Optimised eucalypt domestication: an example using *E. cladocalyx,* a species for low rainfall environments

David James Bush

A thesis submitted for the degree of

Doctor of Philosophy

of

The Australian National University

December 2014

Australian
National
University

Abstract

Eucalyptus cladocalyx (F. Muell.) is endemic to South Australia and has been planted extensively throughout southern Australia and in dry, Mediterranean climates overseas. Its wood is hard, strong and naturally durable, and is suitable for fuelwood and solid-wood applications. Domestication in Australia commenced in 2001, when a breeding population was established across 11 sites. The breeding objective was to maximise sawlog production per hectare per year. This thesis examines genetic characteristics of the breeding population relevant to the species' domestication.

The thesis begins (Chapter 1) by arguing that methods and assumptions historically used to identify selections in the first-generation breeding programs of the main commercial tree species can be improved upon by:

(1) Examining wood properties, in addition to growth and form traits, during the first generation, taking advantage of modern labour-saving techniques, rather than delaying until later generations by which time unidentified adverse genetic correlations between traits may have caused problems;
(2) Employing molecular markers to determine population genetic parameters and reconstruct pedigree and inbreeding information from families that have unknown or uncertain ancestry;
(3) Using recently-developed mixed-modelling techniques that allow integration of marker-based pedigree and inbreeding information to model genotype-by-environment (GxE) interactions using large datasets.

Genetic parameters for key traits, including heritability of growth and wood natural durability traits and additive genetic correlations among traits, are reported in Chapter 2. The use of near-infrared reflectance as a low-cost method of screening durability traits such as decay mass loss and wood extractive content is also investigated.

The application of marker-based data to modify traditional assumptions made in analysis of first generation breeding populations is reported in Chapter 3. Previously published growth-trait estimates, together with an earlier isozyme study, indicated that the traditional approach may give upwardly-biased heritability estimates due to high and heterogeneous selfing. Existing isozyme data were reanalysed to reconstruct the pedigree for heterogeneous family-

level selfing. The result was reduced additive variance estimates and re-ranking of breeding values compared with the traditional, half-sib family assumption.

The results of single-nucleotide polymorphism (SNP) genotyping of the breeding population are presented in Chapter 4. Population structure and diversity, family relatedness, inbreeding and inbreeding depression (ID) were investigated. The study revealed that the breeding population is strongly structured with heterogeneous relatedness, inbreeding and ID among wild subpopulations. High outcrossing rates and genetic diversity were indicated for the land-race selections.

Marker-based estimates of relatedness and inbreeding reported in Chapter 4 were then integrated into a quantitative genetic analysis across sites using an extension of the methodology developed in Chapter 3, and the results reported in Chapter 5. Individual-tree mixed models, based on (i) the traditional half-sib family assumption and (ii) a modified model incorporating marker-based data, were compared. Analysis of GxE was performed across the 11 sites using individual-tree, factor analytic mixed models. While the traditional assumptions proved to be unsuitable for parameter estimation in this highly heterogeneous breeding population, marker-based models successfully accounted for the effects of mixed mating. GxE was found to be significant, though stable clusters of sites were evident for each trait.

The overall conclusion of the study, reported in Chapter 6, is that the prospects for genetic improvement of *E. cladocalyx* are favourable. Adverse genetic correlations between growth and wood property traits were not apparent. The substantial subpopulation and additive genetic variance present various options for ongoing breeding population management. The marker-based study enabled resolution of genetic diversity, ancestry of land-race material and pedigree reconstruction, which was indispensable for this breeding population. Integrating this information into mixed-model analyses provided more-realistic estimates of genetic parameters, breeding values and assessment of GxE. It is concluded that early assessment of wood properties and integration of quantitative and molecular techniques is cost-effective and will be widely applicable to domestication of other tree species.

Candidate's Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university. To the best of the author's knowledge, it contains no material previously published or written by another person, except where due reference is made in the text and/or as explicitly detailed on the following two pages.

David Bush

Date: 18 september 2015

Publications included in this thesis

Chapter 2

Bush D, McCarthy K, Meder R (2011) Genetic variation of natural durability traits in *Eucalyptus cladocalyx* (sugar gum). Ann. For. Sci. 68: 1057-1066

Contributor	Statement of contribution
David Bush	Conception and design of project (85%)
	Writing paper (90%)
	Statistical analysis (100%)
Kevin McCarthy	Conception and design of project (10%)
	Writing paper (5%)
	Mycological work (100%)
Roger Meder	Conception and design of project (5%)
	Writing paper (5%)
	Near Infrared Analysis (100%)

This work was completed with assistance of a grant from Rural Industries Research and Development Corporation to David Bush (Principal Investigator)

Chapter 3

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

Contributor	Statement of contribution
David Bush	Conception and design of project (85%)
	Writing paper (90%)
	Statistical analysis (100%)
Dominic Kain	Conception and design of project (5%)
	Writing paper (5%)
Colin Matheson	Conception and design of project (5%)
	Writing paper (2.5%)
Peter Kanowski	Conception and design of project (5%)
	Writing paper (2.5%)

Chapter 4

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

Contributor	Statement of contribution
David Bush	Conception and design of project (95%)
	Writing paper (97.5%)
	Statistical analysis (100%)
Bala Thumma	Conception and design of project (5%)
	Writing paper (2.5%)

Chapter 5

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

Contributor	Statement of contribution
David Bush	Conception and design of project (85%)
	Writing paper (90%)
	Statistical analysis (100%)
Dominic Kain	Conception and design of project (5%)
	Writing paper (5%)
Colin Matheson	Conception and design of project (5%)
	Writing paper (2.5%)
Peter Kanowski	Conception and design of project (5%)
	Writing paper (2.5%)

Acknowledgements

I would like to thank my supervisors, Dominic Kain, Colin Matheson and Peter Kanowski for all of the help, guidance and encouragement they provided during the course of my study. Chris Harwood also deserves special thanks for his deep involvement in setting up the trials reported and for providing useful discussion and encouragement throughout the writing of this thesis. I would also like to thank Cris Brack who has done a great job of managing the administration of my PhD program during the last two years.

I would also like to thank my CSIRO colleagues and co-authors who contributed to the research that forms part of this thesis. I would particularly like to thank Kevin McCarthy and Roger Meder for their work on the wood properties study and Bala Thumma for work on the molecular marker study. Simon Southerton provided a lot of support for the molecular study and it would not have been possible to complete this part of the work without his assistance. Bronwyn Matheson and Charlie Bell are thanked for providing excellent advice and help with DNA extraction – and troubleshooting the results! Luke Bulkeley, Brioni Brammall, Paul Macdonell and David Spencer all provided valuable assistance with field and laboratory measurements. Washy Gapare and Milos Ivkovich provided valuable comments on drafts and discussion on analysis. The staff of the Australian Tree Seed Centre, Sarah Whitfeld, Bronwyn Clarke, Carmel Sinosic and Debbie Solomon all provided a lot of help throughout this study, extending back to the 1990s, when some of the fascinating seedlots that are the subject of study here were collected, up until the present day. I would also like to thank Rowland Burdon, who was the Communicating Editor for three of the four papers presented here, for provision of helpful anonymous reviews as well as his own highly valuable suggestions.

Finally, I would like to thank my family, Karen, Chris and Jeffrey for being patient with me for so many years while this thesis has been 'under construction', and my parents, Wendy and Steve, for their constant encouragement.

Contents

Abstract ..... **ii**
Publications included in this thesis ..... **v**
Acknowledgements ..... **vii**
Contents viii
List of tables ..... **xi**
List of figures ..... **xiii**
List of abbreviations ..... **xiv**
Chapter 1: General introduction **1**
1.1 Tree crops for marginal land 1
1.2 *E. cladocalyx* – species description 2
1.3 Overview of the domestication process 8
1.4 Wood properties and interactions between traits 12
1.5 Accounting for mixed mating and unknown relatedness 13
1.6. The role of molecular markers 15
1.7 Genotype-by-environment interaction 16
1.8 Research objectives and motivation for study 21
1.9 Thesis structure 22
Chapter 2: Genetic variation of natural durability traits in *Eucalyptus cladocalyx* (sugar gum) . 25
Abstract 25
Keywords 26
Introduction 26
Materials and Methods 28
Sample population and procedures 28
NIR spectroscopy and model calibration 29
Accelerated biodeterioration 30
Statistical analysis 31
Results 32
Diameter and heartwood proportion 32
Basic density 33
Methanol extractives 33
Resistance to biodeterioration 33
Phenotypic and genetic correlations 33
NIR Calibration 35
Discussion 37
Conclusion 39
Acknowledgements 39

Chapter 3: Marker-based adjustment of the additive relationship matrix for estimation of genetic parameters – an example using *Eucalyptus cladocalyx* **40**
Abstract 40
Introduction 41
Materials and methods 43
Phenotypic data 43
Calculation of inbreeding rates and coefficients of relationship 44
Modification of the matrix 47
Variance-component estimation and BLUP breeding value prediction 48
Correlations 49
Genetic gain 50
Results 50
Estimates of selfing rates 50
Variance components and heritability 52
BLUP shrinkage and rankings 53
Relationship between rate of selfing, survival and phenotypic measurements 56
Discussion 56
Acknowledgements 60
Chapter 4: Characterising a *Eucalyptus cladocalyx* breeding population using SNP markers **61**
Abstract 61
Key words 61
Introduction 62
Materials and methods 64
Breeding populations and sampling 64
Molecular markers 65
Results 68
SNP marker properties 68
Measures of heterozygosity and diversity 70
Relatedness 72
Relatedness, inbreeding and fitness 74
Discussion 76
Data Archiving 80
Acknowledgements 80
Chapter 5: Genetic parameter estimates informed by a marker-based pedigree: a case study with *Eucalyptus cladocalyx* in southern Australia **81**
Abstract 81
Introduction 82
Methods 85
Breeding population 85
Site and phenotypic data 87
Marker data 87
Genetic analysis 88

Results 94
Phenotypic means and evidence of inbreeding depression 94
Comparison of HS and MM models 95
Provenance variation 96
Trait-trait correlations (r_A) 96
Genotype-by-environment interaction 97
Breeding value correlations 101
Comparison with family model assuming a coefficient of relationship 101
Discussion 101
Acknowledgements 106
Chapter 6: Discussion and conclusion 107
6.1 Drivers for the domestication process 107
6.2 Wood properties 108
6.3 High heritability in Western Australian trials 110
6.4 Modification of the A matrix using isozyme markers: the "pilot study" 111
6.5 SNP study of breeding population 112
6.6 Comparison of mixed-mating and half-sib assumptions on analysis of ALRTIG breeding population 114
6.7. Genotype-by-environment interactions 115
6.8 Why is the species apparently different? 117
6.9 Practical implications of the study 118
Consolidated References 122
Appendix 1 Calibration plot for prediction of decay and extractives 132
Appendix 2.1 Subpopulation/spatial data for the first-generation *E. cladocalyx* breeding population. Summary information for seedlots included in the ALRTIG breeding population at Lismore, Victoria 133
Appendix 2.2 Description of genes (n=39) and SNPs (n=75) located within chromosome scaffolds 134
Appendix 2.3 Detection of SNPs under selection in *E. cladocalyx* first-generation breeding population 136
Appendix 2.4 Family-average relatedness estimates between pairs of trees within families of *E. cladocalyx* 137
Appendix 3.1 Distribution of families among sites from two wild regions of provenance (Kangaroo Island and South Flinders Ranges, South Australia) and cultivated stands in Victoria and South Australia 139
Appendix 3.2a Table of predicted DBH means and fixed effect significance from single-site and across-sites (FA2) analyses 140
Appendix 3.2b Table of predicted HT means and fixed effect significance from single-site and across-sites (FA2) analyses 141
Appendix 3.2c Table of predicted AP (axis persistence) means from single-site and across-sites (FA2) models 142
Appendix 3.3 Variance component and heritability estimates with standard errors for each trait resulting from single-site analyses 143
Appendix 3.4 Tables of correlation, variance and parameter estimates from k=2 factor analytic (FA2) models 144

