Programa de Pós-Graduação em Agronomia, Universidade Estadual Paulista, Av. Brasil Centro 56, CP 31, Ilha Solteira, SP, 15385-000, Brazil.; Alessandro Alves-Pereira, Roland Vencovsky† - Escola Superior de Agricultura "Luiz de Queiroz", Av. Pádua Dias, 11, PO Box 9, 13418-090, Piracicaba, SP, Brazil; José Cambuim, Alexandre M. da Silva, Marcela A. de Moraes, Mário L.T. de Moraes - Faculdade de Engenharia de Ilha Solteira, Universidade Estadual Paulista, Av. Brasil Centro 56, CP 31, Ilha Solteira, SP, 15385-000, Brazil.

Manuscript received August 17, 2017; revised Spetember 13, 2017; accepted September 15, 2017; online first September 15, 2017.

Introduction

The Brazilian savanna is one of the world's biodiversity hotspots and among the top 25 priority areas for conservation (Simon et al. 2009). With more than 7,000 plant species, it is considered the most diverse tropical savanna in the world (Mendonça et al. 2008). Currently, this biome is experiencing extensive deforestation and its area has been drastically reduced and fragmented (Sano et al. 2007). The significant fragmentation of the savanna biome may decrease or even interrupt many reproductive and demographic processes of tree species, such as pollen and seed flow and colonization (Collevatti et al. 2013, Baldauf et al. 2014, Ibanes et al. 2016). Studies have shown that forest fragmentation can lead to decreased population size (genetic bottleneck effect), spatially isolated trees and populations, reductions in gene flow, genetic diversity, and effective population size, and increases in selfing rates, spatial genetic structure, and genetic divergence among populations of remnant tree species (Bittencourt & Sebbenn 2007, Melo & Franceschinelli 2016, Tambarussi et al. 2016). All of these processes increase the probability of population extinction. To mitigate the negative effects of reproductive isolation, reforestation strategies, such as the establishment of populations between remaining fragments or biodiversity corridors, are important to promote gene flow. Studies have also shown the importance of conserving spatially isolated trees in the landscape as they can act as stepping-stones, enabling gene flow between isolated populations in forest fragments (Bittencourt & Sebbenn 2007, Lander et al. 2010, Fuchs & Hamrick 2011, Sebbenn et al. 2011, Rymer et al. 2015). Thus, the *in situ* conservation of trees both in forest remnants and of spatially isolated trees in the landscape are extremely important, as well as seed collection strategies for *ex situ* conservation and environmental restoration.

To establish effective strategies to maintain gene flow among remaining fragmented populations for *in situ* conservation and to guide seed collection for *ex situ* conservation and environmental reforestation, it is essential to understand the pollen and seed dispersal distances and mating patterns of the species. Pollen and seed dispersal distances determine the maximum distance at which populations must be located to maintain gene flow and where populations must be established to ensure genetic connectivity. Mating patterns are especially important for seed collection in the context of *ex situ* conservation and environmental restoration (established populations) because they determine the levels of relatedness, inbreeding, genetic diversity, and effective size of sampled progeny arrays (Sebbenn 2006). In particular, sampling strategies for seed collection should be guided by the effective size of open-pollinated seedling families (Sebbenn 2006).

Assessing gene flow and mating patterns can effectively be carried out using gene markers, such as microsatellite loci, due to their high polymorphism and codominant inheritance (Burczyk et al. 2004, Ashley 2010, Ellstrand

2014). Based on genetic markers, gene dispersal distance and mating patterns can be estimated directly using parentage analyses (Burczyk et al. 2004, Ashley 2010, Leonarduzzi et al. 2012, Ellstrand 2014) and indirectly using models of gene dispersal distance, such as TWOGENER analysis (Austerlitz & Smouse 2001, Leonarduzzi et al. 2012, Robledo-Arnuncio et al. 2007). Parentage analysis requires extensive sampling of adult trees within a defined area and samples of descendant populations, which can be seeds or regenerants. TWOGENER analysis requires only samples of seed trees and their seeds. Both indirect and direct methods produce robust estimates of gene dispersal, but TWOGENER analysis only provides information about pollen dispersal distances (Austerlitz and Smouse 2001, Leonarduzzi et al. 2012, Robledo-Arnuncio et al. 2007).

In the present study, we investigate the potential to include isolated trees in agricultural landscapes in the Brazilian savanna in seed collection strategies. Among the vastly diverse tree species in the biome, we selected *Dipteryx alata* Vog. (Fabaceae) since it is classified as threatened in the endangered tree species list (IUCN 1998). Its biome has been strongly deforested and many populations have gone extinct. The species is economically important; oil is extracted from its seeds, which is consumed or used as raw material for pharmaceutical and oleochemical industries (Takemoto et al. 2001), and the wood is highly durable and widely used for civil and naval construction (Lorenzi 2002). The species is recommended for silvopastoral systems and pasture forestation serving as both food and shade for animals (Oliveira & Sigrist 2008). *Dipteryx alata* reaches 25 m in height and 70 cm in diameter at breast height (dbh) and can live up to 60 years (Lorenzi 2002). The species is hermaphroditic, presents a late acting self-incompatibility, and is pollinated by bees, mainly *Xylocopa suspecta*, *Pseudaugochlora graminea*, and *Apis mellifera* (Oliveira & Sigrist 2008). Fruiting starts when the tree reaches about six years of age and reproduction occurs mainly by outcrossing (Tarazi et al. 2010). Seeds are primarily dispersed by barochory and secondarily by mammals, such as monkeys, bats, and rodents (Sano et al. 2004), which may restrict seed dispersal distance, although seed dispersal can reach up to 8.2 km (Soares et al. 2015).

Molecular genetic studies have indicated that fragmented populations of *D. alata* are strongly genetically structured, presenting low levels of genetic diversity and some inbreeding (Tarazi et al. 2010, Collevatti et al. 2013, Soares et al. 2015). However, many remaining individuals of the species are spatially isolated in pastures and along roads, and there are no studies of the genetic diversity, inbreeding, and pollen dispersal distances for such populations. Due to the need to include this species in genetic conservation and environmental reforestation programs, isolated trees in the landscape may be an alternative for seed collection, especially considering the increased ease of collecting in open areas. Thus, in this study we use six microsatellite markers to describe the mating system, pollen dispersal distance, spatial genetic structure (SGS), genetic diversity, and inbreeding in isolated trees of two *D. alata* populations. The goal of the study is to determine an appropriate sample size for *ex situ* conservation, environmental reforestation, and tree breeding. We address four main questions: i) do the adult tree populations show SGS? ii) is genetic diversity higher and inbreeding lower in adults than in open-pollinated seeds, an indicator of inbreeding depression? iii) what are the rates of population and individual outcrossing, mating among relatives, and correlated mating? and iv) what is the mean pollen dispersal distance to estimate the number of seed trees and the distance between neighboring seed trees to be used for seed collection aiming at *ex situ* conservation and breeding programs.

Material and methods

Site and sampling

The study was carried out in two populations located about 560 km apart, with one in the municipality of Paulo de Farias (PF), São Paulo State, and one in Campo Grande (CG), Mato Grosso do Sul State, Brazil (Figure 1). PF is located on the left bank of the Rio Grande in a transitional zone between semi-deciduous and savanna forests in the northern region of São Paulo State; CG is located within savanna forest remnants surrounding the capital of Mato Grosso do Sul. The climate in both study areas is classified as tropical hot and humid (AW), characterized by a dry season (April to September) with average rainfall of 167 mm, and a rainy season (October to March) with average rainfall of 978 mm. Sampling occurred in 2014 in forest fragments, pastures, and along roads, in an area of 124,911 ha in PF (population density estimated from our data is 0.00055 trees/ha) and 45,365 ha in CG (0.00119 trees/ha). We sampled 69 trees in PF and 54 in CG. All trees were mapped (GPS Garmin Colorado 300, Olathe, KS, USA), sampled (leaf tissue), and measured for total height (mean of 16.0 m) and diameter at breast height (dbh, mean of 50 cm). Fruits were collected from 12 seed trees in PF and from 11 seed trees in CG. Seed trees were randomly selected, without consideration of growth, vigor, or stem form, but were separated by at least 50 m in order to decrease the probability of seed collection from related trees. In PF, the distance between trees ranged from 16 to 51,917 m (mean of 12,801 m) and between seed trees from 69 to 29,466 m (mean of 8,371 m). In CG, the distance between trees ranged from 40 to 26,469 m (mean of 6,124 m) and between seed trees from 208 to 8,550 m (mean of 3676 m). More than 100 fruits were col-

