0.05$) and positive in four of the six analyzed loci. The average fixation index over all loci was 0.310, suggesting inbreeding.

Table 1 Characteristics of six microsatellite loci from *H. courbaril* in adult and offspring

Generation/Locus	n	k	$\hat{A}_e$	$\hat{H}_o$	$\hat{H}_e$	$\hat{F}$	Pr(Ex_1)	Pr(Ex_2)
Seed-trees								
Hc12	63	17	7.87	0.365	0.873	0.581**	0.420	0.265
Hc14	65	11	6.41	0.815	0.844	0.034	0.487	0.319
Hc17	65	8	4.76	0.831	0.790	−0.052	0.589	0.408
Hc33	65	6	3.60	0.292	0.722	0.596**	0.690	0.511
Hc40	64	15	7.81	0.828	0.872	0.050	0.424	0.267
Hc42	63	22	4.46	0.508	0.776	0.345**	0.571	0.387
Means		13.2	5.82	0.607	0.813	0.259**	0.9799	0.9982
95% standard error[a]		2.34	0.71	0.098	0.024	0.113		
Offspring								
Hc12	146	18	6.80	0.390	0.853	0.543**	–	–
Hc14	176	12	5.75	0.824	0.826	0.002	–	–
Hc17	175	9	4.78	0.629	0.791	0.205	–	–
Hc33	172	7	3.82	0.291	0.738	0.606**	–	–
Hc40	174	18	7.69	0.764	0.870	0.122	–	–
Hc42	145	27	5.03	0.497	0.801	0.380**	–	–
Means		15.2	5.64	0.566	0.813	0.310**	–	–
95% standard error[a]		2.89	0.55	0.082	0.019	0.094		

Sample size (n); number of detected alleles per locus (k); effective number of alleles per locus (A_e); observed heterozygosity (H_o); expected heterozygosity (H_e); exclusion probability of the first [Pr(Ex_1)], and second parent [Pr(Ex_2)]

** $P < 0.01$

[a] 95% standard error was calculated by jackknife method

Mating system

Because the employed sample size was small in terms of individuals sampled within families (1–8 individuals per family) and this large variation could affect the estimates (especially the paternity correlation), we carried out a mating system analysis for three different scenarios, namely: (a) total sample with 65 families and one to eight individuals per family; (b) 47 families and two to eight individuals per family, and (c) 28 families and three to eight individual per family (Fig. 2). The estimates of multilocus (t_m) and single-locus outcrossing rate (t_s), mating among relatives ($t_m - t_s$) and paternity correlation (r_p) showed very small differences between the scenarios. Estimates of multilocus outcrossing rates were high and did not significantly differ ($P > 0.05$) among scenarios, ranging from 0.975 (for 28 families) to 0.981 (for 47 families), indicating that *H. courbaril* is a perfect outcrossing species. The estimative of single-locus outcrossing rate was lower than and significantly different from the multilocus outcrossing rate ($P < 0.05$), which ranged from 0.690 (for 65 families) to 0.710 (for 28 families). Differences between multilocus and single-locus outcrossing rates were high and significantly different from zero ($P < 0.05$), indicating mating among relatives (ranging between 26.5 and 28.8%). Paternity correlation was also significantly different from zero ($P < 0.05$), ranging from 0.175 to 0.177 between scenarios. The standard deviation of the estimates of multilocus, single-locus and differences between multilocus and single-locus outcrossing rates (Fig. 3) increased as the sample size moved from 65 to 28 families, except for a substantial reduction in multilocus and single-locus estimates between sampled sizes of 47 to 28 families. This suggests a reduction in the accuracy of estimates. However, for estimates of paternity correlation, a reduction in standard deviation between sample sizes of 65 to 47 and 28 families, suggests, as expected, increasing estimate precision with an increased number of offspring sampled within families. These results indicate that the best strategy would be to estimate outcrossing rates using 65 families and paternity correlation using 28 families. Thus, differences between the scenarios were very small. Therefore, scenario involving 65 families (1–8 individuals per family) was used to estimate the outcrossing rate and the scenario involving 28 families (4–8 individuals per family) was used to estimate the paternity correlation. The combination of these two estimated parameters was used to calculate the number of pollen donors, the coancestry coefficient and effective population size (Table 2).