List of tables

Table 1.1 Edaphic parameters for regions of provenance of *E. cladocalyx*. 4

Table 1.2 Some important domestication events for *E. cladocalyx* and *P. radiata* in Australia 11

Table 2.1 Composition and sampling detail of the 8-year old *E. cladocalyx* provenance-progeny trial at Benalla, Victoria 28

Table 2.2 Trait means, significance of effects (F probabilities), heritability estimates ($\hat{h}^2$) and coefficient of phenotypic (CV_P) variation of an *E. cladocalyx* trial at 8 years of age at Benalla, Victoria 34

Table 2.3 Phenotypic r_P and genetic correlations r_G (standard error) between pairs of traits in 8-year-old *E. cladocalyx* 35

Table 2.4 Calibration statistics for NIR Partial Least Squares (PLS-1) calibration models giving PCs (principal components), R^2 (coefficient of determination of calibration and validation), RMSECV and RMSEP 36

Table 3.1 Additive relationships for elements of **A** incorporating partial selfing (s) and partial selfing and inbreeding in previous generations (f) 48

Table 3.2 Estimates of family single- $(\hat{t}_s)$ and multilocus $(\hat{t}_m)$ outcrossing rates with standard errors computed from arrays of n progeny and eight isozyme loci per family with MLTR.... 51

Table 3.3 Genetic parameter estimates ($\hat{h}^2$ narrow-sense heritability, $\hat{\sigma}^2_A$ additive genetic variance and $\hat{\sigma}^2_e$ residual variance) and standard errors for DBH and height assuming varying levels of ancestral inbreeding and selfing 53

Table 3.4 Pairwise Spearman rank correlations between BLUP height and DBH (individual trees) for models assuming (a) no inbreeding ($f = 0$, $s = 0$), (b) selfing equivalent to population average s=0.36) applied to all trees, (c) variable selfing (Var. s) and (d) variable selfing plus subpopulation equilibrium inbreeding 55

Table 3.5 Mean of selections and genetic gain from the top 15-ranked individuals from each model, expressed in terms of their breeding values from either the maximal model (MAX) or predicted using the other models (OM) 56

Table 4.1 MAF (minor allele frequency), average observed ($\hat{H}_O$) and expected ($\hat{H}_E$) heterozygosity, percentage of polymorphic loci, (%P) relatedness ($\overline{2\theta}$) and inbreeding coefficient ($\bar{f}$) calculated by the Wang (2007) triadic method 65

Table 4.2 Classification of SNP markers used to characterise the *E. cladocalyx* breeding population at Lismore, Victoria 69
Table 4.3 Bias from triadic estimation of relatedness using 75 simulated SNP loci with allele frequencies drawn from the *E. cladocalyx* empirical dataset. Simulated sets of 500 dyads (corresponding to the smallest *E. cladocalyx* subpopulation) and 45 dyads (corresponding to a family of 10 siblings) are presented. The 95 % CI (which apply to single-dyad estimates of 2θ and single-tree estimates of f) are calculated from 1000 bootstraps over loci 72
Table 4.4 From 500 simulated dyads (allele frequencies distribution drawn from Dirichlet distribution): (1) individual-dyad bias of 2θ and single-tree bias of f (deviation from expected values, $E[2\theta]$ and $E[f]$) and 95 % CI obtained from 100 bootstraps (over varying numbers of loci) for different relationships and (2) bias and CI of family-average relatedness ($\overline{2\theta_f}$) and inbreeding coefficient $\bar{f}_f$ obtained from bootstraps over a family of 10 siblings 45 dyads..... 73
Table 4.5 Summary of regression analysis of family-mean relatedness ($\overline{2\theta_f}$) and inbreeding coefficient ($\bar{f}_f$) against phenotypic trait means (DBH, height and percent survival at various ages) with percent variance accounted for by regression and probability that the fitted line is different to zero for groups of subpopulations in *E. cladocalyx* (where significant, the regression coefficients were all negative) 76
Table 5.1 Summary of seedlots included in the ALRTIG breeding population 86
Table 5.2 Summary site data for the ALRTIG breeding population 86
Table 5.3 Single-site estimates of heritability $\hat{h}^2$ (with approximate standard errors) from mixed mating (MM) and half-sib (HS) family assumptions, phenotypic mean and coefficient of variation (CV_p) for height (HT), diameter at breast height (DBH) and axis persistence (AP) traits 94
Table 5.4 Linear regression statistics of growth and form traits on standardised across-sites data on $\hat{f}$ 94
Table 5.6 HS and MM variance component estimates for genotype ($\hat{\sigma}_t^2$), site-by-genotype interaction $\hat{\sigma}_{t.s}^2$) and residual ($\hat{\sigma}_e^2$) terms and narrow-sense heritability (..... 98
Table 5.7 Log likelihoods from G+GxE and factor analytic models for half-sib (HS) and mixed-mating (MM) assumptions 98
Table 6.1 heat-mapped table of type-B correlations averaged across the DBH, HT and AP traits 117

List of figures

Fig. 1.1 Natural range of *E. cladocalyx* with main populations in the South Flinders Ranges, Eyre Peninsula and Kangaroo Island regions of provenance 2

Fig. 1.2 Phenogram for natural and cultivated populations of *E. cladocalyx* reproduced from McDonald et al. (2003). C10 and C11 are western Victorian cultivated stands, F*x* are South Flinders Ranges provenances, K*x* are Kangaroo Island provenances, E*x* are from Eyre Peninsula and C9 is a cultivated, dwarf form. 6

Fig. 3.1 Path diagrams for deriving coefficients of co-ancestry 47

Fig. 3.2 Narrow-sense heritability estimates ($\hat{h}^2$) modeled using the small dataset as a function of selfing proportion assuming no ancestral inbreeding (*solid lines*) and population-wide inbreeding equilibrium with $f_{eq.}$=0.25 (*dashed lines*) 52

Fig. 3.3 BLUP point (individual-tree) histograms showing the degree of shrinkage for models using individual-family selfing for height. Each *pane* compares models assuming (a) no inbreeding, (b) selfing equivalent to the population average applied to all parents, (c) variable selfing (s = var.) for each family and (d) variable selfing plus variable subpopulation equilibrium inbreeding ($f_{eq.}$ = var.). The *top pane* shows the entire population (318 trees).. ... 54

Fig. 4.1 Box and whisker diagram showing MAF from a simulated Dirichlet distribution, the South Flinders Ranges (SFR), Kangaroo Island (KI) and planted stand groups of *E. cladocalyx* subpopulations 69

Fig. 4.2 Intra-gene (from within 39 genes) LD (R^2) for wild (Cygnet River) and cultivated (Kersbrook) *E. cladocalyx* subpopulations. Logarithmic trend lines indicate rapid decay of LD 70

Fig. 4.3 UPGMA phenogram for the *E. cladocalyx* breeding population at Lismore, Victoria based on $\hat{F}_{ST}$ uncorrected genetic distances 71

Fig. 4.4 Relationship between family-average relatedness ($\overline{2\theta_f}$), inbreeding ($\bar{f}$) and fitness trait, DBH, at age 7.8-years 75

Fig. 5.1 Graphical representation of family-average relatedness ($2\hat{\theta}$) within and between families 83

Fig 5.2 Relationship between family-average inbreeding ($\hat{f}$) and fitness trait, DBH 93

Fig. 5.3 Vector plot showing principal component-transformed loadings ($\Lambda 1^*$ and $\Lambda 2^*$) from the FA2, mixed mating model for DBH and AP. 100

Fig. 5.4 Plot of family scores for the first and second loading of the FA2 MM model for HT 100

List of abbreviations

Abbreviation	Term
ALRTIG	Australian Low Rainfall Tree Improvement group
AMMI	Additive main effects multiplicative interaction
AMOVA	Analysis of molecular variance
ANOVA	Analysis of variance
AP	Axis persistence
ATSC	Australian Tree Seed Centre
AVE	Avenue
BLAST	Basic local alignment search tool
BLUP	Best linear unbiased predictor
BTN	Bordertown
COR	Corowa
DBH	Diameter at breast height
DBHOB,	Diameter at breast height over bark
DBHOB	Diameter at breast height under bark
DNA	Deoxyribonucleic acid
ESP	Esperance
FA	Factor analytic
FAO	Food and Agriculture Organization of the United Nations
GxE	Genotype-by-environment interaction
HAM	Hamilton
HS	Half sib
HT	Height
ID	Inbreeding depression
IUFRO	International Union of Forest Research Organizations
KI	Kangaroo Island
KOJ	Kojonup
LD	Linkage disequilibrium
LIS	Lismore
MAF	Minor allele frequency
MAR	Mean annual rainfall
MAT	Mean annual temperature
MM	Mixed mating
MOR	Moora
NIR	Near-infrared reflectance
NP	National Park
PCA	Principal component analysis
PGG	Predicted genetic gain
RFLP	Restriction fragment length polymorphism
RMSECV	Root mean square error of cross validation
RMSEP	Root mean square error of prediction
ROP	Region of provenance
SE	Standard error
SF	State Forest
SFR	South Flinders Ranges
SNP	Single-nucleotide polymorphism
SPA	Seed production area
SSR	Simple sequence repeat
WEL	Wellstead
WGW	Wagga Wagga

Chapter 1: General introduction

1.1 Tree crops for marginal land

As the world's population grows towards a projected 9.6 billion during the next 40 years (United Nations 2013), the demand for food, fibre and energy is expected to grow significantly. The land currently used to produce these staples will come under pressure from urban encroachment, new uses such as bioenergy crops, and be reduced and degraded by the impact of climate change (Stern 2007). Elements of solutions include increased crop productivity, and bringing land that is presently considered marginal or non-productive into use either by adaptation of existing crops or introduction of new crop species and varieties (Fischer et al. 2014).

In southern Australia, forest plantations have been established mainly where rainfall is relatively plentiful and reliable, often in landscapes co-habited by productive agriculture, and increasingly, urban development. Though the impacts of climate change in the coming decades cannot be predicted with certainty, some widely-accepted models (e.g. CSIRO 2014) predict that southern Australia's climate is very likely to become warmer and possibly drier. In recent years, water availability in the irrigated farmlands of the Murray-Darling River system has reduced dramatically (Murray Darling Basin Commission 2007). If the trend of the early part of the new millennium continues, the demand for rain-fed, arable land for food production in places such as Tasmania, the Green Triangle and other important plantation forestry centres may increase, as will competition for suitable forest plantation land in southern Australia. Meanwhile, there is heightened interest in sequestration of carbon by planting trees and addressing the dryland salinity problem (National Land & Water Resources Audit 2008) that is placing further pressure on Australia's fresh water resources: both activities need to be carried out at a very wide scale to have significant impact. There is a requirement for plantation species that are adapted to the drier climate (350 -650 mm mean annual rainfall [MAR]) of the southern sheep-wheat belt where land is relatively plentiful. Such species may also find application in existing planting areas should some climate-change predictions eventuate. One of the most prospective of a small number of species that have been identified to fill this need is *Eucalyptus cladocalyx* (sugar gum) (Harwood et al. 2007).