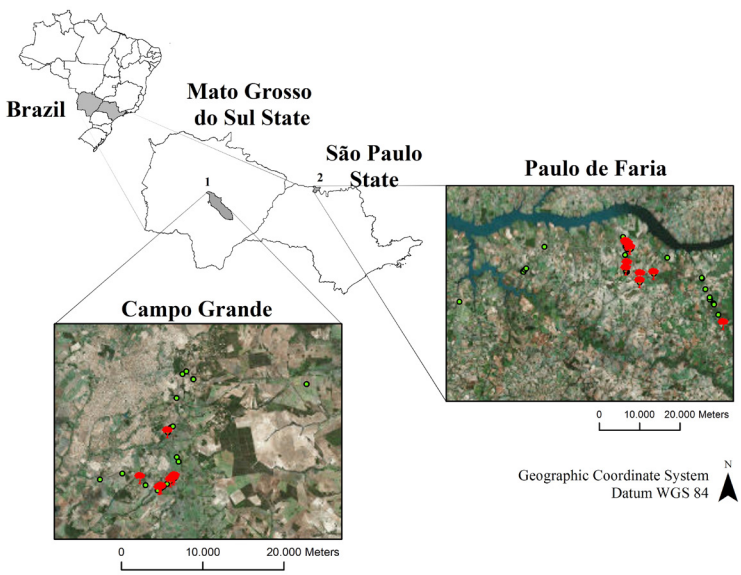

Figure 1 Spatial distribution of sample trees in Paulo de Farias (PF) and Campo Grande (CG) populations. Red circles - seed trees, green circles - adult trees from which seeds were not sampled.

lected per tree due to the fact that each fruit contains a single seed. Seeds were germinated in the nursery at the Education, Research, and Extension Farm (EREF) of FEIS/UNESP, located in the municipality of Selviria, Mato Grosso do Sul, Brazil. The first leaf pair from 20 to 30 day old seedlings was collected for DNA extraction. Leaves were sampled from 30 seedlings of each seed tree, for a total of 690 individuals. However, 91 seedlings were excluded from the analysis due to missing data for two or more loci; thus only 298 and 301 seedlings from PF and CG, respectively, were used in the statistical analyses. In March 2015, the genotyped seedlings were used to establish a provenance and progeny test at EREF, using a random block design, with 23 treatments (seedling families), six repetitions (blocks), five plants per plot, and spacing of 3 x 2 m, for a total of 30 seedlings per family.

Microsatellite analysis

DNA was extracted from 100 mg of fresh leaf tissue from adults and seedlings using the method described by Doyle & Doyle (1990). We used five specific microsatellite loci (DaE12, DaE41, DaE46, DaE63, and Dod08) developed by Soares et al. (2012) and BM164 transferred from *Phaseolus vulgaris* (Gaitán-Solís et al. 2002). Amplification reactions were performed in an ABI Veriti 96 thermocycler, under the following thermal profile: 2 min at 96 ºC, followed by 33 cycles of 1 min at 94 °C, annealing at 56 °C for 1 min and at 72 °C for 2 min, followed by 5 min at 72 °C. The total reaction volume was 25 µL, as specified for the Qiagen 10 Multiplex PCR Kit. Amplification products were subjected to electrophoresis in an ABI3100XL– D Filter (FAM-HEX-NED-ROX) sequencer. Genotypic data generated were exported using Genotyper® software (Applied Biosystems Inc., Foster City). Data were analyzed with the software TANDEM 9.0 (Matschiner & Salzburger 2009), which provides the best statistical approximation for discrete allele classes, considering microsatellite motifs (di-, tri-nucleotide, etc.).

Analysis of genetic diversity and structure

The genotypic disequilibrium between pairwise loci was estimated only for adult samples. Seedlings were excluded because gene frequencies of open-pollinated seedling families are biased due to the fact that all seedlings from the same family receive at least one maternal allele, thus genotypic disequilibrium is very likely to be observed as an artifact of the maternal contribution. Statistical significance of genotypic disequilibrium was tested by permuting alleles among individuals, associated with a Bonferroni correction for multiple tests (95%, α= 0.05). Genetic diversity was estimated for adults and seedling families based on: total number of alleles across loci (k), allelic richness (R), and observed (H_o) and expected heterozygosity (H_e). All estimates were calculated using the FSTAT 2.9.3.2 software (Goudet 2002). To check for inbreeding in adults (F), seed trees (F_m), and seedling families (F_o), we estimated the fixation index F as follows. For adults, was estimated using FSTAT and the statistical significance was tested by permuting alleles among individuals, associated with a Bonferroni correction. The frequency of null alleles (*Null*) and the fixation index corrected for null alleles (F_{Null}) was estimated for adults in each population using a Bayesian approach (IIM), with 200,000 cycles and a burn-in of 50,000, implemented in INEST 1.1 software (Chybicki & Burczyk 2009). Mean and individual seed tree fixation index (F_m) was estimated using the SPAGEDI 1.3 software (Hardy & Vekemans 2002). For seeds, mean population and family fixation indices [F_o = 1 - (H_o/H_e)] were estimated using H_o calculated in each population and H_e estimated from pol-

len pool gene frequencies of each population, due to the fact that each plant within a family receives at least one maternal allele, which can bias F_o estimates (Tambarussi et al. 2015). The pollen pool gene frequencies were estimated using the MLTR 3.4 software (Ritland 2002). To investigate whether estimates of genetic diversity and the fixation index were significantly different between adults and seedling families, we used an unpaired *t*-test. Genetic differentiation between pairwise adult, family, and adult and family samples was estimated using the standardized G'_{st} statistic (Hedrick 2005). However, these indices were estimated manually between adults and seeds and between seed samples using the gene frequencies in the pollen pool calculated with MLTR, because sampling of family structures may result in overestimates of gene frequencies of maternal alleles, as noted above.

Analysis of spatial genetic structure

The spatial genetic structure (SGS) for adult individuals of each population was estimated for eight distance classes, using the coancestry coefficient (θ_{xy}) described in Loiselle et al. (1995), and the SPAGEDI 1.3 software (Hardy & Vekemans 2002). To compare populations, similar distance classes were used and selected based on CG, where the density was lower, and considering the criteria that each class must present at least 100 pairwise θ_{xy} estimates. The statistical significance of θ_{xy} per distance class was determined by the limits of the confidence interval at 95% probability, calculated with 1,000 Monte Carlo permutations of individuals between distance classes. To compare the extent of SGS between populations, the S_p statistic (Vekemans & Hardy 2004) was used.

Mating system analysis

Mating system indices were estimated at the population and individual seed tree levels (family), using the method of maximum likelihood (Expectation Maximization algorithm) implemented in MLTR software (Ritland 2002). The estimated indices were: multilocus (t_m) and single-locus (t_s) outcrossing rate, mating among related individuals (t_m - t_s), correlation of selfing (r_s), multilocus paternity correlation (r_p), and correlation of selfing among loci (r_{sl}). The 95% confidence interval for each estimate was obtained with 1000 bootstraps, using individuals within seedling families as the re-sampling unit. We also calculated the effective number of pollen donors ($N_{ep} = 1/r_p$, Ritland 2002), proportions of self-sibs ($P_{ss} = s^2$, where $s = 1 - t_m$ is the selfing rate), half-sibs ($P_{hs} = t^2_m(1 - r_p)$), full-sibs ($P_{fs} = t^2_m r_p$), and self-half-sibs ($P_{shs} = 2st_m$), mean coancestry coefficient (Θ), and variance effective size (N_e) as described in Tambarussi et al. (2016). The number of seed trees necessary for seed collection to retain a reference effective population size ($N_{e(r)}$) of 150 (Lacerda et al. 2008) was calculated as: $M = N_{e(r)}/N_e$ (Sebbenn 2006). The estimate of m is based on three assumptions: (i) seed trees are not related, (ii) seed trees do not receive an overlapping pollen pool (each seed tree mates with a different set of pollen donors) and (iii) seed trees do not mate with each other. Thus, related individuals in the whole progeny sample occur only within a progeny array, but not among different progeny arrays. Details for the estimation of 95% confidence intervals for Θ, N_e, and *m* are given in Tambarussi et al. (2016). We used the Spearman's rank correlation coefficient (ρ) to investigate if the sum of the rate of selfing (*s*) and mating among relatives (t_m - t_s), and the sum $s + (t_m - t_s)$, decreased the allelic richness (*R*) and observed heterozygosity (H_o) and increased the fixation index within seedling families (F_o), and if the within family coancestry (Θ) decreased the allelic richness (*R*) and (H_o).

Pollen dispersal analysis

As we did not sample all trees in the very large sampling areas, the true pollen donors for many seedlings may remain undetected by paternity analysis. Thus, we opted to use an indirect

method to estimate pollen dispersal distance. Pollen dispersal distance was estimated using the POLDISP software (Robledo-Arnuncio et al. 2007). Because this method assumes an absence of selfing, we used the CERVUS 3.1 software (Kalinowski et al. 2007) to identify selfed seeds (PF= 154; CG= 34) and excluded them from the analysis. The KINDIST module was used to estimate the paternity correlation for each family (r_p) and pairwise r_p and spatial distance (D) between seed trees. Mean r_p per family was used to calculate the mean effective number of pollen donors ($N_e = 1/r_p$). Pairwise r_p and D between seed trees was used to test if r_p declines as D increases, using the Spearman ranking correlation (ρ). Since (ρ) did not indicate a significant decrease in r_p for PF (-0.111; $P = 0.421$) and CG (-0.004, $P = 0.977$), pollen dispersal distance was estimated using the TWOGENER module, based on normal and exponential power distributions. The dispersal distribution that best explained the pollen dispersal was selected based on the lowest standard error. We also estimated the effective reproductive pollen donor density (D_e), pollen pool genetic differentiation among seed trees (ϕ_{ft}), and mean, axial variance (δ), scale (a) and shape (b) of pollen dispersal distance.

