



How important are provenance and pedigree in later cycles of tree breeding? an example using *Acacia auriculiformis*

David Bush · Christopher E. Harwood ·
Hoa Thanh Duong · Nguyen Duc Kien

Received: 15 December 2024 / Accepted: 3 April 2025 / Published online: 29 April 2025
© The Author(s) 2025

Abstract Low input breeding programs, based primarily on open-pollinated mating, have been operating for *Acacia* and *Eucalyptus* species over the last thirty years. These programs have now progressed to later-generation breeding cycles after one or more rounds of selection, breeding and infusion of additional wild or exchanged later-generation selections. They typically lack the capacity for advanced genetic analysis that incorporates extensive marker-based genotyping. Using an *Acacia auriculiformis* breeding program in Vietnam as a case study, we compare simple models of advanced-generation breeding populations with an individual-tree model that incorporates all available pedigree information. We also examine the importance of including ancestral provenance in these models. Although laborious, assembling the pedigree was worthwhile in this example: we advocate commencing this work as soon as possible after domestication commences while records are more

readily accessible. A key benefit is more realistic estimation of breeding values and genetic parameters. Additionally, access to the pedigree allows calculation of group co-ancestry and status number, enabling better management of inbreeding and the accumulation of relatedness in the breeding population. These benefits will likely increase in subsequent generations as kinship accumulates in the breeding population. In the absence of pedigree records, inclusion of ancestral provenance in family models can give acceptable estimates of heritability.

Keywords Co-ancestry · Coefficient of relationship · Inbreeding · Mixed mating

Introduction

Several generic issues have arisen in low-input (e.g. Lindgren 2003) open-pollinated (OP) tree breeding programs for *Acacia* and *Eucalyptus* managed by governmental forest research agencies and private companies in Asia, Africa and Latin America. Since the 1980s, species of these two genera have been introduced as exotics to several countries and then improved by successive cycles of open-pollinated breeding. Initial breeding populations have typically comprised a total of 100–200 OP families from a range of natural provenances, together with additional progenies from non-native seed sources such as land races, plantations and seed orchards. Low input

D. Bush (✉)
CSIRO Australian Tree Seed Centre, GPO Box 1700,
Canberra 2600, Australia
e-mail: david.bush@csiro.au

C. E. Harwood
School of Natural Sciences, University of Tasmania,
Private Bag 55, Hobart, TAS 7000, Australia

H. T. Duong · N. D. Kien
Institute of Forest Tree Improvement and Biotechnology,
Vietnamese Academy of Forest Sciences, Duc Thang,
North Tu Liem, Hanoi 100000, Vietnam

breeding was adopted when the breeding team did not have the skills or resources for intensive breeding, or when control-pollinated (CP) breeding was not practical (e.g. for acacias; Harwood et al. (2015)). Analysis of the provenance-progeny trials established in the first cycle of breeding was typically conducted using a straightforward analysis of variance, usually implemented in the linear mixed model framework. Such analyses provide estimates of family variance, from which useful genetic parameters such as narrow-sense heritability can be estimated. They rank the genetic performance of the families under test to make selections for the next cycle of OP breeding and for seedling or clonal seed orchards. It is usual to declare provenance or seed source as an effect in the mixed model, typically fixed (Williams et al. 2024), or less commonly random (e.g. Stackpole et al. 2010). This is necessary, because significant provenance-level genetic effects, if not declared, will inflate the estimate of additive variance, leading to over-estimation of heritability and predicted gain (Williams et al. 2024). There is typically an additional degree of relatedness within families arising from some proportion of selfing, neighbourhood inbreeding and correlated paternity (Eldridge et al. 1993), and this needs to be accounted for when estimating heritability (Williams et al. 2024). Provenance rankings provide the tree breeder with guidance on interim seed sources that might be used until such time as seed orchards become productive, as well as identifying prospective sources of infusion for later generations (Squillace 1974; Eldridge et al. 1993).

Following genetic assessment, it is common practice to selectively thin provenance-progeny trials, converting them to seedling seed orchards (SSOs). Inferior trees are removed by within-family thinning with complete removal of very poor families. *Eucalyptus* and *Acacia* are predominantly insect-pollinated, although birds, other animals and wind may play a role in some instances (Stone et al. 2003; Eldridge et al. 1993). Thus, distances of pollen transfer are typically short (e.g. Le et al. 2019), so in SSOs of these genera most pollen parents will be within the orchard, but a proportion of progeny may derive from external pollen donors.

Genetic analysis becomes more complex in the second and subsequent cycles. Typically, individual tree selections are made from the better-performing families identified in the first generation of testing,

and OP progenies are collected from them. Within-family relatedness arising from inbreeding and correlated paternity may be reduced relative to the wild, and progenies from the SSOs may include a proportion of inter-provenance crosses. Additional families may be infused into second-generation progeny tests, either from natural provenances, or from other breeding programs. Many breeding populations have become a "rolling front" (Borrallho and Dutkowski 1998) with advanced-generation progeny trials incorporating progenies from the wild, first-, second- and later-generation orchards, and from other breeding programs as they become available. The commonly used family-based model with a term for the ancestral seed source or provenance is no longer entirely appropriate, given the complex mix of interbred natural provenances, secondary seed sources and new infusions in the second and subsequent breeding cycles.

Open-pollinated breeding presents analytical challenges, as only the maternal side of the pedigree is known with any certainty. If accurate pedigree records are maintained, maternal pedigrees can be tracked back to the original seed collections. Some information on the paternal side of the pedigree can be provided by modelling the pollen clouds of individual seed sources as fixed effects or genetic groups (Westell et al. 1988) in the pedigree model. However, many analyses do not account for this pedigree information, because it has not been recorded, and/or because the analysts cannot deal with the added complexity. Modelling proceeds largely as it did in the first generation, using only maternal identities of the families in the current generation under test. Advanced-generation progeny are simply assumed to be unrelated among families, and to either be half-sibs or to possess a homogeneous level of mixed mating within families. In some cases, the ancestral provenance of later-generation families is ignored (e.g. Nirsatmanto et al. 2023; Zhang et al. 2020), while in others, each family is assigned via the maternal pedigree to its ancestral provenance, which is specified as a fixed effect in the model (e.g., Swain et al. 2015; Luangviriyasaeng and Pinyopusarerk 2002), despite the likelihood of the progeny being the result of mating among provenances in the first generation.

Molecular markers provide an alternative way of recovering pedigree information. Marker data

has long been used to estimate breeding system parameters including levels of inbreeding at the species and/or provenance level. More recently, it has become feasible to estimate the relatedness of individuals within and among families, or to recover the pedigree for every tree in a trial or series of trials. While estimation of population-level parameters has already been undertaken for many species and is relatively affordable, pedigree reconstruction typically requires more genotyping and somewhat higher marker density. The concept of breeding without breeding (El-Kassaby et al. 2011), i.e. using markers to infer kinship, is certainly worthwhile at a small scale when no traditional pedigree exists, but it would be very expensive to repeat the process for every new stage of a recurrent selection process. Although the price of genotyping has fallen in recent years, especially on a per locus basis, the overall cost of the technology is significant (Estopa et al. 2023) and extensive marker-based genotyping of breeding and/or trial populations remains out of reach for many researchers, especially in developing countries. Assembly of a traditional pedigree is an alternative or complementary activity that makes good practical and financial sense.

A complication for OP breeding is that the level(s) of outcrossing in natural populations, plantations and seed orchards is not known with certainty. For genetic analysis, an average coefficient of relationship (ρ) among individuals within open-pollinated families is usually assumed. For eucalypt species, a value of 1/2.5 is often selected, reflecting a degree of selfing and neighbourhood inbreeding in wild eucalypt populations (Eldridge et al. 1993), while for tropical acacias, which are known to display high outcrossing rates in most natural populations and seed orchards (Butcher et al. 2004), values of 1/3 (*A. auriculiformis*, Hai et al. 2008b) and 1/3.3 (*A. mangium*, Williams et al. 2024) have been used. Usually, the chosen coefficient of relationship is assumed to be constant across different provenances and seed sources. However, in cases where investigations of breeding systems have been conducted using molecular markers, this assumption was shown to be flawed (Bush et al. 2011; Klápště et al. 2018).

Acacia auriculiformis breeding in Vietnam

Acacia auriculiformis Cunn. ex Benth. is an important species for wood production in south-east Asia, both in its own right, and as a parent in interspecific hybridization between *A. auriculiformis* and *A. mangium* Willd. (hereafter termed *Acacia* hybrid). By 2014, at least 100,000 ha of pure-species *A. auriculiformis* and 500,000 ha of clonal *Acacia* hybrid plantations were established in Vietnam, where they are grown in short rotations of 5–10 years for production of pulpwood and small-diameter sawlogs (Nambiar et al. 2015).

The natural range of *A. auriculiformis* extends across three geographically separate regions: (i) Western Province, New Guinea (PNG) and an adjacent area in West Papua Province, Indonesia, (ii) Cape York in north Queensland, Australia (QLD), and (iii) the far north of the Northern Territory of Australia (NT). The species occurs in riverine gallery forests, run-on sites and coastal landforms where its roots can access permanent or semi-permanent soil water (Boland et al. 1990).

Breeding of *A. auriculiformis* in Vietnam, which is managed by the Vietnam Academy of Forest Science (VAFS), commenced in the mid-1990s. Initial breeding populations in the north and south of Vietnam were established using a total of 200 family seedlots from a number of natural provenances from PNG and QLD and progenies collected from seed orchards in Thailand and NT (Hai et al. 2008b). The approach was to establish provenance-progeny trials in different regions of the country, followed by selective thinning to convert them to seedling seed orchards (SSOs) (Harwood et al. 2015).

A clonal trial testing 121 *A. auriculiformis* clones sourced from the best trees of superior families in the Chon Thanh progeny trial in the south of Vietnam was established at Bau Bang, Vietnam, in 2001. After thinning, this clonal trial functioned as a clonal seed orchard (CSO). The clones were selected from a total of 54 OP families, with 37 of these families each contributing two or more clones.

Second-generation progeny trials testing OP progenies of selections made in the first-generation SSO at Ba Vi and the CSO at Bau Bang, together with infusions from elsewhere were planted in 2008, and the breeding program for *A. auriculiformis* in Vietnam has now entered a third cycle of progeny

trials. The best-performing *A. auriculiformis* clones captured and ranked in clone trials have been certified as production clones but their commercial planting is not yet widespread.