Coancestry coefficient and effective variance size

Combining the estimate of the paternity correlation ($r_{p(m)}$) with the estimated multilocus outcrossing rate (t_m) it is possible to determine the proportion of different kinship types within families (Table 2). The families were composed predominantly of half-sibs (80.6%), following by full-sibs (17.2%) and some self-sibs (2.2%), although the outcrossing rate was

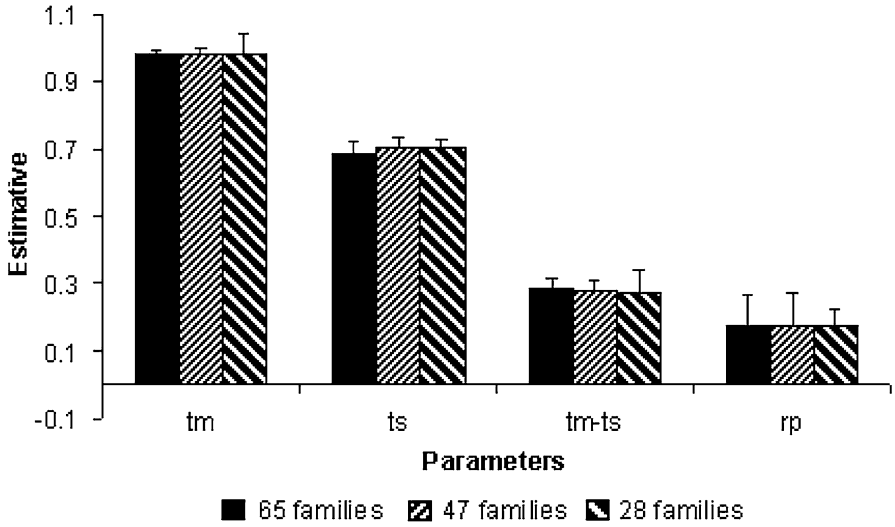

Fig. 2 Estimates of multilocus outcrossing rate (t_m), single-locus outcrossing rate (t_s), mating among relatives ($t_m - t_s$), and paternity correlation (r_p) for three different sample size. *Vertical lines* are maximum the 95% standard error

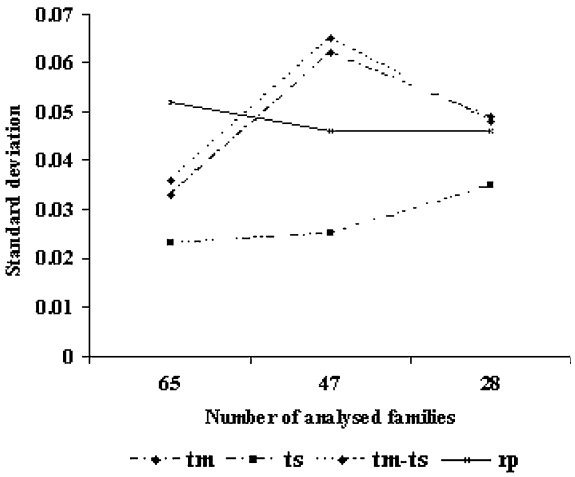

Fig. 3 Estimates of standard deviation of the multilocus outcrossing rate (t_m), single-locus outcrossing rate (t_s), mating among relatives ($t_m - t_s$), and paternity correlation (r_p) for three different number of sampled families

not significantly different from unity (1.0). The coancestry coefficient ($\hat{\Theta}_{xy} = 0.190$) was higher than expected for half-sib families ($\Theta_{xy} = 0.125$). The variance effective size was also low ($\hat{N}_{e(v)} = 2.63$) compared to the expected value for half-sib families ($N_{e(v)} = 4$). The total effective variance size of the sample was estimated to be 170.7.

Discussion

Sample size, mating system estimates and number of SSR loci

Correct estimates of the genetic diversity index of populations require at least 60 sampled individuals, unless the population itself is smaller than that value (Kirst et al. 2005). As this study uses at least 65 individuals sampled from the populations; thus, we believe our sample to be sufficient. For a good estimate of outcrossing rate (standard error of 0.03–0.05), at least 200 offspring must be assayed for several co-dominant loci (Ritland 2004). Though our sample (176 offspring) is smaller, for the hyper-variable microsatellite loci used here, the standard error of the estimates will be close to $\sqrt{t(1-t)}/n$, and sample sizes over 50 individuals should be sufficient (Ritland 2004). For the sample used here, the expected standard error is 0.0008, indicating that our data is sufficient to estimate the outcrossing rate. The number of individuals sampled per family is also important for good estimates of paternity correlation to be obtained. A low number of individuals sampled per family, or a large variation in that number can affect estimates of paternity correlation. Although no great variation was observed in the average paternity correlation estimated from the different scenarios (65 families with 1–8 individuals per family; 48 families with 2–8 individuals per family; 28 families with 3–8 individuals per family), the standard deviation of the estimates was reduced in scenarios with less variation in the number of individuals sampled per family (3–8 vs. 1–8) (Fig. 2). Thus, to obtain a reasonable estimate of paternity correlation we use the 28 family scenario. However, it is important to note that good estimates of paternity correlation require a relatively large number of individuals to be sampled from each family (e.g. 20 individuals). Sampling fewer than 10 individuals per family can underestimate the paternity correlation via sampling error. Future studies

