CrossMark

# Are tree breeders properly predicting genetic gain? A case study involving *Corymbia* species

**Evandro Vagner Tambarussi** ⓘ · **Fernanda Bortolanza Pereira** ⓘ ·
**Paulo Henrique Müller da Silva** ⓘ · **David Lee** ⓘ · **David Bush** ⓘ

**Abstract** The estimation of quantitative genetic parameters in breeding programs is important to ensure efficient selection. In this context, knowledge of the mating system is critical, as it underpins assumptions about inter-relatedness on which variance component estimation depends. However, proper account of the breeding system is not always taken, either because it is unknown and/or because it is ignored. That eucalypts have a mixed-mating system is well-established, however many breeders use models that assume outcrossed mating with an infinite number of male parents (i.e. allogamous mating), from which genetic parameter estimates are then used to predict genetic gains. First-generation, open-pollinated progeny tests of *Corymbia citriodora* subsp. *citriodora* and *C. citriodora* subsp. *variegata*, being managed for seed production, were used to investigate the likely bias, resulting in overestimation of genetic parameters under the assumption of allogamous mating. When we assumed allogamous mating, we observed inflated predictions of additive variance and narrow-sense heritability ($\hat{h}_a^2$) of diameter at breast height and height. The overestimate of $\hat{h}_a^2$ was approximately 32% for *C. citriodora* subsp. *citriodora* and 21% for *C. citriodora* subsp. *variegata*. Inappropriate modelling of relatedness in eucalypts that assumes panmictic outcrossing when in fact these species have a mixed-mating system results in overestimates of the population genetic gain with selection.

E. V. Tambarussi · F. B. Pereira
Programa de Pós-Graduação em Ciência Florestal,
Faculdade de Ciências Agronômicas, UNESP –
Universidade Estadual Paulista, Rua José Barbosa de
Barros, 1780, Portaria II: Rodovia Alcides Soares, Km 3,
Botucatu, SP 18610-307, Brazil
e-mail: tambarussi@gmail.com

F. B. Pereira
e-mail: fbp.engflorestal@gmail.com

P. H. M. da Silva
Instituto de Pesquisas e Estudos Florestais (IPEF),
Avenida Pádua Dias 11, Caixa Postal 530, Piracicaba,
SP CEP 13400-970, Brazil
e-mail: paulohenrique@ipef.br

D. Lee
University of the Sunshine Coast,
Locked Bag 4, Maroochydore DC, QLD 4558, Australia
e-mail: dlee@usc.edu.au

D. Bush (✉)
CSIRO Australian Tree Seed Centre, Clunies Ross Street,
Acton, ACT 2601, Australia
e-mail: david.bush@csiro.au

🖄 Springer

## Introduction

Evaluation of breeding programs based on predictions of genetic and phenotypic parameters obtained by studies in quantitative genetics is important both for conservation and to assure efficient selection, especially for perennial plants (Pires and Resende 2011; Bernardo 2010). Studies on many species have been carried out assuming that they have either a predominantly autogamous reproductive system, whose outcrossing rate ($t$) ranges from 0.00 to 0.05, or an allogamous one, with $0.95 \leq t < 1.00$ (Fuchs et al. 2015). However, rates of $t$ in eucalypt species (*Eucalyptus* and *Corymbia*) range widely, for example Eldridge et al. (1993) document rates between 0.45 and 0.96. The assumption that the eucalypt reproductive system is predominantly allogamous is therefore inappropriate.

Species with a mixed-mating system are problematic for geneticists and plant breeders, as modelling relationships within breeding populations and making selections is complicated (Namkoong 1966; Vencovsky et al. 2001). Considering species as either completely autogamous or allogamous when they in fact have a mixed-mating system will lead to errors in prediction of breeding values, and consequently, in the selection of superior genotypes.

To avoid overestimation of genetic parameters such as heritability, genetic variances (i.e. $\sigma_A^2$ = additive variance, $\sigma_D^2$ = dominance variance) and genetic gains in breeding programs, the genetic variance of progeny ($\sigma_p^2$) or covariance of any two relatives ($\varsigma_{XY}$) based on a test of open-pollinated families should be estimated (Cockerham and Weir 1984) as:

$$
\begin{aligned}
\hat{\sigma}_p^2 = {} & 2\theta_{XY}\sigma_A^2 + 2(\Delta_{\ddot{X}+\ddot{Y}} - \delta_{\ddot{X}\ddot{Y}})\sigma_D^2 + 2(\gamma_{\ddot{X}Y} + \gamma_{X\ddot{Y}})D_1 \\
& + \delta_{\ddot{X}\ddot{Y}}D_2^* + (\Delta_{\ddot{X}\cdot\ddot{Y}} - F_X F_Y)H^* \\
& + (\tilde{\Delta}_{XY} - F_X F_Y)(H^2 - H^*)
\end{aligned} \tag{1}
$$

where $F_X$, $F_Y$, $\theta_{XY}$, $\Delta_{\ddot{X}+\ddot{Y}}$, $\Delta_{\ddot{X}-\ddot{Y}}$, $\delta_{\ddot{X}\ddot{Y}}$, $\gamma_{\ddot{X}Y}$, $\gamma_{X\ddot{Y}}$ measures the identity for a locus for individual $X$ with alleles $x_1$ and $x_2$, and individual Y with alleles

$y_1$ and $y_2$ (see Cockerham and Weir 1984), $D_1$ is the covariance between the additive effects and homozygous dominance effects, $D_2^*$ is the total variance of homozygous dominance effects, $\tilde{\Delta}_{XY}$ is the probability for two individuals $X$ and $Y$ that two genes at the *ith* locus of one and the same two genes at the *jth* locus of the other are identical by descent, and $H$–$H^*$ are the correlation of inbreeding at different loci within individuals.

The assumption that $\sigma_p^2 = 1/4\sigma_A^2$, which is used in the case of true half-sib seedling families ($t = 1$) is a misleading over-simplification. The traditional approach to dealing with the mixed-mating system in open-pollinated progeny tests of forest trees has been to modify estimates of additive genetic variance by assuming a coefficient of relationship ($\rho$) of less than 1/4 (e.g. Squillace 1974). The coefficient of relationship is a measure of the proportion of genes that two individuals have inherited directly from a common ancestor (Wright 1922). This is to account for an assumed proportion of selfed and other inbred individuals. The coefficient of relationship between two individuals is given by:

$$
\rho_{XY} = \frac{2\theta_{XY}}{\sqrt{(1 + F_X)(1 + F_Y)}} \tag{2}
$$

where $F_X$ and $F_Y$ are their inbreeding coefficients and $\theta_{XY}$ is their coefficient of co-ancestry.

