Estimating coancestry within open-pollinated progenies of a dioecious species: the case study of Myracrodruon urundeuva

By M. A. Moraes¹, A. P. S. Gaino¹, M. L. T. Moraes¹, M. L. M. Freitas² and A. M. Sebenn²,³,*

(Received 20th March 2012)

Abstract
Understanding the coancestry coefficient within open-pollinated progenies has long been an area of interest because of the implications of coancestry on estimates of additive genetic variation, variance effective size and the number of seed trees required for seed collection for ex situ conservation, tree breeding and environmental restoration. This study compares three methods to calculate the coancestry coefficient within open-pollinated progenies of the dioecious tree species, Myracrodruon urundeuva, using six microsatellite loci. The methods compared were: i) correlated mating model (CMM) (Ritland, 1989); ii) TwoGENER method (S Mouse et al., 2001) to estimate the differentiation among pollen pools (TGM); and iii) HARDY et al’s 2004 method using the estimate of coancestry from LOISELLE et al. (1995) (HLM) and from Ritland (1996) (HRM). The data analysis was based on four data sets: two populations were composed of 12 progenies, two of 24 progenies, and all progenies consisted of 15 plants. The coancestry estimated using CMM ranged among populations from 0.145 to 0.158, using TGM it ranged from 0.153 to 0.181, using HLM from 0.153 to 0.162, and HRM from 0.144 to 0.147. To investigate the bias of the estimates of true relatedness within open-pollinated progenies, Myracrodruon urundeuva, was CMM because the estimated value presented no bias for true half- or full-sib progenies (0.125 and 0.25, respectively). These results have significant implications for breeding and conservation programs because coancestry-within-progenies is a key parameter in assessing the variance effective size.

Key words: additive genetic variation; effective population size; kinship; mating system.

Introduction
Open-pollinated progenies have been widely used in tree breeding, ex situ conservation and environmental reforestation programs due to the ease, low cost and speed of their production. Because little is known about the true relatedness within open-pollinated progenies, some estimates of the additive genetic variance ($\sigma^2_A$) in progeny tests of tree species assume that progenies are derived from random mating and are half-sibs (Squillace, 1974). In such a situation, the heritability coefficients can be estimated based on the assumption that the genetic covariance between plants within progenies ($COV_{A} = 2 \theta_{A} \sigma^2_A$) is $\frac{1}{4}$ of $\sigma^2_A$ (Squillace, 1974). In this hypothetical case, measurable relatedness only occurs within progenies and the coancestry coefficient would be 0.125. However, the assumptions of random mating and unrelated reproductive trees are rarely encountered in nature and the true genetic covariance between trees within progenies is usually greater than $\frac{1}{4}$ (Namkoong, 1966; Squillace, 1974; Ritland, 1989; Sebenn, 2006).

These various mating processes result in mixtures of different relationships in open-pollinated progenies, such as: i) self-sibs, ii) half-sibs, iii) full-sibs and iv) self- and outcrossing-sibs (self-half-sibs) (Squillace, 1974; Ritland, 1989). Thus, the coancestry within progenies is likely to be higher than expected in true half-sib progenies ($\theta = 0.125$) and the genetic variance among progenies will be overestimated. However, genetic marker-based estimates of coancestry within progenies can properly identify the relationship coefficients.

Understanding the relationships between plants within progenies is fundamental to ex situ conservation and environmental restoration programs that are based on seed collection strategies. Greater degrees of relatedness within progenies will result in higher frequencies of identical by descent (IBD) alleles. Consequently, larger sample sizes are needed in order to establish artificial populations that will enable the conservation and restoration of tree species.

Coancestry ($\theta_{xy}$) is the probability that two homologous alleles, randomly selected from two individuals, $x$ and $y$, are identical by descent (Lynch and Walsh, 1998). In random mating, the coancestry is equivalent to inbreeding in the offspring ($F_{m} = \theta_{xy}$) (Lindgren et al., 1997). This definition includes cases where the two individuals are identical, self-coancestry ($\theta_{xx}$), which is equivalent to inbreeding in the offspring after self-pollination ($F_{m} = \theta_{xx} = 0.5 + F_{x}$), where $F_{x}$ is the inbreeding coefficient for the mother (Lindgren et al., 1997). Thus, the average coancestry of a population ($\theta$) includes self-coancestry and coancestry among all pairs of individuals.

¹) Faculdade de Engenharia de Ilha Solteira/UNESP, Ilha Solteira, SP, Brazil, Caixa Postal, 31, 15385-000.
²) Instituto Florestal de São Paulo, CP 1322, São Paulo, SP, 01059-970, Brasil.
³) Corresponding author: ALEXANDRE M. SEBBENN.
Instituto Florestal de São Paulo, CP 1322, São Paulo, SP, 01059-970, Brazil. Telefone: 19 34351681.
E-Mail: alexandresebenn@yahoo.com.br
Inbreeding \( (F) \) is the probability of randomly withdrawing two alleles at a locus for individual \( x \) and these alleles are identical by descent (Lynch and Walsh, 1998). The probability that the alleles in \( x \) are IBD depends on the self-coancestry of the mother (if \( x \) is the result of selfing), or coancestry between parents (when \( x \) is the result of outcrossing) (Lindgren et al., 1997).