Results

Genetic diversity and structure

For the adult sample, no genotypic disequilibrium between pairwise loci was detected after Bonferroni correction (data not shown). A total of 48 alleles were scored for adults and seedling families of both populations. Based on an unpaired t-test, the allelic richness (R), observed (H_o) and expected heterozygosity (H_e) were not significantly different between adults, seedling families, and between adults and seedling families together, of both populations (Table 1). The uncorrected fixation index for null alleles (F) was significantly higher than zero ($P < 0.05$) in adults and seedling families of both populations (range 0.19-0.33), suggesting inbreeding. The estimated frequency of null alleles for adults in PF ranged among loci from 0.005 (DaL12) to 0.129 (DaE46) and from 0.06 (DaL12) to 0.048 (DaE41) in CG. The fixation index corrected for null alleles (F_{Null}) was not significantly greater than zero for the adult populations. The F value was significantly lower ($P < 0.05$) in adults than in seedling families. Genetic differentiation G'_{st} was significantly higher than zero ($P < 0.05$) for all pairwise sample comparisons between adults (0.268: 0.199-0.336, mean: 95% confidence interval), PF and CG seedling families (0.340: 0.271-0.409), and between adults and seedling families of PF (0.018: 0.005-0.031) and CG (0.153: 0.095-0.210).

Spatial genetic structure

Spatial genetic structure (SGS) was detected in both populations, with the mean pairwise coancestry coefficient (θ_{xy}) decreasing with increasing distance between adults (Figure 2). SGS was significantly ($P < 0.05$) higher than

Table 1 Genetic diversity for adults and seedling families of PF and CG populations

Sample	n	R (95% CI)	H_o (95% CI)	H_e (95% CI)	F (95% CI)	F_{Null}
PF: adults	69	3.4 (2.4/4.3)	0.45 (0.36/0.54)	0.55 (0.44/0.66)	0.19 (0.07/0.31)*	0
CG: adults	54	3.8 (2.9/4.8)	0.43 (0.27/0.60)	0.59 (0.44/0.74)	0.27 (0.01/0.53)*	0
PF: seedling families	298	3.1 (2.2/4.0)	0.36 (0.31/0.42)	0.56 (0.54/0.67)[A]	0.33 (0.23/0.43)*	-
CG: seedling families	301	5.5 (3.0/8.0)	0.50 (0.30/0.70)	0.64 (0.52/0.77)[A]	0.25 (0.03/0.46)*	-

Note. Abbreviations: n - sample size, R - allelic richness for 51 genotypes, H_o and H_e observed and expected heterozygosity, respectively; F and F_{Null} - fixation index uncorrected and corrected for null alleles, respectively; 95% CI: 95% confidence interval; *$P < 0.05$. A: was estimated from pollen pool gene frequencies.

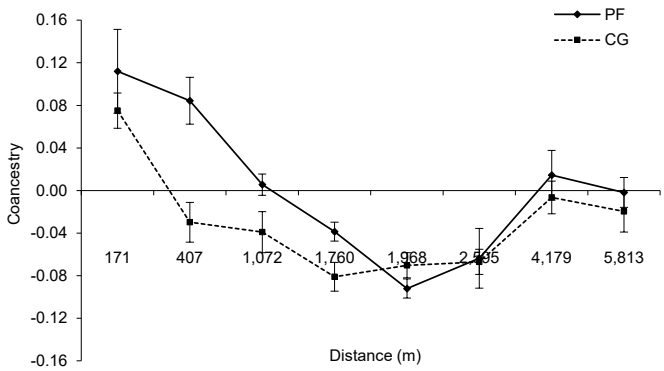

Figure 2 Correlogram with the mean pairwise coancestry coefficient (lines) and 95% confidence interval for adult trees of PF and CG populations

expected for the hypothesis of an absence of SGS at distances up to 275 m in PF and 700 m in CG, indicating that individuals located within these distances may be related. Although the mean pairwise θ_{xy} in the first distance class was higher in PF (0.112) than in CG (0.075), no significant differences were detected between populations. The extent of SGS inferred by the S_p-statistic (up to 1300 m) was not significantly different between PF (0.037: 0.018-0.056, mean: min. and max. 95% CI) and CG (0.043: 0.030-0.056).

Mating system

Maternal fixation index (F_m) was variable among trees of both populations (-0.56-0.70), but the means of both populations were not significantly different from zero (Tables 2 and 3). Outcrossing rate (t_m) was significantly lower than unity (1) in both populations (PF = 0.45, CG = 0.90), in 11 and four seedling families from PF and CG, respectively, as well as significantly lower in PF than in CG. At the population level, mating among related individuals (t_m - t_s) was significantly greater than zero in both PF (0.12) and CG (0.26), but at the individual level, it was significantly greater than zero in seven seedling families of CG ranging from 0.12 to 0.37. Mating among relatives was also significantly lower in PF than CG, suggesting biparental inbreeding. The selfing correlation (r_s) was significantly higher than zero, indicating individual variation in t_m within populations. Correlation of selfing among loci ($r_{s(l)}$) was significantly higher than zero in PF (0.62), but it was not significantly different from zero in CG (0.09). Paternity correlations (r_p) were significantly higher than zero in both populations (PF = 0.59; CG = 0.47) and in all seedling families of PF (0.10-0.65) and CG (0.10-0.53), indicating a low effective number of pollen donors (N_{ep}) fertilizing the seed trees of both populations (PF: mean of 1.7, range: 1.5–9.8; CG: mean 2.1, range: 1.9 to 10.3). PF seedling families were composed mainly of self-half-sibs (P_{shs} = 49.5%) and self-sibs (P_{ss} = 30.3%), while CG seedling families were composed mostly of half-sibs (P_{hs} = 43.3%) and full-sibs (P_{fs} = 38.2%). These results produced a significantly higher coancestry and lower effective size (N_e) in PF (Θ = 0.322, N_e = 1.47) than in CG (Θ = 0.200, N_e = 2.26). The number of seed trees (m) for seed collection was significantly higher in PF (102) than in CG (66). The indices F_o, R and H_o were also variable among seedling families (F_o: -0.18-0.65, R: 1.5–3.9, H_o: 0.21-0.75); while the indice s, t_m - t_s, [s + (t_m - t_s)] and (Θ) significantly decreased R and H_o, as suggested by the negative and significant (P < 0.05) Spearman's rank correlation coefficient (ρ) values between s vs R (ρ = -0.726), t_m - t_s vs R (ρ = -0.663), s + (t_m - t_s) vs R (ρ = -0.527),Θ vs R (ρ = -0.620), s vs H_o (ρ = -0.488), t_m - t_s vs H_o (ρ = -0.413), [s + (t_m - t_s)] vs H_o (ρ = -0.488), and Θ vs H_o (ρ = -0.513).

Table 2 Mating system and pollen dispersal indices based on normal and exponential distributions in PF and CG populations

	PF (SE)	GC (SE)
Mating system		
Fixation index of seed trees: F_m	0.07 (-0.28/0.13)	-0.01 (-0.37/0.16)
Multilocus outcrossing rate: t_m	0.45 (0.39/0.51)	0.90 (0.87/0.94)
Mating among relatives: t_m - t_s	0.12 (0.11/0.12)	0.26 (0.23/0.26)
Correlation of selfing: r_s	0.14 (0.07/0.22)	0.12 (0.06/0.23)
Correlation of selfing among loci: $r_{s(l)}$	0.616 (0.39/0.91)	0.09 (-0.09/0.26)
Correlation of paternity: r_p	0.59 (0.44/0.74)	0.47 (0.38/0.55)
Effective number of pollen donors: N_{ep}	1.7 (1.3/2.3)	2.1 (1.8/2.6)
Frequency of pairwise self-sibs: P_{ss}	0.303 (0.242/0.367)	0.009 (0.004/0.017)
Frequency of pairwise half-sibs: P_{bs}	0.082 (0.067/0.087)	0.433 (0.399/0.466)
Frequency of pairwise full-sibs: P_{fs}	0.120 (0.068/0.191)	0.382 (0.288/0.480)
Frequency of pairwise self-half-sibs: P_{sbs}	0.495 (0.478/0.500)	0.175 (0.116/0.229)
Coancestry within family: U	0.322 (0.296/0.348)	0.200 (0.184/0.236)
Variance effective size: N_e	1.47 (1.37/1.60)	2.26 (1.94/2.46)
Number of seed trees for seed collection: *m*	102 (94/110)	66 (61/77)
Pollen dispersal		
Sample size for outcrossed seedlings: *n*	144	267
Correlation of paternity: r_p	0.39 (0.21/0.53)	0.37 (0.28/0.46)
Effective number of pollen donors: N_{ep}	2.7 (0.4/5.0)	2.7 (1.9/3.4)
Global pollen pool differentiation: ϕ_{ft}	0.260	0.241
Normal dispersal distance: $\delta \pm \sigma$ (m)	6,572 ± 5,244	1,395 ± 1,113
Exponential dispersal distance: δ (m)	7,402	1,585
Scale parameter for exponential: *a*	3,701	793
Error: normal dispersal distribution	1.885	0.832
Error: exponential dispersal distribution	1.788	0.729

Note. Abbreviations: PF - Paulo de Faria population and CG - Campo Grande population, SE: standard error; σ: squared rooted axial variance.