The coefficient of relationship can then be used to appropriately scale the estimate of additive variance ($\sigma_A^2$) in the calculation of narrow sense heritability $\hat{h}_a^2$ thus:

$$
\hat{h}_a^2 = \frac{(1/\rho)\hat{\sigma}_A^2}{\hat{\sigma}_P^2} \tag{3}
$$

where $\sigma_P^2$ is the estimated phenotypic variance component. Eldridge et al. (1993), reiterated by Bush et al. (2011), recommended a value of $\rho = 1/2.5$ as appropriate for eucalypts, though values between 1/4 and 1/1.85 have been used in various eucalypt studies (Hodge et al. 1996). However, a recent trend has seen several eucalypt studies published that simply assume that $\rho = 1/4$ (e.g. Apiolaza et al. 2011; Cane-Retamales et al. 2011; Denis et al. 2013; Henson et al. 2008; Mora et al. 2009; Pelletier et al. 2008; Raymond et al. 2008; Rojas et al. 2017; Vargas-Reeve et al. 2013), noting that some of these studies have acknowledged that their heritability estimates may

be upwardly biased as a result. A possible justification for setting $\rho = 1/4$, even when a mixed-mating system is likely or known to be in operation, is that actual the degree of inbreeding is unknown.

In the case of the ($\varsigma_{XY}$ or $\sigma_p^2$) expression proposed by Cockerham and Weir (1984) the level of kinship ($F_X, F_Y, \theta_{XY}, \Delta_{\ddot{X}+\ddot{y}}, \delta_{\ddot{X}\ddot{Y}}, \gamma_{\ddot{X}Y}, \gamma_{X\ddot{Y}}, \Delta_{\ddot{X}\ddot{Y}}, \tilde{\Delta}_{XY}$ and $\delta_{\ddot{X}\ddot{Y}}$) is increased by the mixture of different genetic constitutions of individuals. However, the dominance components, in particular, cannot be predicted in open-pollinated progeny tests. Thus, to simplify the expression we have: $\hat{\sigma}_p^2 = 2\frac{1}{4}(1+\hat{F})\hat{\sigma}_A^2 + \hat{V} = \frac{(1+\hat{F})\hat{\sigma}_A^2}{2} + \hat{V}$, we would have an approximate estimate of $\sigma_A^2$ where $\hat{V}$ is a bias [showed in Eq. (1)], that is not easily estimated in tree species. Thus, the effect of selection (genetic gain) is complex in this type of population.

The use of molecular markers to provide information on forest-tree breeding systems is now becoming more prevalent, and the opportunity to provide a better-informed estimate of the degree of selfing for a breeding population is now possible. In the case of *Corymbia* species, a number of marker-based determinations of outcrossing have been made. Selfing rates for *Corymbia citriodora* subsp. *citriodora* are estimated as $s = 0.15$ based on an isozyme markers (Yeh et al. 1983), and for *C. citriodora* subsp. *variegata* as $s = 0.10$ using microsatellite markers (Bacles et al. 2009).

Commercial eucalypt plantations are important around the world, however only a few species and their hybrids are widely planted, especially species that belong to the genus *Eucalyptus* L'Hér. Within *Eucalyptus* the subgenus *Symphyomyrtus* Schauer in J.G.C. Lehmann contributes the most widely-planted species (Harwood 2011; Potts and Dungey 2004). However, *Corymbia* K.D. Hill & L.A.S. Johnson species, particularly *C. citriodora* (Hook.) K.D. Hill & L.A.S. Johnson and hybrids with *C. torelliana* (F. Muell.) K.D. Hill & L.A.S. Johnson, have shown promise in tropical and subtropical regions (Lee 2007; Lee et al. 2009).

The *Corymbia* genus includes the spotted gum complex of species that is distributed naturally on the east coast of Australia. The complex has a distribution from tropical north Queensland (*C. citriodora* subsp. *citriodora*) through subtropical latitudes in southern Queensland and northern NSW (*C. citriodora* subsp.

*variegata* and *C. henryi*) to temperate Victoria in south-east Australia (*C. maculata*). In Brazil, *Corymbia* taxa are of interest for planting in regions that may become unsuitable for *Eucalyptus* species in future due to abiotic and biotic stress driven by climate change. In addition, pest and disease incidences are also expected to increase, due to the expansion of large areas of genetically- and silviculturally-homogeneous forest plantations (Wingfield et al. 2008). The *Corymbia* spp. have shown differences in resistance and tolerance to heat, water stress and pest and disease resistance relative to *Eucalyptus* species (Brawner et al. 2011; Dianese et al. 1986). These differences may be used to advantage when *Corymbia* species and hybrids are planted as components of a more-diverse mix of plantation species.

The area planted to spotted gums has recently expanded in many countries of the world, especially sub-tropical and tropical plantation regions, due to their edaphic and climatic adaptation, potential for fast growth and useful wood properties (e.g. Lee 2007; Gardner et al. 2007; Morais et al. 2010; Brawner et al. 2012; Lin et al. 2017). Pure-species breeding programs of *C. citriodora* are not being progressed in Brazil, though interest in *Corymbia* hybridization for environmental-stress-resilient commercial plantations (Lee et al. 2009) with improved wood quality and resistance to pests and diseases (Segura and Silva Jr. 2016) is now emerging. Brazil is also one of the largest producers of *C. citriodora* essential oil in the world (Reis et al. 2013), and this may provide further impetus for renewed interest in genetic improvement.

This study aimed to answer the follow questions regarding two *Corymbia* spp.: (1) is there enough genetic variability in silvicultural traits in open-pollinated families, evaluated at ages 18 and 36 months, to provide a sufficiently broad base populations?; (2) how does the marker-based estimate of inbreeding ($s$) influence the estimated genetic parameters at these ages?

## Materials and methods

### Experimental location and design

In August 2013, two seedling-based *Corymbia* spp. progeny trials were established using open-pollinated, seed-orchard seedlots. For *C. citriodora* subsp.