1.2 *E. cladocalyx* – species description

1.2.1 Distribution

Eucalyptus cladocalyx is endemic to South Australia where it occurs in four disjunct areas: two on the Eyre Peninsula and one each in the South Flinders Ranges and Kangaroo Island (Fig. 1.1). The natural range is restricted to within approximately 3 degrees each of latitude and longitude. McDonald et al. (2003), using allozyme markers, demonstrated clustering of subpopulations in three distinct regions: Eyre Peninsula, South Flinders Ranges and Kangaroo Island (see later, Fig. 1.2, section 1.2.4). These have been considered to be regions of provenance in a recent study (Callister et al. 2008).

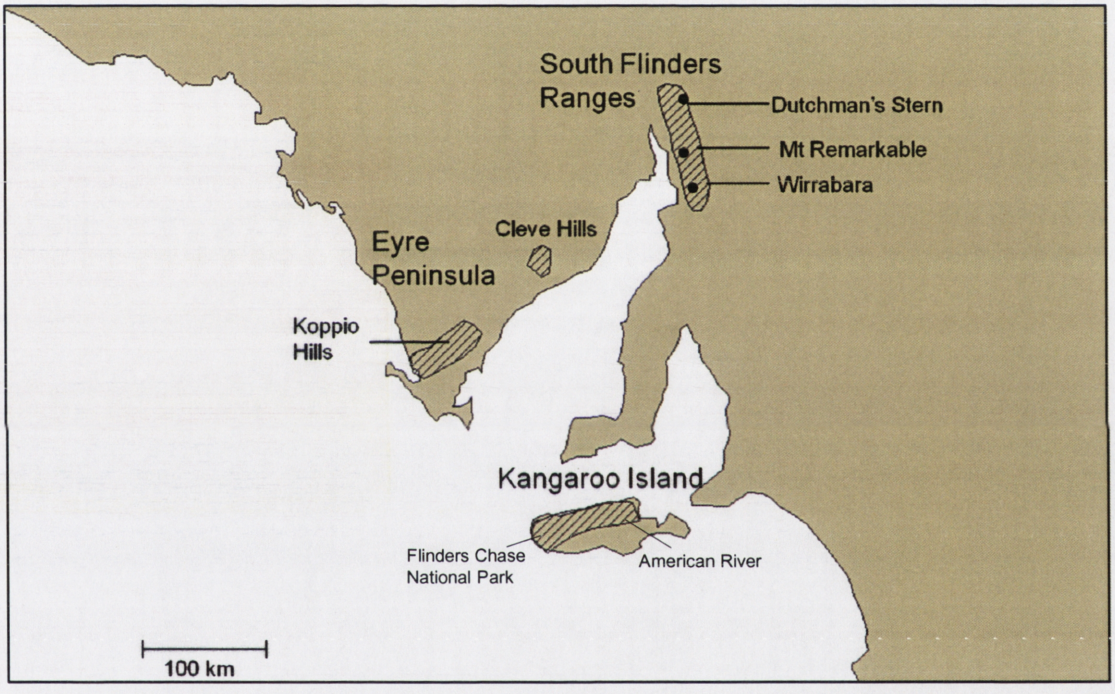


Fig. 1.1 Natural range of *E. cladocalyx* with main populations in the South Flinders Ranges, Eyre Peninsula and Kangaroo Island regions of provenance

It is likely that *E. cladocalyx* had a wider and more contiguous distribution during the Pleistocene (Crocker and Wood 1947). McDonald et al. (2003) contend that the population structure indicated by their study supports the Crocker and Wood (ibid.) theory that the existing natural range represents relict populations remaining after a larger population contracted inland as a result of increasing aridity during the early Holocene. Gene flow between the present-day regions of provenance is likely to be very low, as they are separated by long distances, and the main pollen vector is assumed to be insects. However 18 000 years ago the coastline of southern Australia was just to the south of Kangaroo Island, which was then connected to the mainland (Hope 1994), and this, in combination with a wetter climate,

may have made gene-flow between the extant populations possible via a more continuous distribution of the species between Kangaroo Island and the South Flinders Ranges.

Fragmentation of the three subpopulations is very likely to have accelerated markedly from the time of colonisation of South Australia, particularly from the 1830s onwards. Clearing of land for grazing and harvesting of timber for building probably had a major impact on Eyre Peninsula and South Flinders Ranges populations in particular. Early South Australian Woods and Forests Department maps also show that significant stands containing *E. cladocalyx* were cleared for *Pinus radiata* plantations around Wirrabara (Ednie Brown 1882), and significant disturbance of parts of the native forests of the South Flinders Ranges as evidenced by a high frequency of stumps was noted by Boomsma (1946) in the process of carrying out vegetation mapping. Though there are significant contiguous extant forests on Kangaroo Island, the species tends to be naturally confined to small stands in gullies there. In the South Flinders Ranges, major stands are now confined to the Wirrabara, Mt Remarkable and Dutchman's Stern/Mt Brown conservation areas. These areas are now separated by agricultural land, but there may have been greater geographic linkage before the 1870s when this region was colonised.

1.2.2 Taxonomy

Eucalyptus cladocalyx F. Muell. is classified within subgenus *Symphyomyrtus* and is the only species within Section *Sejunctae* (Brooker and Kleinig 2006). It has no close relatives and is distinctive within the State of South Australia for being the only eucalypt with strongly discolorous adult leaves (Nicolle 2013). It has recently been reclassified by Nicolle (2013) as three separate subspecies corresponding to region of occurrence: *E. cladocalyx* subsp. *cladocalyx* (Eyre Peninsula), *E. cladocalyx* subsp. *crassa* (Kangaroo Island) and *E. cladocalyx* subsp. *petila* (South Flinders Ranges). This classification has been made on the basis of distinctive combinations of fruit, growth habit and leaf characteristics.

1.2.3 Ecology

In its natural range, *E. cladocalyx* tends to occur on sites where the index of annual evapotranspiration is favourable (Specht 1972) relative to surrounding areas and often in hilly or mountainous terrain, except on Kangaroo Island where it is often situated near creeks (Edaphic data are summarised in Table 1.1). It occurs in two main Land Systems (*sensu* Specht 1972): *Sclerophyll*, where soils are less fertile and which have a low shrub understorey, and *Savannah*, where soils are more fertile and that support tussock and perennial grass understorey (Boomsma 1946; Specht 1972).

In the South Flinders Ranges, where substantial extant forest is situated in conservation reserves, the species occurs on ridges and hillsides of skeletal rocky soils of quartzite lithography, most commonly above 300 m altitude (Boomsma 1969). This region produces the tallest trees on average, which sometimes occur as pure stands or grow in association with *E. goniocalyx*, *E. microcarpa* and *E. leucoxylon*. This part of the species' natural distribution receives between 500 mm and up to 630 mm MAR at the Wirrabara (southern) end of the ranges and 400-500 mm at Mt Brown to the north (Schwerdtfeger and Curran 1996).

Table 1.1 Edaphic parameters for regions of provenance of *E. cladocalyx* (from Bureau of Meterology climate records, and information in Wood (1930), Boomsma (1946), Specht (1972), Schwerdtfeger (1985) and Schwerdtfeger and Curran (1996) and Australian Tree Seed Centre provenance data (CSIRO unpublished).

Region of provenance	Description of soil and/or geology	Associated spp.	Altitude (m)	MAR (mm)	Pan evapor-ation (mm)	MAT (°C)
South Flinders Ranges (upper slopes)	Quartzite-derived skeletal podsols in uplands (Sclerophyll land system)	*E. goniocalyx* or pure stands, *Acacia, Hakea, Xanthorrhoea* understorey	600	600-630	~1600	16
South Flinders Ranges (uplands)	Terra-rossa (limestone-derived) in uplands (Sclerophyll land system)		400-600	460	~1600	16
South Flinders Ranges (lower rainfall)	Yellow podzolic (Savannah land system)	*E. microcarpa, E. leucoxylon* Grass, shrubs	300-600	350	~1600	16
Eyre Pa. (Cleve Hills)	Yellow duplex (Sclerophyll land system)		200	350-375	>1000	17.5
Eyre Pa. (Koppio Hills, Wanilla)	Lateritic podzolic (Sclerophyll land system)	Often pure, *E. diversifolia, Xanthorrhoea* understorey	100-200	400-550	>1000	17.5
Kangaroo Island	Coherent sandy (skeletal) (Savannah land system)	*E. diversifolia Dodonaea, Bursaria*	5-175	480-650	1000	15

MAR mean annual rainfall *MAT* mean annual temperature

Populations on the Eyre Peninsula occur in two upland regions: the Koppio Hills to the north of Port Lincoln and in the hills to the north of Cleve. Trees from the Eyre Peninsula are generally stunted and often multi-stemmed, leading Crocker (1946) to conclude that the species is at the edge of its edaphic range there. In fact, rainfall and soils in the Koppio Hills are not significantly poorer than those in parts of the South Flinders Ranges and Kangaroo Island, receiving between about 400 and 550 mm MAR (Schwerdtfeger 1985). The Eyre

Peninsula phenotype has since been shown to be under significant genetic control (Bulman et al. 2000; Harris 2000) with low, heavily branched and forked trees being produced on favourable sites in western Victoria and the Mt Lofty Ranges. A dwarf form, (commonly designated *E. cladocalyx* var. *nana*) often cultivated for windbreaks, also originates from the Cleve Hills region (see Fig. 1.1), which receives between about 300 and 375 mm MAR (Schwerdtfeger 1985). The populations on Eyre Peninsula are now very fragmented due to clearing for agriculture.

On Kangaroo Island, the species is mainly confined to gullies along creek lines and at lower altitudes than in the rest of its range. Stands can be found very close to the sea, for example, at Cape Borda and American River. At the western end of the island, it is often situated on stabilised sand ridges in association with *E. diversifolia,* and occasionally fringes swamps in an open forest formation, while in the east and in wetter situations along protected watercourses, it can form a dominant overstorey with an almost impenetrable understorey of *Acacia, Dodonaea, Bursaria* and other shrubs (Wood 1930). Trees on Kangaroo Island can achieve considerable size, the tallest known specimen (45.5 m) from the entire natural range being located near the Ravine des Casoars at the western end of the island (Nicolle 2013).

1.2.4 Genetic diversity

McDonald et al. (2003) showed that *E. cladocalyx* is characterised by low overall levels of diversity and high levels of genetic divergence among populations. Expected panmictic heterozygosity (H_E), a common, sample-size-insensitive measure of genetic diversity, is similarly low ($H_E = 0.191$) to that of other eucalypts with disjunct distributions, and lower than is typical for those with widespread ranges (compared with data collected from a range of published diversity measures in eucalypts by Potts and Wiltshire 1997). McDonald et al. (2003) observed that there is a general tendency in *E. cladocalyx* for decreasing diversity with increasing latitude. The Wirrabara provenance, in the South Flinders Ranges, had the highest diversity estimate ($H_E = 0.221$), while Cygnet River from Kangaroo Island had the lowest ($H_E = 0.070$). An estimated 26 % of genetic diversity identified was due to differences among subpopulations, which is another common finding in eucalypts with regional and disjunct distributions (Potts and Wiltshire 1997).