Genetic gain from the first cycle of breeding of *A. auriculiformis* in Vietnam was estimated by Hai et al. (2008a). Three realised genetic gain trials were established in central, central-north and northern Vietnam, comparing 5–6 treatments including a mix of the best ten OP families from a seedling seed orchard in northern Vietnam, bulk seed from the orchard and a control (a mix of wild families from known best-performing natural provenances) in five replicates of 49-tree plots. Across the three trials, the mix of best orchard family displayed an increase of 15% over the control in stem diameter at breast height (DBH) and significant improvements in stem straightness and axis persistence scores, while the orchard bulk was only 7% above the control in DBH. The realised gains from breeding were broadly consistent with gain predicted from estimates of genetic parameters and selection intensities (Hai et al. 2008a).

Breeding system and genetic structure inferred from molecular marker studies

Isozyme marker studies provided important insights into the *A. auriculiformis* breeding system and population structure. Wickneswari and Norwati (1993) found that the three regions of natural occurrence were genetically distinct, with greater genetic difference between NT and the other two regions than between PNG and QLD, and that PNG provenances had the highest diversity. Moran et al. (1989), using 12 loci, found that two natural provenances, Coen, QLD and Bensbach, PNG, both had mean single locus outcrossing rates (t) of 0.88. A similar mean multi-locus outcrossing rate, ($t=0.89$), was identified for 12 *A. auriculiformis* progenies in an SSO in northern Vietnam (Vuong and McDonald 2002). However, Wickneswari and Norwati (1995), using six loci, identified varying rates of outcrossing among seedling progeny samples from a total of seven natural provenances in PNG, QLD and the NT, with multi-locus estimates ranging between $t=0.76$ and $t=0.95$.

Outcrossing rates in six Vietnamese *A. mangium* seed orchards varied, based on microsatellite assays

of parents and progeny. Four orchards had high mean multi-locus outcrossing rates ranging from $t=0.93$ to 1 (Butcher et al. 2004). However, two of the orchards had much lower mean outcrossing rates; one orchard with patchy, light flowering had $t=0.13$, whereas a collection of late-seeding trees from the final orchard had $t=0.49$. Field testing of the progeny under study established that selfed seedlings displayed significantly lower early growth than did outcrossed seedlings (Harwood et al. 2004) providing evidence for inbreeding depression in tropical acacias.

Because pollen of *A. auriculiformis* occurs in sticky, heavy polyads of 16 grains which do not disperse readily from the anthers, long-distance cross-pollination by wind is considered unlikely. Reproductive biology studies (Nghiem et al. 2010; Sedgley et al. 1991) suggested that the species was insect-pollinated and that bee species were likely effective pollinators. Pollen transfer, as indicated by the frequency of occurrence of interspecific hybrid progeny in adjacent CSOs of *A. auriculiformis* and *A. mangium* at Bau Bang declined rapidly with increasing distance from the inter-orchard boundary and it was concluded that very little pollen transfer occurred over distances greater than 100 m (Le et al. 2019). A recent marker-based study of *A. crassiparva* (Martins et al. 2025) showed that pollen parents in an adjacent orchard around 150 m distant from an elite seed orchard (ESO) sired around 13% of the orchard progeny produced by the ESO.

In addition to outcrossing, correlated paternity (r_p) is a significant consideration for management and genetic analysis of OP breeding populations. Correlated paternity gauges the proportion of pairs of outcrossed siblings that are full siblings. Although r_p has not been assessed in *A. auriculiformis*, Butcher et al. (2004) found variable but generally high rates of correlated paternity in Vietnamese *A. mangium* seed orchards, with r_p ranging between 0.34 and 0.82. A recent SNP-based study of a large progeny array of *A. crassiparva* from an ESO showed high levels of correlated paternity, together with very high outcrossing ($t>0.99$) (Martins et al. 2025).

Study objectives

The VAFS breeding program for *A. auriculiformis* provides a case study to examine the utility of pedigree records in OP breeding. We analysed the

genetic control of growth in two second-generation progeny trials under a range of assumptions. The simplest was the assumption of half-sib relatedness within families and no relationship between families. More complex models used a pedigree that accounted for provenance origins and maternal histories through previous seed orchards, correlated paternity, and different assumptions about within family relatedness. Impacts of the different sets of assumptions on genetic parameter estimates and selection outcomes were examined. The prospects for integrating pedigree data and molecular marker analysis to improve the precision of genetic analysis in OP breeding were then considered.

Materials and methods

Genetic material

The first generation of the VAFS breeding program comprised two large progeny trials that incorporated:

- OP families collected by the CSIRO Australian Tree Seed Centre (ATSC) from many provenances

in the species' natural range. These were collected using ATSC guidelines that prescribe collection of trees spaced at least two tree lengths apart to minimise neighbourhood inbreeding within the seed fall zone of selected trees (Gunn 2001)

- OP progenies of superior trees selected in the first generation of the Royal Thai Forestry Department breeding program in Thailand (Luangviriyasaeng and Pinyopusarerk 2002)
- OP progenies of three superior trees from a conservation stand based on approximately 65 families from Bensbach, PNG provenance (Harwood et al. 1991).

As noted above, selected individual trees from one of the progeny trials were cloned and established in a clonal trial that tested 100 clones, which was subsequently converted to a CSO that contributed progenies to the second cycle of testing. The two trials analysed in our study, Duc Co and O Toi, contained a subset of progenies in the VAFS second cycle of testing, as detailed in Table 1.

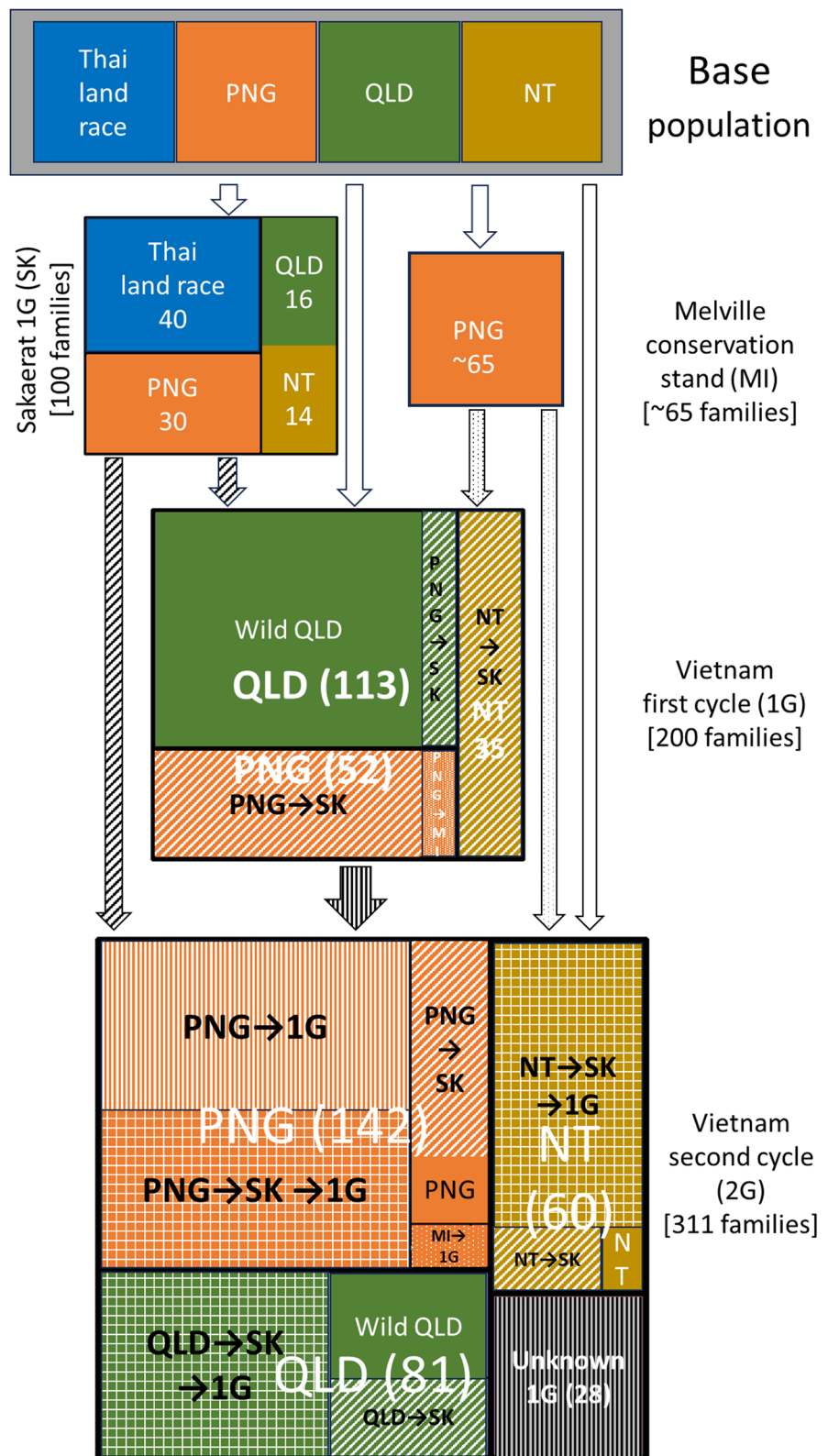
Given the inclusion of progenies from the first-generation Thai program, the VAFS pedigree is actually mixed-generation in both its first and second

Table 1 Summary of families included at the Duc Co and O Toi sites grouped by origin of selection in the present cycle and ancestral sources (regions of provenance of maternal founders including PNG, QLD and NT)

Ancestral region of provenance (source of selection in present cycle)	Duc Co	O Toi	Common to both	Total unique
Site				
PNG	25	34	5	54
Wild (ATSC)	0	7	0	7
Melville conservation stand	0	1	0	1
Sakaerat	19	0	0	19
VAFS 1G*	0	0	0	0
Sakaerat via VAFS 1G	6	23	5	24
Melville conservation stand via VAFS 1G	0	3	0	3
QLD	33	51	12	72
Wild (ATSC)	4	12	1	15
Sakaerat	14	0	0	14
VAFS 1G	10	24	8	26
Sakaerat via VAFS 1G	5	15	3	17
NT	12	27	1	38
Wild (ATSC)	1	2	0	3
Sakaerat	8	0	0	8
VAFS 1G	0	5	0	5
Sakaerat via VAFS 1G	3	20	1	22
Total	70	112	18	164

*1G—first generation

Fig. 1 Treemaps show constitutions of the Vietnam first- (1G) and second-cycle (2G) breeding populations. The Vietnam 1G population was based on (i) selections from the wild, (ii) selections from the 1G Thai breeding program at Sakaerat (SK) which incorporated wild and land race selections; and (iii) a small number of selections from a conservation stand at Melville Island, Australia based on Bensbach, PNG. The Vietnam 2G population was predominantly based on selections from the Vietnam 1G, further infusions from the Thai 1G program and some new wild selections. The treemaps for the 1G and 2G orchards show major blocks (bold borders), the colour of which correspond to the ancestral region of provenance, and the number of selections in each. Hatching of the minor blocks denotes their origins as follows: *solid colours* wild sources; *diagonal hatching* Thai 1G; *vertical hatching* Vietnam 1G; *crosshatching* Thai 1G selections via the Vietnam 1G; *stippling* Melville Island conservation stand



cycles of progeny testing. A further complication was that some of the natural-provenance families included in the first generation of the VAFS breeding population were also included in the first generation of the Thai program. Further Thai families, some of which were included in the VAFS first cycle, were again infused directly into the second VAFS cycle. A set of treemaps (Fig. 1) created using the R language Treemap library shows the rather complicated pattern of introductions. Here, we examine the performance of two progeny trials representing the second cycle of the VAFS breeding program.