*citriodora* (CCC), family seedlots were sourced from plus trees located at Ouricangas, Bahia State, (ex Zimbabwe), while *C. citriodora* subsp. *variegata* (CCV) originated from Anhembi, São Paulo State (based on wild families from Woondum and Wondai—CSIRO seedlots 14426 and 14434). The commercial controls (checks) used in the trials were seedlots from Anhembi (ex Lat 15°00′S to 17°00′S), Barra Bonita (ex. Rio Claro) and Bauru (ex Rio Claro). The trials were established in the "Estação Experimental de Ciências Florestais" owned by Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo (ESALQ/USP), located in Itatinga Municipality, São Paulo State, (23°03′S, 48°22′W) at 840 m altitude and with climate classified as Cwa under the Koeppen Climate Classification (CEPAGRI 2018). The experimental designs were randomized complete block with 48 and 32 families of CCC and CCV, respectively, with three-tree linear plots and 8 or 10 replicates per trial (Table 1).

Measurement and mating system assumptions

In July 2016, 36 months after planting, diameter at breast height (DBH) and height (H) of the individual trees were evaluated. The estimates of genetic parameters were obtained using different selfing rates ($s$), given by $s = 1 - t$, first considering $s = 0$, then considering $s = 0.15$ for CCC based on an isozyme marker estimate (Yeh et al. 1983) and $s = 0.10$ for CCV based on a microsatellite marker estimate (Bacles et al. 2009).

Linear model and genetic parameter estimation

For this analysis, we used a linear mixed model of the form:

$$y = \mathbf{X}r + \mathbf{Z}a + \mathbf{W}b + e \tag{4}$$

where $y$ is the phenotypic data vector for a trait; $\mathbf{X}$ is the incidence matrix of fixed effects; $r$ is the fixed effects vector (general mean and complete-block *replicate* effects); $\mathbf{Z}$ is the incidence matrix of genetic random effects; $a$ is the vector of *family* genetic random effects; $\mathbf{W}$ is the incidence matrix of *plot* effects; $b$ is the vector of *plot* effects (random) and $e$ is the vector of residuals (random).

We fitted a univariate normal plot function that uses both the numerical Shapiro–Wilks test and graphical methods to test for normality of residuals, using R statistical language (R Core Team 2018).

We used the Restricted Maximum Likelihood procedure (REML) to estimate the variance components from Eq. (1): genetic variance among families ($\hat{\sigma}_p^2$), environmental variance among plots ($\sigma_{plot}^2$) and residual variance ($\hat{\sigma}_e^2$). Phenotypic individual variance ($\hat{\sigma}_P^2$) is given as

$$\hat{\sigma}_P^2 = \hat{\sigma}_p^2 + \hat{\sigma}_{plot}^2 + \hat{\sigma}_e^2 \tag{5}$$

We estimated narrow-sense heritability ($\hat{h}_a^2$), the determination coefficient of plot effects ($\hat{c}_{plot}^2$) and selection accuracy ($Ac_{prog}$). Narrow-sense heritability was estimated using Eq. (3) with $\rho = 1/4$ corresponding to true half-sib relationships among open-pollinated family members.

We estimated the coefficient of experimental variation ($\hat{C}V_{exp}(\%)$) and the genetic variation coefficient ($\hat{C}V_g\%$), using the expressions $\hat{C}V_{exp}(\%) = \left(\sqrt{\hat{\sigma}_e^2}/m\right) \times 100$ and $\hat{C}V_g(\%) = \left(\sqrt{\hat{\sigma}_p^2}/m\right) \times 100$, where $m$ is the general mean of the original population. For all these estimates we considered the species as perfectly allogamous ($s = 0$).

To examine the effects of mixed mating on heritability estimates, $h_a^2$ was again estimated with values of $\rho$ updated to reflect the published marker-

Table 1 Number of families, replicates and total trees evaluated per species in the experimental trials at Ouricangas, Bahia State Brazil

| Species | Total number of genetic entries [a] | Number of families | Number of blocks | Total trees |
|---|---|---|---|---|
| *Corymbia citriodora* subsp. *citriodora* | 50 | 48 | 8 | 1152 |
| *Corymbia citriodora* subsp. *variegata* | 34 | 32 | 10 | 930 |

[a]Sum of families and control treatments

based selfing rates for CCC and CCV. From Resende et al. (1995), the coefficient used to correct the value of variance among families considering the mixed-mating system was

$$\frac{(1 + \hat{s})^2}{2(2 - \hat{s})} \qquad (6)$$

that is derived following Wright and Cockerham (1986). Thus, for a selfing rate of $s = 0.1$ applicable to CCV this corresponds to $\rho = 1/3.3$; and for a selfing rate of $s = 0.15$, applicable to CCC, this corresponds to $\rho = 1/3.0$. It also corresponds to Squillace (1974) correction for heritability, that is given in this situation by $\hat{h}_a^2 = \frac{\hat{\sigma}_p^2}{r_{OO} \times \hat{\sigma}_P^2}$ when parents are uncorrelated (artificial stands), and considering the proportion of effective natural selfing and correlation among the families $(r_{OO})$. In our case, for $s = 0.10$ and $s = 0.15$, the correlations among families were $r_{OO} = 0.30$ and $r_{OO} = 0.33$, respectively and correspond to the $\rho$ values previously stated.

All computations were implemented with model 110 of the SELEGEN-REML/BLUP Software (Resende 2007). This model has in-built functionality that allows the user to apply various estimates of $s$ and automatically applies Eq. (6).

The significance of the family variance component $(\hat{\sigma}_p^2)$ was tested by comparison with a separate model, where the term had been omitted, using the likelihood ratio test (see for example Resende 2007; Gilmour et al. 2009) implemented using SELEGEN-REML/BLUP Software Model 5 (Resende 2007).

## Results and discussion

### Genetic parameter estimates

The residuals were normally distributed, and we found significant family effects $(\hat{\sigma}_p^2)$ for all traits (Table 2), indicating genetic variability among the families studied, making selection and breeding for growth a viable option (Namkoong 1966; Zimback et al. 2011). Other authors have also observed significant variability for growth and silvicultural traits in *C. citriodora* at ages between 13 months and 25 years (Morais et al. 2010; Brawner et al. 2011).