Table 2 Parameters of mating system for 47 families of *H. courbaril*

Parameter	Estimation
Multilocus outcrossing rate: t_m	0.978 (0.959–1.200)
Single-locus outcrossing rate: t_s	0.690 (0.664–0.760)
Mating among relatives rate: $\hat{t}_m - \hat{t}_s$	0.288 (0.086–0.258)
Multilocus paternity correlation: $r_{p(m)}$	0.176 (0.028–0.211)
Average number of pollen donors: N_{ep}	5.68 (4.74–35.71)
Proportion of self-sibs: P_{ss}	0.022 (0.0–0.041)
Proportion of half-sibs: P_{hs}	0.806 (0.757–0.972)
Proportion of full-sibs: P_{fs}	0.172 (0.027–0.211)
Average coancestry coefficient within families: Θ_{xy}	0.190 (0.148–0.300)
Variance effective size: $N_{e(v)}$	2.63 (1.67–3.37)
Total variance effective size: $\hat{N}_{e(v)t} = m\hat{N}_{e(v)}$	170.7 (108.5–219.0)

95% CI Confidence interval for 95% error probability is shown in brackets

will aim to sample more individuals per family, ideally more than 20 individuals per family.

Although the number of loci used in this study is small, high levels of polymorphism make SSR markers useful for population genetic studies. In addition, the combined probabilities of paternity exclusion were very high (Table 1), demonstrating that these markers have enough resolution to be used for mating system analysis in *H. courbaril*.

Genetic diversity

The *H. courbaril* germplasm bank was established with a large number of families (75), ensuring that the *ex situ* procedure would preserve the population's genetic diversity. Our results show small differences in gene frequencies among seed-trees and offspring. The effective number of alleles and expected heterozygosity were similar in offspring and seed-trees. The F_{st} statistic was null ($\hat{\theta}_p = -0.0002$, following Weir and Cockerham 1984) between seed-trees and offspring, indicating that the genetic diversity of the original population was retained in the germplasm bank. The largest genetic changes were observed in genotypic frequencies: between adult and offspring generations, observed heterozygosity was reduced and fixation index increased (Table 2), indicating inbreeding. Mating system analyses did not detect self-fertilization, but found a high rate of mating among relatives ($t_m - t_s = 0.288$). Thus, the inbreeding in offspring can be explained by biparental inbreeding. However, the high and significant fixation index observed in adult trees ($\hat{F} = 0.259$), cannot be due to inbreeding, but rather, it is an artifact of the sampling strategy, which produces a Wahlund effect due to the wide distance between sampled seed-trees (Fig. 1). The sampled seed-trees probably occur in different subpopulations, producing an artificial increase in the fixation index determined here.

Low frequency private alleles were observed in both generations. Among the adult trees, three private alleles were found. However, as only 1–3 offspring were sampled for some families, and all families have unsampled individuals which remain in the bank, it is possible that these alleles have been retained. In a diploid genotype, there is a 50% chance of transmitting a given allele at a locus to the next generation, showing a perfect 1:1 Mendelian segregation ratio. Fifteen low frequency, private alleles are found in the offspring, and probably come from other, unsampled reproductive individuals in the population, further indicating that the bank retains allelic diversity.