The pedigree for the entire breeding population incorporating 15 trials was assembled in 2015 from import and export records retained by VAFS, ATSC and Royal Forestry Department Thailand. Fortunately, virtually all necessary records were available—noting that 28 OP progenies of trees selected in a multi-provenance unpedigreed seed production area in central Vietnam were brought into the second generation without full maternal pedigree (Fig. 1), although none of these were included in the two trials analysed here. Assembling the pedigree was a considerable task that took hundreds of hours and careful checking of historical records held by the institutions involved.

Trial sites

Progeny trials were planted at Duc Co and O Toi, two sites with similar soil on gently sloping terrain, 2 km apart, at Ba Vi Research Station in northern Vietnam (21° 8' N; 105° 28' E). The station is situated at approximately 60 m above sea level and receives about 1700 mm of rainfall annually, with a wet season between May and September and dry season from October to April. The soil type is a degraded ferrallitic

clay loam. Mean annual temperature is 23 °C; mean daily maximum temperature of the hottest month is 32 °C and mean daily minimum temperature of coldest month is 14 °C (Kien et al. 2009). The sites were prepared by mechanised ripping of planting rows to a depth of 50 cm and hand cultivation of 40×40×40 cm planting positions, with 200 g NPK fertilizer (5:10:3) and 3 kg composted cattle manure mixed into each planting position during cultivation. Planting took place in late 2008. Manual weeding was carried out twice per year for the first three years.

A total of 164 open-pollinated families were tested. Seventy families were planted at Duc Co and 112 planted at O Toi, with 18 families common to both sites (Table 1). Each trial had 8 complete-block replicates incorporating single, 3-tree family row plots, with incomplete blocking structures within the replicates (Table 2). The trial designs were generated using design software program CycDesign (VSN International, UK). Initial spacing was 3 m between rows and 2 m between trees within rows.

Trials were measured annually, with height and diameter at breast height (DBH) of all surviving trees recorded. Here we focus on the 2012 (4-year) measure of DBH.

Analysis

Pedigree information can be included into mixed-model analyses using an individual-tree model, a forestry analogue of the animal model (Mrode 1996). Our objective was to compare the genetic parameter estimates and breeding values that are obtained when the pedigree (represented schematically in Fig. 1) is included, with those from “traditional models” that assume no kinship among OP families. We included analyses both with and without ancestral provenance

Table 2 Design details of two *A. auriculiformis* progeny trials at Ba Vi Station

Trial site	Duc Co	O Toi
Number of families under trial	70	112
Number of replicates	8	8
Number of trees per plot	3	3
Incomplete blocking of plots within replicates	7 rows×10 columns	15 rows×8 columns (one column with only 7 plots)
Spacing between rows (m)	3	3
Spacing between trees within rows (m)	2	2

effects modelled as fixed terms, and which assumed various levels of homogeneous, within-open-pollinated-family relatedness resulting from mixed-mating. These assumptions would traditionally be implemented using a family model with the mixed-mating effects handled by post hoc application of a suitable coefficient of relationship to the additive variance component estimate. Rather than comparing individual-tree and family model outputs, all “traditional model” approaches were analysed using their near-identical individual-tree model analogues.

Mixed-model equations (Henderson 1953; Searle 1968) were implemented for prediction of breeding values and genetic parameters. These models were defined as;

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix} \quad (1)$$

where y is the vector of trait observations for DBH, b and u are vectors of fixed- and random- effect estimates respectively, and X and Z are incidence matrices for fixed and random model terms. Also included are terms R for the variance/covariance matrix of the residuals. If the random-effect factors are assumed to be independent, G is the direct sum of the variance/covariance matrices of each of the factors. It is also possible to assume that the levels of random effects are not independent. For breeding values, the additive relationship matrix (A) can then be used. In this case,

$$G = \sigma_a^2 A = \sigma_a^2 \begin{bmatrix} 1+f_i & 2\theta & 2\theta & 2\theta \\ 2\theta & 1+f_i & 2\theta & 2\theta \\ 2\theta & 2\theta & 1+f_i & 2\theta \\ 2\theta & 2\theta & 2\theta & 1+f_i \end{bmatrix} \quad (2)$$

where 2θ is the coefficient of relatedness of pairs of individuals and θ is the coefficient of co-ancestry. The diagonal is effectively the relatedness of an individual to itself: if the individual is not inbred this will be unity, but for inbred individuals it will be one plus the inbreeding coefficient f . For half-sibs, $2\theta=0.25$, for full sibs and parent–offspring relationships $2\theta=0.5$ and for pairs of selfed offspring, (self-full-sibs), $2\theta=0.66$.

Equation 1 was used to separately model four scenarios:

(i) Full pedigree model

Firstly, A^{-1} (the inverse additive relationship matrix) was formed directly using pedigree records for each tree: this was done using Asreml version 3 software (Gilmour et al. 2009) which implements the algorithm of Meuwissen and Luo (1992). The inverse relationship matrix was modified to include 43 genetic groups (see conceptual explanation in Westell et al. 1988) corresponding to ancestral provenance and seed orchard origins of the included families. In this model, *site* and complete-block *replicate* were included as fixed effects in vector b , while u contained sub-vectors for the random effects of *individual tree*, *plot* and in exploratory model runs, *site-by-individual tree* interaction. The latter interaction term was found to be non-significant, so the term was dropped from subsequent model runs.

(ii) Half-sib (100% outcrossing, no correlated paternity) model ($t=1$, $r_p=0$)

Equation 1 was then used to model the simple assumption that each family in the current-generation trials at Ba Vi was unrelated to all other families, i.e. members of different families are not *close* relatives, even though, in reality, many families actually share an ancestor and/or belong to the same provenance or subpopulation. Close relatives were defined as individuals that share alleles identical by descent, currently detectable for up to about ten previous generations at most (Donnelly 1983; Edge and Coop 2020). In this scenario, the progeny of later-generation mothers are modelled as half-sibs in A^{-1} with no inter-relatedness via grand-maternal or earlier ancestors. Model variants that either included the ancestral provenance origins of each seedlot modelled as a fixed effect in vector b or omitted provenance were implemented. Referring to Eq. 2, diagonal elements of A were all one, off-diagonal elements for trees in the same family were 0.25 and all other elements were zero (Table 3). The half-sib model would be expected to give an additive variance estimate four times greater than a traditional family model using the same data without a pedigree. We assumed no correlated paternity in this model ($r_p=0$).

Table 3 Summary of model variants including values for outcrossing (t), correlated paternity (r_p) elements of **A** including diagonal ($1+f$) and off-diagonal elements for relatedness

	Model	Fixed genetic effects	t	r_p	2θ (within family)	$1+f$	ρ (within families)
i	Full pedigree	Genetic groups for maternal provenances and orchard pollen clouds	1	0	Variable	1	Variable within and between families
ii	Homogeneous half-sib	Fixed effects for maternal provenances	1	0	0.25	1	1/4 within families
ii-a		<i>Nil</i>					
iii	Homogeneous, 90% outcrossing with 20% full-sib progeny	Fixed effects for maternal provenances	0.9	0.2	0.35	1.05	1/2.98 within families
iii-a		<i>Nil</i>					
iv	Homogeneous, 75% outcrossing, no correlated paternity	Fixed effects for maternal provenances	0.75	0	0.39	1.13	1/2.88 within families
iv-a		<i>Nil</i>					

(iii) Homogeneous, 90% outcrossing with 20% full-sib progeny ($t=0.9$, $r_p=0.2$)

The isozyme study of Moran et al. (1989) indicated that wild *A. auriculiformis* OP families may include 10% selfed progeny ($s=0.1$; outcrossing, $t=0.9$). It is also likely that some percentage of the progeny will share pollen parents. In this scenario we assumed that 90% of progeny were outcrossed and 20% of relationships were full sib resulting from common pollen parents i.e., correlated paternity (r_p)=0.2. Referring to Eq. 2, diagonal elements of **A** were all 1.05, off-diagonal elements for trees in the same family were 0.35 and all other elements were zero (Table 3). This model is analogous, in terms of narrow-sense heritability estimation, to a traditional family model with a coefficient of relationship of 1/2.98 applied post hoc to the additive variance component.

(iv) Homogeneous, 75% outcrossing model with no correlated paternity ($t=0.75$, $r_p=0$)

A variation of the half-sib model, assuming within-family relatedness resulting from a homogeneous maternal outcrossing rate of 75% ($t=0.75$), resulting from mixed mating with higher rates of inbreeding (as indicated by Wickneswari and Norwati (1995) and no correlated paternity ($r_p=0$). In this final scenario, the seedlots were again treated as unrelated except within families. The elements of **A**⁻¹ were adjusted directly following the method of

within families (2θ) and the equivalent coefficient of relationship (ρ) that would be used to adjust the variance from a traditional family model

Dutkowski and Gilmour (2001). Referring to Eq. 2, diagonal elements of **A** would be 1.13, off-diagonal elements for trees in the same family would be 0.39 and all other elements zero. This model is analogous, in terms of narrow-sense heritability estimation, to a traditional family model with a coefficient of relationship of 1/2.88 applied post hoc to the additive variance component. The coefficient of relationship can be calculated for individuals i and j as:

$$\rho_{ij} = \frac{2\theta_{ij}}{\sqrt{(1+f_i)(1+f_j)}} \quad (3)$$

where f_i and f_j are their inbreeding coefficients and θ_{ij} is the coefficient of co-ancestry.