The high $\hat{C}V_{exp}(\%)$ values that we observed may be due to within-trial environmental variance not accounted for by the experimental design (Pimentel-Gomes and Garcia 2002), which is reflected by the large proportion of $\hat{\sigma}_e^2$ to $\hat{\sigma}_P^2$ (Table 3), greater than 76% for all traits. The $\hat{C}V_{exp}(\%)$ values were classified as high (ranging from 24.2 to 30.2%) for DBH and as high (ranging from 16.8 to 21.2%) to very-high (> 21.2%) (Mora and Arriagada 2016). However, $\hat{C}V_{exp}(\%)$ with high values are not unusual in progeny testing of eucalypt species (Moraes et al. 2007; Berti et al. 2011; Costa et al. 2015).

In contrast to the $\hat{C}V_{exp}(\%)$ classification, the estimated determination coefficient of plot effects $(\hat{c}_{plot}^2)$ was considered low (Table 3), as gauged by the relation $c_{plot}^2 \big/ h_a^2 \leq 1/3$, which indicates relatively little environmental variation among plots, within blocks, and low environmental correlation among observations within plots (Resende and Sturion 2003). These results mean that we had high experimental precision and low environmental variability within plots, contributing to good accuracy of genetic parameters estimates (Resende 2002; Pagliarini et al. 2016).

The $\hat{C}V_g(\%)$ estimates were much higher than what has been reported in the literature for open-pollinated families of other eucalypt species including *Eucalyptus tereticornis* Sm. at age 25 years (Macedo et al. 2013), *Corymbia maculata* (Hook.) K.D. Hill & L.A.S. Johnson at age 4 years (Sato et al. 2010) and *E. camaldulensis* at age 19 years (Moraes et al. 2007), which ranged between 2.8 and 5.39% for DBH and H, whereas we observed values between 6.79 and 11.92%. We observed high values of $Ac_{prog}$ for DBH and H at both ages $(0.70 \leq Ac_{prog} < 0.90)$ (Resende and Duarte 2007). These results allowed us to infer high precision in genetic variation of observed phenotypic traits. Indeed, $Ac_{prog}$ refers to the correlation between predicted genetic values and the true genetic values for each tree (Resende 2002; Moraes et al. 2007; Costa et al. 2015).

The $\hat{h}_a^2$ estimates, related to the component of inheritance that is effectively transmitted to the next generation (Falconer and Mackay 1996; Bernardo 2010), were classified as moderate $(0.15 < \hat{h}_a^2 \leq 0.50)$ to high $(0.50 < \hat{h}_a^2)$ (Resende 2002) for all the traits

**Table 2** Deviance analysis and likelihood ratio tests for families (LRT$_{Prog}$), and control seedlots (LRT$_{GM}$) for each trait at ages 18 and 36 months

| Species | Deviance[b] | Traits[a] | | | |
|---|---|---|---|---|---|
| | | DBH$_{18}$ | H$_{18}$ | DBH$_{36}$ | H$_{36}$ |
| *C. citriodora* subsp. *citriodora* | LRT$_{Prog}$ | 35.68** | 73.02** | 47.36** | 33.81** |
| | LRT$_{GM}$ | 0.08[ns] | 3.47[ns] | 0.01[ns] | 0.34[ns] |
| *C. citriodora* subsp. *variegata* | LRT$_{Prog}$ | 46.10** | 67.16** | 73.51** | 73.77** |
| | LRT$_{GM}$ | 0.22[ns] | 2.91[ns] | 0.02[ns] | 0.14[ns] |

[a]DBH$_{18}$ = diameter at breast height at 18 months old (cm); H$_{18}$ = height at 18 months old (m); DBH$_{36}$ = diameter at breast height at 36 months old (cm); H$_{36}$ = height at 36 months old (m)

[b]ns = not significant; * = significant to 5% of error probability; ** = significant to 1% of error probability

**Table 3** Estimates of variance components and variation coefficients for *Corymbia* spp. traits at 18 and 36 months old, assuming open-pollinated families comprise entirely of half-sib, within-family relationships

| Parameter estimate[a] | Traits[b] | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | *C. citriodora* subsp. *citriodora* | | | | *C. citriodora* subsp. *variegata* | | | |
| | DBH$_{18}$ | H$_{18}$ | DBH$_{36}$ | H$_{36}$ | DBH$_{18}$ | H$_{18}$ | DBH$_{36}$ | H$_{36}$ |
| $\hat{\sigma}_p^2$ | 0.06 | 0.15 | 0.52 | 0.44 | 0.11 | 0.13 | 0.42 | 0.31 |
| $\hat{\sigma}_{plot}^2$ | 0.01 | 0.05 | 0.02 | 0.09 | 0.01 | 0.03 | 0.02 | 0.02 |
| $\hat{\sigma}_e^2$ | 0.61 | 0.69 | 4.83 | 4.47 | 0.74 | 0.52 | 2.45 | 1.52 |
| $\hat{\sigma}_P^2$ | 0.67 | 0.89 | 5.38 | 5.01 | 0.87 | 0.68 | 2.88 | 1.84 |
| $\hat{h}_a^2(SE)$ | 0.35 (0.10) | 0.68 (0.14) | 0.39 (0.11) | 0.35 (0.10) | 0.52 (0.13) | 0.74 (0.16) | 0.58 (0.14) | 0.67 (0.15) |
| $\hat{c}_{plot}^2$ | 0.01 | 0.05 | 0.00 | 0.02 | 0.01 | 0.05 | 0.01 | 0.01 |
| $Ac_{prog}$ | 0.83 | 0.90 | 0.85 | 0.83 | 0.90 | 0.93 | 0.91 | 0.92 |
| $\hat{C}V_g(\%)$ | 7.93 | 9.49 | 9.67 | 6.79 | 10.27 | 8.43 | 11.92 | 7.99 |
| $\hat{C}V_{exp}(\%)$ | 25.48 | 20.19 | 29.49 | 21.65 | 26.43 | 17.12 | 28.86 | 17.65 |
| $\bar{y}_p$ | 3.05 | 4.12 | 7.46 | 9.77 | 3.26 | 4.20 | 5.42 | 6.98 |