Based on data from genetic markers, there are various methods available to estimate the coancestry or relatedness between pairwise individuals (see for example, Queiller and Goodnight (1989), Loselle et al. (1995), Ritland (1996), Lynch and Ritland (1999), Wang (2002), Milligan (2003), Kalinowski et al. (2006) and Wang (2007)). These approaches can be used to estimate the coancestry or relatedness within open-pollinated progenies, using the Mark (Ritland, 2000), SPMITH (Hardy and Vekemans, 2002) and COANCESTRY programs (Wang, 2011a). A different method to estimate the coancestry within progenies is based on mating system parameters (Ritland, 1989). The Ritland (1989) method estimates the coancestry within progenies based on parameters of the mating system, such as outcrossing rate, individual variation in outcrossing rate and paternity correlation, and progenies may include mixtures of: i) half-sibs; ii) self-sibs; iii) self-sibs and outcrossing; and iv) full-sibs. Using the MLTR program (Ritland, 2002), the outcrossing rate and correlated mating have been estimated for many trees species (Burgess et al., 1996; Sampson, 1998; James et al., 1998; Millar et al., 2000; Collevatti et al., 2001; Butcher et al., 2002; Fuchs et al., 2003; El-Kassaby et al., 2003; Degen et al., 2004; Moraes et al., 2004; Jones et al., 2005; Azevedo et al., 2007; Lemes et al., 2007; Lacerda et al., 2008; Naito et al., 2008; Tamaki et al., 2009; Gago et al., 2010; Moraes and Sebenn, 2011; Feres et al., 2012). These studies have shown that many of the species analysed present a mixed mating system and open-pollinated seeds include various levels of relatedness, such as self-sibs, half-sibs, full-sibs and self-half-sibs. In contrast, in dioecious species, analyses have shown that open-pollinated progenies are composed of varying proportions of half-sibs and full-sibs (Moraes et al., 2004; Bittencourt and Sebenn, 2008; Silva et al., 2008). Another method to estimate the coancestry within progenies is based on TWOGENER analysis (Austinlitz and Smouse, 2001; Smouse et al., 2001); this method separates the maternal and paternal gametic contribution of the seeds and uses analysis of variance to estimate the differentiation in the pollen pools \( (P_{\Phi}) \) received by different seed trees. From the parameter \( P_{\Phi} \), it is possible to estimate the paternity correlation \( \left( F_{\Phi} = 2P_{\Phi} \right) \) (Hardy et al., 2004). This paternity correlation can then be combined with other mating system parameters (outcrossing rate) and the Ritland (1989) method to estimate the coancestry within progenies. One further method to estimate the coancestry within progenies is based on the approach proposed by Hardy et al. (2004) which disregards the mating process. This method directly estimates the coancestry among paternal gametes within progenies \( (F) \) as an average relatedness between all pairs of individuals within each progeny. From the parameter \( F_{\Phi} \) it is also possible to estimate the paternity correlation \( \left( F_{\Phi} = 2F_{\Phi} \right) \) (Hardy et al., 2004) and then the coancestry within progenies. If highly polymorphic loci are used in the analysis it is expected that the estimate of coancestry will determine with some accuracy the true relationship between plants within progenies (half-sibs and full-sibs). However, the problem with this method is that it only estimates the proportion of half-sibs and full-sibs and it is only suitable for dioecious species, not for species with a mixed mating system.

To date, no study has compared these different methods to estimate coancestry within progenies. Therefore, in this study we investigate the estimates of coancestry within progenies using the methods described above, the dioecious species Myracrodruon urundeuva as a model, and six microsatellite loci. A dioecious species was chosen for this study as the Hardy et al. (2004) method does not accommodate selfing and is only suitable for outcrossing species. The three methods used to estimate the coancestry coefficients within progenies were: i) CMM – the correlated mating model (Ritland, 1989); ii) TGM – measurements of differentiation between the pollen pools received by different seed trees based on TWOGENER analysis (Smouse et al., 2001); iii) HLM – the method outlined by Hardy et al. (2004), using estimates of coancestry based on Loselle et al. (1995); and HRM from Hardy et al. (2004) using estimates of coancestry based on Ritland (1996). Specifically, the goal of this study was to answer the following questions: i) Are there significant differences between the methods in estimating coancestry within progenies? ii) Which method is most appropriate to estimate the coancestry within progenies for dioecious species? iii) What are the implications of the estimate of coancestry within progenies using these different methods to calculate the variance effective size (Cockerham, 1969) within progenies?

**Material and Methods**

**The study species**

Myracrodruon urundeuva F.F.M.F. & Allemão (aroeira) is a dioecious tropical tree species belonging to the Anacardiaceae family that is widely distributed in Brazil, also occurring in Argentina, Bolivia and Paraguay (3°30’S, Brazil to 25°S, Argentina). The species is found in the following biomes: Seasonal Semi-Deciduous Forest, Seasonal Deciduous Forest, Savanna, Chaco-Sul-Matogrossense and Pantanal Matogrossense (Carvalho, 2003). The population density of the species varies widely, ranging from 2 to 115 trees per hectare. The flowers are unisex and pollinated by insects, especially bees. Seeds are dispersed by wind and animals, particularly birds (Carvalho, 2003). Generally the fruit produced by the species contains only one seed and the wood has a high economic value due to its quality and durability (Carvalho, 2003). Trees can reach 30 m in height and 100 cm in diameter at breast height (Carvalho, 2003).

**Sampling**

The data set used in this study was obtained from tissue samples of open-pollinated *M. urundeuva* progenies...
from four populations, established in progeny tests at the Fazenda de Ensino, Pesquisa e Extensão (FEPE), UNESP Ilha Solteira Campus, located in Selvíria, Mato Grosso do Sul State, Brazil. Two of the sampled populations are from São Paulo State (Baurú and Aramina), one from Mato Grosso do Sul State (Selvíria), and one from Rio Grande do Norte State (Seridó). Each sampled forest is a forest fragment that has undergone significant anthropogenic disturbance. From each forest fragment, seeds were collected from 12 to 28 seed trees. All populations were evaluated using a panel of six microsatellite loci developed specifically for the species (Caetano et al., 2005). To reduce sources of variation and to ensure balanced data sets, all progenies consisted of 15 plants; the Baurú and Seridó populations were composed of 12 progenies and Aramina and Selvíria were composed of 24 progenies. All individuals were genotyped for the six loci.