Mating system

Our results for outcrossing rate ($\hat{t}_m = 0.978$, $P > 0.05$, Table 2) agree with previous mating system studies of *H. courbaril* and indicate an absence of self-fertilization. Dunphy et al. (2004), studying a *H. courbaril* population in Costa Rica, and Lacerda et al. (2008a) studying a population in Brazil's Amazon region, described this species as perfectly outcrossed ($\hat{t}_m = 1$) and probably self-incompatible. Self-incompatible systems have been described in many tropical tree species (Cascante et al. 2002; Fuchs et al. 2003; Lobo et al. 2005); this system of mating prevents inbreeding due self-fertilization and, consequently, inbreeding depression. Self-fertilization is the strongest mechanism producing inbreeding in plants, responsible for at least 50% of inbreeding (homozygosis for identical per descending alleles) in the offspring of each generation. Another mechanism producing inbreeding is mating among relatives (biparental inbreeding), but the degree of inbreeding created through this mechanism depends on the coancestry coefficient between the outcrossed individuals. For example, the mating between two half-sibs ($\theta_{xy} = 0.125$) will produce 12.5% biparental inbreeding, while mating between two full-sibs ($\theta_{xy} = 0.25$) will produce 25% inbreeding. In summary, *H. courbaril* seems to be perfect outcrossed and self-incompatible; inbreeding probably occurs only by mating among relatives.

It was observed that part of the mating in the population occurred between related trees ($\hat{t}_m - \hat{t}_s = 0.288$, $P < 0.05$), indicating biparental inbreeding. As previously mentioned, mating among relatives can explain the high and significant fixation index observed in offspring (Table 1). This result also suggests that the parental population, in which seed were collected, presents a spatial genetic structure, possibly due to near-neighbor relative trees originating from seed dispersal close to the parent seed-tree. Lacerda et al. (2008b) studied this aspect of a *H. courbaril* population in Brazil's Amazon and found spatial genetic structures in different ontogenic stages of the population (seedlings, juveniles, adults). Biparental inbreeding is also an indicative of a deviation from random mating.

Deviation of random mating was also indicated by the levels of correlated mating ($\hat{r}_{p(m)} = 0.176$, $P < 0.05$), suggesting that part of the open-pollinated offspring are full-sibs. Correlated mating may result from the fragmentation of Atlantic forest which reduced both the number of populations of the species and the density of reproductive trees, limiting gene flow between populations and individuals via isolation. Alternatively, changes in the behavior of the trees' pollinating species, i.e., bats, could be responsible. Bats tend to forage in the same area (Handley and Morrison 1991), sometimes visiting just a few trees. Thus, the low-population density and pollinator behavior are likely to be the cause of correlated matings.

From paternity correlations it was estimated that about 5.68 (N_{ep}) pollen donors visit each sampled seed-tree. This value is comparable to the patterns observed for other tropical tree species, with values from 2 to 7 (James et al. 1998; Rocha and Aguilar 2001; Dunphy et al. 2004; Moraes et al. 2004; Azevedo et al. 2007; Carneiro et al. 2007), and is similar to those detected for other populations of *H. courbaril,* where the number of pollen donors, estimated by paternity analysis, was determined to be 6.9 in Costa Rica (Dunphy et al. 2004) and 4 in the Brazilian Amazon (Lacerda et al. 2008a).

Effective population size and conservation genetics

We estimated the variance effective size of *H. courbaril* in the germplasm bank (BG/USP/RP). This value (170.1) was lower than the census number of individuals retained in the bank ($N = 464$); this is to be expected, considering that the preserved population is structured in progeny arrays. As progeny arrays are compounded by relatives, there are many genes identical per descend. This causes allelic redundancy, reducing the effective population size as compared to census numbers. The average coancestry coefficient within families (Θ_{xy}) was estimated at 0.190, 34% higher than expected for seeds collected from panmictic populations (0.125: half-sibs). Our analysis of mating systems suggests that about 81% of offspring are half-sibs and 18% are full-sibs (Table 2). Thus, the calculated average variance effective size within families ($N_{e(v)}$) is only 2.63. This value, equal to the coancestry coefficient, was about 34% lower than expected in half-sibs families ($N_{e(v)} = 4$).

For loci with two alleles in outcrossing species with discrete generation, it has been determined that an effective population size of 50 is sufficient to retain 95% of the actual level of genetic variation (heterozygosity) over approximately ten generations (Frankel and Soulé 1981). However, in species with overlapping generations, a larger population is necessary (Nunney and Campbell 1993). For these species, Nunney and Campbell (1993) suggest increasing the values of 50 two or three fold (100 or 150) to compensate. Therefore, an effective population size on the order of 170 can retain a substantial part of the present genetic diversity of *H. courbaril,* at least over short terms (<10 generations). This may represent, in this long-lived tree species, at least 500 years. In agreement with this estimate, Lacerda et al. (2008a) show that to retain an effective population size of 150 in *H. courbaril* seed pool samples, it is necessary to collect seeds from at least 56 trees. Thus, the present results confirm the hypothesis that the GB/USP/RP germplasm bank retains substantial genetic variability of *H. courbaril* populations. This study validates that the existing *H. courbaril* germplasm collection is representative, and helps delineate future strategies for the conservation of this species.