Pollen dispersal

According to the Spearman correlation, the individual paternity correlation (r_p) of seed trees estimated using MLTR and POLDISP (Tables 2 and 3) were highly associated (0.810, $P < 0.001$). Mean r_p population values were not different between MLTR and POLDISP estimates and between populations (Tables 2 and 3). The effective number of pollen donors (N_{ep}) estimated using POLDISP ranged among seed trees in PF from 1 to 14.5 (mean of 2.7) and from 1.6 to 5.2 (mean of 2.7) in CG. Global pollen pool differentiation among seed trees (ϕ_{ft}) was similar for PF (0.260) and CG (0.241). The dispersal distribution was obtained with the normal and exponential dispersal kernel; the exponential power dispersal distribution did not converge in both populations (Table 2). The best dispersal distribution fit was obtained by the exponential dispersal kernel in both populations (lowest standard error). The mean pollen dispersal distance estimates were 6,572 m in PF and 1,395 m in CG.

Table 3 Mating system indices estimated using MLTR and POLDIS for seed trees of PF and CG populations

		MLTR (SE)										POLDISP		
	n	F_m	t_m	t_m - t_s	r_p	N_{ep}	U	F_o	N_e	R	H_o	n	r_p	N_{ep}
PF														
9	25	0.28	0.38 (0.20)	-0.30 (0.08)	0.10 (0.04)	9.8	0.422	0.26	1.26	2.5	0.43	9	0.15	6.7
14	25	0.70	0.33 (0.18)	-0.15 (0.08)	0.11 (0.04)	9.3	0.593	0.22	0.90	2.1	0.42	8	0.28	3.5
27	25	0.20	0.55 (0.22)	-0.07 (0.12)	0.17 (0.10)	5.9	0.325	0.10	1.63	2.7	0.51	11	0.31	3.2
32	25	0.05	0.01 (0.01)	-0.24 (0.01)	0.11 (0.01)	9.5	0.516	0.40	1.04	2.3	0.30	0	NE	NE
33	24	-0.10	0.66 (0.22)	-0.14 (0.14)	0.18 (0.12)	5.7	0.235	0.52	2.25	2.4	0.26	16	0.25	4.0
36	25	-0.37	0.47 (0.20)	-0.13 (0.12)	0.13 (0.06)	7.8	0.297	0.35	1.78	2.7	0.35	11	0.31	3.2
37	25	-0.02	0.38 (0.20)	-0.23 (0.10)	0.12 (0.04)	8.4	0.329	0.50	1.61	2.3	0.30	10	0.26	3.8
42	24	-0.38	0.46 (0.20)	-0.19 (0.14)	0.41 (0.16)	2.4	0.307	0.65	1.73	2.3	0.21	11	1.06	1.0
51	24	-0.34	0.46 (0.24)	-0.34 (0.16)	0.11 (0.04)	9.3	0.300	0.33	1.77	2.5	0.38	12	0.07	14.8
57	25	-0.50	0.92 (0.18)	0.02 (0.14)	0.65 (0.08)	1.5	0.215	0.38	2.45	1.5	0.28	23	0.56	1.8
66	26	0.06	0.59 (0.22)	-0.15 (0.12)	0.36 (0.18)	2.8	0.281	0.12	1.87	2.8	0.50	15	0.50	2.0
71	25	-0.45	0.71 (0.22)	-0.09 (0.12)	0.30 (0.18)	3.3	0.228	0.19	2.31	2.6	0.43	18	0.35	2.9
CG														
79	27	0.05	1.00 (0.01)	0.15 (0.06)	0.46 (0.16)	2.2	0.191	0.29	2.40	3.3	0.46	27	0.43	2.3
81	29	0.00	0.84 (0.14)	0.20 (0.10)	0.29 (0.16)	3.4	0.194	0.24	2.71	3.2	0.53	24	0.26	3.9
89	24	-0.56	0.66 (0.20)	-0.03 (0.14)	0.25 (0.18)	3.9	0.238	0.25	2.22	3.3	0.47	16	0.35	2.8
113	23	-0.27	0.93 (0.10)	0.14 (0.10)	0.23 (0.20)	4.3	0.169	0.16	3.10	3.8	0.54	21	0.42	2.4
116	28	0.26	1.00 (0.01)	0.12 (0.06)	0.10 (0.08)	10.3	0.174	0.18	3.03	3.9	0.57	28	0.19	5.2
120	28	0.43	1.00 (0.01)	0.19 (0.08)	0.48 (0.24)	2.1	0.265	0.29	1.99	3.3	0.49	28	0.64	1.6
122	29	-0.23	0.97 (0.03)	0.02 (0.06)	0.24 (0.16)	4.1	0.161	-0.18	3.26	3.2	0.75	28	0.24	4.1
123	30	0.05	0.80 (0.16)	0.09 (0.10)	0.26 (0.14)	3.9	0.209	0.43	2.52	2.9	0.40	23	0.20	5.0
131	28	0.33	0.90 (0.12)	0.37 (0.08)	0.53 (0.20)	1.9	0.273	0.33	1.94	3.7	0.44	26	0.62	1.6
139	25	-0.07	0.83 (0.16)	0.09 (0.10)	0.21 (0.12)	4.9	0.189	0.32	2.78	2.6	0.46	20	0.34	2.9
142	30	-0.08	0.86 (0.14)	0.23 (0.10)	0.37 (0.20)	2.7	0.198	0.37	2.66	2.9	0.39	26	0.35	2.9

Note. Abbreviations: n - sample size, F_m and F_o - fixation index for mother and family, respectively; t_m - outcrossing rate, $t_m - t_s$ - mating among relatives, r_p - paternity correlation, N_{ep} - effective number of pollen donors, U - coancestry coefficient, N_e - effective size, R - allelic richness for 19 genotypes, H_o - observed heterozygosity, *SE* - standard error, *NE* - not estimated as the selfing rate was practically 1.0.

Discussion

Mixed mating system

In both populations, we found a non-random mating pattern due to the occurrence of self-fertilization, mating among related individuals (t_m - t_s), and correlated mating (r_p). These results indicate genetic drift in the studied reproductive events. Deviation from random mating has also been reported for another *D. alata* population (Tarazi et al. 2010), along with many other tropical tree species (Degen and Sebbenn 2014). Our estimates of outcrossing rate (t_m) at the population and individual tree levels show that *D. alata* has a mixed mating system, producing seeds through combinations of self-fertilization and outcrossing, especially in PF (Tables 2 and 3). We also found wide variation between populations (PF = 0.45, CG = 0.90) and individual seed trees within PF (0.01-0.92) and CG (0.66-1.00). These results are similar

to a previous study in one population of the species located close to Sao Paulo state by Tarazi et al. (2010), who detected a population outcrossing rate of 0.711 and high individual variation among trees (0.338-0.998). Since the species has been determined as self-incompatible (Oliveira & Sigrist 2008), our results suggest that there is individual and population variation for self-compatibility. The self-incompatibility was determined using hand pollination in 20 trees located close to our CG population, within Campo Grande municipality, Mato Grosso do Sul state (Oliveira & Sigrist 2008), where we detected the highest outcrossing rate and low variation for outcrossing rate (t_m). These results indicate that individuals of the PF population have lower frequencies of self-incompatibility or deleterious alleles than the CG population to avoid seed production from selfing. However, the results from both our analysis and that of Tarazi et al. (2010) may be overestimates of outcrossing rates due to the likely occurrence of inbreeding depression between the fertilization and seedling stages. Inbreeding depression can result in either a lack of seed germination or seedling mortality of some inbred individuals. Therefore, analyses based only on germinated seeds underestimate the rates of selfing and mating among related individuals. Oliveira and Sigrist (2008) detected a late acting self-incompatibility system and a high rate of fruit abortion (45%) for the species, which may be a result of inbreeding depression. Inbreeding depression between fertilization and seed, seedling, juvenile and adult stages has been observed for many tropical trees (Hufford and Hamrick 2003; Chaves et al. 2011; Ismail et al. 2014; Tambarussi et al. 2017), which supports the idea that it may also occur in *D. alata.*

The lower t_m in PF than CG may also be explained by the lower population density in PF (0.00055 trees/ha) than CG (0.00119 trees/ha), which may affect pollinator foraging behavior. The species is pollinated by the bee species *Xylocopa suspecta*, *Pseudaugochlora graminea*, and *Apis mellifera*, which generally visit many flowers of the same tree for at least 60 minutes before flying to the next tree (Oliveira & Sigrist 2008). The effect of population density on t_m estimates has been evaluated for other tropical trees, and populations with low densities tend to have lower t_m than populations with higher densities, such as *Pachira quinata* (Jacq.) W.S. Alverson (Rymer et al. 2015), *Guaiacum sanctum* L. (Fuchs and Hamrick 2011), *Tabebuia rosea alba* (Ridl.) Sand. (Feres et al. 2012), *Swietenia macrophylla* King (Breed et al. 2012), and *Cariniana legalis* (Mart.) Kuntze (Tambarussi et al. 2016). Alternatively, the difference in t_m between populations, as well as between individual trees, can derive from differences in genetic load or number and frequency of identical by decent deleterious alleles (IBDA) among populations and individuals. If the frequency of IBDA causing inbreeding depression in CG was extensive, a large number of inbred seeds would not have germinated nor would seedlings have survived at the nursery stage. This could result in an even greater underestimation of selfing and mating among related individuals in CG than PF, as discussed above.