**Estimate of coancestry within progenies**

i) Estimates of the coancestry coefficient from correlated mating model (CMM)

The mixed-mating and correlated mating models were initially used to characterize the mating system of the four sample populations using the multilocus MLTR program version 3.4 (Ritland, 2002). The calculated parameters were: multilocus outcrossing rate \( t_m \); single-locus outcrossing rate \( t_i \); outcrossing rate among related trees \( t_m - t_i \); and multilocus paternity correlation \( r_{pm} \). Due to our relatively small sample size in two populations (Baurú and Seridó), the estimates were carried out using the Expectation-Maximization (EM) numerical method. The 95% confidence interval (95% CI) of each parameter was calculated from 1,000 bootstrap replicates (plants within progenies), using the 95% quartile of the ranked bootstraps. These parameters were used to estimate the average coancestry coefficient \( \Theta \) among plants within progenies, calculated using the expression: \( \Theta = 0.125(1 + F_p)(1 + F_{pm}) \) (Souza et al., 2005), where \( F_p \) is inbreeding in the parental populations (assumed to be zero, \( F_p = 0 \)).

ii) Estimates of the coancestry coefficient by TwoGENER analysis (TGM)

The coefficient of coancestry within progenies was also calculated based on differentiation in the pollen pools among seed trees \( \Phi_p \) using TwoGENER analysis (Austerlitz and Smouse, 2001; Smouse et al., 2001). TwoGENER analysis was performed using the multilocus genotype of seed trees and their open-pollinated seeds. The parameter \( \Phi_p \) was calculated as an intra-class correlation of male gametes (pollen) within progenies: \( \Phi_p = \sigma_p^2 / (\sigma_p^2 + \sigma_y^2) \), where \( \sigma_p^2 \) is the genetic variation among pollen gametes received by different seed trees, and \( \sigma_y^2 \) is the genetic variation among pollen gametes within progenies (Smouse et al., 2001). From parameter \( \Phi_p \) we estimated the paternity correlation \( (\hat{r}_p = 2\Phi_p) \) (Hardy et al., 2004). The coancestry coefficient within progenies \( \Theta \) was calculated by: \( \Theta = 0.125(1 + F_p)(1 + 2\Phi_p) \) (Bittencourt and Sebenn, 2008). We assumed an absence of inbreeding in the parental populations \( (F_p = 0) \).

iii) Hardy et al.’s (2004) estimates of coancestry coefficient (HLM and HRM)

The estimate of the coancestry coefficient within progenies using the method outlined by Hardy et al. (2004) requires an analysis of the coancestry coefficient between all pairs of paternal gametes (pollen) within progenies \( F_i \). To do this, paternal gametes are converted into homozygous diploid genotypes. For this analysis, we subtracted the paternal gametes from the offspring genotype; this analysis is unambiguous, except in cases where both maternal and offspring genotypes are heterozygous for the same alleles. In such ambiguous cases, the genotype of offspring was assumed as the paternal gamete (Hardy et al., 2004). For example, if an offspring has the alleles \( A_1 \) and \( A_2 \) and its mother has the alleles \( A_1 \) and \( A_3 \), the allele \( A_3 \) is paternal. However, if the mother has the alleles \( A_4 \) and \( A_5 \), the father may have contributed alleles \( A_1 \) or \( A_2 \) (ambiguous); thus, following Hardy et al. (2004) we maintained the diploid heterozygote genotype of the offspring as the paternal origin. The haplotype gametes were then diploidized and used to calculate the coancestry among all pairwise paternal gametes (Hardy et al., 2004). This determination of paternal gametes was carried out using MS Excel. The coancestry coefficient between all pairs of plants within progenies was calculated using the SPAGEDI 1.3 program (Hardy and Vekemans, 2002). The coancestry among parental gametes within progenies \( F_i \) was estimated using two methods: the method based on Lonseille et al. (1995) (HLM) and that based on Ritland (1996) (HRM). The expected value for the coancestry coefficient is 0.5 if a genotype pair has the same paternal parent and zero if the pair is different paternal parents (Hardy et al., 2004). From the parameter \( F_i \) we estimated the paternity correlation \( (r_p = 2F_i) \) (Hardy et al., 2004). The coancestry within progenies \( (\Theta) \) was calculated by: \( \Theta = 0.125(1 + F_i)(1 + F_p) = 0.125(1 + 2F_i)(1 + 2F_p) \), assuming an absence of inbreeding in the parental populations \( (F_p = 0) \).

iv) Estimates of other parameters dependent on the coancestry coefficient

After assessing coancestry, we also estimated other parameters such as the correlation of relatedness within progenies \( (\hat{r}_e = 2\Theta) \), the variance effective size and number of seed trees required for seed collection in order to retain the effective size of 150, as suggested by Lacerda et al. (2008). The correlation of relatedness within progenies was assessed to determine the genetic covariance between plants within progenies which is used in the calculation of the additive genetic variance \( \sigma_r^2 \) in open-pollinated progenies: \( \sigma_r^2 = \sigma_p^2 / \hat{r}_p \), where \( \sigma_r^2 \) is the genetic variation among progenies. The variance effective size \( (N_{ve}) \) was calculated using the expression: \( N_{ve} = \hat{r}_p N_{ref} / (1 + \hat{r}_p) \) (Sebenn, 2006). This approach assumes: i) an absence of relatedness among seed trees; and ii) no overlap in the pollen gene pools among seed trees.