Acknowledgements This study was supported by FAPESP (A.L.A.M. #03/04199-4), Provost for Research of São Paulo University (A.L.A.M.) and FAEPA (M.A.M.) grants. A.L.A.M. was supported by Researcher assistantship from FAPESP. J.M.F. was supported by FAPESP and M.C.G. by a CAPES fellowships. The authors thank Maria C. T. Canas and Ana L. Pimentel for technical support during microsatellite analyses, as well as two anonymous reviewers who provided important suggestions and constructive criticism for revising the early version of the manuscript.

References

Azevedo VCR, Kanashiro M, Ciampi AY, Grattapaglia D (2007) Genetic structure and mating system of *Manilkara huberi* (Ducke) A.Chev., a heavily logged Amazonian timber species. J Hered 98(7):646–654. doi:10.1093/jhered/esm074

Bawa KS (1974) Breeding systems of tree species of a lowland tropical community. Evolution 28:85–92. doi:10.2307/2407241

Carneiro F, Sebbenn AM, Kanashiro M, Degen B (2007) Low inter-annual variation of mating system and gene flow of

Symphonia globulifera in the Brazilian Amazon. Biotropica 39:628–636. doi:10.1111/j.1744-7429.2007.00314.x

Cascante A, Quesada M, Lobo JJ, Fuchs EA (2002) Effects of dry tropical forest fragmentation on the reproductive success and genetic structure of the tree *Samanea saman*. Conserv Biol 16:137–147. doi:10.1046/j.1523-1739.2002.00317.x

Chase M, Kesseli R, Bawa K (1996) Microsatellite markers for population and conservation genetics of tropical trees. Am J Bot 83:51–57. doi:10.2307/2445953

Ciampi AY, Azevedo VCR, Gaiotto FA, Ramos ACS, Lovato MB (2008) Isolation and characterization of microsatellite loci for *Hymenaea courbaril* and transferability to *Hymenaea stigonocarpa*, two tropical timber species. Mol Ecol Res. doi:10.1111/j.1755-0998.2008.02159.x

Cockerham CC (1969) Variance of gene frequencies. Evolution 23:72–84. doi:10.2307/2406485

Cowan RS, Polhill RM (1981) Detarieae. In: Polhill RM, Raven PH (eds) Advances in legume systematics, 1st edn. Royal Botanic Gardens, Kew, UK, pp 117–134

Crestana CSM, Mariano IS (1985) Pollination ecology of *Hymenaea stilbocarpa* Hayne, the jatobá. Silvicultura 17:31–37

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15

Dunphy BK, Hamrick JL, Schwagerl J (2004) A comparison of direct and indirect measures of gene flow in the bat-pollinated tree *Hymenaea courbaril* in the dry forest life zone of southwestern Puerto Rico. Int J Plant Sci 165:427–436. doi:10.1086/382802

FAO (1984) A Guide to in situ conservation of genetic resources of tropical woody species. Based on the work of L. Roche Durojeanni. Rome (FORGEN/MISC/84/2)

Frankel OH, Soulé MS (1981) Conservation and evolution. Cambridge University Press, Cambridge

Fuchs EJ, Lobo JA, Quesada M (2003) Effects of forest fragmentation and flowering phenology on the reproductive success and mating patterns of the tropical dry forest tree *Pachira quinata*. Conserv Biol 17:149–157. doi:10.1046/j.1523-1739.2003.01140.x

Gibbs PE, Oliveira PE, Bianchi MB (1999) Postzygotic control of selfing in *Hymenaea stignocarpa* (Leguminosae-Caesalpinioideae) a bat pollinated tree of the Brazilian cerrados. Int J Plant Sci 160:72–78. doi:10.1086/314108

Handley CO Jr, Morrison WM (1991) Foraging behaviour. In: Handley CO Jr, Wilson DE, Gardner AL (eds) Demography and natural history of the common fruit bat, *Artebeus jamaicensis*, on Barro Colorado Island, Panama. Smithsonian Contributions to Zoology, Washington, pp 137–140

James T, Vege S, Aldrich P, Hamrick JL (1998) Mating systems of three tropical dry forest tree species. Biotropica 30:587–594. doi:10.1111/j.1744-7429.1998.tb00098.x