Pollen dispersal

Our indirect estimates of pollen dispersal distance (δ) based on the TWOGENER method (Austerlitz and Smouse 2001) represent a novel result for *D. alata* (Table 2) as they are based on spatially isolated trees in pastures, along roads, and in agricultural fields. Pollen dispersal distribution was better explained by the exponential dispersal kernel for both populations, showing long-distance pollen dispersal, which is greater in PF (6,572 m) with low population density (0.000552 trees/ha) than in CG (1,395 m) with a higher population density (0.00119 trees/ha). Estimates of δ based on TWOGENER are dependent on the assumed effective population density (D_e) and a low D_e results in high δ (Austerlitz and

Smouse 2001), which can explain the longer dispersal distance estimated for PF. However, it is important to examine the pollen and seed dispersal distances using a direct method such as parentage analysis, to directly assess gene flow patterns without depending on density dispersal models.

Genetic differentiation

Genetic differentiation (G'_{st}) between populations for adults (0.268) and pollen pool measured in seedling families (0.340) indicate important differences in gene frequencies between populations. High levels of genetic differentiation among *D. alata* populations were also detected in three populations from central Brazil (G'_{st} = 0.803, Tarazi et al. 2010) and 25 populations from central and northern Brazil (F_{st} = 0.267, Collevatti et al. 2013). The detected genetic differentiation can be explained by the distance between study populations (560 km in this study, at least 219 km in Tarazi et al. (2010), and ranging from 20 to 1350 km in Collevatti et al. (2013), associated with an isolation by distance (IBD) gene dispersal pattern. Collevatti et al. (2013) also found that historic gene immigration was generally very low ($N_em < 0.5$), with few populations presenting more than one migrant per generation. Our results support these findings, showing that contemporary pollen dispersal follows the IBD pattern, with strong decreases in mating frequency as the distance among trees increases, and strong genetic drift due to selfing, mating among relatives, and correlated mating. The results presented herein and in Tarazi et al. (2010) indicate that IBD is caused by the behavior of the *D. alata* pollinators, which tend to visit near-neighbor trees.

The high levels of G'_{st} and inbreeding detected within populations for seedlings in our study, and for adults and seeds in Tarazi et al. (2010) and Collevatti et al. (2013), reinforced the need for the conservation of *D. alata* genetic resources. Increased inbreeding ultimately leads to a loss of heterozygosity in populations, which may result in a reduction of fitness due to the negative effects of inbreeding depression, thus compromising the long-term survival of populations (Tambarussi et al. 2017). In order to obtain a satisfactory representation of allelic diversity for *ex situ* conservation of *D. alata* genetic resources, seed collection must sample from both populations, given the genetic divergence between them. Future breeding programs aiming at genetic gains and the maintenance of genetic diversity may benefit from crossings between highly divergent individuals selected for economic traits. On the other hand, if the genetic structure found between the *D. alata* populations evaluated in this study partly reflects selection for differential adaptive variation, mating between individuals from different populations may result in out-breeding depression. This may also cause a reduction in fitness of inter-specific hybrid seedlings in natural environments or progeny trials due to the breakup of locally adapted genomic combinations (Allendorf et al. 2013).

Mating among relatives

Mating among related individuals (t_m - t_s) was detected in both populations, but it was lower in PF (0.12) than in CG (0.26). Mating among relatives indicates the occurrence of related individuals or intrapopulation spatial genetic structure (SGS) within the populations. We detected SGS in both populations, which can explain the observed rates of mating among relatives. Tarazi et al. (2010) also detected t_m - t_s (0.135) and SGS (196 m) in their study of the species, while Soares et al. (2008) found SGS in several *D. alata* populations. SGS in plants is caused by the dispersal of seeds near to mother trees, vegetative propagation, as well as isolation by distance pollen dispersal patterns (Vekemans & Hardy 2004, Degen & Sebbenn 2014, Duminil et al. 2016). *Dypteryix alata* seeds are dispersed primarily by barochory and secondarily

by monkeys, bats, and rodents (Sano et al. 2004), which may restrict seed dispersal distances and result in SGS (Soares et al. 2008). Furthermore, pollen dispersal is mediated by bees, which generally restricts pollen foraging to near-neighbor individuals. If such trees are genetically related due to SGS, mating among related individuals is inevitable.

The high t_m - t_s in CG can be explained by the combination of the distance at which SGS was detected, the behavior of pollinators, and probable differences in genetic load between the studied populations, resulting in inbreeding depression. Relatedness estimates based on gene markers are robust, but can be subject to estimation errors due to the number of gene markers used and assumptions about gene frequencies in parental populations (Hardy & Vekemans 2002, Moraes et al. 2012). We assumed that the gene frequencies of the current adult population represent the gene frequencies of their parental populations. Very likely, this assumption is incorrect, as the studied adults originated from various reproductive events and were subject to genetic drift. It is possible that this assumption resulted in an error in the estimation of the pairwise coancestry between adults. However, based on Ismail et al. (2014), related individuals may present pairwise estimates of $\theta_{xy} > 0.099$. The number of pairwise related individuals with $\theta_{xy} > 0.099$ in the first distance class (171 m) was lower in PF (31) than in CG (49); it was also lower within the second distance class, where SGS was significant in CG (275 m; PF = 49, CG = 67). Furthermore, bee species that pollinate *D. alata* generally forage among near-neighbor trees (Oliveira & Sigrist 2008). Therefore, the higher t_m - t_s in CG than PF may be explained by the greater frequency of near-neighbor related trees in CG and near-neighbor foraging behavior of pollinators. Even though significant differences were not detected between populations, the extent of SGS was higher in CG (S_p = 0.043) than in PF (S_p = 0.037), suggesting a greater probability of mating occurring among related individuals in CG. The S_p values also indicate strong SGS in both populations, and are at least two times higher than the general estimate for animal pollen dispersed species (0.0171) and four times greater than for animal seed dispersed species (0.0088) (Vekemans and Hardy, 2004). The mean θ_{xy} in the first (171 m) and second (275 m) distance class was higher in PF (θ_{xy} = 0.112 and θ_{xy} = 0.103) than CG (θ_{xy} = 0.075 and θ_{xy} = 0.049). These results suggest that there is a greater probability of inbreeding occurring through random mating of related individuals within these two distance classes (as well as a greater probability of inbreeding depression) in PF than CG. The level of inbreeding from mating among related individuals is equal to the θ_{xy} between parents and a higher θ_{xy} will result in increased rates of inbreeding and inbreeding depression in descendants. Thus, the greater θ_{xy} in PF probably resulted in higher inbreeding depression thus affecting seedling survival. As such, some individuals were not analysed, resulting in an underestimation of t_m - t_s. Pollen dispersal distance may also explain the higher t_m - t_s in CG, as mean pollen dispersal distance was higher and t_m - t_s lower in PF than that detected in CG.

Correlated mating

Based on the MLTR software, correlated mating was detected in both populations (PF = 0.59, CG = 0.47) and all seed trees of PF and CG (0.10 - 0.65), although the results were not significantly different between populations. These results indicate that there are some full-sibs within seedling families and that a low effective number of pollen donors (N_{ep}) fertilized the seed trees (1.5 - 10.3). Similarly, r_p estimates using POLDIS were strongly correlated with MLTR results, supporting the idea that seed trees of both populations were effectively fertilized by a low number of pollen donors (1 - 14.5). Global pollen pool differentiation among seed trees (ϕ_{ft}) was also similar for PF (0.260) and CG (0.240), suggesting that the

pollen pool that fertilized the seed trees was equally diverse in the two populations. The pollinators of *D. alata* likely pollinate flowers with a limited set of pollen collected from previously visited conspecifics, producing many offspring from the same seed tree that are full-sibs. Correlated mating may also be the result of a small number of potential pollen parents, as is the case in both populations (low population density), and due to the foraging behavior of the pollinators (Oliveira & Sigrist 2008). Together, these factors can explain correlated mating and mating among related individuals. Due to the differences in rates of selfing and correlated mating between populations, the mean coancestry coefficient (Θ) was higher and effective size (N_e) lower within seedling families from PF (Θ = 0.322, N_e = 1.47) than seedling families from CG (Θ = 0.200, N_e = 2.26). For *ex situ* conservation, tree breeding, and environmental reforestation, the number of seed trees (m) required for seed collection must be at least 102 trees in populations with low density (approximately 0.00055 trees ha^{-1}), such as PF, and 66 in populations with a density around 0.00119 trees ha^{-1}.