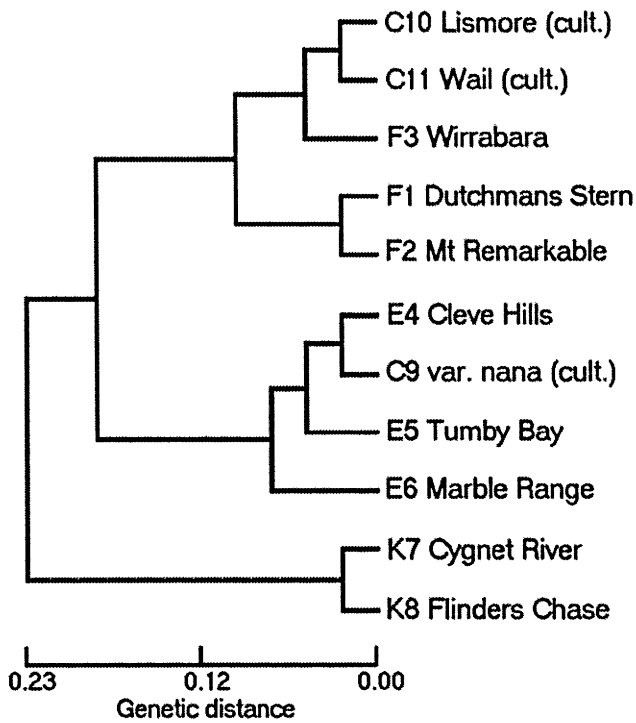


Fig. 1.2 Phenogram for natural and cultivated populations of *E. cladocalyx* reproduced from McDonald et al. (2003). C10 and C11 are western Victorian cultivated stands, F*x* are South Flinders Ranges provenances, K*x* are Kangaroo Island provenances, E*x* are from Eyre Peninsula and C9 is a cultivated, dwarf form.

McDonald et al. (2003) showed that western Victorian cultivated populations from Wail and Lismore are derivative of the Wirrabara area (Fig. 1.2) and that diversity estimates for these were slightly higher than those of the natural populations. Such differentiation is consistent with the definition of a land race (e.g. White et al. 2007, p. 188). McDonald et al. (2003) speculated that a possible cause for the higher allelic diversity in the planted stands might be that the forest was more extensive and genetically diverse in the period between the 1870s and 1900 when the seed was likely to have been collected.

1.2.5 Reproductive biology

As is commonly the case with eucalypts (Pryor 1976), individual *E. cladocalyx* flowers are morphologically bisexual and protandrous, i.e. pollen is shed before the stigma becomes receptive. Most eucalypts that have been formally investigated have been found to have a mixed-mating system (Eldridge et al. 1993), i.e. each fertilisation event may be either the result of outcrossing or self-fertilisation. Geitonogamous pollination occurs despite the protandry of individual flowers because individual flowers within trees and even with individual umbellasters develop asynchronously. The breeding system of *E. cladocalyx* including floral morphology and pistil cytology was studied by Ellis and Sedgley (1992). Their study of controlled crosses on three trees indicated that they ranged from self-

compatible to self-incompatible. By analogy to eucalypts with flowers of a similar size and colour, the species is presumed to be predominantly insect pollinated, though birds and mammals may also play a role (Harwood and Bulman 2001). Flowering occurs between January and April and seed matures around 12 months later. Seed is assumed to be dispersed in a manner similar to many other eucalypts (Florence 1996), i.e. only over short distances with the majority of germinants being situated within two tree-lengths of their mother (Potts and Wiltshire 1997). This is an important characteristic giving rise to clusters of close relatives.

1.2.6 Uses and planting

E. cladocalyx was one of the first species recognised as having potential for low-rainfall forestry in South Australia, where it occurs naturally and is planted in areas that receive between 350 and 650 mm MAR. Its plantation performance was noted soon after the Colony's establishment of the Woods and Forests Department (Ednie Brown 1882). Seed was collected from natural stands in South Australia, and it was dispersed widely throughout the British colonies and other parts of the world. General literature reviews of the species' performance (e.g. Doran 2000) indicate that it is well adapted to a wide range of site types in temperate and Mediterranean climates. This, coupled with its relatively high growth rate within the target planting region and useful wood properties, led State and Commonwealth research organisations to commence provenance and limited progeny testing of the species during the late 1980s and 1990s (Bulman et al. 2000; Harris 2000; Harwood and Bulman 2001). A more systematic approach to domestication was commenced by the Australian Low Rainfall Tree Improvement Group (ALRTIG), a consortium of Australian state and Commonwealth government forestry and tree breeding agencies, to commence a low-input breeding program for the species in 1999. The broad objective was to develop a breed of trees adapted to southern Australia's sheep-wheat belt that will produce small sawlogs within about 25 years, thereby underpinning a future industry and achieving environmental objectives such as salinity abatement and greenhouse gas sequestration (Bush et al. 2001).

The heartwood of *E. cladocalyx* is of yellow-brown to dark-brown colour while the sapwood is pale brown. Mature plantation trees are suitable for timber production including appearance grade and furniture applications (Washusen and Waugh 2000). The timber also has high hardness and wear-resistance making it suitable for flooring. Wood of mature plantation trees is of high basic density, approximately 750 kgm^{-3} (Blakemore 2004) , therefore making wood at sawlog rotation age and probably thinnings unsuitable (i.e. too dense) for pulpwood production. However the timber is naturally durable, having an in-ground heartwood durability rating of Class 1, the most durable class (Standards Australia 2005). This is a useful

attribute for exposed decking and farm or vineyard poles, especially for farms that grow organic produce where chemical preservative treatment is undesirable (Bush 2011). Production of small vineyard posts might be a useful end-product achievable within 8-10 years, with small sawlogs becoming available at around 25 years. The species is also a good nectar producer (Doran 2000) and planted stands in South Africa are considered to be a valuable resource for apiarists, even though the species has become a weed there (de Lange et al. 2013). Domestication of the species has also commenced in Chile where it is being planted in semiarid environments (Cane-Retamales et al. 2011; Mora et al. 2009) with an emphasis on nectar and pollen production for apiary, though its growth and general form traits are also being considered (Vargas-Reeve et al. 2013). The species is also undergoing preliminary field trialling in New Zealand under the auspices of the New Zealand Dryland Forestry Initiative [www.nzdfi.co.nz and see also Walker (2011)] for its potential to produce naturally-durable vineyard posts in drier planting regions.

1.3 Overview of the domestication process

Domestication describes the anthropogenic process by which the characteristics (and genetic composition) of a wild species are changed to better suit human needs. The process can be characterised by a number of steps, some of which may be unplanned or informal, typically followed by various stages of scientifically-based genetic improvement. Both informal and formal processes will change the genetic composition of the domesticated population relative to that of the wild, ancestral population. The forest tree domestication process might include some or all of the following aspects:

Informal processes

- Recognition of the potential usefulness of a wild species for a particular purpose and/or its apparent adaptation to particular regions, including exotic locations.
- Development of plantations from collected wild seeds followed by harvesting and seed collection from the plantations. Phenotypes with desirable characteristics may be selected 'by eye' both from the wild and from the derivative plantations.
- Natural selection pressures in the exotic, target planting environment which may be colder, drier, be free from pests etc. will also result in selection pressure that may lead to genetic differentiation of the population under domestication.

Such introductions, under a combination of anthropogenic and natural selection in exotic and/or modified plantation environments have in many cases resulted in the formation of *land races*: populations genetically distinct from the wild progenitors with specific adaptation to the local environment (White et al. 2007).

Formal processes (see also Namkoong et al. 1988; White et al. 2007)

- Systematic provenance testing, perhaps including a greater range of provenances than were originally introduced to a region, as well as subpopulations for which domestication is already underway via the informal processes described above.
- Systematic testing of a common set of genetic materials (e.g. provenances, families, clones) across a range of sites.
- Resolution of the genetic structure of the species (identification and definition of genetically distinct subpopulations).
- Formation of breeding populations as part of a formal breeding program with specific target traits and breeding objectives.
- Selection and controlled mating involving genetic materials selected for their superiority in one or more traits.
- Estimation of genetic parameters such as heritability, genetic correlation among traits and genetic diversity as a basis for breeding program decision-making.
- Studies of generative processes such as mating system, flowering phenology, outcrossing rates and other factors that will impact on the breeding program and the overall rate of genetic change that drives domestication.

The initial steps of formal domestication might be compared to laying the keel of a ship: correct actions taken early will maximise gain in later generations, whereas inclusion of an inadequate diversity of material or taking ill-informed or incorrect decisions will take considerable effort to remedy later. However there are several challenges when commencing formal domestication that make life for the tree breeder difficult, the biggest of which is lack of prior information such as:

1. Details of the reproductive system (e.g. flowering synchronicity, inbreeding/outbreeding propensity) of the species may not be well understood;
2. The genetic structure (e.g. definition of genetically distinct subpopulations) of the species may not be known;
3. The origins and genetic diversity of land-race materials are unlikely to be known;
4. Range-wide testing of provenances may not have occurred; and
5. Traits to focus on may not be known or industry may change or diversify their product needs and/or plantation management practices in the future.

Obtaining such information should be a high priority, because it will inform techniques that can be used to maximise genetic gain. When formal domestication of *Pinus radiata*

commenced in Australia (see Table 1.2 for timeline), tree breeders did not have all of the relevant information available to them, though they were in a position to be relatively well informed about the origins of the land race which was in fact quite genetically diverse and formed a sound initial breeding population (Wu et al. 2007). Selections of plus trees on the basis of growth and tree form from this resource provided significant genetic gain. While some efforts were made to study wood properties, traits relating to structural properties were not given the emphasis that they should have received (Burdon 2010), requiring later modification of breeding strategies (Jayawickrama and Carson 2000; Johnson 1989; Powell et al. 2004). A significant problem that radiata breeders in Australia and New Zealand now face is that of an adverse genetic correlation between growth and traits related to structural wood properties (Gapare et al. 2009; Jayawickrama 2001). Unfortunately, selection for rapid growth in the domestication phase combined with a change in management to shorter rotation lengths (partly enabled by the more-rapid growth), and in New Zealand, wider spacing (Burdon 2010), has led to an undesirably large proportion of low-strength, juvenile wood or corewood. Considerable effort is now being expended to remedy this (Gapare et al. 2012; Wu et al. 2007).

Of considerable advantage to today's tree breeders is the availability of more sophisticated and cheaper wood-property determination techniques. The problem of adverse genetic correlations between growth and wood property traits in *P. radiata* did not occur because it wasn't recognised when formal domestication commenced that wood structural properties would be important in future: *P. radiata* has always been used for structural timber in Australasia. The problem occurred because it was difficult and expensive to screen large populations for wood structural properties. The domestication and breeding plan for *E. cladocalyx* (Harwood and Bulman 2001) also focuses entirely on growth, form and apparent environmental adaptation, despite the fact that the species is likely to be used for structural purposes. This is less likely to be a problem for *E. cladocalyx* than for *P. radiata,* even if negative correlations exist, because the timber strength is likely to be extremely high by current commercial standards. Nevertheless, the relative ease of measuring wood properties now means that even a low-input program can afford to make basic determinations of key wood-related traits.

Another advance in tree breeding technology has been the development of readily available molecular markers. These can be used in a number of useful ways to provide information that will inform the domestication process. One application is the study of 'provenance' collections of seed from the wild range to define genetically distinct subpopulations. Another use is analysis of land races to determine their genetic diversity and ancestral origins. These

determinations were made retrospectively for *P. radiata,* as the technology became available (Moran and Bell 1987), and commenced soon after the establishment of the first generation breeding populations for *E. cladocalyx* (McDonald et al. 2003).