**Simulated data**

To investigate the bias of the estimates of true relatedness within progenies we simulated two half-sib populations (12 and 24 progenies, each with 15 individuals)
and two full-sib populations (12 and 24 progenies, each with 15 individuals), using the Coancestry program (WANG, 2011a). The simulations were based on gene frequencies of the six loci of the Aramina population (chosen arbitrarily). In the simulations we considered no inbreeding in the populations. The identical by descent coefficient ($\Delta$) for half-sibs is $\Delta = 0.0, 0.0, 0.0, 0.0, 1/2, 1/2$ and for full-sibs $\Delta = 0.0, 0.0, 0.0, 0.0, 1/4, 1/2, 1/4$ (WANG, 2011a). These four simulated populations were analysed using the methods described above. The bias of coancestry estimates was measured by mean square error (MSE), calculated following WANG (2011b) as:

$$\sum_{i=1}^{k}(\theta_i - \Theta)^2$$

where $k$ is the number of replicates (bootstraps or pairwise estimates), $\Theta$ is the estimate of coancestry in the replicate $i$ (1–$k$), and is the simulated value of the coancestry (0.125 or 0.25). As the number of pairwise estimates for Hardy’s methods (HLM and HRM) was 1,260 for populations of 12 progenies ($m(n(n-1)/2) = 12[15(15-1)/2]$) and 2,520 for 24 progenies ($m(n(n-1)/2) = 24[15(15-1)/2]$), we ran the 1,260 and 2,520 bootstraps using MLTR and TwoGENE analyses.

Results

Mating system

The multilocus outcrossing rate ($t_m$) was not significantly different from unity (1.0) in any of the studied populations (Table 1). The single-locus outcrossing rate ($t_s$) was significantly lower than unity only for the Seridó population (Table 1), indicating mating among relatives. The multilocus paternity correlation ($r_{p,ml}$) was significantly greater than zero for all populations (Table 1), ranging from 0.163 (Aramina) to 0.265 (Bauru), demonstrating that progenies were composed of mixtures of full- and half-sibs.

Pollen pool heterogeneity

The differentiation in the pollen pools among seed-trees ($\Phi_p$) was significantly different from zero for all populations, ranging from 0.113 (Seridó) to 0.223 (Bauru), suggesting genetic structure in the pollen pools (Table 1). The paternity correlation ($r_p$) estimated from $\Phi_p$ (TGM; Table 1) ranged among populations from 0.226 (Seridó) to 0.446 (Bauru) and was not significantly different from that estimated using the correlated mating model (CMM).

Estimate of coancestry using Hardy’s methods

The coancestry among paternal gametes within progenies ($F_s$) estimated using Hardy’s methods was also significantly different from zero for all populations, ranging from 0.110 (Selvivria) to 0.146 (Aramina) for the HLM method and from 0.076 (Selvivria) to 0.088 (Seridó) for HRM (Table 1). The paternity correlation ($r_p$) estimated from $F_s$ ranged among populations from 0.220 to 0.292 for HLM and from 0.152 to 0.176 for HRM.

Coancestry coefficient within progenies

The average coancestry coefficient within progenies (Table 1) estimated using CMM ranged among the populations from 0.145 to 0.153. Using TGM it ranged from 0.153 to 0.181, using HLM it ranged from 0.153 to 0.162, and using HRM from 0.144 to 0.147. No significant differences were observed between the results from CMM and TGM and between TGM and HLM (Table 1). However, the estimates based on HLM were significantly higher than the estimates from CMM in the Seridó and Aramina populations. Similarly, the estimates from HLM were all significantly higher than those provided by HRM in all populations. The estimate based on HRM was significantly lower than the estimates from CMM and TGM in the Bauru population.

Coefficient of relatedness, effective size and number of seed-trees for seed collection

As we assumed an absence of inbreeding in the parental population, the coefficient of relatedness is twice the coefficient of coancestry within progenies ($r_p = 29$); consequently, the results of these estimates were similar to that observed for coancestry (Table 2).