Inbreeding from selfing and mating among relatives

Selfing and mating among related individuals can produce inbreeding and decrease genetic diversity in terms of allelic richness, heterozygosity, and effective size within seedling families. Self-fertilization ($s = 1 - t_m$) and mating among related individuals ($t_m - t_s$) were detected in both populations and in almost all seed trees, indicating that there is inbreeding within seedling families (Table 2). These results can explain the inbreeding detected within seedling families (F_o) from PF (0.10-0.65) and CG (0.16-0.43). In PF, 11 of the 12 seed trees presented self-fertilization; in CG, four of the 11 seed trees presented self-fertilization and seven seed trees presented mating among relatives. The correlation of selfing among loci ($r_{s(l)}$) in PF (0.62) indicates that inbreeding in seedlings mainly originated from selfing, where in CG inbreeding originated from mating among relatives (0.91 = 1 - 0.09). The higher $t_m - t_s$ value in CG (0.26) is likely related to SGS, pollinators behavior, and pollen-dispersal distances, as well as the likely low genetic load producing inbreeding depression in this population, as discussed above. Selfed individuals are expected to present higher levels of inbreeding [$F_0 = 0.5(1 + F_m)$], where F_m is the inbreeding coefficient of the mother] than individuals resulting from mating among relatives. In the latter case, inbreeding will be equal to the coancestry coefficient between parents ($F_0 = xy$), which for a mixed mating system is expected to vary between the degree of coancestry expected for second cousins (θ_{xy} = 0.03125) and full-sibs (θ_{xy} = 0.25). Mating between two selfed individuals from the same mother (θ_{xy} = 0.5) are unlikely to occur in tree populations due to the expected inbreeding depression, which removes inbred individuals between the fertilization and adult stages, as discussed below. Although the estimated inbreeding in PF (F_0 = 0.33) was higher than in CG (F_0 = 0.25), these values were not significantly different. The results, however, did follow the expected pattern of higher values for inbreeding originating from selfing than mating among relatives. Furthermore, although not statistically different, the values of allelic richness (R), observed (H_o) and expected heterozygosity (H_e) were lower in seedlings from PF than CG, supporting the idea that selfing decreases genetic diversity more than mating among relatives. The Spearman correlation also showed that higher levels of the indices s, $t_m - t_s$, [s+($t_m - t_s$)] and Θ decreased R and H_o within seedling families.

The fixation index corrected for null alleles (F_{Null}) in adults was significantly lower than the fixation index estimated for seedling families (F_0) of both populations (Table 1), indicating selection against inbred individuals between seedling and adult stages (inbreeding depression). The same result was detected by

Tarazi et al. (2010) and is a common pattern in tropical tree species (Degen & Sebbenn 2014), with the levels of inbreeding being more pronounced in the initial life stages and adults generally presenting lower levels of inbreeding than seeds, seedlings, and juveniles (Bittencourt & Sebbenn 2007, Duminil et al. 2016, Tambarussi et al. 2016).

Genetic diversity

Our results support previous studies that report low levels of genetic diversity and the presence of inbreeding in *D. alata* populations (Tarazi et al. 2010, Collevatti et al. 2013). However, our results for mating system and pollen dispersal provide evidence that explains the low genetic diversity and inbreeding in the populations. We found a maximum observed heterozygosity (H_o) of 0.5 and a minimum fixation index (F) of 0.19 in adults. However, the fixation index corrected for null alleles (F_{Null}) for adults was zero, indicating an absence of inbreeding (Table 1). Using nine microsatellite loci, Tarazi et al. (2010) studied adults from three *D. alata* populations in Brazil and found H_o ranging from 0.32 to 0.35 and significant positive F values (range: 0.08 - 0.16), suggesting inbreeding. Collevatti et al. (2013), using eight microsatellite loci, studied adult trees from 25 populations across the natural geographic distribution of the species in Brazil. They found H_o ranging from 0.16 to 0.64, with 24 populations presenting H_o lower than 0.5, and significant positive F values for 16 populations (range: 0.12 - 0.51), which also indicates inbreeding. The low levels of H_o and inbreeding in the populations may be partly attributed to the fact that the species presents a mixed mating system, as detected in this study and by Tarazi et al. (2010). As such, seeds are produced by mixtures of self-fertilization and outcrossing, and part of the outcrossing occurs among related individuals, some seeds present endogamy. Combined with correlated mating (full-sibs produced by the mating between the same mother and father trees), the mixed mating system can decrease genetic diversity, thus explaining the low genetic diversity and high inbreeding detected in seedling populations of the species.

Conclusions and implications for genetic conservation and breeding

The studied *D. alata* populations present a mixed mating system and mating was not random, as outcrossed seedlings were produced in part by mating among related individuals and correlated mating. Pollen dispersal distances indicate an isolation by distance dispersal pattern due to the behavior of pollinators, resulting in high genetic differentiation among populations and contributing to SGS. Thus, open-pollinated seedling families may present inbreeding from both selfing and mating among related individuals and are composed of mixtures of self-, half-, full-, and self-half-sibs. Because of this mating pattern, the coancestry coefficient (Θ) was higher and the effective size (N_e) within seedling families was lower than that expected for panmictic populations ($\Theta = 0.125$, $N_e = 4$). This mating pattern explains the low genetic diversity and the detected inbreeding, as well as the high levels of genetic differentiation found both in our study and in previous studies (Tarazi et al. 2010, Collevatti et al. 2013). These results have practical implications for the genetic conservation and breeding of the species. In terms of conservation, the low levels of genetic diversity within populations, as well as the presence of inbreeding at adult and seed stages, highlight the importance of both *in situ* and *ex situ* genetic conservation of *D. alata* genetic resources. Increased inbreeding ultimately leads to a loss of heterozygosity in populations, which may result in reduced fitness due to the negative effects of inbreeding depression, and may compromise the long-term survival of these populations.

As the Brazilian savanna biome is currently experiencing strong deforestation rates, *D. alata* exhibits ecological and economic importance. Because it is a threatened tree species (IUCN 1998), all remaining populations are potentially important for in situ and ex situ conservation. For in situ conservation, our results indicate that spatially isolated populations located more than 5 km apart must be linked via biodiversity corridors, or in other words, by establishing new populations through reforestation between remnant populations. Oliveira and Sigrist (2008) also emphasize the importance of *D. alata* in silvopastoral systems and pasture recovery, as it is an evergreen tree with a leafy canopy that can provide shade for cattle and fruit for their consumption during the dry season. Silvopastoral systems can play an important role in the establishment of biological corridors, which enable the exchange of genes via pollination and seed dispersal, thus connecting scattered forest fragments and isolated populations across a landscape.

Furthermore, the high genetic divergence and differences in selfing rate, which suggests differences in gene frequencies for self-incompatibility alleles or genetic load between the populations studied herein, reinforces the importance of these populations for genetic conservation. The genetic diversity of these populations can be recombined in germplasm banks or seed orchards with other provenances and used for tree improvement and environmental reforestation. To obtain a satisfactory representation of the allelic diversity for *ex situ* conservation, tree breeding, and environmental conservation of *D. alata* genetic resources, our results indicate that the number of seed trees (*m*) required for seed collection for *ex situ* conservation must be determined based on population density and extent of SGS. For populations with a density of about 0.00119 trees/ha, our results indicate that seeds must be collected from at least 66 seed trees to retain an effective size in progeny arrays of 150. For populations with a lower population density, or 0.00055 trees/ha, our results indicate that collection must occur from at least 102 seed trees. Furthermore, our results for SGS indicate that seed trees identified for collection must be at least 275 m apart to avoid collecting seeds from related trees, which can decrease the effective size of progeny arrays due to the occurrence of related seedlings among seedling families. In populations from PF region, but as well as from other regions, special attention must be taken in the nursery stage. For ex situ conservation and environmental reforestation only seedlings without indication of inbreeding such as albinisms, low growth, presence of diseases, bad stem form should be selected do avoid mortality in the newly established populations.

In terms of population breeding of the species, populations with mixed mating systems differ from outcrossing and self-fertilizing populations because they are composed of individuals with different degrees of inbreeding and relatedness. The selection process for breeding programs is therefore more complex. Previous studies have assessed mixed mating systems in the context of breeding programs (Burgess et al. 1996, Miranda et al. 2013) and such systems can have a significant impact on the performance of individuals with regard to economically important silvicultural traits (Tambarussi et al. 2016). Our results for *D. alata* show deviations from random mating, with open-pollinated seedling families presenting different levels of relatedness as well as inbreeding. Thus, to avoid overestimations of heritability and expected genetic gains in selections, the additive genetic variance ($\sigma^2_A = \sigma^2_p/2\Theta$, Cockerham & Weir 1984) of progeny tests based on open-pollinated seedling families must be estimated by at least $\sigma^2_A = \sigma^2_p/0.4$) ($2\Theta = 2 \cdot 0.2$) and not by $\sigma^2_A = \sigma^2_p/0.25$, which is used in the case of true half-sib seedling families. Inbreeding has an impact on individual phenotypic values and genetic variances and as such the effect of selection is more complex in this type of species.