Table 1.2 Some important domestication events for *E. cladocalyx* and *P. radiata* in Australia (Harwood and Bulman 2001; Wu et al. 2007 and see also specific references given in table)

Domestication event	Comment	*E. cladocalyx*	*P. radiata*
Dispersal of seed	Value of both species for plantations recognised in about the 1860s-70s. *P. radiata* was only sourced from the Californian provenances, *E. cladocalyx* only from the South Flinders Ranges. Quite large amounts of *P. radiata* and *E. cladocalyx* were exported from the wild, though in both cases the range of collections was restricted.	1870-1900	1860-1880
Plantations established	*E. cladocalyx* from the S. Flinders Ra. was planted as hundreds of km of farm windbreaks in western Victoria. Many of these were then managed on coppice rotations, or seed was collected from the plantings for re-establishment. Natural and some human selection in these new environments probable, commencing development of land race. *P. radiata* plantation seed was still being sourced from descendants of original imports until 1960. There were 30 000 ha of plantations of radiata in South Australia by the 1930s	1870-1900	1880 onwards
New provenances identified	The Mexican island provenances of *P. radiata* were recognised nearly 100 years after the Californian ones were introduced to Australia. Foresters did not recognise that *E. cladocalyx* came from the Eyre Pa. and Kangaroo Island until after 1890 (Ednie Brown 1882), before which time much seed had been dispersed. Even after their recognition, these provenances were generally not widely dispersed or used for plantations until much later (probably mid-20th C) when Eyre Pa. material became popular for farm windbreaks.	1880	circa 1960
Formal domestication commences	Provenance testing of *E. cladocalyx* in Australia indicated the poor suitability of Eyre Pa. provenances for solid wood production (Bulman and Fairlamb 1998; Bulman et al. 2000). In Australia, tests of *P. radiata* cuttings in 1937 demonstrated the importance of genetics.	1980	1937
Progeny testing and commencement of formal breeding programs	The *E. cladocalyx* breeding program comprised 137 families across 11 sites. *P. radiata* plus trees were derived from the original importations made in the 1870s. The first *P. radiata* orchard was established in 1957, with 300 progeny tests established by the 1970s. Testing of 600 *P. radiata* plus trees commenced in the 1980s	2001	1950s, 1980s
Genotype by environment testing	Genotype-by-environment testing in *P. radiata* is now an ongoing part of the breeding program	2001	1969-70
Infusion from the wild	For *P. radiata* a large collection was made in 1978 - before this time the breeding program was based on material introduced in the 19th Century (Wu et al. 2007). This material is only slowly being integrated into the breeding program.	Not done	1978
Molecular analysis of population structure	Initial analysis of diversity in wild and land race *E. cladocalyx* was carried out by McDonald et al. (2003) and of *P. radiata* by Moran and Bell (1987). *P. radiata* analysis ongoing, e.g. Dillon et al. (2013)	2003	1987
Determination of wood property data and integrating into the breeding program	In *P. radiata* the importance of wood quality was recognised early, but the problem of adverse correlations between growth and wood quality was not tackled in earnest until 2003. Wood property data collection for *E. cladocalyx* commenced in 2006 (Chapter 2 of this thesis).	Data collected 2006 but not yet integrated into breeding program	Progressively after 1990, became priority after 2003
Second generation		2015?	1992
Third generation		2025?	2003

A further application now becoming accessible, as the price of genotyping individual trees has fallen, is the provision of information about the co-ancestry of progeny derived from unpedigreed mothers (or pedigreed only as far as the identification of their subpopulation) (Blouin 2003; Pemberton 2008). This allows verification and/or reconstruction of the traditional records-based pedigree.

1.4 Wood properties and interactions between traits

A common breeding objective for virtually all forest tree species is improved growth – larger trees are wanted in less time. Growth is a 'general purpose' trait that makes sense for the majority of products, ranging from essential oil to sawlogs. This together with the fact that physical size attributes of trees such as height and stem diameter (often along with stem and branch form traits) are cheap and easy to measure explain why growth is a major focus in first-generation breeding programs. One commonly-taken approach is to focus on growth in the first generation and then to look at other traits in addition, especially those that are more expensive to measure such as wood chemical and physical properties, in second and subsequent generations (e.g. Namkoong et al. 1980). A possible problem with this approach is the potential existence of adverse genetic correlations between growth and other traits. Trait-trait correlations, also termed Type-A correlations are calculated as (Burdon, 1977) :

$$r_A = \frac{\hat{\sigma}_{t_x t_y}}{\sqrt{(\hat{\sigma}^2_{t_x}\hat{\sigma}^2_{t_y})}} \qquad (1.1)$$

where $\sigma_{t_x t_y}$ is the additive genetic covariance component between traits, and $\sigma^2_{t_x}$ and $\sigma^2_{t_y}$ are the additive variance components for traits *x* and *y* respectively. If an adverse genetic correlation exists, selection for improvement of one trait will result in loss of genetic gain in the other. A good example is the adverse type-A correlation between growth and strength-related traits caused by higher proportions of juvenile core wood in fast-grown, genetically-improved *P. radiata* (Gapare et al. 2009). Rapid and relatively low-cost techniques (see Downes et al. 1997) including non-destructive sampling of trees using powered coring equipment and wood physico-chemical analysis using tools such as near-infrared reflectance (NIR) have improved and become cheaper in recent years and have the potential to help address the problem of early screening of wood properties. This should allow assessment of the magnitude and direction of type-A correlations in wood property traits likely to be important in later-stage breeding programs. The relatively low costs of these technologies make such analyses achievable even for quite modestly-resourced programs.

1.5 Accounting for mixed mating and unknown relatedness

In first generation breeding programs, including that of ALRTIG for *E. cladocalyx*, open-pollinated progeny of trees from wild and planted stands are typically assembled as a breeding population in a series of progeny trials (Namkoong et al. 1988). The pedigree is very shallow in that only the identities, but not ancestry, of mothers common to open-pollinated families are known. However most eucalypts including *E. cladocalyx* have a mixed-mating system; i.e. varying proportions of their progeny will be the product of outcrossing between individual trees with others that are unrelated, as well as an inbred proportion attributable to selfed mating. In fact, due to the presence of clusters of near relatives, there will also be another form of inbreeding among relatives termed 'neighbourhood inbreeding' that can be modelled as 'effective selfing' (Ritland 1984). Another factor that will increase relatedness among progeny, but not inbreeding, is a low effective number of pollen parents that will give rise to significant proportions of full-sibs. Though eucalypts are generally thought to be preferentially outcrossing (Eldridge et al. 1993), the presence of inbred progeny causes problems in estimation of genetic parameters such as narrow-sense heritability and breeding values because the usual assumption that open-pollinated progeny will share one quarter of their genes (e.g. Falconer 1981) is flawed.

The strategy chosen by ALRTIG of testing open-pollinated families of *E. cladocalyx* is very commonly used in tree domestication programs (White et al. 2007; Williams et al. 2002). Typically such trials are measured and statistical models are used to partition the total phenotypic variance as:

$$\sigma^2{}_P = \sigma^2{}_A + \sigma^2{}_E \qquad (1.2)$$

Where $\sigma^2{}_A$ and $\sigma^2{}_E$ are additive and environmental genetic variance components, respectively. Though non-additive components of genetic variance will also be present, they are not separately estimable from OP progeny tests. In traditional family models, $\sigma^2{}_A$ will be equal to the variance among families (σ_f^2) multiplied by the inverse of the coefficient of relationship $(1/\rho)$. The coefficient of relationship is a measure of the proportion of genes that two individuals have inherited directly from a common ancestor (Wright 1922). For half-sib families, $1/\rho = 4$ and $\sigma_A^2 = 4\sigma_f^2$. Narrow sense heritability is then estimated as:

$$\hat{h}^2 = \frac{(1/\rho)\hat{\sigma}_f^2}{\hat{\sigma}_P^2} \qquad (1.3)$$

In fact, the value of $1/\rho$ will be below four if some proportion of the progeny are selfs or result from matings with near relatives. Another likelihood is that there will be a proportion of full-sibs present, for which $1/\rho = 2$. It is therefore usual to apply a value of $1/\rho$ that will effectively correct for the likelihood of relationships that are closer than half-sib (e.g. Squillace 1974). A value ranging between 3, which is commonly used for pines (e.g. Adams and Joyce 1990; Squillace 1974) and 2.5 which is commonly used for eucalypts (Eldridge et al. 1993) is applied. However, a review by Hodge et al. (1996) showed that values between 4 and 1.85 have been used for eucalypts, and a value of 4 has been selected by some authors in recent eucalypt domestication studies (e.g. Apiolaza et al. 2011; Cane-Retamales et al. 2011; Denis et al. 2013; Henson et al. 2007; Mora et al. 2009; Pelletier et al. 2008; Raymond et al. 2008; Vargas-Reeve et al. 2013). A value of 2.5 corresponds to progeny within families that are 70% outcrossed (Griffin and Cotterill 1988; Volker et al. 1990). Values of $(1/\rho) < 4$ will effectively adjust heritability estimates downward for the whole population. This may result in acceptable heritability and breeding value estimates, provided that selfing and inbreeding levels are quite heterogeneous within and among subpopulations.

The homogeneity of outcrossing rates among and within families of eucalypts is not always a valid assumption, especially when the mating has occurred in the wild. McDonald et al. (2003) showed that there is a large amount of variation in selfing and inbreeding levels among families within provenances in *E. cladocalyx* (and that selfing incidence is, overall, quite high in the species). Choosing a globally applied coefficient of relationship may therefore lead to inflation of breeding values for some families and provenances, and underestimation in others. This may in turn lead to bias and error in estimation of genetic parameters such as heritability, type-A and type-B correlations and breeding value ranks.

The development of more sophisticated models and computing power in the last decade has presented an opportunity for more detailed modelling of relatedness via defined pedigrees. The individual-tree model, which is a forestry analogue of the animal model (Mrode 1996), allows estimation of the additive genetic variance among individual trees (σ^2_t) in populations where heterogeneous relatedness is present. It is implemented within the best linear unbiased prediction (BLUP) model framework (Henderson 1949; Henderson 1975) and estimation is typically by restricted maximum likelihood (REML) (Patterson and Thompson 1971). The individual-tree model has been used in estimation of breeding parameters of later-generation breeding populations, for which traditional pedigrees can be constructed from

existing records (Baltunis et al. 2008; Borralho 1995; Costa e Silva et al. 2006; Gapare et al. 2008; Jarvis et al. 1995).

In the first generation, pedigree information is nearly always limited. If wild families of open-pollinated progeny are being tested, the only available information will be that each family has a common mother, and the identity of the provenance to which she belongs. It is generally assumed that the mothers are unrelated, which may be a valid assumption provided the spatial sampling of wild mothers has been carried out to avoid inbreeding and coancestry. In the case of land-race selections, there are no assurances about relatedness, and even quite extensive stands may be derivative of only one or a small number of ancestors. Under these circumstances, it is not possible to construct a records-based pedigree, and it is often assumed, as an analytical starting point, that the progeny of each mother are half siblings, and that no two mothers are related.

1.6. The role of molecular markers

One potential way to resolve absent pedigree information in first-generation breeding populations is to estimate relatedness using molecular markers (Blouin 2003). Molecular markers have been used effectively to determine missing pedigree information in a number of tree species (Gaspar et al. 2009; Gea et al. 2007; Kumar and Richardson 2005; Lambeth et al. 2001). El-Kassaby et al. (2011) have gone so far as to propose that it may be possible to carry out "breeding without breeding", whereby artificially manipulated matings are substituted for by naturally occurring offspring whose relatedness and parentage is resolved using molecular markers. A potential advantage of molecular markers over traditional, records-based pedigrees is the opportunity to identify and correct pedigree errors, which will almost certainly be present in breeding populations (e.g. Adams et al. 1988; Doerksen and Herbinger 2008; Kumar and Richardson 2005; Munoz et al. 2014) due to the complexity of their establishment and large numbers of trees. A further advantage is that if markers are sufficiently dense, variations in the proportion of genes that are identical by descent among individuals due to Mendelian sampling can be accounted for (Hill and Weir 2011). Though the price of genotyping has fallen considerably in recent years, the cost of screening the large numbers of trees associated with breeding populations appears to have precluded the widespread use of molecular markers for pedigree reconstruction. Though the use of markers for resolving relatedness in first-generation tree breeding populations has been quite feasible for some time, they do not appear to have been employed for this purpose outside of this study.