The model variants are summarised in Table 3.

The narrow-sense heritability of the DBH trait in each scenario was estimated as

$$\hat{h}^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_p^2 + \sigma_e^2} \quad (4)$$

where σ_A^2 is the individual-tree (additive) genetic variance, σ_p^2 is the plot variance and σ_e^2 is the residual variance. Individual-tree models, unlike family models, do not include the coefficient of relationship (ρ) in Eq. 3, as inbreeding and relatedness are modelled for each tree and pair of trees, respectively, via the elements of **A**⁻¹ and hence σ_A^2 requires

no adjustment. The standard errors of $\hat{h}^2$ were calculated using a first-order Taylor series expansion to approximate the variance of a ratio of variances implemented in *ASReml* (Gilmour et al. 2009; Lynch and Walsh 1998).

The inverse additive relationship matrices ($\mathbf{A}^{-1}$) associated with the full pedigree and half-sib model variants were inverted to give $\mathbf{A}$, so that (i) the relatedness in the pedigree could be visualised by heat mapping the matrices and (ii) the contribution of founder mothers to the current generation as represented by the two trials could be examined. This was done by summing the off-diagonal elements of $\mathbf{A}$, associating each founder with the current generation.

Group co-ancestry (Θ) is the probability that two genes taken from a population gene pool are identical by descent (Cockerham 1967) or alternatively, the average pairwise co-ancestry over all pairs of population members given by:

$$\Theta = \frac{\sum_{i=1}^n \sum_{j=1}^n \theta_{ij}}{N^2} \quad (5)$$

where N is the census number of individual genotypes in the breeding population and θ_{ij} is the coefficient of kinship between individuals i and j . Note that the numerator of Eq. 5 includes estimates of coefficients of self-kinship for each individual which is given by $0.5 + f/2$ as well as the pairwise co-ancestries of all individuals.

The status number (N_s) is equivalent to the census number of a population of unrelated, non-inbred individuals, and can be calculated as:

$$N_s = \frac{0.5}{\Theta} \quad (6)$$

Practical breeding scenarios

Individual-tree and family best linear unbiased predictions (BLUP) from the model scenarios were compared using the Kendall rank correlation coefficient. Two scenarios to further explore the practical implications of BLUP re-ranking were made. The first scenario involved selections of individual trees from the breeding population at the two sites for a 30-clone CSO. The top-ranking individuals were selected from the breeding population with a limit of one tree per second-cycle

family for the traditional model variants and a group co-ancestry among selections of zero for the pedigree model (i.e. no close kinship among selections). The impact of re-ranking associated with co-ancestry management when selecting from the pedigree model was gauged by comparing the full pedigree model BLUP average over the 30 selected clones with that of each traditional model variant. The co-ancestry of the selections made in the traditional models was also calculated, bearing in mind that while there was only one individual per second-cycle family, selections were not limited by ancestral relatedness.

The second scenario examined family BLUP averages from the second-cycle families under each of the models. Comparisons were made for the bottom 20% of families, since these might be culled as a step in converting the progeny trials to seedling seed orchards.

Results

Additive relationship matrix from pedigree

The additive relationship matrices ($\mathbf{A}$) associated with the half sib (model ii) and full pedigree (model i) are shown as heatmaps in the lower and upper triangles, respectively, in Fig. 2a and b. Inclusion of pedigree information results in numerous ancestral connections among families, as presaged by the composition of the breeding population shown in Fig. 1. The 164 families included in the half-sib model assumption are effectively reduced to 92 extended families formed by ancestral relationships in the full pedigree model.

Variance component estimates

Although there was a significant difference ($p < 0.001$) in DBH between the two sites (predicted means of 12.6 cm at Duc Co and 9.7 cm at O Toi), the estimated site-by-genotype interaction variance was close to zero for all model variants. There was a significant difference among complete block replicates within sites ($p < 0.001$), but plot variance was non-significant, as gauged by the Akaike Information Criterion, and close to zero in all model variants (Table 4). Row variance (within sites), fitted as incomplete blocks at each site, was non-significant for

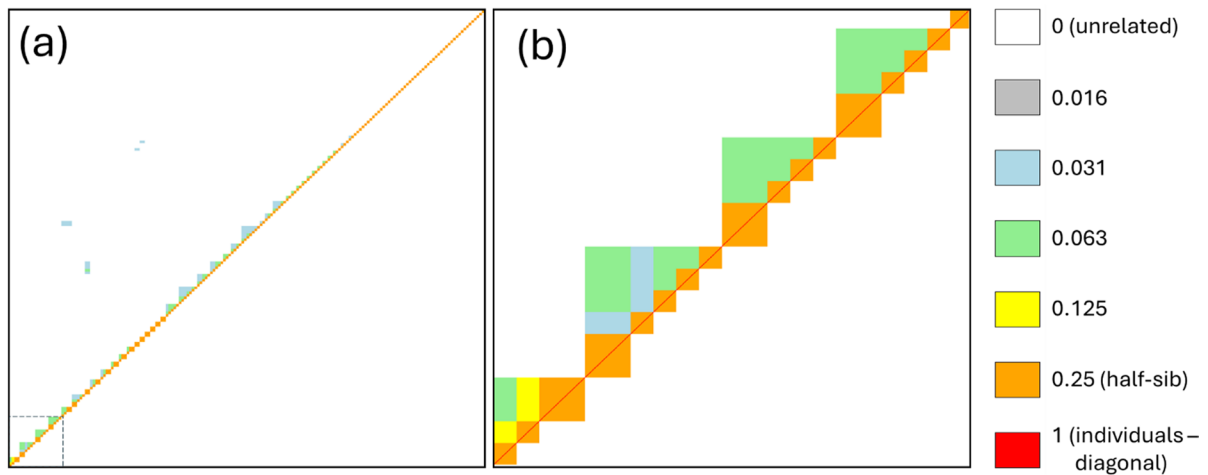


Fig. 2 Heat-mapped **A** matrices showing the relatedness (2θ) of all individuals (pane a) and an enlargement showing the first 500 individuals (inside dashed area of pane a) (pane b) included in the two second-cycle trials at Ba Vi. The diagonals in pane a and b, respectively, comprise 4638 and 500 individual blocks, each with a value of $2\theta=1$. These represent the relatedness of each individual tree to itself (unity in these two models). The off-diagonal elements of the lower triangles of each pane show blocks representing pairwise relationships between half-sib progeny ($2\theta=0.25$) from individual mother

trees that would be assumed using a traditional modelling approach. In pane b, smaller orange off-diagonal blocks represent half-sib families comprising 24 individuals while the larger orange blocks represent families comprising 48 individuals, that were planted at both sites. The upper triangles of pane a and b show the additional relationships (extended families) included in **A** when the full pedigree is used. These result from individuals that share common ancestors, effectively extending the size of many, but not all, families

Table 4 Variance components for DBH including σ_A^2 (additive) σ_p^2 (plot) and σ_e^2 (residual) associated with random terms, narrow-sense heritability estimates with standard errors for the model variants and equivalent coefficients or relationship (ρ) for each model

Model	Variance component/fixed term	σ_A^2	σ_p^2	σ_e^2	$\hat{h}^2$	$\hat{h}^2$ s.e	ρ
i	Provenances and pollen clouds as genetic groups (full pedigree model)	0.34	0.04	1.38	0.22	0.05	N/A
ii	$t=1$ (HS) with ancestral provenance fixed	0.46	0.06	1.31	0.26	0.06	0.25
ii-a	$t=1$ (HS) with no ancestral provenance	1.09	0.06	0.84	0.56	0.08	0.25
iii	$t=0.9$, $r_p=0.2$ with ancestral provenance fixed	0.36	0.04	1.41	0.20	0.04	0.34
iii-a	$t=0.9$, $r_p=0.2$, with no ancestral provenance	0.76	0.04	1.13	0.40	0.05	0.34
iv	$t=0.75$ with ancestral provenance fixed	0.30	0.06	1.44	0.17	0.04	0.35
iv-a	$t=0.75$ with no ancestral provenance	0.70	0.06	1.15	0.38	0.06	0.35

N/A not applicable

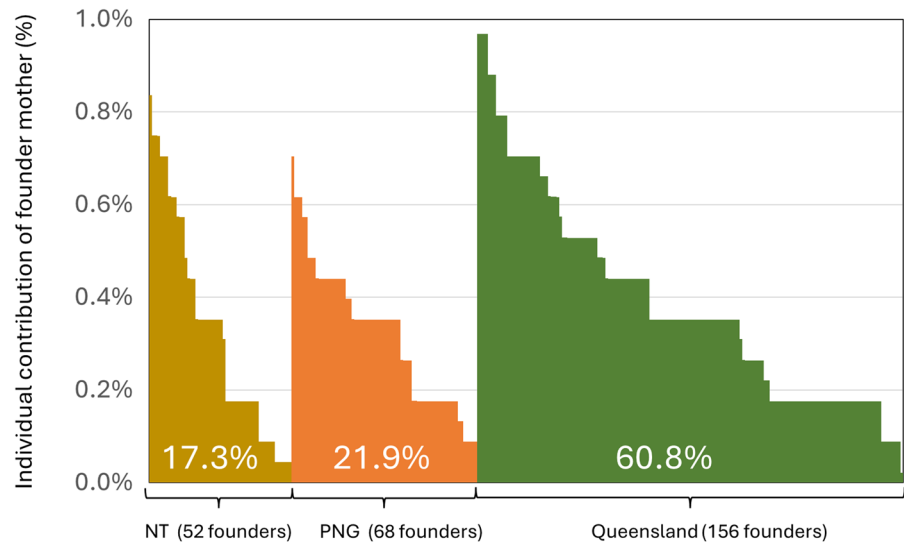
all model variants. Column variance, though statistically significant, was low, and of little practical significance, relative to the additive and residual components, in all models.

There were marked differences in genetic variance among model variants. Ancestral provenance was a significant source of variation ($p < 0.001$) in all model variants that included the fixed term, even after

one or two generations of selection and breeding. The additive variance component more than doubled relative to models ii, iii and iv in models ii-a (half-sib), iii-a ($t=0.9$, $r_p=0.2$) and iv-a ($t=0.75$, $r_p=0$), respectively, when the provenance term was dropped (Table 4).