Table 1. – Mating system parameters, pollen pool differentiation and coancestry coefficient among paternal gametes within progenies with 95% confidence interval (95% CI) for four Myracrodruon urundeuva populations.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Bauru (95% CI)</th>
<th>Seridó (95% CI)</th>
<th>Aramina (95% CI)</th>
<th>Selvivria (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMM - Mixed and correlated mating model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of progenies (total of trees)</td>
<td>12 [180]</td>
<td>12 [180]</td>
<td>24 [360]</td>
<td>24 [360]</td>
</tr>
<tr>
<td>Multilocus outcrossing rate: $t_m$</td>
<td>1.000 (1.000 to 1.000)</td>
<td>1.000 (1.000 to 1.000)</td>
<td>1.000 (1.000 to 1.000)</td>
<td>0.990 (0.990 to 1.000)</td>
</tr>
<tr>
<td>Single-locus outcrossing rate: $t_s$</td>
<td>0.000 (0.000 to 0.000)</td>
<td>0.880 (0.860 to 0.910)</td>
<td>1.000 (0.990 to 1.000)</td>
<td>0.990 (0.990 to 1.000)</td>
</tr>
<tr>
<td>Mating among relatives: $t_s - t_m$</td>
<td>0.000 (0.000 to 0.000)</td>
<td>0.120 (0.090 to 0.140)</td>
<td>0.000 (0.000 to 0.000)</td>
<td>0.066 (0.000 to 0.000)</td>
</tr>
<tr>
<td>Multilocus paternity correlation: $r_{p,ml}$</td>
<td>0.265 (0.191 to 0.321)</td>
<td>0.164 (0.103 to 0.210)</td>
<td>0.163 (0.115 to 0.197)</td>
<td>0.175 (0.130 to 0.212)</td>
</tr>
<tr>
<td>TGM - TwoGENE analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Differentiation in pollen pool: $\Phi_p$</td>
<td>0.223 (0.110 to 0.407)</td>
<td>0.113 (0.072 to 0.142)</td>
<td>0.116 (0.084 to 0.167)</td>
<td>0.123 (0.073 to 0.158)</td>
</tr>
<tr>
<td>Paternity correlation: $r_p$</td>
<td>0.446 (0.221 to 0.814)</td>
<td>0.226 (0.145 to 0.285)</td>
<td>0.232 (0.167 to 0.335)</td>
<td>0.246 (0.145 to 0.317)</td>
</tr>
<tr>
<td>HLM and HRM - Hardy analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coancestry of paternal gametes: HLM - $F_s$</td>
<td>0.121 (0.113 to 0.129)</td>
<td>0.139 (0.131 to 0.147)</td>
<td>0.146 (0.138 to 0.154)</td>
<td>0.110 (0.102 to 0.118)</td>
</tr>
<tr>
<td>Coancestry of paternal gametes: HRM - $F_s$</td>
<td>0.081 (0.074 to 0.088)</td>
<td>0.088 (0.081 to 0.095)</td>
<td>0.085 (0.078 to 0.092)</td>
<td>0.076 (0.069 to 0.083)</td>
</tr>
<tr>
<td>Paternity correlation: HLM - $r_p$</td>
<td>0.242 (0.226 to 0.258)</td>
<td>0.278 (0.262 to 0.294)</td>
<td>0.292 (0.276 to 0.308)</td>
<td>0.229 (0.204 to 0.243)</td>
</tr>
<tr>
<td>Paternity correlation: HRM - $r_p$</td>
<td>0.162 (0.148 to 0.176)</td>
<td>0.176 (0.162 to 0.190)</td>
<td>0.170 (0.156 to 0.184)</td>
<td>0.152 (0.138 to 0.166)</td>
</tr>
</tbody>
</table>
All estimates were significantly different from zero (Table 2).

The estimate of the variance effective size \( (N_{\text{e}}) \) using CMM was significantly different from that estimated using HRM in the Baurú population and from HLM in the Seridó and Araminha populations (Table 2). For the Baurú population, the estimate of \( N_{\text{e}} \) using TGM was significantly lower than the result using HRM. The estimates of \( N_{\text{e}} \) using HLM were significantly lower than the results from HRM in all populations.

The estimated number of seed trees \( (m) \) required for seed collection was not significantly different between CMM and TGM in any of the studied populations. Significant differences were observed between CMM and HRM in the Baurú population and between CMM and HLM in the Seridó and Araminha populations. Significant differences were also observed between TGM and HRM in the Baurú population and between HLM and HRM for all populations.

Simulated data

The multilocus \( (t_m) \) and single-locus \( (t_s) \) outcrossing rates were not significantly lower than unity \((1.0)\) in any of the simulated populations (Table 3). The multilocus paternity correlation \( (r_{p,m}) \) was not significantly greater than zero for the half-sib populations (Table 3) and was significantly different from zero for the full-sib populations, with values close to unity, as expected for true full-sib progenies \( (r_{p,m} = 1) \).

The differentiation in pollen pools among seed-trees \( (\Phi_9) \) was significantly different from zero for all populations, with the exception of the simulated 12 half-sib progenies (Table 3). Although the parameter \( \Phi_9 \) was significantly different from zero in the population of 24 half-sib progenies, the estimated value was very low \( (\Phi_9 = 0.012) \). The paternity correlation \( (r_s) \) was close to zero in the half-sib populations \((12 \text{ and } 24 \text{ progenies})\) and higher than unity in the full-sib progenies, indicating a strong bias in the full-sibs progeny structure.

The estimates of coancestry, using Hardy’s methods \( (P) \) were all different from zero, although the estimates were close to zero in the half-sib progenies and greater than 0.5 for full-sib progenies \((\text{ranging from } 0.624 \text{ to } 0.691) \). Consequently, the estimates of paternity correlation \( (r_p = 2P) \) were all significantly different from zero and higher than unity for the full-sib progenies.

The estimates of the average coancestry within progenies (Table 3) for the 12 half-sib population were not significantly different across the methods used; however, only the CMM and TGM methods included the true value \( (\Theta = 0.125) \) in the 95% confidence interval. For the 24 half-sib population, no significant differences were observed between CMM and TGM and between HLM and HRM but significant differences were observed when comparing the results of CMM and TGM to the results provided by HLM and HRM. Again, only the 95% confidence interval for CMM and TGM included the true value \( (\Theta = 0.125) \). The other methods (HLM and

### Table 2. – Estimates of the relatedness coefficient, variance effective size and number of seed trees necessary for seed collection with 95% confidence interval (CI; 95%) for four Myracrodruon urundeuva populations.