[a]$\hat{\sigma}_p^2$: genetic variance among families; $\hat{\sigma}_{plot}^2$: environmental variance among plots; $\hat{\sigma}_e^2$: residual variance (environmental + not additive); $\hat{\sigma}_P^2$: phenotypic individual variance; $\hat{h}_a^2(SE)$: heritability in the narrow sense (standard errors of heritability estimates); $\hat{c}_{plot}^2$: determination coefficient of plot effects; $Ac_{prog}$: selective accuracy; $\hat{C}V_g(\%)$: coefficient of genetic variation; $\hat{C}V_{exp}(\%)$: coefficient of experimental variation; $\bar{y}_p$: general mean

[b]DBH$_{18}$ = diameter at breast height at 18 months old (cm); H$_{18}$ = height at 18 months old (m); DBH$_{36}$ = diameter at breast height at 36 months old (cm); H$_{36}$ = height at 36 months old (m)

evaluated (Table 3) with estimates for CCV higher than those for CCC. Hung et al. (2016) observed estimates of $\hat{h}_a^2$ for *CCV* at age 3 years of 0.34 ± 0.03 and 0.42 ± 0.04 for DBH and H, respectively, which was similar to what we found in *CCC* and lower than what we observed for the same species. Though Hung et al. (2016) postulated that their results may be biased due to the use of one-third rather than 1/2.5 for the

coefficient of relationship among families and the heavy thinning in some trials, their assumption is similar to our assumption of 1/3.3. Moreover, is probable that our own estimates of additive variance are inflated, due to (1) confounding of genotype-by-environment interaction (GxE) variance with additive variance, due to the single-site trials that do not allow estimation of GxE and (2) lack of pedigree

information including provenance of ancestral origin information for the families. Although these trials are based on later-generation selections, persistent provenance effects (that include non-additive genetic variation) have been observed in other, later-generation eucalypt trials (Swain et al. 2015).

### Genetic gains

The importance of considering the mixed reproductive system to obtain genetic estimates and predicting genetic gains with the selection of the best individuals is evidenced by the results that we observe in the present study.

When we use $\hat{t} = 0.85$ and $\hat{t} = 0.90$ for CCC and CCV, respectively, the estimates of $\hat{h}_a^2$ at both ages were reduced (Table 4) relative to the estimates from the half-sib model. Open-pollinated families of plants with mixed-mating system will comprise at least three types of relative (selfs, half-sibs and full-sibs) (Namkoong 1966), and more if near-relative inbreeding and ancestral inbreeding are considered. This can result in overestimates of heritabilities and, consequently, inflation of predicted genetic gain if we consider that all within-family relationships are half-sibs, as has been done by some forest tree breeders. Inbreeding depression and dominance effects are also likely to be complicating factors associated with mixed mating.

The inattention of breeders and researchers to the importance of the reproductive system in estimates of genetic parameters and gain by selection is easily observed with only a few studies (e.g. Zanata et al. 2010; Berti et al. 2011; Macedo et al. 2013; Bush et al. 2011) dealing with the effect of a mixed breeding system on the prediction of genetic gain. In the present research, the overestimates of $\hat{h}_a^2$ range from 20.93 to 21.01% and 32.22 to 32.25% for CCV and CCC, respectively (Table 4). These values are lower than those estimated ($\hat{h}_a^2 = 46.5\%$) for DBH and H in *Eucalyptus pellita* (F. Muell.) families at age 23 years (Zanata et al. 2010). These authors used a mean $\hat{t} = 0.557$ obtained from three populations of *E. pellita* (House and Bell 1996). Similarly, Berti et al. (2011) obtained overestimates of $\hat{h}_a^2$ between 40.38 and 64.31% for 24-year-old *Eucalyptus cloeziana* (F. Muell.), using a mean $\hat{t} = 0.754$ obtained from 18 species of *Eucalyptus*.

The predicted genetic gain by selection, in percentage, is given by $Gs\% = \left\{ \left[ (\bar{y}_s - \bar{y}_p) \times \hat{h}_a^2 \right] / \bar{y}_p \right\} \times 100$, where $\bar{y}_s$ is the selected families average and $\bar{y}_p$ is the families general average. Then, the lower estimates of $\hat{h}_a^2$ when we consider the mixed reproductive system leads to lower values of genetic gain ($Gs\%$), and these underestimates will be greater than the observed to $\hat{h}_a^2$, since the mean of the selected population tends to be reduced too. Goodwillie et al. (2005) presented updated evidence suggesting that mixed reproductive systems are more frequent than previously thought, showing the importance of studying the evolutionary and biometric scope for these species.

**Table 4** Estimates of narrow-sense heritability ($\hat{h}_a^2$) considering different outcrossing rates (*t*) for *Corymbia* spp. traits at ages 18 and 36 months

| Parameter estimate[a] | Traits[b] | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | *C. citriodora* subsp. *citriodora* | | | | *C. citriodora* subsp. *variegata* | | | |
| | $DBH_{18}$ | $H_{18}$ | $DBH_{36}$ | $H_{36}$ | $DBH_{18}$ | $H_{18}$ | $DBH_{36}$ | $H_{36}$ |
| $\hat{h}_a^2(SE)$ (HS) | 0.35 (0.10) | 0.68 (0.14) | 0.39 (0.11) | 0.35 (0.10) | 0.52 (0.13) | 0.74 (0.16) | 0.58 (0.14) | 0.67 (0.15) |
| $\hat{h}_a^2(SE)$ (MMS) | 0.26 (0.08) | 0.52 (0.11) | 0.29 (0.08) | 0.27 (0.08) | 0.43 (0.11) | 0.61 (0.13) | 0.48 (0.11) | 0.56 (0.12) |
| Overestimate (%) | 32.25 | 32.24 | 32.24 | 32.22 | 20.93 | 21.01 | 21.01 | 21.01 |

[a] $\hat{h}_a^2(SE)$: heritability in the narrow sense (standard errors of heritability estimates); HS = estimates considering half-sibs; MMS = estimates considering mixed-mating system

[b] $DBH_{18}$ = diameter at breast height at age 18 months (cm); $H_{18}$ = height at age 18 months (m); $DBH_{36}$ = diameter at breast height at age 36 months (cm); $H_{36}$ = height at age 36 months (m)