The narrow-sense heritability estimate for DBH (Table 4) was low-to-moderate from the pedigree

Fig. 3 Contributions of founder mothers to the group co-ancestry (Θ) of the breeding population sample at Duc Co and O Toi grouped by region of ancestral provenance from the full pedigree model. Each bar represents the individual percentage contribution to the total value of $\Theta=0.00212$



(model i) ($\hat{h}^2=0.22$). The two models (iii and iv) that included ancestral provenance and some inbreeding produced lower estimates with model iii ($t=0.9$, $r_p=0.2$) producing an estimate very close to the pedigree model ($\hat{h}^2=0.22$). Model ii-a (half-sib/no provenance) produced a high heritability estimate ($\hat{h}^2=0.56$) whereas model iii-a ($t=0.9$, $r_p=0.2$ /no provenance) and model iv-a ($t=0.75$ /no provenance) produced moderate estimates ($\hat{h}^2=0.40$ and 0.38 respectively). Precision was high from all model variants ($\hat{h}^2$ s.e. between 0.04 and 0.08).

Genetic contribution of founder mothers

Analysis of **A** gave the relative genetic contributions of maternal ancestors to the current generation at O Toi and Duc Co (Fig. 3). The current cycle of the breeding population's maternal ancestry was dominated by contributions from QLD ancestors

with approximately equal contributions from NT and PNG. No individual mother contributed more than 1%, although the range of contributions is nearly ten-fold. There was no contribution from the Thai land race via maternal selections, but it is likely that genes from this source were transmitted by pollen parents in the first cycle. Likewise, there were 55 founder selections that appeared in the complete Vietnam *A. auriculiformis* pedigree that did not form part of the maternal pedigree of material represented at these two sites, but which may be pollen-parent ancestors.

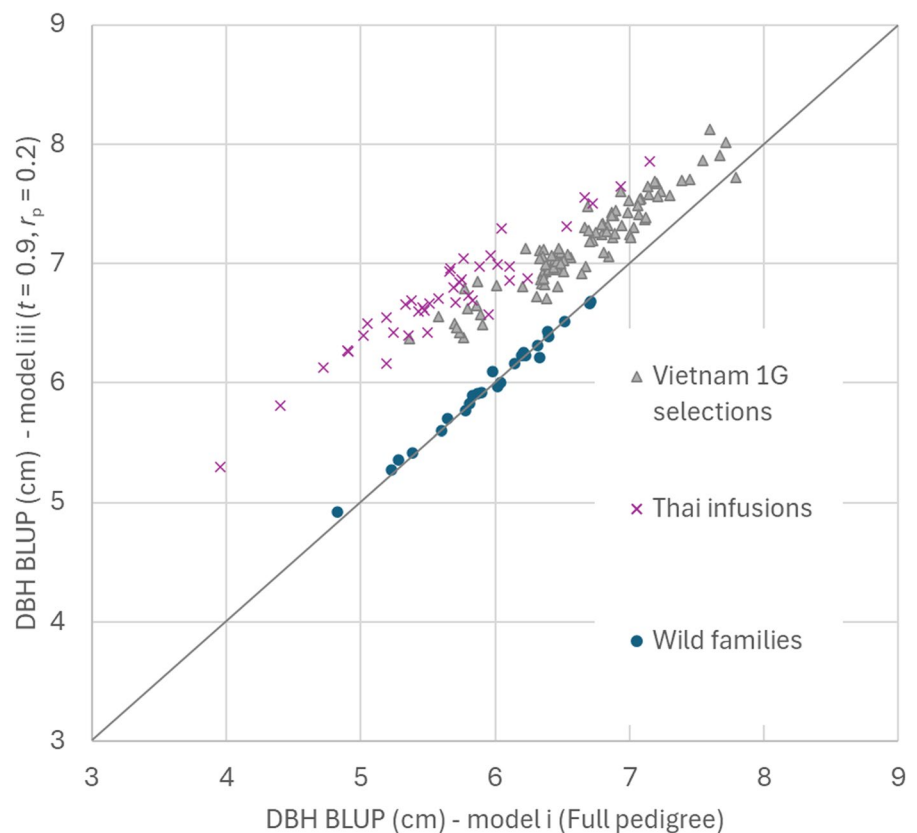
Group co-ancestry (Θ) for the breeding population at Duc Toi and O Toi, based on the maternal pedigree, was low (Table 5). The half-sib (model ii) and 75% outcrossing (model iv) assumptions resulted in estimates of Θ that were 123% and 83% of the pedigree (model i), resulting in status numbers (Ns) that were 83% and 123% of the pedigree model, respectively. Model iii ($t=0.9$, $r_p=0.2$) produced estimates that were intermediate to models i and iv.

Table 5 Genetic parameter comparison between pedigree and traditional model assumptions with different levels of outcrossing (t) and correlated paternity (r_p)

Parameter	Symbol	From pedigree	HS families ($t=1$, $r_p=0$)	$t=0.9$, $r_p=0.2$	$t=0.75$, $r_p=0$
Individuals in breeding population	N	4326	4326	4326	4326
Group co-ancestry	Θ	0.00212	0.00170	0.00201	0.00254
Status number	Ns	235.8	294.2	248.5	196.7
Relative status number	Nr	0.05	0.07	0.06	0.05

Table 6 Kendall rank correlation coefficients for DBH BLUPs from paired model variants for all families (above diagonal) and individual trees (below diagonal)

Model variant	Full pedigree	$t=1$ (HS), with prov	$t=1$ (HS), no prov	$t=0.9$, $r_p=0.2$, no prov	$t=0.9$, $r_p=0.2$, no prov	$t=0.75$, with prov	$t=0.75$, no prov
i Full pedigree	1	0.65	0.65	0.65	0.68	0.68	0.65
ii $t=1$, $r_p=0$ (HS), with prov	0.90	1	0.76	0.94	0.94	0.94	0.94
ii-a $t=1$, $r_p=0$ (HS), no prov	0.69	0.94	1	1.00	0.93	0.93	1.00
iii $t=0.9$, $r_p=0.2$, with prov	0.92	0.91	0.68	1	0.76	0.93	1.00
iii-a $t=0.9$, $r_p=0.2$, no prov	0.77	0.84	0.90	0.93	1	1.00	0.93
iv $t=0.75$, $r_p=0$, with prov	0.91	0.90	0.66	0.99	0.75	1	0.77
iv-a $t=0.75$, $r_p=0$, no prov	0.78	0.85	0.88	0.78	0.98	0.93	1

Fig. 4 Two-way plot of DBH family BLUPs from the full pedigree (model i) and model iii ($t=0.9$, $r_p=0.2$). Each point is an open-pollinated family in the Vietnam second-cycle breeding population, identified by its maternal source in the previous generation. For reference, a line with unity slope is plotted

Correlations between BLUP breeding values from different models

Both the family and individual tree DBH BLUPs from the different model variants were highly correlated and all correlations were significantly greater than zero ($p < 0.001$) (Table 6). For individual tree BLUPs, the pedigree model was most strongly

correlated with the traditional models that included ancestral provenance.

Figure 4 shows the family DBH BLUPs for model iii ($t=0.9$, $r_p=0.2$, with provenance) compared with the full pedigree (model i). The figure shows the better average performance, in both models, of families selected from the first cycle of breeding in Vietnam, especially relative to the wild

families infused into the second cycle. The figure also shows the significant divergence of BLUPs between the two model variants which is especially marked for poorer-performing families originating from the Thai infusions (first-generation orchard selections from Sakaerat). As expected, differences between the models were minimal for the wild families, as pedigree depth is low under both model assumptions, i.e. few of the infused families have recorded ancestral relationships with each other or other families in the breeding population. Families from the Thai infusions and the first cycle of breeding in Vietnam have significantly higher BLUPs under the traditional model than the full pedigree model. These groups of families differ from the wild families in that they (i) have deeper pedigrees with many families sharing common ancestors, leading to closer relatedness in the second-cycle and (ii) in the full pedigree, but not the traditional models, the male side of the pedigree has been modelled by including genetic groups representing the pollen clouds at Sakaerat and Ba Vi first-cycle seedling seed orchards, for the Thai infusions and the Vietnam first-cycle selections, respectively.

This difference in model assumptions is responsible for the differential in BLUPs, predicted by the traditional model, evident in Fig. 4. The pattern observed for model iii ($t=0.9$, $r_p=0.2$) was similar to those of models ii and iv ($t=1$ (HS), $t=0.75$, respectively) when compared with the full pedigree (model i) (corresponding figures not presented).

Model iv-a ($t=0.75$, $r_p=0$, no provenance) (Fig. 5) has the widest spread of individual-tree BLUPs. For example, this model predicts that several trees have DBH breeding values greater than 8.5 cm with lower estimates for these trees in all other models, including model iii ($t=0.9$, $r_p=0.2$, with provenance) also shown in Fig. 5. Each model that excluded provenance resulted in a wider spread of BLUPs than the corresponding model that included provenance.

Breeding scenario 1. Restricted selection of 30 top-ranking individuals for a clonal seed orchard

The effects of the various models on BLUP ranking of the top 30 selections (limited to one selection per second-cycle family) were minor. Table 7 shows

Fig. 5 Individual-tree DBH BLUP comparison between models I (full pedigree), iii ($t=0.9$, $r_p=0.2$ with provenance) and iv-a ($t=0.75$, $r_p=0$, no provenance) models. Each point is an individual tree in the Vietnam second-cycle breeding population. For reference, a line with unity slope is plotted

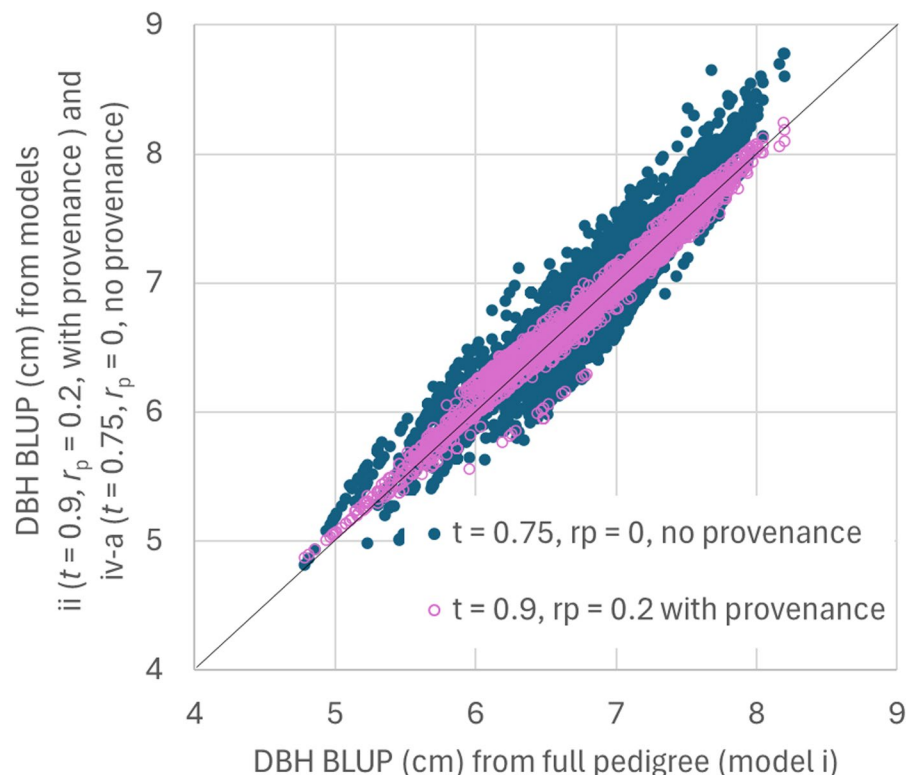


Table 7 Comparison of DBH BLUP averages for 30-tree selections for different model variants with breeding values compared on (i) the scale of the full pedigree model and (ii) the scales of the other model variants