<table>
<thead>
<tr>
<th>Model/Parameter</th>
<th>Baurú (95% CI)</th>
<th>Seridó (95% CI)</th>
<th>Araminha (95% CI)</th>
<th>Sclerifa (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coancestry within progenies: ( \Theta )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMM (Ritud, 1989)</td>
<td>0.158 (0.149 to 0.165)</td>
<td>0.146 (0.138 to 0.151)</td>
<td>0.145 (0.139 to 0.150)</td>
<td>0.147 (0.141 to 0.152)</td>
</tr>
<tr>
<td>TGM (Smouse et al., 2001)</td>
<td>0.181 (0.153 to 0.227)</td>
<td>0.153 (0.143 to 0.161)</td>
<td>0.154 (0.146 to 0.167)</td>
<td>0.156 (0.143 to 0.165)</td>
</tr>
<tr>
<td>HLM (Loiselle et al., 1995)</td>
<td>0.155 (0.152 to 0.157)</td>
<td>0.160 (0.158 to 0.162)</td>
<td>0.162 (0.160 to 0.164)</td>
<td>0.153 (0.151 to 0.155)</td>
</tr>
<tr>
<td>HRM (Ritud, 1996)</td>
<td>0.145 (0.144 to 0.147)</td>
<td>0.147 (0.145 to 0.149)</td>
<td>0.146 (0.145 to 0.148)</td>
<td>0.144 (0.142 to 0.146)</td>
</tr>
<tr>
<td>Relatedness within progenies: ( r_s )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMM (Ritud, 1989)</td>
<td>0.316 (0.298 to 0.330)</td>
<td>0.291 (0.276 to 0.303)</td>
<td>0.290 (0.278 to 0.300)</td>
<td>0.294 (0.282 to 0.304)</td>
</tr>
<tr>
<td>TGM (Smouse et al., 2001)</td>
<td>0.362 (0.365 to 0.345)</td>
<td>0.307 (0.286 to 0.321)</td>
<td>0.308 (0.292 to 0.334)</td>
<td>0.311 (0.286 to 0.329)</td>
</tr>
<tr>
<td>HLM (Loiselle et al., 1995)</td>
<td>0.311 (0.307 to 0.315)</td>
<td>0.320 (0.316 to 0.324)</td>
<td>0.323 (0.319 to 0.327)</td>
<td>0.305 (0.301 to 0.309)</td>
</tr>
<tr>
<td>HRM (Ritud, 1996)</td>
<td>0.291 (0.287 to 0.294)</td>
<td>0.294 (0.291 to 0.298)</td>
<td>0.295 (0.289 to 0.296)</td>
<td>0.288 (0.285 to 0.292)</td>
</tr>
<tr>
<td>Variance effective size: ( N_{\text{e}} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMM (Ritud, 1989)</td>
<td>3.16 (3.03 to 3.36)</td>
<td>3.42 (3.31 to 3.62)</td>
<td>3.45 (3.33 to 3.60)</td>
<td>3.40 (3.29 to 3.55)</td>
</tr>
<tr>
<td>TGM (Smouse et al., 2001)</td>
<td>2.76 (2.20 to 3.27)</td>
<td>3.27 (3.11 to 3.50)</td>
<td>3.25 (2.99 to 3.42)</td>
<td>3.21 (3.03 to 3.50)</td>
</tr>
<tr>
<td>HLM (Loiselle et al., 1995)</td>
<td>3.22 (3.18 to 3.26)</td>
<td>3.13 (3.09 to 3.17)</td>
<td>3.10 (3.06 to 3.13)</td>
<td>3.28 (3.14 to 3.32)</td>
</tr>
<tr>
<td>HRM (Ritud, 1996)</td>
<td>3.44 (3.40 to 3.48)</td>
<td>3.40 (3.36 to 3.44)</td>
<td>3.42 (3.38 to 3.46)</td>
<td>3.47 (3.41 to 3.51)</td>
</tr>
<tr>
<td>Number of seed trees: ( m )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMM (Ritud, 1989)</td>
<td>47 (45 to 50)</td>
<td>44 (41 to 45)</td>
<td>44 (42 to 45)</td>
<td>44 (42 to 46)</td>
</tr>
<tr>
<td>TGM (Smouse et al., 2001)</td>
<td>54 (46 to 68)</td>
<td>46 (43 to 48)</td>
<td>46 (44 to 50)</td>
<td>47 (43 to 50)</td>
</tr>
<tr>
<td>HLM (Loiselle et al., 1995)</td>
<td>47 (46 to 47)</td>
<td>48 (47 to 49)</td>
<td>44 (48 to 49)</td>
<td>46 (45 to 46)</td>
</tr>
<tr>
<td>HRM (Ritud, 1996)</td>
<td>44 (43 to 44)</td>
<td>44 (44 to 45)</td>
<td>44 (43 to 44)</td>
<td>43 (43 to 44)</td>
</tr>
</tbody>
</table>
HRM) significantly overestimated the true value. For full-sib progenies, CMM estimated the expected value (0.25) for 12 full-sib progenies, without any bias, but underestimated the expected value for the 24 full-sib progenies. For both full-sib progenies, the estimated coancestry values provided by CMM were significantly lower than that estimated by TGM, HLM and HRM. For the 24 full-sib progenies, HRM also provided significantly lower values than TGM and HLM.

According to MSE, among the four methods compared, CMM is the most accurate and HRM is the least accurate (Table 4). CMM presents no bias while TGM presented a small bias in the estimate of the coancestry (Table 4). HLM and especially HRM presented the greatest bias. The results for HLM and HRM also suggest that estimates for half-sib progenies were less accurate than estimates for full-sibs.