Though we have used molecular marker-based estimates of outcrossing to better estimate the coefficient of relationship in this study, there are a number of complicating factors in connection with the eucalypt mixed-mating system that should also be considered. Firstly, the estimation of outcrossing rates from seedling populations may lead to an overestimation of $s$ in the adult population of trees. Selection against inbred individuals is normally very high in eucalypts, such that in some species, very few inbred individuals survive to reproductive maturity (e.g. Costa e Silva et al. 2010; Griffin and Cotterill 1988; Hardner and Potts 1997), though this does not appear to be the case for all eucalypt populations (Bush and Thumma 2013). A further, often-overlooked complication is the likelihood of full-sib progeny within families (i.e. individuals that share a common pollen parent). The assumption of panmixia in eucalypt seed orchards is unrealistic, especially in the quite-common case that a few, unrelated individuals flower synchronously with each other but not with the majority of other genotypes in the stand. As full-sibs are more-closely related than are half-sibs, but are not inbred, they are likely to survive to reproductive maturity. Their presence is a reason to choose a coefficient of relationship reflecting slightly greater average relatedness than would be chosen on the basis of correcting for inbreeding alone. In addition to these considerations, the issue of non-homogenous rates of inbreeding and relatedness among families and provenances can be significant (Bush et al. 2015), especially where breeding populations contain a mix of wild (first generation), landrace and later-generation materials.

In addition, breeders of these species may be misrepresenting the number of trees ($m$) required to retain a suitable effective population size ($N_e$). If breeders do not properly model the relationship between progeny within a species with a mixed-mating system, they neglect the intraclass correlation that will exist between these individuals. Indeed, mixed-mating populations differ from allogamous and autogamous ones because they are likely to be composed of a mixture of individuals with different degrees of inbreeding. Inbreeding influences individual phenotypic values and genetic variances. As an example, if we consider an open-pollinated population of a species without intraclass correlation and that the species reproduces by panmictic outcrossing, the value of coancestry ($\theta_{xy}$) will be 0.125, inbreeding ($f$) will be 0 and $m$ can be estimated in a straightforward manner. However, if there is 10% selfing, one in ten trees will have a self-half-sib relationship with the other nine, and on average, each family will have a higher intraclass correlation than expected in a panmictic, allogamous population. To retain the same effective population size, the breeder needs to maintain a larger number of trees, in a sample (n) taken from a large population in equilibrium $N_e = \frac{n}{1+F} = n(1 - 0.5s)$, where n is the number of generations and $F$ is the fixation index (inbreeding coefficient).

## Concluding remarks and future considerations for eucalypt breeding

The study populations exhibited moderate genetic variability, with potential to become breeding base populations. Estimation of variance components and prediction of genetic gain under the assumption of complete outcrossing in a species which has a mixed-mating system results in overprediction of genetic gain, leading to significant errors in the selection of superior trees and/or families.

Our results show that when we model eucalypt species with complete allogamy there is some upward bias in additive genetic variance estimates and therefore of genetic parameters that depend on this quantity. Thus, seeds for this species can be produced with deviations from random mating, with open-pollinated seedling families presenting various levels of relatedness.

With the recent emergence of next-generation molecular techniques, studying the breeding system and making estimates of relatedness among individual trees has become faster and cheaper. This allows a much more accurate characterisation of a species' mating system, in which case there is no justification for using an approximation that leads to error in selection of superior genotypes, as illustrated by our results. Though not routinely carried out, a number of studies of eucalypt genetic parameters informed by marker-based estimates of relatedness and population structure have been made in recent years, either based on marker data taken directly from the breeding populations (e.g. Bush et al. 2015; Cappa et al. 2016; Klápště et al. 2017), or like in the present study, making use of previously determined marker-based

estimates of outcrossing (Varghese et al. 2017). The advent of high-density markers that can be obtained through genotyping-by-sequencing or "SNP-chip" technologies has opened the door for detailed estimation of inter-tree relatedness. In some cases, it may even be possible to estimate dominance variance in populations where sufficient full-sib relationships exist (e.g. Doerksen and Herbinger 2010; Klápště et al. 2017), opening the door for significantly more-realistic estimation of quantitative genetic parameters in species with mixed-mating systems.

## References

Apiolaza L, McConnochie R, Millen P, Van Ballekom S, Walker J (2011) Introducing durable species to New Zealand drylands: genetics of early adaptation of *Eucalyptus bosistoana*. In: Walker J (ed) Developing a eucalypt resource: learning from Australia and elsewhere. University of Canterbury, Christchurch, pp 137–146

Bacles CFE, Brooks J, Lee DJ, Schenk PM, Lowe AJ, Kremer A (2009) Reproductive biology of *Corymbia citriodora* subsp. *variegata* and effective pollination across its native range in Queensland, Australia. South For 71(2):125–132

Bernardo R (2010) Breeding for quantitative traits in plants. Stemma Press, Woodbury

Berti CLF, Freitas MLM, Zanatto ACS, Morais E, Moraes MLT, Sebbenn AM (2011) Variação genética, herdabilidades e ganhos na seleção para caracteres de crescimento e forma em teste de progênies de polinização aberta de *Eucalyptus cloeziana*. Rev Inst Flor 23(1):13–26

Brawner JT, Lee DJ, Hardner CM, Dieters MJ (2011) Relationships between early growth and *Quambalaria* shoot blight tolerance in *Corymbia citriodora* progeny trials established in Queensland, Australia. Tree Genet Genomes 7:759–772

Brawner JT, Meder R, Dieters M, Lee DJ (2012) Selection of *Corymbia citriodora* for pulp productivity. South For 74:121–131

Bush D, Thumma B (2013) Characterising a *Eucalyptus cladocalyx* breeding population using SNP markers. Tree Genet Genomes 9:741–752

Bush D, Kain D, Matheson C, Kanowski P (2011) Marker-based adjustment of additive relationship matrix for estimation of genetic parameters—an example using *Eucalyptus cladocalyx*. Tree Genet Genomes. 7(1):23–35

Bush D, Kain D, Kanowski PJ, Matheson AC (2015) Genetic parameter estimates informed by a marker-based pedigree: a case study with *Eucalyptus cladocalyx* in southern Australia. Tree Genet Genomes 11:798

Cane-Retamales C, Mora F, Vargas-Reeve F, Perret S, Contreras-Soto R (2011) Bayesian threshold analysis of breeding values, genetic correlation and heritability of flowering intensity in *Eucalyptus cladocalyx* under arid conditions. Euphytica 178:177–183