1.7 Genotype-by-environment interaction

That certain plant varieties and animal breeds will thrive in some environments while in others, alternative breeds should be substituted, has been known since ancient times. The formal development of the theory that the phenotype of an individual is influenced by both its genetic make-up and the environment to which it is exposed gained some early momentum in the nineteenth century, for example Galton's (1874) recognition of 'nature and nurture' in connection with the inheritance and social and environmental influences on human intellect and achievement, and solidified in the 1930s with statistical constructs, building on the discovery of ANOVA (Fisher 1918), for example Yates and Cochrane (1938) who presented the theory for partitioning variance among varieties, sites and years. In forestry, interest in GxE started in the mid-1960s and early 1970s (e.g. King 1965; Morgenstern and Teich 1969; Shelbourne 1972) as new methods developed by agricultural scientists (e.g. Eberhart and Russell 1966; Finlay and Wilkinson 1963; Wricke 1962) became available. The establishment of coordinated, international species-provenance testing programs in the 1970s by organisations such as FAO and IUFRO (e.g. Burley and Kemp 1972; Doran and Boland 1978; Greaves and Kemp 1977) also generated increased interest in the subject and need for improved techniques.

There are two important aspects associated with GxE: *identifying* whether or not it exists, and then *characterising* the nature of the interaction, whether, for example, certain genotypes or environments are interactive and whether responses can be extrapolated to other environments and/or genetic material. GxE can be of two main forms (i) scale-change GxE, where genotypes perform differentially on different sites, but genotype ranks do not change appreciably; and (ii) rank-change interaction, where genotype ranks do change (White et al. 2007). In the latter case, it is possible to sort genotypes into those that are stable (ranking performance for a particular trait does not vary widely from site to site) and those that are unstable.

GxE is an important consideration in the domestication process because its characteristics might be used to better manage the breeding and deployment populations. Of particular practical importance is whether genetically-improved material can be widely deployed across the entire target planting zone, or alternatively, specific breeds might need to be developed for specific site types (Annicchiarico 2002). It is important to recognise that domestication programs typically involve testing of provenances, as well as families nested within these, across a range of sites. GxE can then be further classified as provenance-by-site (PxE) and family-by-site interaction. Strong PxE as indicated by rank changes in provenances are usually minimal unless associated with large changes in climate or soils (Matheson and

Raymond 1986). When many families are planted across sites within a region and GxE is statistically significant for growth rate it is commonly found that (i) the GxE is usually due to scale effect as opposed to dramatic rank changes (ii) if there are rank changes these are very often due to a small number of unstable families, or, (iii) sometimes only one or two of the locations are very different and these cause much of the observed interaction (Matheson and Cotterill 1990; White et al. 2007).

GxE can be quantified and characterised by numerous methods that vary in complexity. Methods have evolved in the last 20 years from traditional ANOVA approaches to more sophisticated mixed models (Smith et al. 2005). The latter can be implemented within the BLUP/REML framework, used with unbalanced data including an incomplete and partially unconnected set of genotypes across sites which is very common in forestry trials, and can integrate individual-tree models involving complex pedigrees (see section 1.5).

1.7.1 Traditional GxE model

The traditional model used to characterise GxE in forestry within the framework of the family model (Kempton 1984; Williams et al. 2002) is:

$$y_{ij} = \mu + g_i + e_j + g.e_{ij} + \varepsilon_{ij} \qquad (1.4)$$

where y_{ij} is the i^{th} family on the j^{th} site, μ is the overall mean, g_i and e_j are family-additive and site main effects, respectively, $g.e_{ij}$ is the interaction effect between site and family and ε_{ij} is the residual or error term. This approach can be used in the traditional ANOVA framework, where all terms would be considered fixed, or in a mixed model framework, where some or all of the terms might be considered random (Smith et al. 2005). In the domestication context, for example, the families are likely to be considered random as they will likely be quite numerous (>50) and randomly sampled to represent the species (or subpopulations), whereas sites may be relatively few and selected to represent specific zones and are therefore best represented by a fixed term. This model assumes that all GxE effects have the same variance, GxE effects for different genotypes are uncorrelated and GxE effects for different pairs of sites have the same variance, covariance and therefore correlation. This simple variance structure is sometimes referred to as the compound symmetric structure (Smith et al. 2005). A further assumption is homogeneity of residual variance, though this is not likely to hold true requiring analysts to rescale data to meet this condition. Another drawback is that it provides no insight into the nature of the GxE interaction (Kempton 1984).

To attain insight into the nature of the GxE interaction, a variety of more complicated approaches have been taken. In general reviews of GxE analysis, Smith (2005) and Kempton (1984) summarise methods of GxE pattern analysis as involving either (i) principal components analysis (PCA) or (ii) regression onto either an environmental variable or onto the marginal mean of the GxE table. Responses are described as a series of multiplicative terms, each being a product of a genotype and an environment effect. Both regression and principle component approaches have been taken in analysis of forestry GxE.

1.7.2 Regression models for GxE

The most commonly used regression method for partitioning GxE in forestry (commonly referred to as "joint regression") was introduced by Yates & Cochrane (1938), made popular by Finlay & Wilkinson (1963) and further developed by Eberhart and Russell (1966). Digby (1979) provided an iterative strategy for fitting a joint regression model to data where some genotypes are missing from some sites, and this strategy was refined by Ng and Williams (2001). Joint regression models the $g.e_{ij}$ effects as a function of the environment main effect (e_j):

$$y_{ij} = \mu + \beta_i e_j + g_i + \varepsilon_{ij} \qquad (1.5)$$

where β_i is the regression coefficient of the genotype *i*. If there is no GxE interaction, the β coefficients will all be equal to one, a family with a regression coefficient close to one having average stability. When coupled with a high family mean, the family is considered to have good general adaptability whereas a low family mean indicates poor adaptation to all sites. A family with $\beta_i > 1$ responds to improvements in site quality (and has low stability). $\beta_i < 1$ indicates greater resistance to change in environments (high stability) and therefore greater specificity of adaptability to unfavourable environments. A further advantage of this method is that an informative graphical representation is available by plotting genotype means against genotype regression coefficients. This approach has been taken in a large number of forestry studies over the past four decades (e.g. Isik et al. 2000; Matheson and Raymond 1984; Morgenstern and Teich 1969; Owino et al. 1977). Matheson and Raymond (1986) noted that forestry experiments tend to differ to those of agricultural crops in that β_i usually increases with g_i, and there are usually no genotypes specifically adapted to favourable environments.

1.7.3 AMMI model

The most popular application of PCA to GxE analysis is known as AMMI (additive main effects and multiplicative interactions), teaming an ANOVA of main effects with double-centred PCA (Gauch 1992). PCA (or singular value decomposition) is applied to the matrix of residuals (including the interaction) from the additive, main effects ANOVA model. The underlying idea is that most of the interaction variance can be explained by the first few terms of the PCA (Smith et al. 2005). The model extends from Eq. 1.4 as:

$$y_{ij} = \mu + g_i + e_j + \Sigma_k \lambda_k \gamma_{ik} \delta_{jk} + \rho_{ij} + \varepsilon_{ij} \qquad (1.6)$$

where λ_k is the singular value for axis k (for $k...n$ axes), γ_{ik} is the genotype eigenvector value for axis k, δ_{ik} is the environment eigenvector value for axis k, ρ_{ij} is the residual and ε_{ij} is the error with $\varepsilon_{ij} = y_{ij} - \mu_{ij}$ as PCA is applied to the data minus the genotype and environment means, so the grand mean must be added back. This method has several advantages and disadvantages. A major advantage is that the results can be graphically represented in biplots or variations thereof (Gauch 2006). These allow the analyst to visualise the nature of the interaction between sites and genotypes. A serious drawback is that the model cannot be implemented in the modern mixed-modelling environment, and data must therefore be close to completely balanced – a rare occurrence in forestry trial series. This, and the relative complexity of the method, appears to have limited its widespread use in forestry with only a few examples having been produced (Campbell 1992; Falkenhagen 1996; Wu and Ying 2001).

1.7.4 Type-B correlations

Another important way to quantify GxE that is commonly employed in tree breeding is to treat a single trait measured at different sites as if they were different traits. This idea was originally proposed by Falconer (1952) and further developed by Burdon (1977) who termed the correlation between a trait at two different sites *Type-B correlation* which is calculated as:

$$r_B = \frac{\hat{\sigma}_{g_{xy}}}{\sqrt{(\hat{\sigma}^2_{g_x} \hat{\sigma}^2_{g_y})}} \qquad (1.7)$$

Where $\hat{\sigma}_{g_{xy}}$ is the covariance between the same trait measured at sites x and y and $\hat{\sigma}^2_{g_x}$ and $\hat{\sigma}^2_{g_y}$ are the genetic variances of the trait at sites x and y respectively. The type-B correlation can be used to gauge whether GxE effects are likely to be of practical importance. If r_B is one

then there is no GxE between sites. Values of $r_B < 0.67$, which correspond to interaction variance that is >50% of the genotypic variance, are often considered to indicate GxE that will be of practical consequence to tree breeders (Shelbourne 1972).

An advantage of modern mixed-modelling approaches is that an additive variance-covariance matrix with separate estimates of additive genetic variance for each site and covariance for each pair can be simultaneously calculated for all sites and genotypes. This is called the fully unstructured (US) model (Gilmour et al. 2009; Smith et al. 2005) and it results in a complete table of type-B correlations among sites being readily calculable. This method of assessing GxE is now commonly employed in forestry studies (e.g. Araújo et al. 2012; Brawner et al. 2011; Chen et al. 2014; Míguez-Soto and Fernández-López 2012) as it has all of the advantages of the BLUP model framework discussed in previous sections. A disadvantage relative to the joint regression and AMMI approaches is that the method is not as informative about the interaction of specific site and genotype combinations. A further disadvantage is that the US model involves estimation of a large number of parameters and thus is computationally demanding, especially if an individual-tree model is being fitted. A positive-definite variance-covariance matrix is unlikely if there are more than a few sites, especially if they have high genetic correlations, and thus convergence may not be achievable (Meyer 2009; Smith et al. 2005).

1.7.5 Factor analytic models

The factor analytic (FA) mixed model was developed around 15 years ago (Piepho 1998; Smith et al. 2001; Smith 1999) and provides extended utility for analysis of GxE in a mixed-model setting. In the most usual case, the FA model postulates dependence of the genotypes on a set of k random hypothetical factors where k is less than the total number of sites: this results in fewer parameter estimates and converges more readily than the US model (Smith et al. 2001 and see further explanation of theory in Chapter 5). The FA model is thus described as a multiplicative model of environment and variety coefficients that are referred to as loadings and scores, respectively (Smith et al. 2005). Alternatively, it is possible to relate the environments to the factors (Piepho 1998). The FA approach is, in some respects, an analogue of AMMI (Smith et al. 2005), though it has the considerable advantage of being implemented in the mixed-model/BLUP environment (de Resende and Thompson 2004). The outputs include both a complete table of Type-B correlations that can be calculated from the factor loadings and the opportunity for graphical analysis: for example a biplot of the genotype scores and factor loadings corresponding to the first two factors can be interpreted to give information on the reactivity of genotypes with respect to certain sites. The FA technique has been used in a number of forestry GxE analyses (Costa e Silva and Graudal 2008; Costa e

Silva et al. 2006; Cullis et al. 2014 ; de Resende and Thompson 2004; Hardner et al. 2011; Ogut et al. 2014). Its use is explored further in Chapter 5.