	Model	Av. BLUP on full pedigree model scale (30 trees)	Av. BLUP on model scale (30 trees)	Selections in common with full pedigree (model i)	Population average (4368 trees)	Group co-ancestry of selections based on the pedigree
i	Full pedigree	7.90	N/A	30	6.87	0
ii-a	$t=1$ (HS), no provenance	7.84	8.05	21	6.92	0.034
iii	$t=0.9$, $r_p=0.2$, with provenance	7.89	7.87	24	6.87	0.034
iv	$t=0.75$ with provenance	7.89	7.84	23	6.87	0.034

that using model ii-a (half-sib/no provenance) (which had the poorest rank correlation to the full pedigree model) resulted in an average DBH breeding value 0.02 cm lower than the full pedigree model, with 22 of the 30 selections in common. However, part of this small difference was caused by the constraint of group co-ancestry (Θ)=0 placed on the full pedigree model, meaning that some trees with highly-ranked DBH BLUPs were not selected due to their ancestral kinship with included selections. Group co-ancestry of selections measured on the full pedigree scale (no inbreeding) was low in the traditional models ($\Theta=0.0340$), although nine and 12 of the 30 selected trees were related via a common great grandmother ($2\theta=0.0625$) to at least one other in the model i (HS) and iv ($t=0.75$) scenarios, respectively.

Breeding scenario 2: Culling the worst 20% of families

From Fig. 4, it is evident that re-ranking of family BLUPs is strongest among families with low DBH. Consequently, there were marked differences among model variants in the culling scenario (removal of the worst 20% of families) (Table 8). On the full pedigree BLUP scale, use of the traditional model variants (models ii-iv) results in cull population BLUP means that are around 0.1 cm higher than the full pedigree cull population. Fewer than half of the worst-performing families identified in the full pedigree model would be selected for culling using the traditional models.

Table 8 Comparison of population DBH BLUP averages (cm) after culling the 32 poorest families (20% of families removed) for different model variants with breeding values compared on

(i) the scale of the full pedigree model and (ii) the scales of the other model variants

	Model	Av. BLUP on full pedigree model scale (32 families)	Av. BLUP on model scale (32 families)	Culls in common with full pedigree model
i	Full pedigree	6.53	6.53	32
ii	$t=1$ (HS), with ancestral provenance	6.45	7.10	14
ii-a	$t=1$ (HS), no ancestral provenance	6.44	7.24	13
iii	$t=0.9$, $r_p=0.2$ with ancestral provenance	6.44	7.17	13
iii-a	$t=0.9$, $r_p=0.2$, no ancestral provenance	6.45	7.08	15
iv	$t=0.75$, with ancestral provenance	6.45	7.08	15
iv-a	$t=0.75$, no ancestral provenance	6.44	7.17	13

Discussion

This analysis of a second-cycle *A. auriculiformis* trial at two sites illustrates the benefits, in terms of genetic parameter estimation and selection, of considering ancestral provenance and pedigree in later-generation breeding populations. *Acacia auriculiformis* breeding in Vietnam is representative of low input tree breeding programs that commence with a founder population of families from multiple provenances or subpopulations. Under OP, only the ancestral provenance and maternal side of the pedigree are known with any certainty. As noted above, the assembly of the pedigree data from available records was very time consuming in this example. Adding to this impost was the complexity of the analysis that required specialised software, sophisticated methodology, and additional expertise relative to traditional methods based on family-level analysis. The question then arises: was the extra effort worth it?

Modelling the full pedigree (model i; Table 3) resulted in an additive variance estimate ($\sigma_A^2=0.34$) that was: very similar to model iii that included ancestral provenance and assumed 90% outcrossing with 20% full-sib relationships ($\sigma_A^2=0.36$); lower than the traditional model (ii) that included fixed provenance effects and assumed half-sib families ($\sigma_A^2=0.46$); and higher than model iv that included provenances and assumed 75% outcrossing ($\sigma_A^2=0.30$). These assumptions resulted in narrow-sense heritability estimates of $\hat{h}^2=0.22$, 0.20, 0.26 and 0.17 for models i (full pedigree); iii ($t=0.9$, $r_p=0.2$); ii ($t=1$, half-sib); and iv ($t=0.75$) (all with provenance included), respectively. All were low-moderate and within the established range for growth trait heritability (Cornelius 1994). It might therefore be concluded, from this example, that traditional models that include provenance of origin as a fixed effect give satisfactory estimates of heritability, even in advanced-cycle progeny trials. Other analyses that have taken such an approach have also produced reasonable estimates. Swain et al. (2015) found that predicted genetic gain from a second-generation *Eucalyptus nitens* family trial, modelled to include ancestral provenances and with $\rho=1/3$, was quite similar to realised genetic gain. They hypothesised that application of $\rho>1/4$ might help account for a proportion of full-sibs in later generations, even

though inbreeding should be lower than in the first generation because seed orchard design and management would be expected to increase the rate of outcrossing relative to that in natural provenances.

The full pedigree model included both ancestral provenances and orchard pollen clouds as genetic groups. The traditional models that did not include provenances (ii-a, iii-a, iv-a) gave high additive variance and heritability estimates, much higher than expected for growth traits (Cornelius 1994). Some studies of later-generation progeny trials that assumed half-sib families without ancestral provenance structure, where it might be expected to exist, gave moderate-to-high estimates, whereas other studies with these assumptions gave low-moderate heritability estimates (Baskorowati et al. 2022; Pagliarini et al. 2016). For traits where provenances differ in their performance, inclusion of the ancestral provenance term is likely to significantly improve the estimation of additive variance, with its omission resulting in inflated heritability estimates. (e.g. Zhang et al. 2020).

Our study demonstrates that it is important to account for ancestral provenance even after two to three generations of breeding. Although we might implicitly assume panmictic mating among provenances, which may diminish provenance effects apparent in the first generation, growth-trait provenance effects were still marked in our case study. This was not unexpected, as Luangviriyasaeng and Pinyopusarerk (2002) also found strong ancestral provenance effects in their analysis of the second-generation *A. auriculiformis* breeding program in Thailand. Moreover, others reported similar results in other hardwood species (Brawner et al. 2010; Bandara and Arnold 2017; Page et al. 2017; Swain et al. 2015). Breeding barriers among provenances including factors such as variations in peak flowering time might contribute to this. In a clonal orchard of *A. auriculiformis* in Thailand, clones of NT origin commenced flowering months earlier than those from QLD and PNG (Jiwarawat et al. 1996). This justified the inclusion of ancestral origins in the analysis of the Thai second generation *A. auriculiformis* population (Luangviriyasaeng and Pinyopusarerk 2002). However, ancestral subpopulation effects are not always detected in later generation trials. For example, significant growth trait differences among *Eucalyptus globulus* races detected in the first

generation had declined to non-significance by the second generation, although ancestral race effects in height to leaf phase change were still evident (Salas et al. 2014).

The assumption of homogeneous, 75% outcrossing with ancestral provenances included (model iv) resulted in a lower heritability estimate than the full pedigree (model i). The $t=0.75$ assumption accounts for homogeneous levels of elevated co-ancestry within families attributed to inbreeding. Model iii ($t=0.9$, $r_p=0.2$) also resulted in reduced heritability relative to model i, attributable to mild inbreeding and correlated paternity. Model i did not include any special assumptions about mixed mating effects (i.e. 100% panmictic outcrossing was assumed in each generation). In reality, we know that in first-generation breeding populations, family outcrossing rates are not necessarily homogeneous (Klápště et al. 2017; Bush and Thumma 2013), nor are they likely to be in subsequent generations. Additionally, there is often natural and artificial selection against inbred offspring due to the negative effects of inbreeding depression; outcrossed mating should be higher in managed orchards than in the wild; and correlated paternity is highly likely to exist and its levels will be heterogeneous in both natural and exotic stands. For advanced-generation analysis, adjustment of the relatedness coefficient (2θ) in the additive relationship matrix required for an individual-tree model, or analogously, choice of a coefficient of relationship (ρ) to apply to variance component estimates from a family model, will involve speculative assumptions. Different levels of relatedness within families are likely for each generation in the pedigree. This becomes increasingly complicated to model for open-pollinated families, being neither practical nor beneficial if the assumptions are incorrect.

It is not clear whether the inclusion of arbitrary rates of elevated relatedness would have given more realistic estimates of narrow-sense heritability, however, the model iv estimate assuming $t=0.75$ might represent the upper bound to likely relatedness, applicable to sensitivity analyses. Model iii, assuming $t=0.9$ and $r_p=0.2$, i.e., 20% full-sib relationships within families, may be more realistic, based on our knowledge of the breeding system of *A. auriculiformis*, and gave parameter estimates intermediate between the model iv ($t=0.75$) and model i ($t=1$, half-sib) assumptions.

Although the study's main objective was to gauge the impacts on genetic parameter estimation and selection decisions of including the pedigree compared with models that use simpler assumptions, the results also demonstrate that individual-tree models have some additional advantages over family models. This is particularly the case when selecting individual trees for purposes such as clonal seed orchards. Individual-tree models give a predicted breeding value (BLUP) for each tree in the current generation as well as their mothers. In our example, the re-ranking of the best 30 individuals attributable to the pedigree model variants was minimal and of no practical significance. There may be more significant re-ranking for traits that are under stronger genetic control than is DBH, or where other factors including design variables cause greater differential BLUP shrinkage (i.e. regression towards the mean).