Acknowledgements

We thank the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) for funding this Project (grant number 2012/18747-2). Alexandre M. Sebbenn and Mario L.T. Moraes are the recipients of research fellowships from Conselho Nacional de Desenvolvimento Cientifico e Tecnologico (CNPq). The authors are very grateful to the anonymous referees for their important suggestions in the data analysis. We also thank Dr. Evelyn R. Nimmo for editing the English of the manuscript. Alessandro Alves-Pereira thanks FAPESP for a scholarship (grant number 2013/11137-7).

References

Allendorf, F.W., Luikart, G.H., Aitken, S.N., 2013. Conservation and the genetics of populations. Oxford Wiley-Blackwell. 2nd ed. 602 p.

Ashley, M.V., 2010. Plant parentage, pollination, and dispersal: How DNA microsatellites have altered the landscape. Critical Reviews in Plant Sciences 29(3): 148–161. DOI: 10.1080/07352689.2010.481167

Austerlitz, F., Smouse, P.E., 2001. Two-generation analysis of pollen flow across a landscape. II. Relation between ΦH', pollen dispersal, and inter-female distance. Genetics 157(2):851–857.

Baldauf, C., Ciampi-Guillardi, M., Aguirra, T.J., Corrêa, C.E., Santos, F.A.M., Souza, A.P., Sebbenn, A.M., 2014. Genetic diversity spatial genetic structure and realized seed and pollen dispersal of *Himatanthus drasticus* (Apocynaceae) in the Brazilian savanna. Conservation Genetetics 15(5): 1073–1083. DOI: 10.1007/s10592-014-0600-5

Bittencourt, J.M., Sebbenn, A.M., 2007. Patterns of pollen and seed dispersal in a small fragmented population of a wind pollinated *Araucaria angustifolia* in southern Brazil. Heredity 99(6):580–591. DOI: 10.1038/sj.hdy.6801019

Breed, M.F., Gardner, M.G., Ottewell, K.M., Navarro, C.M., Lowe, A.J., 2012. Shifts in reproductive assurance strategies and inbreeding costs associated with habitat fragmentation in Central American mahogany. Ecology Letters 15(5): 444–452. DOI: 10.1111/j.1461-0248.2012.01752.x

Burgess, I.P., Williams, E.R., Bell, J.C., Harwood, C.E., 1996. The effect of outcrossing rate on the growth of selected families of *Eucalyptus grandis*. Silvae Genetica 45: 97–100.

Burczyk, J., DiFazio, S.P., Adams, W.T., 2004. Gene flow in forest trees: how far do genes really travel. Forest Genetics 11(3-4): 179–192.

Chaves, L.J., Vencovsky, R., Silva, R.S.M., Telles, M.P.C., Zucchi, M.I., Coelho, A.S.G., 2011. Estimating inbreeding depression in natural plant populations using quantitative and molecular data. Conservation Genetics 12(2): 569–576. DOI: 10.1007/s10592-010-0164-y

Chybicki, I.J., Burczyk, J., 2009. Simultaneous estimation of null alleles and inbreeding coefficients. Journal of Heredity 100(1): 106–113. DOI: 10.1093/jhered/esn088

Cockerham, C.C., Weir, B.S., 1984. Covariances of relatives stemming from a population undergoing mixed self and random mating. Biometrics. 40(1): 157–164. DOI: 10.2307/2530754

Collevatti, R.G., Telles, M.P.C., Nabout, J.C., Chaves, L.J., Soares, T.N., 2013. Demographic history and the low genetic diversity in *Dipteryx alata* (Fabaceae) from Brazilian Neotropical savannas. Heredity 111(2): 97–105. DOI: 10.1038/hdy.2013.23

Degen, B., Sebbenn, A.M., 2014. Genetics and Tropical Forests. In: Pancel L., Kohl M. (eds.) Tropical Forestry Handbook, 2nd ed. Springer Verlag, pp. 1–30. DOI: 10.1007/978-3-642-41554-8_75-1

Doyle, J.J., Doyle, J.L., 1990. Isolation of plant DNA from fresh tissue. Focus 12(1):13–15.

Duminil, J., Abessolo, D.T.M., Bourobou, D.N., Doucet, J-L., Loo, J., Hardy, O., 2016. High selfing rate, limited pollen dispersal and inbreeding depression in the emblematic African rain forest tree *Baillonella toxisperma* – Management implications. Forest Ecology and Management 379(1): 20–29. DOI: 10.1016/j.foreco.2016.08.003

Duminil, J., Dainou, K., Kaviriri, D.K., Gillet, P., Loo, J., Doucet, J.L., Hardy, O.J., 2016. Relationships betwenn population density, fine-scale genetic structure, mating system and pollen dispersal in a timber tree from African rainforests. Heredity 116(3): 295–303. DOI: 10.1038/hdy.2015.101

Ellstrand, N.C., 2014. Is gene flow the most important evolutionary force in plants? American Journal of Botany101(5): 737–753. DOI: 10.3732/ajb.1400024

Feres, J.M., Sebbenn, A.M., Guidugli, M.C., Mestriner, M.A., Moraes, M.L.T., Alzarte-Marin, A.L., 2012. Mating system parameters at hierarchical levels of fruits, individuals and populations in the Brazilian insect-pollinated tropical tree, *Tabebuia roseo-alba* (Bignoniaceae). Conservation Genetics 13(2):393–405. DOI 10.1007/s10592-011-0292-z

Frankhan, R., 2012. How closely does genetic diversity in finite populations conform to predictions of neutral theory? Large deficits in regions of low recombination. Heredity 168(3): 167–178. DOI: 10.1038/hdy.2011.66

Fuchs, E.J., Lobo, J.A., 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*. Conservation Biology 17: 149–157. DOI: 10.1046/j.1523-1739.2003.01140.x

Fuchs, E.J., Hamrick, J.L., 2011. Mating system and pollen

flow between remnant populations of the endangered tropical tree, *Guaiacum sanctum* (Zygophyllaceae). Conservation Genetics 12(1):175–185. DOI 10.1007/s10592-010-0130-8

Gaitán-Solís, E., Duque, M.C., Edwards, K.J., Tohme, J., 2002. Microsatellite in common bean (*Phaseolus vulgaris*): Isolation, characterization, and cross-species amplification in *Phaseolus sp*. Crop Science 42(6): 2128–2136. DOI:10.2135/cropsci2002.2128

Goudet, J., 2002. FSTAT (Version 2.9.3.2.): a computer program to calculate F-statistics. Journal of Heredity 86(6): 485–486. DOI: 10.1093/oxfordjournals.jhered.a111627

Hardy, O.J., Vekemans, X., 2002. SPAGeDI: a versatile computer program to analyze spatial genetic structure at the individual or population levels. Molecular Ecology Notes 2(4): 618–620. DOI: 10.1046/j.1471-8286.2002.00305.x

Hedrick, P.W., 2005. A standardized genetic differentiation measure. Evolution 59(8): 1633–1638. DOI: 10.1111/j.0014-3820.2005.tb01814.x

Hufford, K.M., Hamrick, J.L., 2003. Viability selection at three early life stages of the tropical tree, *Platypodium elegans* (Fabaceae, Papilionoideae). Evolution 57(3): 518–526. DOI: 10.1111/j.0014-3820.2003.tb01543.x

Ibanes, B., Sebbenn, A.M., Azevedo, V.C.R., Moreno, M.A., Gandara, F.B., Tambarussi, E.V., Ferraz, E.M., Damasceno-Silva, K.J., Lima, P.S.C., Carvalhaes, M.A., 2016. Genetic diversity and spatial genetic structure in populations of *Orbignya phalerata* Mart. under different exploitation intensities in the Brazilian savanna. Silvae Genetica 64: 201–211.

Ismail, S.A., Ghazoul, J., Ravikanth, G., Kushalappa, C.G., Shaanker, R.U., Kettle C.J., 2014. Fragmentation genetics of *Vateria indica*: implications for management of forest genetic resources of an endemic dipterocarp. Conservation Genetics 15(3): 533–545. DOI: 10.1007/s10592-013-0559-7

IUCN (1998). 1998 IUCN Red List of Threatened Species. Web: http://www.iucnredlist.org/details/32984/0. Accessed: 08.2017.