1.8 Research objectives and motivation for study

This thesis aims to explore the options for maximising potential genetic gain in the early stages of formal domestication using *E. cladocalyx* as a model species. When originally conceived, the study had two foci. The first of these was investigation of the genetic control of wood properties, especially those related to natural wood durability. This research was motivated by a need to find potential economic uses for pole-sized thinnings and/or short-rotation harvesting options, as it was recognised by the authors of the ALRTIG breeding strategy (Harwood and Bulman 2001) that the geographic situation of the target plantation zone (i.e. inland, distant from ports) and likelihood of unsuitable, high-density wood would make pulpwood production unfeasible. The second was assessment of GxE for growth and form traits using the network of open-pollinated provenance/progeny trials planted in 2001 by ALRTIG's partners in Western Australia, South Australia, Victoria and New South Wales.

The publication of a preliminary study of GxE for growth and form traits on three ALRTIG sites in south-west Western Australia by Callister et al. (2008) raised some serious questions about traditional approaches to genetic parameter estimation for this species. The study revealed unusually-high narrow-sense estimates of heritability for growth traits ($\hat{h}^2 = 0.40 - 0.85$), despite the authors' inclusion of fixed-effect terms to remove variance attributable to subpopulations and use of a coefficient of relationship of $\rho = 1/2.5$ to correct for inbreeding and full-sib relationships resulting from mixed mating. It was suggested that differential inbreeding among families and provenances, that had been identified by McDonald et al. (2003), may be a possible cause of the apparently highly inflated heritability estimates.

The findings of Callister et al. (2008) prompted an expansion of the objectives of the study to investigate and, if possible, address the potential impact of heterogeneous relatedness and inbreeding arising from mixed mating on quantitative genetic parameter estimates. The general approach conceived was to use molecular markers to make estimates of selfing and relatedness for individual subpopulations and families and to use these to modify the traditional assumption of half-sib relatedness within each family. This work was conducted in two stages. Firstly, a pilot-scale study was implemented. This study involved re-analysing the scored isozyme assays of McDonald et al. (2003) that included a small number of families common to the ALRTIG breeding program. This allowed development of the necessary theoretical framework for the second-stage analysis. This second, larger study firstly entailed

genotyping mature trees from the ALRTIG breeding population to estimate relatedness within and among families, as well as estimating other population parameters for the breeding population and providing information on the origins and genetic diversity of the land-race selections. This was followed by an analysis of GxE across 11 sites for the entire breeding population, using both the traditional assumption that open-pollinated families are half-sib and an alternative model that made use of marker-based estimates of relatedness and inbreeding to reconstruct the pedigree and account for inbreeding depression.

1.9 Thesis structure

The thesis is presented as six chapters, four of which are facsimiles of published, peer-reviewed journal papers. They appear here in their published form. Small changes have been made to the formatting and some symbols etc. for consistency between the different journal styles and the rest of the thesis. The formatting of the thesis complies with The Australian National University Fenner School of Environment and Society guidelines for a Thesis by Compilation. A brief description of each chapter follows:

Chapter 1: General introduction, research objectives and chapter outline

Chapter 1 provides background and context for the following four chapters. The species is described and features of its biology, uses and the Australian domestication and breeding program are introduced. Important aspects of the domestication process are described. These include the examination of wood properties, with reference to the domestication of *P. radiata*; accounting for departures from half-sib relatedness in analysis of progeny trials that comprise open-pollinated families; the use of molecular markers to provide information to underpin quantitative analysis, and; methods for undertaking GxE analysis. Chapter 1 also explains the underlying motivation and the development of the research project.

Chapter 2: Genetic variation of natural durability traits in *Eucalyptus cladocalyx* (sugar gum) (Bush et al. 2011b)

This chapter focuses on genetic variation in *E. cladocalyx* wood property and growth traits at age 8 years from a single provenance/progeny site that was established at Benalla, Victoria prior to the ALRTIG trial series. It was sampled because the trees were of the correct size for pole production at the time the study was carried out. The relationship between wood properties related to durable pole production including heartwood proportion, extractive content, resistance to biodeterioration and density and growth traits were determined. The use of near-infrared reflectance for rapid screening of some of these traits is also examined. Type-A genetic correlations between pairs of traits are calculated. Overall implications for the domestication and breeding program are discussed.

Chapter 3: Marker-based adjustment of the additive relationship matrix for estimation of genetic parameters—an example using *Eucalyptus cladocalyx* (Bush et al. 2011a)

This chapter applies estimates of family and provenance-level selfing from a previously published (McDonald et al. 2003) isozyme assay of *E. cladocalyx* natural populations composed of open-pollinated families and planted stands to construct an additive relationship matrix ('**A** *matrix*'). The modified **A** matrix is then used to estimate individual-tree breeding values for a subset of progeny belonging to provenances and families common to both the ALRTIG breeding population and the McDonald et al. (2003) study. The results are compared with additive variance estimates from a 'standard' analysis using an **A** matrix reflective of half-sib relationships and modified using various coefficients of relationship. The implications of BLUP breeding value re-ranking between the alternative models is assessed.

Chapter 4: Characterising a *E. cladocalyx* breeding population using SNP markers (Bush and Thumma 2013)

Population genetic parameter estimates for the majority of families in the ALRTIG breeding program sampled at age five years are made using SNP (single-nucleotide polymorphism) markers. Determination of relatedness and inbreeding among individuals and families is carried out using the *Coancestry* program of Wang (2010). Population genetic parameter estimates are compared with those previously made by McDonald et al. (2003) using isozyme markers to assess very young seedling material. The loss of homozygous individuals in the breeding population normally expected due to selection against deleterious recessive alleles in eucalypts by five years of age (Potts and Wiltshire 1997) and heterozygosity-fitness relationships are investigated. The underlying genetic structure of the wild and land-race constituents of the ALRTIG breeding population are elucidated.

Chapter 5: Genetic parameter estimates informed by a marker-based pedigree: a case study with *Eucalyptus cladocalyx* in southern Australia (2015)

Though indications prior to the establishment of the ALRTIG breeding program were that *E. cladocalyx* is widely adaptable, the actual plantation performance within the sheep-wheat planting zone were largely untested and unknown beyond information that can be gleaned from scattered plantings of various ages with often unknown history of management. Testing the species throughout the planting region was therefore a priority for ALRTIG. The sheep-wheat belt itself is a large geographic region, covering circa 13 degrees of latitude, and its western and eastern extents are separated by some 37 degrees of longitude. Within this region, there is significant variation in rainfall seasonality, which ranges between winter-dominant and uniform, and also a wide range of soil types. This chapter quantifies and characterises the genotype-by-environment interaction of growth and form traits across 11 sites in southern

Australia. Provenance-level variation as well as single- and across-sites estimates of heritability as well as genetic correlations between traits are explored. The calculations are made using both the traditional pedigree, which assumes families are half-sib (the HS model), and with a modified linear model that accounts for the effects of mixed mating (the MM model), including a pedigree constructed using the method introduced in Chapter 3 and the marker data introduced in Chapter 4. The MM model also accounts for inbreeding depression that was identified in some sections of the breeding population. The GxE model is implemented within the factor analytic mixed model framework, allowing calculation of type-B genetic correlations and visualisation of the interaction of sites and genotypes.

Chapter 6: Discussion and conclusions

The final chapter of the thesis draws together the findings of the four separately published papers in the context of the overall study. The impact of the findings of each research chapter are discussed in terms of their implications for the *E. cladocalyx* domestication program, and makes concluding general recommendations for future domestication and tree improvement strategies.

Chapter 2: Genetic variation of natural durability traits in *Eucalyptus cladocalyx* (sugar gum)

Annals of Forest Science (2011) 68:1057–1066

David Bush[1,2*] Kevin McCarthy[1] and Roger Meder[1]

1. Commonwealth Scientific and Industrial Research Organisation, Australia

2. Australian National University

* Corresponding author: CSIRO Black Mountain Laboratories, Clunies Ross Street Black Mountain Canberra ACT, Phone +61 2 6246 4829, david.bush@csiro.au

Received: 6 December 2010 / Accepted: 10 May 2011 / Published online: 25 August 2011

Handling Editor: Christophe Plomion

Abstract

- ***Introduction*** We present a study on genetic variation in natural durability traits of young-aged *Eucalyptus cladocalyx,* a species adapted to temperate, low rainfall regions. Our motivation was production of naturally durable posts for applications such as vine trellises, a sector dominated by heavy metal preservative-treated wood in some of the world's main wine-producing countries.
- ***Methods*** Stem diameter at breast height over- and under-bark, heartwood proportion, wood density, methanol extractives and fungal decay were assessed in a progeny test on a set of 48 families from 8 provenances nested within three regions-of-provenance (ROP) from the species' natural range. Near- infrared reflectance (NIR) was examined as an efficient assessment method..
- ***Results*** Strong variation among ROP was indicated for all traits, with low-moderate narrow-sense heritability for growth traits and extractives content and moderate-high heritability of basic density and decay-resistance to two of the three fungi. Trait-trait correlations ranged from low to high, with basic density and extractives content being negatively correlated to decay mass-loss.
- ***Discussion*** NIR was an effective predictor of methanol extractives, moderately effective for basic density, but unsuccessful for fungal decay. Generally, there were no practically adverse correlations between growth and durability traits.
- ***Conclusion*** Substantial genetic variation in natural durability traits is indicated, with wide scope for genetic improvement.

Keywords
Natural durability, *Eucalyptus cladocalyx*, Near infrared reflectance, Genetic variation

Introduction

Natural durability is a wood property allowing it to resist biodeterioration caused by bacteria, fungi, termites, borers and marine organisms without treatments of preservative chemicals or coatings. Certain species are highly resistant to biodeterioration and are recommended for use in untreated applications, while others are known to be highly susceptible. Natural durability is increasingly regarded as a desirable wood property obviating the need for costly and potentially harmful chemical preservative treatments. This study was motivated by the desire to produce untreated vineyard trellis posts. In Australia, Chile New Zealand and South Africa, copper-chrome-arsenate-treated softwood posts are typically used: untreated softwood[1] is not sufficiently durable to last the 25 years in-ground that is required of a vineyard post. Similarly motivated research is also underway in Europe to examine the natural durability of eucalypts suited to Mediterranean climates (Lorenzo et al. 2007; Palanti et al. 2010) and softwood species for use in durable applications.

There is often significant diversity in resistance to biodeterioration within species (Hillis 1987) and a component of this variation may be genetic, and possibly amenable to improvement through tree breeding. Genetic variation in natural durability traits has been studied for a number of taxa including *Larix* spp. (reviewed by Gierlinger et al. 2004), *Picea glauca* (Yu et al. 2003), *Pinus sylvestris* (e.g. Ericsson and Fries 1999), *P. taeda* (Schmidtling and Amburgey 1982), *Quercus* spp. (Mosedale et al. 1996) and *Tectona grandis* (Kjaer et al. 1999). Until recently, *Eucalyptus* species, which are very variable in their natural durability properties (Rudman 1966), have been relatively unstudied except in relation to heartwood/sapwood ratio (e.g. dos Santos et al. 2004; Nicholls and Matheson 1980) and extractive content (Poke et al. 2006; Washusen et al. 2001), the latter usually in connection with species suited to pulping. Recently Palanti et al. (2010) found significant variation in fungal decay resistance among a small number of *E. grandis* and *E. camaldulensis* x *botryoides* clones.