Jansen DH (1983) Costa rican natural history. University of Chicago Press, Chicago

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how computer program Cervus accommodates genotyping error increase success in paternity assignment. Mol Ecol 16:1099–1106. doi:10.1111/j.1365-294X.2007.03089.x

Kirst M, Cordeiro CM, Rezende GDSP, Grattapaglia D (2005) Power of microsatellite markers for fingerprinting and parentage analysis in *Eucalyptus grandis* breeding populations. J Hered 96:161–166. doi:10.1093/jhered/esi023

Lacerda AB, Sebbenn AM, Kanashiro M (2008a) Long-pollen movement and deviation of random mating in a low-density continuous population of *Hymenaea courbaril* in the Brazilian Amazon. Biotropica 40:462–470. doi:10.1111/j.1744-7429.2008.00402.x

Lacerda AB, Kanashiro M, Sebbenn AM (2008b) Effects of reduced impact logging on genetic diversity and spatial genetic structure of a *Hymenaea courbaril* population in the Brazilian Amazon Forest. For Ecol Manag 255:1034–1043. doi:10.1016/j.foreco.2007.10.009

Lee YT, Langenheim JH (1974) Additional new taxa and new combinations in *Hymenaea* (Leguminosae, Caesalpinioideae). J Arnold Arbor 55:441–452

Lee YT, Langenheim JH (1975) Sistematics of the genus *Hymenaea* L. (Leguminosae, Caesalpinioideae, Detarieae), vol 69. University of California Press, Berkeley, pp 1–109

Leite EJ (2007) State-of-knowledge on *Hymenaea courbaril* L. var. *stilbocarpa* (Hayne) Lee et Lang. (Leguminosae: Caesalpinoideae) for genetic conservation in Brazil. Res J For 1:1–26. doi:10.1080/17496530701237092

Lewis PO, Zaykin D (2000) GDA—Genetic Data Analysis: version 1.1 for Windows 95/NT. http://www.lewis.eeb.uconn.edu/lewishome/

Lobo JA, Quesada M, 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. doi:10.3732/ajb.92.2.370

Moraes MLT, Kageyama PY, Sebbenn AM (2004) Correlated mating in dioecious tropical tree species, *Myracrodruon urundeuva* Fr. All. For Genet 11:53–59

Myers N, Mittermeier RA, Mittermeier CG, Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. Nature 403:853–858. doi:10.1038/35002501

Nunney L, Campbell KA (1993) Assessing minimum viable population size: demography meets population genetics. Trends Ecol Evol 8:234–239. doi:10.1016/0169-5347(93)90197-W

Reis AMM, Grattapaglia D (2004) RAPD variation in a germplasm collection of *Myracrodruon urundeuva* (Anacardiaceae), an endangered tropical tree: recommendations for conservation. Genet Resour Crop Evol 51:529–538. doi:10.1023/B:GRES.0000024159.48644.e6

Ritland K (1989) Correlated matings in the partial selfer *Mimulus guttatus*. Evolution 43:848–859. doi:10.2307/2409312

Ritland K (2002) Extensions of models for the estimation of mating systems using n independent loci. Heredity 88:221–228. doi:10.1038/sj.hdy.6800029

Ritland K (2004) Multilocus mating system program MLTR. Version 3.1. University of British Columbia, Canada. Free program distributed by the authors from http://kritland@.interchange.ubc.ca. Accessed 20 Jan 2007

Rocha O Jr, Aguilar G (2001) Variation in the breeding behavior of the dry forest tree *Enterolobium cyclocarpum* (Guanacaste) in Costa Rica. Am J Bot 89:1600–1606. doi:10.2307/3558404

Sanguinetti CJ, Dias-Neto E, Simpson AJG (1994) RAPD silver staining and recovery of PCR products separated on polyacrylamide gels. Biotechniques 17:914–921

Tilman D, May RM, Lehman CL, Nowak MA (1994) Habitat destruction and the extinction debt. Nature 371:65–66. doi:10.1038/371065a0

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. Evolution 38:1358–1370. doi:10.2307/2408641

White GM, Powell W (1997) Cross-species amplification of SSR loci in the Meliaceae family. Mol Ecol 6:1195–1197. doi:10.1046/j.1365-294X.1997.00297.x

Zorzetto R, Fioravanti C, Ferroni MA (2003) Floresta renasce. Pesquisa Fapesp 91:48–52