Cappa EP, Klápště J, Garcia MN, Villalba PV, Marcucci Poltri SN (2016) SSRs, SNPs and DArTs comparison on estimation of relatedness and genetic parameters' precision from a small half-sib sample population of *Eucalyptus grandis*. Mol Breed 36:97

CEPAGRI (2018) Clima dos municípios paulistas – Itatinga [Internet]. Centro De Pesquisas Meteorológicas E Climáticas Aplicadas à Agricultura. http://www.cpa.unicamp.br/outras-informacoes/clima_muni_271.html. Accessed 18 Jan 2018

Cockerham CC, Weir BS (1984) Covariances o relatives stemming from a population undergoing mixed self and random mating. Biometrics 40:157–164

Costa e Silva J, Hardner C, Tilyard P, Pires AM, Potts BM (2010) Effects of inbreeding on population mean performance and observational variances in *Eucalyptus globulus*. Ann For Sci 67:605

Costa RB, Marínez DT, Silva JC, Almeida BC (2015) Variabilidade e ganhos genéticos com diferentes métodos de seleção em progênies de *Eucalyptus camaldulensis*. Rev Cien Agra 58(1):69–74

Denis M, Favreau B, Ueno S, Camus-Kulandaivelu L, Chaix G, Gion JM, Nourrisier-Mountou S, Polidori J, Bouvet JM (2013) Genetic variation of wood chemical traits and association with underlying genes in *Eucalyptus urophylla*. Tree Genet Genomes 9:927–942

Dianese JC, Haridasan M, Moraes TSA (1986) Screening *Eucalyptus* species for rust resistance in Bahia, Brazil. Trop Pest Manag 32:292–295

Doerksen TK, Herbinger CM (2010) Impact of reconstructed pedigrees on progeny-test breeding values in red spruce. Tree Genet Genomes 6:591–600

Eldridge K, Davidson J, Harwood C, van Wyk G (1993) Eucalypt domestication and breeding. Claredon Press, Oxford

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics. Pearson Prentice Hall, Suffolk

Fuchs MCP, Tambarussi EV, Lurenção JC, Nogueira LM, Bortoloto TM, Gonzalez ER, Oda S, Marino CL (2015) Molecular marker associated with a deleterious recessive anomaly in *Eucalyptus grandis* seedlings. Ann For Sci 72:1043–1052

Gardner RAW, Little KM, Arbuthnot A (2007) Wood and fibre productivity of promising new eucalypt species for coastal Zululand, South Africa. Aust For 70(1):37–47

Gilmour AR, Gogel BJ, Cullis BR, Thompson R (2009) ASReml user guide: release 3. VSN International Ltd, Hemel Hempstead

Goodwillie C, Kalisz S, Eckert CG (2005) The evolutionary enigma of mixed mating systems in plants: occurrence, theoretical explanations, and empirical evidence. Annu Rev Ecol Evol Syst 36:47–79

Griffin AR, Cotterill PP (1988) Genetic variation in growth of outcrossed, selfed and open-pollinated progenies of *Eucalyptus regnans* and some implications for breeding strategy. Silvae Genet 37:124–131

Hardner CM, Potts BM (1997) Postdispersal selection following mixed mating in *Eucalyptus regnans*. Evolution 51:103–111

Harwood C (2011) New introductions - doing it right. In: Walker J (ed) Developing a eucalypt resource: learning from Australia and elsewhere. Wood Technology Research Centre, University of Canterbury, Christchurch

Henson M, Smith HJ, Boyton S (2008) *Eucalyptus longirostrata*: a potential species for Australia's tougher sites? New Zeal J For Sci 38(1):227–238

Hodge GR, Volker PW, Potts BM, Owen JV (1996) A comparison of genetic information from open-pollinated and control-pollinated progeny tests in two eucalypt species. Theor Appl Genet 92:53–63

House APN, Bell JC (1996) Genetic diversity, mating system and systematic relationships in two redmahoganies, Eucalyptus pellita and E. scias. Aust J Bot 44(2):157–174

Hung DT, Brawner JT, Lee DJ, Meder R, Dieters MJ (2016) Genetic variation in growth and wood-quality traits of *Corymbia citriodora* subsp. *variegata* across three sites in south-east Queensland, Australia. Australia. South For 78(3):225–239

Klápště J, Suontama M, Telfer E, Graham N, Low C, Stovold T, McKinley R, Dungey H (2017) Exploration of genetic architecture through sib-ship reconstruction in advanced breeding population of *Eucalyptus nitens*. PLoS ONE 12:e0185137

Lee DJ (2007) Achievements in forest tree improvement in Australia and New Zealand 2: development of Corymbia species and hybrids for plantations in eastern Australia. Aust For 70(1):11–16

Lee DJ, Huth JR, Brawner JT, Diskinson GR (2009) Comparative performance of *Corymbia* hybrids and parental species in subtropical Queensland and implications for breeding and deployment. Silvae Genet 58:205–212

Lin Y, Liu S, Luo J, Liu X, Lu W, Wang C, Arnold RJ (2017) Landrace origins and phenotypic diversity through seedling morphology in *Corymbia citriodora* subsp. *citriodora*. Aust. Forestry 80(1):43–56

Macedo HR, Freitas MLM, Moraes MLT, Zanata M, Sebbenn AM (2013) Variação, herdabilidade e ganhos genéticos em progênies de *Eucalyptus tereticornis* aos 25 anos de idade em Batatais-SP. Sci For 41(100):533–540

Mora F, Arriagada O (2016) A classification proposal for coefficients of variation in *Eucalyptus* experiments involving survival, growth and wood quality variables. Bragantia 75:263–267

Mora F, Gleadow R, Perret S, Scapim CA (2009) Genetic variation for early flowering, survival and growth in sugar gum (*Eucalyptus cladocalyx* F. Muell) in southern Atacama Desert. Euphytica 169:335–344

Moraes MA, Zanatto ACS, Moraes E, Sebbenn AM, Freitas MLM (2007) Variação genética para caracteres silviculturais em progênies de polinização aberta de *Eucalyptus camaldulensis* em Luiz Antônio-SP. Rev Inst Flor 19(2):113–118