Kalinowski, S.T., Taper, M.L., Marshall, T.C., 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecular Ecology 16(5): 1099–1106. DOI: 10.1111/j.1365-294X.2007.03089.x

Lacerda, A.E.B., Kanashiro, M., Sebbenn, A.M., 2008. Long-pollen movement and deviation of random mating in a low-density continuous population of a tropical tree *Hymenaea courbaril* in the Brazilian Amazon. Biotropica. 40(4): 462–470. DOI: 10.1111/j.1744-7429.2008.00402.x

Lander, T.A., Boshier, D.H., Harris, S.A., 2010. Fragmented but not isolated: Contribution of single trees, small patches and long distance pollen flow to genetic connectivity for *Gomortega keule*, and endangered tree. Biological Conservation 143(11): 2383–2590. DOI. org/10.1016/j.biocon.2010.06.028

Leonarduzzi, C., Leonardi, S., Menozzi, P., Piotti, A., 2012. Towards an optimal sampling effort for paternity analysis in forest trees: what do the raw numbers tell us? iForest 5: 18–25. DOI: 10.3832/ifor0606-009

Loiselle, B.A., Sork, V.L., Nason, J., Graham, C., 1995. Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). American Journal of Botany 82(11): 1420–1425. DOI: 10.2307/2445869

Lorenzi, H., 2002. Árvores brasileiras: manual de identificação e cultivo de plantas arbóreas nativas do Brasil [Brazilian Trees: manual for identification and cultivation of native tree plants in Brazil]. 4 ed. Nova Odessa: Instituto Plantarum, 384 p.

Matschiner, M., Salzburger, W., 2009. Tandem: integrating automated allele binning into genetics and genomics workflows. Bioinformatics 25(15):1982-1983. DOI: 10.1093/bioinformatics/btp303

Melo, A.T.O., Franceschinelli, E.V., 2016. Gene flow and fine-scale spatial genetic structure in *Cabralea canjerana* (Meliaceae), a common tree species from the Brazilian Atlantic forest. Journal of Tropical Ecology 132(2): 135–135. DOI: 10.1017/S0266467416000067

Mendonça, R.C., Fefili, J.M., Walter, B.M.T., Silva Júnior, M.C., Resende, A.V., Filgueiras, T.S., Nogueira, P.E., 2008. Flora vascular do Cerrado [Vascular flora of the brazilian savanna]. In: Sano S.M., Almeida S.P. (eds.), Cerrado: ambiente e flora [Brazilian savanna: environment and flora] (1st ed). Planaltina, Brazil. Embrapa-CPAC.

Miranda, A.C., Moraes, M.L.T., Tambarussi, E.V., Furtado, E.L., 2013. Heritability for resistance to *Puccinia psidii* Winter rust in *Eucalyptus grandis* Hill ex Maiden in Southwestern Brazil. Tree Genet Genomes 9(2): 321–329. DOI.org/10.1007/s11295-012-0572-x

Moraes, M.A., Gaino, A.P.S., Moraes, M.L.T., Freitas, M.L.M., 2012. Estimating coancestry within open-pollinated progenies of a dioecious species: the case study of *Myracrodruon urundeuva.* Silvae Genetica 61: 256–264.

Oliveira, M.I.B., Sigrist, M.R., 2008. Fenologia reprodutiva, polinização e reprodução de *Dipteryx alata* Vogel (Leguminosae-Papilionoideae) em Mato Grosso do Sul, Brasil. Revista Brasiliena Botanica 31(2): 195–207. DOI: 10.1590/S0100-84042008000200002

Ritland, K., 2002. Extensions of models for the estimation of mating systems using n independent loci. Heredity 88(4): 221–228. DOI: 10.1038/sj.hdy.6800029

Robledo-Arnuncio, J.R., Austerlitz, F., Smouse, P.E., 2007. POLDISP: a software package for indirect estimation of contemporary pollen dispersal. Molecular Ecology 7: 763–766. DOI: 10.1111/j.1471-8286.2007.01706.x

Rymer, P.D., Sandiford, M., Harris, S.A., Billingham, M.R., Boshier, D.H., 2015. Remnant *Pachira quinata* pasture trees have greater opportunities to self and suffer reduced reproductive success due to inbreeding depression. Heredity 115(2): 115–124. DOI: 10.1038/hdy.2013.73

Sano, E.E., Rosa, R., Brito, J.L.S., Ferreira, L.G., 2007.

Mapeamento de cobertura vegetal do bioma Cerrado. Planaltina: Embrapa Cerrados. 60 p.

Sano, S.M., Ribeiro, J.F., Brito, M.A., 2004. Baru: biologia e uso [Baru: biology and use]. Planaltina: Embrapa Cerrados. Embrapa Cerrados. Documentos. 52 p.

Sebbenn, A.M., 2006. Sistema de reprodução em espécies arbóreas tropicais e suas implicações para a seleção de árvores matrizes para reflorestamentos ambientais [Mating system in tropical tree species and its implications for the selection of seed trees reforestation]. In: Higa A.R., Silva L.D. (eds.), Pomares de sementes de espécies florestais nativas [Orchards of seeds of native forest species]. FUPEF, Curitiba, pp. 93- 138.

Sebbenn, A.M., Carvalho, A.C.M., Freitas, M.L.M., Moraes, S.M.B., Gaino, A.P.S.C., Silva, J.M., Jolivet, C., Moraes, M.L.T., 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(1): 134–145. DOI: 10.1038/hdy.2010.33

Silva, C.R.S., Albuquerque, P.S.B., Ervedosa, F.R., Figueira, A., Sebbenn, A.M., 2011. Understanding the genetic diversity, spatial genetic structure and mating system at the hierarchical levels of fruits and individuals of a continuous *Theobroma cacao* population from the Brazilian Amazon. *Heredity* 106(6): 973–985. doi: 10.1038/hdy.2010.145

Simon, M.F., Grether, R., de Queiroz, L.P., Skema, C., Pennington, R.T., Hughes, C.E., 2009. Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by in situ evolution of adaptations to fire. Proceedings of the National Academy of the United States of America 106(48): 20359–20364. DOI: 10.1073/pnas.0903410106

Soares, T.N., Chaves, L.J., Telles, M.P.C., Diniz-Filho, J.A.F., Resende, L.V., 2008. Landscape conservation genetics of *Dipteryx alata* ("baru" tree: Fabaceae) from Cerrado region of central Brazil. Genetica 132(1): 9–19. DOI: 10.1007/s10709-007-9144-7

Soares, T.N., Diniz-Filho, J.A.F., Nabout, J.C., Telles, M.P.C., Terribile, L.C., Chaves, L.J., 2015. Patterns of genetic variability in central and peripheral populations of *Dipteryx alata* (Fabaceae) in the Brazilian Cerrado. Plant Systematics and Evolution 301: 1315–1324. DOI: 10.1007/s00606-014-1155-0

Soares, T.N., Melo, D.B., Resende, L.V., Vianello, R.P, Chaves, L.J., Collevatti, R.G., Telles, M.P.C., 2012. Development of microsatellite markers for the Neotropical tree species *Dipteryx alata* (Fabaceae). American Journal of Botany e72–e73. DOI: 10.3732/ajb.1100377

Surles, S.E., Arnold, J., Schnabel, A., Hamrick, J.L., Bongarten, B.C. 1990. Genetic relatedness in open pollinated families of two leguminous tree species, *Robinia pseudoacacia* L. and *Gleditsia triacanthos* L. Theoretical and Applied Genetics 80(1): 49–56.

Takemoto, E., Ohada, I.A., Garbelotti, M.L., Tavares, M., Aued-Pimentel, S., 2001. Composição química da semente e do óleo de baru (*Dipteryx alata* Vog.) nativo do Município de Pirenópolis, Estado de Goiás [Chemical composition of the seed and oil of Baru (*Dipteryx alata* Vog.) native of the Municipality of Pirenópolis, State of Goiás]. Revista do Instituto Adolfo Lutz 60(2): 113–117.

Tambarussi, E.V., Boshier, D., Vencovsky, R., Freitas, M.L.M., Sebbenn, A.M., 2015. Paternity analysis reveals significant isolation and near neighbor pollen dispersal in small *Cariniana legalis* Mart. Kuntze populations in the Brazilian Atlantic Forest. Ecology and Evolution 5(23): 5588–5600. DOI: 10.1002/ece3.1816

Tambarussi, E.V., Boshier, D.H., Vencovsky, R., Freitas, M.L.M. Di-Dio, O.J., Sebbenn, A.M., 2016. Several Small: How inbreeding affects conservation of *Cariniana legalis* Mart. Kuntze (Lecythidaceae) the brazilian Atlantic Forest's largest tree. International Forestry Review 18(4): 502–510. DOI: 10.1505/146554816820127550

Tambarussi, E.V., Boshier, D.H., Vencovsky, R., Freitas, M.L.M., Sebbenn, A.M., 2017. Inbreeding depression from selfing and mating between relatives in the Neotropical tree *Cariniana legalis* Mart. Kuntze. Conservation Genetics 18(1):225–234. DOI: 10.1007/s10592-016-0896-4

Tarazi, R., Moreno, M.A., Gandara, F.B., Ferrazi, E.M., Moraes, M.L.T., Vinson, C.C., Ciampi, A.Y., Vencovsky, R., Kageyama, P.Y., 2010. High levels of genetic differentiation and selfing in the Brazilian cerrado fruit tree *Dipteryx alata* Vog. (Fabaceae). Genetics and Molecular Biology 33(1): 78–85. DOI: 10.1590/S1415-47572010005000007

Vekemans, X., Hardy, O.J., 2004. New insights from fine-scale spatial genetic structure analyses in plant populations. Molecular Ecology 13(4): 921–935.