**Table 3.** Mating system parameters, pollen pool differentiation and genetic parameters with 95% confidence interval (CI: 95%) for four simulated populations (two half-sib and two full-sib).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Half-sibs</th>
<th>Full-sibs</th>
<th>Half-sibs</th>
<th>Full-sibs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multilocus outcrossing rate: $t_m$</td>
<td>1.000 (1.000 to 1.000)</td>
<td>1.000 (1.000 to 1.000)</td>
<td>1.000 (1.000 to 1.000)</td>
<td>1.000 (1.000 to 1.000)</td>
</tr>
<tr>
<td>Single-locus outcrossing rate: $t_s$</td>
<td>1.000 (1.000 to 1.000)</td>
<td>1.000 (1.000 to 1.000)</td>
<td>1.000 (1.000 to 1.000)</td>
<td>1.000 (1.000 to 1.000)</td>
</tr>
<tr>
<td>Mating among relatives: $t_m - t_s$</td>
<td>0.000 (0.000 to 0.000)</td>
<td>0.000 (0.000 to 0.000)</td>
<td>0.000 (0.000 to 0.000)</td>
<td>0.000 (0.000 to 0.000)</td>
</tr>
<tr>
<td>Multilocus paternity correlation: $r_{ps}$</td>
<td>0.000 (0.000 to 0.000)</td>
<td>0.999 (0.999 to 0.999)</td>
<td>0.000 (0.000 to 0.000)</td>
<td>0.974 (0.971 to 0.984)</td>
</tr>
</tbody>
</table>

**Table 4.** Coancestry coefficient and mean square error (MSE) for four simulated populations (two half-sib and two full-sib).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Half-sibs (MSE)</th>
<th>Full-sibs (MSE)</th>
<th>Half-sibs (MSE)</th>
<th>Full-sibs (MSE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coancestry CMM: $\Theta$</td>
<td>0.125 (0.125 to 0.125)</td>
<td>0.250 (0.250 to 0.250)</td>
<td>0.125 (0.125 to 0.125)</td>
<td>0.247 (0.246 to 0.248)</td>
</tr>
<tr>
<td>Coancestry TGM: $\Theta$</td>
<td>0.128 (0.124 to 0.134)</td>
<td>0.302 (0.283 to 0.322)</td>
<td>0.128 (0.125 to 0.130)</td>
<td>0.293 (0.286 to 0.299)</td>
</tr>
<tr>
<td>Coancestry HLM: $\Theta$</td>
<td>0.134 (0.131 to 0.138)</td>
<td>0.298 (0.296 to 0.300)</td>
<td>0.135 (0.130 to 0.135)</td>
<td>0.291 (0.291 to 0.300)</td>
</tr>
<tr>
<td>Coancestry HRM: $\Theta$</td>
<td>0.136 (0.131 to 0.140)</td>
<td>0.283 (0.280 to 0.286)</td>
<td>0.138 (0.132 to 0.138)</td>
<td>0.281 (0.278 to 0.284)</td>
</tr>
</tbody>
</table>

**Discussion**

In the present study we investigated various methods of estimating the coancestry coefficient within progenies. Under the conditions of perfectly random mating among a large number of unrelated and non-inbred trees, the minimum value that the coancestry coefficient within open-pollinated progenies should be is at least 0.125. In the case of dioecious species, correlated mating and mating among related individuals increases the IBD alleles within progenies and thus increases the estimate of the coancestry coefficient. Therefore, as we observed the occurrence of both mating among relatives and correlated mating in the studied populations, we expect to find values of coancestry higher than 0.125 within progenies.

We observed significant differences in the estimates among populations as well as between the various meth-
ods used. As a result of correlated mating, all methods presented a coancestry coefficient ($\Theta$) that was higher than expected for panmictic populations ($\Theta = 0.125$). The coancestry estimate used here is in fact a weighted average of the different types of relatedness that exist within the progeny. In the case of dioecious species, which may contain mixtures of half- and full-sibs, this estimate can be summarized by the expression: $\Theta = \theta_{WS} + \theta_{FS}$, where $\theta_{WS}$ is the proportion of half-sibs, and $\theta_{FS}$ is the proportion of full-sibs, respectively. The differences in the values for coancestry among the methods were expected because the calculations are based on very different approaches. CMM and TGM estimate coancestry within progenies based on mating system parameters, such as outcrossing rate and paternity correlation. In contrast, HLM and HRM disregard the process of mating and estimate an average coancestry between all pairs of paternal gametes. If all paternal gametes come from different pollen donors, the expected coancestry is zero and if all gametes come from the same pollen donor, the expected coancestry is 0.5.

The adequacy of the method used to estimate the coancestry coefficient within progenies is of fundamental importance because it is from this value one can calculate other parameters, such as the correlation of relatedness within progenies, the variance effective size and the number of seed trees required for seed collection in order to retain a certain effective population size. Thus, for example, if the coancestry coefficient is based on an underestimate of the coancestry within progenies, it will result in an underestimation of the correlations of relatedness within progenies and an overestimation of the variance effective size. This in turn will result in an underestimation of the number of seed trees from which it is necessary to collect seeds. In our case study, if we compare the results from CMM and HRM, the HRM method underestimated the coancestry at a rate of between 31 to 39% among the studied populations; in comparing TGM and HRM, the underestimation ranged from 34 to 46%. Thus, based on the comparison of the three methods analyzed here, we suggest using CMM or TGM.