Morais E, Zanotto A, Freitas M, Moraes M, Sebbenn AM (2010) Variação genética, interação genótipo solo e ganhos na seleção em teste de progênies de *Corymbia citriodora* Hook em Luiz Antonio, São Paulo. Sci For 38(85):11–18

Namkoong G (1966) Inbreeding effects on estimation of genetic additive variance. For Sci 12(1):8–13

Pagliarini MK, Konrad ECG, Silva FC, Silva MSC, Moreira JP, Sato AS, Machado JAR, Freitas MLM, Aguiar AV, Moraes MLT, Sebbenn AM (2016) Variação genética em caracteres de crescimento em progênies de *Dipteryx alata* Vog. Sci For 44(112):920–935

Pelletier M, Henson M, Boyton S, Thomas D, Vanclay J (2008) Genetic variation in shrinkage properties of *Eucalyptus pilularis* assessed using increment cores and test blocks. N Z J For Sci 38:194–210

Pimentel-Gomes F, Garcia CH (2002) Estatística aplicada a experimentos agronômicos e florestais: exposição com exemplos e orientações para uso de aplicativos. Fealq, Piracicaba

Pires IE, Resende MDV (2011) Genética Florestal. SIF, Viçosa

Potts BM, Dungey HS (2004) Genetic Interspecific hybridization of *Eucalyptus*: key issues for breeders and geneticists. New For 27(2):115–138

R Core Team (2018) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; https://www.R-project.org/. Accessed 19 June 2018

Raymond CA, Henson M, Pelletier MC, Boyton S, Joe B, Thomas D, Smith H, Vanclay J (2008) Improving dimensional stability in plantation-grown *Eucalyptus pilularis* and *E. dunnii*. Forest and Wood Products Australia; https://epubs.scu.edu.au/cgi/viewcontent.cgi?article=1480&context=esm_pubs. Accessed 16 Jan 2018

Reis CAF, Assis TF, Santos AM, Paludzyszyn Filho E (2013) *Corymbia citriodora*: estado da arte de pesquisas no Brasil. Documentos. 255. Dados eletrônicos – Colombo: Embrapa Florestas

Resende MDV (2002) Genética Biométrica e Estatística no Melhoramento de Plantas Perenes. Embrapa Informação Tecnológica, Brasília

Resende MDV (2007) Selegen-REML/BLUP: sistema estatístico e seleção genética computadorizada via modelos lineares mistos. Embrapa Florestas, Colombo

Resende MDV, Duarte JB (2007) Precisão e controle de qualidade em experimento de avaliação de cultivares. Pesq Agrop Trop 37(3):182–194

Resende MDV, Sturion JA (2003) Análise estatística espacial de experimentos via modelos mistos individuais com erros modelados por processos Arima em duas dimensões. Rev Mat Estat 21(1):7–33

Resende MDV, Vencovsky R, Fernandes JSC (1995) Selection and genetic gains in populations of *Eucalyptus* with a mixed mating system. In: IUFRO conference, 1995, Hobart. Proceedings Hobart: IUFRO, pp 191–193

Rojas PV, Ipinza RC, Gutierrez B, Molina MP, Arnold RJ (2017) Breeding *Eucalyptus globulus* for lower rainfall sites in the Bío-Bío Region of Chile. Aust For 80:105–112

Sato AS, Freitas MLM, Lima IL, Zimback L, Toniato MTZ, Sebbenn AM (2010) Genetic variation among and within provenances and progenies of *Corymbia maculata* (Hook.)

K. D. Hill and L. A. S. Johnson, in Pederneiras, SP. Cerne 16(1):60–67

Segura TES, Silva Jr. FG (2016) Hybrids of *C. citriodora* and *C. torelliana* for kraft pulp production. In: TAPPI's PEERS conference. https://www.researchgate.net/publication/308765503_Hybrids_of_C_citriodora_and_C_torelliana_for_kraft_pulp_production. Accessed 27 July 2017

Squillace AE (1974) Average genetic correlations among offspring from open-pollinated forest trees. Silvae Genet 23(5):149–156

Swain TL, Verryn SD, Laing MD (2015) An investigation of assumptions made in estimating genetic parameters and predicting genetic gain in a *Eucalyptus nitens* breeding programme in South Africa. New For 46:7–21

Vargas-Reeve F, Mora F, Perret S, Scapim CA (2013) Heritability of stem straightness and genetic correlations in *Eucalyptus cladocalyx* in the semi-arid region of Chile. Crop Breed Appl Biotechnol 13:107–112

Varghese M, Harwood CE, Bush DJ, Baltunis B, Kamalakannan R, Suraj PG, Hegde D, Meder R (2017) Growth and wood properties of natural provenances, local seed sources and clones of *Eucalyptus camaldulensis* in southern India: implications for breeding and deployment. New For 48:67–82

Vencovsky R, Pereira MB, Crisóstomo JR, Ferreira MAJ (2001) Genética e melhoramento de populações mistas. In: Nass LL, Valois ACC, Melo IS, Valadares-Inglis MC (eds) Recursos genéticos e melhoramento. Fundação MT, Rondonópolis

Wingfield MJ, Slippers B, Hurley BP, Coutinho TA, Wingfield BD, Roux J (2008) Eucalypt pests and diseases: growing threats to plantation productivity. South For 70(2):139–144

Wright S (1922) Coefficients of inbreeding and relationship. Am Nat 56:330–338

Wright AJ, Cockerham CC (1986) Covariances of relatives and selection response in generations of selfing from an outcrossed base population. Theor Appl Genet 72(5):689–694. https://doi.org/10.1007/BF00289010

Yeh FC, Brune A, Cheliak WM, Chipman DC (1983) Mating system of *Eucalyptus citriodora* in a seed-production area. Can J For Res 13(6):1051–1055

Zanata M, Freitas MLM, Silva MT, Morais E, Zanatto ACS, Sebbenn AM (2010) Parâmetros genéticos e ganhos na seleção em teste de progênies de polinização aberta de *Eucalyptus pellita*, em Batatais—SP. Rev Inst Flor 22(2):233–242

Zimback L, Mori ES, Brizolla TF, Chaves R (2011) Correlações entre caracteres silviculturais durante o crescimento de *Eucalyptus grandis* Hill ex Maiden. Rev Inst Flor 23(1):57–67