Further benefits of assembling the pedigree and forming the **A** matrix accrue from the management of close kinship, which may result in inbreeding, at least for the maternal side of OP breeding populations. As the initial number of founders was comparatively large, with additional infusions, the group co-ancestry in our example was low ($\Theta=0.002$) and status number, based on the maternal pedigree, was high ($N_s=252$) compared with other low-input breeding programs, some of which have group co-ancestry estimates in the range 0.01 to 0.06 and status numbers between 10 and 70 (Ferreira et al. 2024; Park et al. 2017; Kang and Lindgren 1998; Machanská et al. 2013; Fedorkov et al. 2005). Nevertheless, the pedigree can be used to aid the selection of trees for activities such as elite tree selection for a CSO. In our example, we were able to use the **A** matrix to select a 30-tree CSO population with no close kinship. In contrast, despite being constrained to include only one selection per family from the traditional models, a significant proportion of trees were ancestrally related. Although the effects of inbreeding depression between distantly related individuals (in this case sibs that share a common grandmother or great-grandmother) is hard to quantify, it should be accounted for rather than focusing solely on the BLUP breeding values, which were only marginally different among the 30-tree populations selected using the different models. Operationally, group co-ancestry data derived from the pedigree can

be used to weight breeding values to achieve an acceptable trade-off between diversity and genetic gain (Rosvall and Mullin 2003).

We note that consideration of the pedigree prior to the assembly of families in the generation analysed herein would probably have been beneficial. From Fig. 1, two separate introductions from Saekarat, Thailand, are evident; one to the first cycle in VN and another to the second cycle. While this broadened the genetic base, many progeny in the later introduction were resultant of mothers already represented in the Vietnam breeding population. The effects of this introduction are not profound in the case of this breeding program, which, thanks to a relatively broad genetic base has a moderately-high status number. Consequences could be more serious in smaller breeding populations. This effect would be exacerbated in later generations, as selection and breeding narrows the breeding population, increasing kinship, and resulting in status number decline, as evidenced in the smaller eucalypt breeding populations studied by Ferreira et al. (2024). The task of assembling the pedigree will be easier if commenced at the start of the breeding program, with subsequent records added as breeding progresses, avoiding the need to track down historical information after years or decades.

Molecular markers can be used in various ways to inform breeding programs. In this example we made use of population-level estimates to guide selection of an appropriate coefficient of relatedness. However, markers are increasingly being used to estimate realised relatedness between individual trees in breeding populations (Ferreira et al. 2024; Klápště et al. 2021), and their benefits could be significant when applied judiciously to a low-input breeding program, particularly if pedigree records are not available. A major benefit is that relatedness accruing from both male and female ancestors can be accounted for. In our example, widespread genotyping might not be necessary, as the status number of the breeding population is quite high. However, in smaller low-input breeding populations with fewer founders, it may be worthwhile to genotype a larger, targeted set of key individuals (for example, every female tree contributing progeny to the current cycle of breeding) to ascertain which members are close relatives through the male side of the pedigree. A further benefit of marker-based relatedness

assessments is to check that the recorded pedigree is correct: where this has been done, significant errors are commonly identified (Vaillancourt et al. 1998; Doerksen and Herlinger 2008) and can be corrected. Notwithstanding these benefits, the cost of genotyping large numbers of individual trees is substantial, despite falling processing costs. The additional expertise required to accurately integrate the marker-based information into quantitative analyses is also not widely available, particularly in the case of low-input breeding programs. Although marker-informed analyses have much to commend them, traditional quantitative analyses based on available pedigree records are likely to continue to be produced for some years to come.

Conclusion

On balance, the benefits of assembling the pedigree from records are tangible and seem worthwhile. Although it takes significant effort to assemble a pedigree, commencing the job as early as possible will reduce the overall effort. Benefits accrue from more realistic modelling of relatedness in terms of additive variance and narrow-sense heritability estimation, breeding value estimation and management of ancestral relatedness and group co-ancestry. If the additional complexity of assembling the pedigree and its inclusion in analysis is too resource-intensive or the pedigree records are simply not available, satisfactory estimates of heritability can be obtained by using a simpler, family-based model and ancestral provenance as a fixed effect in a traditional family model, a result previously noted by Swain et al. (2015) that was corroborated by this study. Further benefits of assembling the pedigree accrue from the management of close kinship, at least on the maternal side of OP breeding populations. Selection based on the pedigree of elite populations for seed production allows management of ancestral relatedness that might lead to some level of inbreeding depression if ignored. While the pedigree can significantly improve the modelling of OP breeding populations, it cannot account for relatedness through pollen flow. Molecular markers are increasingly being used to estimate realised relatedness in breeding populations, and their benefits could be significant when applied

judiciously to a low-input breeding program, particularly if pedigree records are not available.

Acknowledgements We thank Mr Vitoon Luangviriyasaeng (retired from Royal Thai Forestry Department) who helped obtain detailed records relating to the Thai breeding program. We thank Mr Khongsak Pinyopusarerk (CSIRO) for useful discussions on the introductions of *Acacia auriculiformis* to Thailand and Vietnam. We thank researchers at the Institute of Forest Tree Improvement and Biotechnology, Vietnam, for providing data and maintaining pedigree information.

Author contributions DB carried out analysis and drafted the manuscript; CEH carried out analytical checks, assembled the pedigree and co-wrote the manuscript; NDK managed the breeding program, supervised and checked the collection of data and co-wrote the manuscript; HTD assembled pedigree data and co-wrote the manuscript.

Funding Open access funding provided by CSIRO Library Services. The study was jointly funded by the Vietnamese Academy of Forest Sciences and Commonwealth Scientific Industrial Research Organisation.

Data availability The data that support the findings of this study are not openly available due to reasons of sensitivity but are available from the corresponding author upon reasonable request. Data are located in controlled access data storage at CSIRO.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bandara KMA, Arnold RJ (2017) Genetic variation of growth and log end-splitting in second-generation *Eucalyptus grandis* in Sri Lanka. *Aust For* 80:264–271. <https://doi.org/10.1080/00049158.2017.1347982>
- Baskorowati L, Purwanto, Hendrati RL, Setiahadi R, Susanto M, Nurtjahjaningsih ILG, Mashudi, Kurniawan A, Pudjiono S, Setiadi D, Sumardi (2022) The approach in selecting the best genetic resistance against invasive aphid for indigenous tropical *Pinus merkusii* Jungh. et de Vriese in Indonesia. *Forests* 13:451. <https://doi.org/10.3390/f13030451>
- Boland D, Pinyopusarerk K, McDonald M, Jovanovic T, Booth T (1990) The habitat of *Acacia auriculiformis* and probable factors associated with its distribution. *J Trop For Sci* 3:159–180
- Borralho NMG, Dutkowski GW (1998) Comparison of rolling front and discrete generation breeding strategies for trees. *Can J For Res* 28:987–993. <https://doi.org/10.1139/cjfr-28-7-987>
- Brawner JT, Bush DJ, Macdonell PF, Warburton PM, Clegg PA (2010) Genetic parameters of red mahogany breeding populations grown in the tropics. *Aust For* 73:177–183. <https://doi.org/10.1080/00049158.2010.10676324>
- Bush D, Thumma B (2013) Characterising a *Eucalyptus cladocalyx* breeding population using SNP markers. *Tree Genet Genomes* 9:741–752. <https://doi.org/10.1007/s11295-012-0589-1>
- Bush D, Kain D, Matheson C, Kanowski P (2011) Marker-based adjustment of the additive relationship matrix for estimation of genetic parameters—an example using *Eucalyptus cladocalyx*. *Tree Genet Genomes* 7:23–35. <https://doi.org/10.1007/s11295-010-0312-z>
- Butcher P, Harwood C, Quang TH (2004) Studies of mating systems in seed stands suggest possible causes of variable outcrossing rates in natural populations of *Acacia mangium*. *For Genet* 11:303–309
- Cockerham CC (1967) Group inbreeding and coancestry. *Genetics* 56:89–104
- Cornelius J (1994) Heritabilities and additive genetic coefficients of variation in forest trees. *Can J For Res* 24:372–379. <https://doi.org/10.1139/x94-05>
- Doerksen TK, Herbinger CM (2008) Male reproductive success and pedigree error in red spruce open-pollinated and polycross mating systems. *Can J For Res* 38:1742–1749. <https://doi.org/10.1139/X08-025>
- Donnelly KP (1983) The probability that related individuals share some section of genome identical by descent. *Theor Popul Biol* 23:34–63. [https://doi.org/10.1016/0040-5809\(83\)90004-7](https://doi.org/10.1016/0040-5809(83)90004-7)
- Dutkowski GW, Gilmour AR (2001) Modification of the additive relationship matrix for open pollinated trials. In: Developing the eucalypt of the future, valdivia, Chile, 10–15 September 2001. Instituto Forestal, Chile, p 71
- Edge MD, Coop G (2020) Donnelly (1983) and the limits of genetic genealogy. *Theor Popul Biol* 133:23–24. <https://doi.org/10.1016/j.tpb.2019.08.002>
- Eldridge K, Davidson J, Harwood CE, van Wyk G (1993) *Eucalypt domestication and breeding*. Clarendon Press, Oxford
- El-Kassaby YA, Cappa EP, Liewlaksaneeyanawin C, Klápště J, Lstibůrek M (2011) Breeding without breeding: is a complete pedigree necessary for efficient breeding? *PLoS ONE* 6:e25737. <https://doi.org/10.1371/journal.pone.0025737>
- Estopa RA, Paludeto JGZ, Müller BSF, de Oliveira RA, Azevedo CF, de Resende MDV, Tambarussi EV,