In this analysis we estimated the paternity correlation ($r_p$) within progenies using the correlated mating models ($r_{PFS}$, RITLAND, 1989), TwoGener analysis ($r_p = 2\Phi_p$, AUSTRÉLITZ and SMOUSE, 2001; SMOUSE et al., 2001), and the methods outlined by HARDY et al. (2004) which estimates the coancestry coefficient between all pairs of paternal gametes (pollen) within progenies ($r_p = 2\Phi_p$) based on LOISELLE et al. (1995) and RITLAND (1996). From these estimates, we calculated the coancestry within progenies. In true half-sib progenies the expected value of $r_p$ is 0; for true full-sib progenies, the expected $r_p$ value is 1. In general, the correlated mating model correctly estimated the paternity correlation for simulated true half- and full-sib progenies (Table 3). The other methods overestimated the $r_p$, especially for full-sib progenies, resulting in values higher than biologically possible ($r_p > 1$). One possible explanation for these high values is the number of loci used (6). If a greater number of loci is used (for example 20), the estimate of $r_p$ using TwoGener analysis ($r_p = 2\Phi_p$, AUSTRÉLITZ and SMOUSE, 2001; SMOUSE et al., 2001) and the methods outlined by HARDY et al. (2004) would present a more accurate result. However, additional analyses are required in order to evaluate the implications of using a greater number of loci, alleles within loci, and gene frequencies on the estimates of $r_p$. Due to the overestimation of $r_p$, the coancestry coefficient estimated by TGM, HLM and HRM were also overestimated, particularly for full-sib progenies (Table 3).

With the exception of CMM, the methods discussed here overestimated the average coancestry within progenies, especially for full-sibs. The difference between the minimum and the maximum value (Table 3) ranged from 7 to 21%. According to MSE, among the four compared methods, CMM is the more accurate and has no bias while TGM presented a low level of bias in the estimate of the coancestry. HLM and particularly HRM presented the greatest bias. The results for HLM and HRM also suggest that estimates for half-sib progenies were less accurate than estimates for full-sibs.

Half-sib progenies are more genetically diverse than full-sibs because half-sibs have one common parent among offspring. Thus, between two half-sibs there is only a 12.5% chance that both receive the same allele from the common parent (IBD alleles). In contrast, two full-sib individuals have two parents in common and a 25% chance of inheriting IBD alleles. Consequently, in half-sib progenies, a larger number of pairwise estimates of coancestry may result in lower (or negative) values than in full-sib progenies. This can explain the low levels of accuracy in the estimates of coancestry in half-sibs. However, it is important to note that this low level of accuracy may be addressed by using a greater number of loci than used here.

Our results are significant for the studied species, _M. urundeuva_ which is widely distributed in South America (LORENZI, 2000). It has economic importance due to the production of excellent quality timber (high density and low deterioration) and because of the high quality of the wood, natural populations of the species have to date been intensely exploited (RIBEIRO, 1989). Thus, _ex situ_ conservation strategies have been adopted (provenance and progeny tests) to conserve many remaining populations of the species (MOARES et al., 2004; GAINO et al., 2010; MOARES et al., 2012). For this purpose, understanding the variance effective size ($N_{sev}$) of open pollinated seeds is important in order to determine the number of seed trees required for seed collection (MOARES et al., 2004; GAINO et al., 2010). Our results suggest that to conserve a reference effective size of 150, seeds must be collected from at least 43 seed trees (m) in each population. This result is similar to a previous study on a fragmented population of _M. urundeuva_ that suggested seeds must be collected from at least 48 seeds trees (GAINO et al., 2010).

The coefficient of relatedness ($r_p$) is also used to understand and genetic control (heritability) of quantitative traits, including diameter at breast height (DBH), tree height, volume and stem form (MOARES et al., 2004; MOARES et al., 2012). The results of the mating system
analysis indicate that, for the studied populations, it is inappropriate to assume that open-pollinated progenies of _M. urundeuva_ are half-sibs. Thus, heritability must be estimated using a coefficient of relatedness within progenies (r_e) that is higher than 0.25. In our case study, and based on the results of CMM, this value should be 0.291. Using this method will avoid an overestimation of the expected genetic gains during selection in the progeny test. Our results also highlight the importance of understanding the coancestry within progenies in programs of genetic conservation and breeding, as well as to collect seeds for reforestation.

Finally, using genetic markers, other studies have compared various methods to estimate coancestry and relatedness among pairwise individuals, including Ritland (1996), Lynch and Ritland (1999), Wang (2002) and Wang (2011a). The conclusion across all studies is that there is no method that performs best in all situations because the calculations are dependent on: the polymorphism of the markers used (number of alleles and frequency of the alleles in the loci); sample size (number of individuals and number of loci); true relatedness to be estimated (unrelated individuals, half-sibs, full-sibs, etc.); and inbreeding in the parental population (Wang, 2011a). Our study did not assess the implications of the quantity of polymorphic loci, the impact of sample size, or the number of plants within progenies on the parameters analysed. However, in estimating the coancestry within progenies, our results suggest that CMM is more accurate than the other methods described here.

**Acknowledgements**

The article is part of the master’s thesis of M.A. Moraes at the Faculdade de Engenharia de Ilha Solteira/UNESP, Ilha Solteira, São Paulo, Brazil. We would like to thank Zosé Cambruim and Alexandre M. da Silva for their assistance in collection samples and Selma M. B. Moraes, Laila T. Cardin and Juliana P. Moreira for their lab work. M.A. Moraes thanks the Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) for granting her a master’s scholarship at Faculdade de Engenharia de Ilha Solteira/UNESP (Brazil). A.M. Sebbenn and M.L.T. Moraes thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for granting them a Research Fellowship. The authors are very grateful to Dr. Evelyn R. Nimmo for her important suggestions and correction of the English in the manuscript. We also would also like to thank one anonymous reviewer for very important suggestions and corrections to the manuscript.

**References**


Butcher, P. A. and E. R. Williams (2002): Variation in outcrossing rates and growth in _Eucalyptus camaldulensis_ from the Petford Region, Queensl;and; Evidence of outbreeding depression. _Silvae Genetica_ 51: 6–12.


Hardy, O. and X. Vekemans (2002): SPAgeDi: a versatile computer program to analyze spatial genetic structure at the individual or population levels. _Molecular Ecology_ 2: 618–620.