- Grattapaglia D (2023) Genomic prediction of growth and wood quality traits in *Eucalyptus benthamii* using different genomic models and variable SNP genotyping density. *New For* 54:343–362. <https://doi.org/10.1007/s11056-022-09924-y>
- Fedorkov A, Lindgren D, David A (2005) Genetic gain and gene diversity following thinning in a half-sib plantation. *Silvae Genet* 54:185–189. <https://doi.org/10.1515/sg-2005-0027>
- Ferreira FM, Grattapaglia D, Bush D, Andrejow GMP, Estopa RA, Tambarussi EV (2024) Marker-based pedigree reconstruction reveals limited diversity within commercial *Eucalyptus benthamii* breeding populations in Brazil. *Tree Genet Genomes* 20:32. <https://doi.org/10.1007/s11295-024-01667-5>
- Gilmour AR, Gogel BJ, Cullis BR, Thompson R (2009) ASReml User Guide: Release 3. VSN International Ltd, Hemel Hempstead
- Gunn B (2001) Australian tree seed centre operations manual. CSIRO, Canberra
- Hai PH, Harwood C, Kha LD, Pinyopusarerk K, Thinh HH (2008a) Genetic gain from breeding *Acacia auriculiformis* in Vietnam. *J Trop For Sci* 20:313–327
- Hai PH, Jansson G, Harwood C, Hannrup B, Thinh HH, Pinyopusarerk K (2008b) Genetic variation in wood basic density and knot index, and their relationship with growth traits for *Acacia auriculiformis* in northern Vietnam. *New Zeal J For Sci* 38:175–192
- Harwood CE, Matheson AC, Gororo N, Haines R (1991) Seed orchards of *Acacia auriculiformis* at Melville Island, Northern Territory, Australia. In: Turnbull JW (ed) *Advances in tropical Acacia research: proceedings of a workshop held in Bangkok, Thailand, 11–15 February 1991*, ACIAR Proceedings No. 35. Australian Centre for International Agricultural Research, Canberra, pp 87–91
- Harwood CE, Ha HT, Tran HQ, Butcher PA, Williams ER (2004) The effect of inbreeding on early growth of *Acacia mangium* in Vietnam. *Silvae Genet* 53:65–69
- Harwood CE, Hardiyanto EB, Yong WC (2015) Genetic improvement of tropical acacias: achievements and challenges. *South For* 77:11–18. <https://doi.org/10.2989/20702620.2014.999302>
- Henderson CR (1953) Estimation of variance and covariance components. *Biometrics* 9:226–252. <https://doi.org/10.2307/3001853>
- Jiwarawat P, Luangviriyasaeng V, Woo KC (1996) Hybrid seeds of *Acacia auriculiformis*. In: Paper presented at the tree improvement for sustainable tropical forestry: QFRI-IUFRO conference, Caloundra, Queensland, Australia, 27 October - 1 November 1996 Caloundra, Queensland
- Kang KS, Lindgren D (1998) Fertility variation and its effect on the relatedness of seeds in *Pinus densiflora*, *Pinus thunbergii* and *Pinus koraiensis* clonal seed orchards. *Silvae Genet* 47:196–201
- Kien N, Jansson G, Harwood C, Thinh H (2009) Genetic control of growth and form in *Eucalyptus urophylla* in northern vietnam. *J Trop For Sci* 21:50–65
- 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. <https://doi.org/10.1371/journal.pone.0185137>
- Klápště J, Suontama M, Dungey HS, Telfer EJ, Graham NJ, Low CB, Stovold GT (2018) Effect of hidden relatedness on single-step genetic evaluation in an advanced open-pollinated breeding program. *J Hered* 109:802–810. <https://doi.org/10.1093/jhered/esy051>
- Klápště J, Ashby RL, Telfer EJ, Graham NJ, Dungey HS, Brauning R, Clarke SM, Dodds KG (2021) The use of “genotyping-by-sequencing” to recover shared genealogy in genetically diverse *Eucalyptus* populations. *Forests* 12:904. <https://doi.org/10.3390/f12070904>
- Le S, Harwood CE, Nghiem CQ, Griffin AR, Vaillancourt RE (2019) Patterns of hybrid seed production in adjacent seed orchards of *Acacia auriculiformis* and *A. mangium* in Vietnam. *Ann For Sci*. <https://doi.org/10.1007/s13595-019-0823-1>
- Lindgren D (2003) Low-input tree breeding strategies. In: Wei R-P, Xu D (eds) *Eucalyptus plantations: research, management and development*. World Scientific Publishing Company, Singapore, pp 149–166
- Luangviriyasaeng V, Pinyopusarerk K (2002) Genetic variation in a second-generation progeny trial of *Acacia auriculiformis* in Thailand. *J Trop for Sci* 14:131–144
- Lynch M, Walsh B (1998) *Genetics and analysis of quantitative traits*. Sinauer Associates Inc., Sunderland, Massachusetts, USA
- Machanská E, Bajcar V, Longauer R, Gömöry D (2013) Effective population size estimation in seed orchards: A case study of *Pinus nigra* ARNOLD and *Fraxinus excelsior* L./F. angustifolia VAHL. *Genetika* 45:575–588. <https://doi.org/10.2298/GENSRR1302575M>
- Martins G, Yulianto M, Melia T, Lakey N, Ordway J, Whetten R, Tambarussi EV, Acosta JJ, Hodge G (2025) Pollination and mating dynamics unveiled by orchard-wide pedigree reconstruction in *Acacia crassicaarpa*. *Tree Genet Genomes*. <https://doi.org/10.1007/s11295-025-01692-y>
- Meuwissen THE, Luo Z (1992) Computing inbreeding coefficients in large populations. *Genet Sel Evol* 24:305–313. <https://doi.org/10.1051/gse:19920402>
- Moran GF, Muona O, Bell JC (1989) Breeding systems and genetic diversity in *Acacia auriculiformis* and *Acacia crassicaarpa*. *Biotropica* 21:250–256. <https://doi.org/10.2307/2388652>
- Mrode RA (1996) *Linear models for the prediction of animal breeding values*. CAB International, Oxford
- Nambiar EKS, Harwood CE, Kien ND (2015) Plantations in Vietnam: research and knowledge application to secure a sustainable future. *South For* 77:1–10. <https://doi.org/10.2989/20702620.2014.999301>
- Nghiem CQ, Harwood CE, Harbard JL, Griffin AR, Ha TH, Koutoulis A (2010) Floral phenology and morphology of colchicine-induced tetraploid *Acacia mangium* compared with diploid *A. mangium* and *A. auriculiformis*: implications for interploidy pollination. *Aust J Bot* 59:582–592. <https://doi.org/10.1071/bt11130>
- Nirsatmanto A, Sunarti S, Putri AI, Haryjanto L, Kartikawati NK, Herawan T, Lestari F, Pudjiono S, Rimbawanto A (2023) Genetic stability observed in third-generation progeny trial of *Acacia mangium*: the importance of

- genotype by environment interaction assessment in advance generation breeding strategy. For Sci Technol 19:285–295. <https://doi.org/10.1080/21580103.2023.2264887>
- Page T, Bush D, Sethy M, Doran J (2017) Variation in early growth in a second-generation whitewood (*Endospermum medullosum*) progeny trial in Vanuatu. Aust For 80:121–126. <https://doi.org/10.1080/00049158.2017.1321462>
- Pagliarini MK, Kieras WS, Moreira JP, Sousa VA, Shimizu JY, Moraes MLT, Furlani E, Aguiar AV (2016) Adaptability, stability, productivity and genetic parameters in slash pine second-generation families in early age. Silvae Genet 65:59–66. <https://doi.org/10.1515/sg-2016-0010>
- Park JM, Kwon SH, Lee HJ, Na SJ, El-Kassaby YA, Kang KS (2017) Integrating fecundity variation and genetic relatedness in estimating the gene diversity of seed crops: *Pinus koraiensis* seed orchard as an example. Can J For Res 47:366–370. <https://doi.org/10.1139/cjfr-2016-0223>
- Rosvall O, Mullin TJ (2003) Positive assortative mating with selection restrictions on group coancestry enhances gain while conserving genetic diversity in long-term forest tree breeding. Theor Appl Genet 107:629–642. <https://doi.org/10.1007/s00122-003-1318-9>
- Salas M, Nieto V, Perafán L, Sánchez A, Borralho NMG (2014) Genetic parameters and comparison between native and local landraces of *Eucalyptus globulus* Labill. ssp. *globulus* growing in the central highlands of Colombia. Ann For Sci 71:405–414. <https://doi.org/10.1007/s13595-013-0342-4>
- Searle SR (1968) Another look at Henderson's methods of estimating variance components. Biometrics 24:749–778. <https://doi.org/10.2307/2528870>
- Sedgley M, Harbard J, Smith RM, Wickneswari R (1991) Development of hybridization techniques for *Acacia mangium* and *Acacia auriculiformis*. In: Carron LT, Aken KM (eds) Breeding technologies for tropical acacias. ACIAR Proceedings No. 37. Australian Centre for International Agricultural Research, Canberra, pp 63–69
- Squillace AE (1974) Average genetic correlations among offspring from open-pollinated forest trees. Silvae Genet 23:149–156
- Stackpole DJ, Vaillancourt RE, Downes GM, Harwood CE, Potts BM (2010) Genetic control of kraft pulp yield in *Eucalyptus globulus*. Can J For Res 40:917–927. <https://doi.org/10.1139/X10-035>
- Stone GN, Raine NE, Prescott M, Willmer PG (2003) Pollination ecology of acacias (Fabaceae, Mimosoideae). Aust Syst Bot 16:103–118
- Swain T-L, Verry 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. <https://doi.org/10.1007/s11056-014-9444-6>
- Vaillancourt RE, Skabo S, Gore PL (1998) Fingerprinting for quality control in breeding and deployment. Aust For 61:207–210. <https://doi.org/10.1080/00049158.1998.10674742>
- Vuong TD, McDonald M (2002) Outcrossing rates in *Acacia auriculiformis* A. Cunn. ex Benth. in a seedling seed orchard at Ba Vi, Vietnam, Technical Report: FST/1998/096. Australian Centre for International Agricultural Research, Canberra
- Westell RA, Quaas RL, Van Vleck LD (1988) Genetic groups in an animal model. J Dairy Sci 71:1310–1318. [https://doi.org/10.3168/jds.S0022-0302\(88\)79688-5](https://doi.org/10.3168/jds.S0022-0302(88)79688-5)
- Wickneswari R, Norwati M (1993) Genetic diversity of natural populations of *Acacia auriculiformis*. Aust J Bot 41:65–77. <https://doi.org/10.1071/bt9930065>
- Wickneswari R, Norwati M (1995) Spatial heterogeneity of outcrossing rates in *Acacia auriculiformis* A. Cunn. ex Benth. in Australia and Papua New Guinea. In: Baradat P, Adams WT, Muller-Starck G (eds) population genetics and genetic conservation of forest trees. SPB Academic Publishing, Amsterdam, pp 329–337
- Williams ER, Matheson AC, Harwood CE (2024) Experimental design and analysis for tree improvement. CSIRO Publishing, Melbourne
- Zhang H, Zhang Y, Zhang D, Dong L, Liu K, Wang Y, Yang C, Chiang VL, Tigabu M, Zhao X (2020) Progeny performance and selection of superior trees within families in *Larix olgensis*. Euphytica 216:60. <https://doi.org/10.1007/s10681-020-02596-9>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.