Standards Australia (2005) assigns durability ratings for timbers in above-ground, in-ground and marine applications, based upon the resistance of the outer heartwood of mature timbers. The sapwood of all species is considered non-durable, and for many, including the eucalypts,

[1] *Chamaecyparis nootkatensis* is the only listed Class-1 durable softwood listed in the Australian standard, though it is not commercially available in quantities or at a cost suitable for vineyard poles

the inner heartwood and pith is also of lower durability than the outer heartwood (Rudman 1966). The Standard lists the in-ground durability of numerous *Eucalyptus* species commonly used for building and construction as well as non-Australian species. Timbers are assigned to one of four classes: Class 1, where timber remains sound in-ground for greater than 25 years, e.g. *Tectona grandis*; Class 2, 15-25 years, e.g. *E. camaldulensis*; Class 3, 5-15 years, e.g. *Thuja plicata*, *E. globulus* and; Class 4, 0-5 years, e.g. *Pinus radiata*, *Pseudotsuga menziesii*.

Amongst the more durable eucalypt species is *E. cladocalyx* F. Muell. (sugar gum), with in-ground durability recently reclassified from Class 2 to Class 1 (Standards Australia 2005). It is an open-forest and woodland species endemic to South Australia. A breeding program has commenced in Australia because of its suitability for low rainfall sites where it might be planted for production of timber and poles as well as environmental remediation (Harwood et al. 2007) and it is also being genetically improved in Chile (Mora et al. 2009). It occurs in three geographically disjunct regions, two situated in mainland Australia and one on Kangaroo Island (see Table 2.1 for geographic coordinates). Clustering of three subpopulations or regions-of-provenance (ROP) within the species, corresponding to these regions has been demonstrated using allozyme markers (McDonald et al. 2003) and Callister et al. (2008) demonstrated that growth and form traits are broadly similar among provenances within regions. It was formerly used in exposed applications requiring high strength and durability such as bridge and railway timbers, though its use declined after the 1940s with the widespread availability of low cost copper-chrome-arsenate treated *Pinus radiata* timber and other materials.

The factors responsible for the differences in durability among and within species are numerous and diverse. High density timbers have relatively low void volume which may inhibit permeation of gases and liquids, though within-species correlation between density and decay resistance has been shown to be sometimes positive (e.g. in *E. delegetensis,* Wong et al. 1983) but also absent or negative in other species (see Yu et al. 2003). The most important factor is the presence of extractives, the quality and quantity of which both influence durability. Toxic polyphenolic extractives act as chemical barriers (e.g. Mohareb et al. 2010) and tyloses act as physical barriers limiting access to the transpiration system (Beckman 2000). The amount of extractives has been shown to be correlated with decay resistance in both softwoods (e.g. Gierlinger et al. 2004) and hardwoods including eucalypts (Wong et al. 1983). An axial core of extractives is deposited as sapwood is transformed into heartwood, and it is therefore only the heartwood of 'durable' species that are actually durable. It is this part of the wood that would be used in pole production.

This study examined the genetic variation in durability properties of *E. cladocalyx* using samples taken from a common-garden, provenance-progeny trial situated at Benalla, Victoria. We have examined durability-related traits including heartwood-sapwood ratio, heartwood extractive content and resistance to biodeterioration from three soil fungi. We also investigated the use of near-infrared reflectance (NIR) to rapidly and cost-effectively screen samples for these traits.

Materials and Methods

Sample population and procedures

Data and samples were collected in August 2006 from an *E. cladocalyx* provenance-progeny trial situated at Benalla, Victoria, established in 1998. The site is temperate (Köppen-Geiger Cfb), receiving 720 mm mean annual rainfall. All families included (Table 2.1) were collected by the Australian Tree Seed Centre (ATSC) as representative random samples from wild populations. The trial incorporates 73 families from seven provenances that came from three regions-of-provenance defined by the findings of McDonald et al. (2003): South Flinders Ranges, Eyre Peninsula and Kangaroo Island.

The trial was established as four replicates of five-tree row-plots in randomised complete blocks. An additional four replicates of randomised five-tree plots including a subset of families were also established. The trial was planted at a density of 1235 stems per hectare.

Table 2.1 Composition and sampling detail of the 8-year old *E. cladocalyx* provenance-progeny trial at Benalla, Victoria

Regions of provenance and provenance	ATSC seedlot number	Latitude	Longitude	No. families in trial	Families sampled	Samples taken (trees)
South Flinders Ranges						
6.7 km NE Wirrabara Forest	16013	33°02'S	138°12'E	11	6	26
Wilmington	19348	32°42'S	138°06'E	10	7	35
9 km S Wilmington	16089	32°43'S	138°06'E	10	10	48
Eyre Peninsula						
Marble Range	19349	34°30'S	135°31'E	10	5	21
6.1 km ESE Wanilla	16018	34°33'S	135°44'E	10	5	20
Kangaroo Island, South Australia						
Flinders Chase National Park 1	16022	35°45'S	136°38'E	12	11	55
Flinders Chase National Park 2	19717	35°57'S	136°44'E	10	5	21
Total	**7**			**73**	**49**	**226**

The diameter of all trees in the trial (n=1440) was measured in August 2006 at which time it was unthinned. It is estimated that the trial had reached the canopy closure stage in circa 2003.

Families were selected for wood property sampling on the basis of their diameter growth ranking, with larger-diameter families ranking highest. Trees were preferentially selected within families on the basis of potential suitability (sufficient diameter and straightness to a height of 3 m where possible) for a vineyard post at the time of sampling. As this trial was based entirely on unselected, unimproved material, it was not always possible to find enough trees with these characteristics, and so trees having sufficient diameter but not stem straightness were selected in some cases.

Coring

Two 12 mm diameter, bark-to-bark cores were removed from each tree at breast height. All cores were taken from the same (north-south) orientation, and the northern end was marked on each core along with the individual tree identity.

Heartwood-sapwood boundary definition and proportion

One core from each pair was stained along its length with dimethyl yellow, an pH indicator that can be used to sharply delineate the heartwood-sapwood boundary of eucalypts (Hillis 1987). The heartwood boundary was traced from the stained core to the unstained core. The core sapwood and heartwood lengths were measured to give heartwood and sapwood proportions of total diameter at breast height under bark (DBHUB). Diameter at breast height over bark (DBHOB) was determined in the field using a diameter tape, and bark diameter and thickness was determined by subtracting DBHUB from DBHOB of the cored stem.

NIR spectroscopy and model calibration

Due to the high cost of extractive and accelerated biodeterioration analysis, NIR was investigated as a potential measurement tool. NIR can be highly effective for determining extractive content in particular (Bailleres et al. 2002; Poke et al. 2005). NIR spectra for each sample were obtained from woodmeal produced by grinding with coarse then fine Wiley Mill screens (200 micron final granule size) heartwood portions of a sample of the second, unstained set of cores. Woodmeal was contained in a quartz vial and spectra acquired using the integrating sphere on a Bruker MPA (Bruker Optics, Ettlingen, Germany, www.brukeroptics.com). Spectra were acquired between 10 000 and 4 000 cm^{-1} (1 000 – 2 500 nm) at 8 cm^{-1} spectral resolution.

Multivariate calibration was performed for each of the wood traits using The Unscrambler v9.8 (www.camo.com).

Basic density

Stained cores were soaked in water for 24 h to achieve fibre saturation. Each core was then partitioned into three sections, two sections of sapwood and the heartwood. The volume of each core section was then accurately determined by the water displacement method (TAPPI 1989). The core segments were then oven dried at 103 ± 1 °C for 24 h and weighed. Basic density for the sapwood and heartwood of each tree's core sample was determined as the oven-dry weight divided by the volume for each segment.

Accelerated biodeterioration

Following basic density determination, the stained heartwood segments were prepared for accelerated decay testing. Prior to exposure to decay fungi, specimens were sterilised by gamma (γ) irradiation. Fungal strains are denoted by the DFP strain number [National Herbarium of Victoria, MEL]. Three fungi, two brown-rot (*Gloeophyllum abietinum [13851]* and *Fomitopsis lilacino-gilva [1109])* and one white-rot (*Perenniporia tephropora [7904]*) were used to evaluate the core specimens. Glass jars (250 ml) were filled with 136 g fresh, damp soil and moistened with 14 ml to 100 % water holding capacity. One poplar sapwood veneer feeder-strip (40 x 45 x 2 mm) saturated in 1 % malt extract solution was placed on top of the soil in each jar. Metal lids were used to close the jars which were then sterilised by autoclaving and allowed to cool. Feeder strips were introduced into the jars and inoculated with actively growing mycelium of the test fungi and incubated at 25°C. When the feeder strips were fully colonised, two sterile test specimens were placed on top. A replicate set of jars was left uninoculated as a sterile control. Specimens were also placed into these jars to determine any mass loss or gain not attributable to fungal attack. All jars were incubated at 25°C for 12 weeks. Specimens were removed from the jars and wiped of any adhering mycelium, then weighed to determine moisture content, and left to air dry for 12 days. Specimens were then dried in vacuum ovens at 40°C and -95 kPa for five days, weighed, adjusted to accommodate any changes recorded in the sterile controls, and percentage mass losses determined.

Extractive percentage weight determination

Methanol extractives (percentage weight) were determined for a subset of 124 cores by the method described in Standards Australia (2003) to enable NIR calibration to be performed. Ground wood samples previously used in the NIR analysis were oven dried at 103 ± 1 °C for 24 hours and placed in an air-tight desiccator until cool. Once cool, 2.000 ± 0.100 g of wood granules from each sample was weighed into a cellulose extraction thimble. Each sample was

placed in a Soxhlet apparatus charged with 180 ml of methanol and extracted for 6 hours. The methanol extract was rotary evaporated under reduced pressure until circa 5 ml of extract remained. It was transferred to a watch glass of known weight, and placed in an oven at 70 °C for 12 hours to fully evaporate the solvent. The watch glass was removed from the oven, allowed to cool in a desiccator and then weighed to determine the mass of extractive as a percentage of the heartwood mass.

Statistical analysis

Restricted maximum likelihood (REML) analysis of variance in each measured trait was carried out using a general linear mixed model of the form:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e} \qquad (2.1)$$

where **y** is the vector of observations on *n* traits, **b** and **u** are vectors of fixed and random effects respectively, **X** and **Z** are incidence matrices for fixed and random model terms and **e** is a vector of random residual terms. Variants of this model were applied as follows:

1. For estimation of significance of effects and effect means

The vector **b** (Eq. 2.1) contained sub-vectors for fixed effects of replicate, region-of-provenance, and provenance effects, and **u** contained sub-vectors for the random effects of families nested within provenances, plots and incomplete blocks (rows and columns of trees superimposed within blocks) in the whole-trial DBHOB analysis. The incomplete block and plot terms were dropped from analyses of traits where variance components were small, negative or with large standard errors. In practice this was virtually all of the wood property traits. Wald tests were performed for fixed effects with approximate F statistics and corresponding numbers of residual degrees of freedom as implemented in Genstat version 12 (VSN International, Hemel Hempstead, UK).

2. For estimation of variance components and functions

ASReml version 2.0 (VSN International) was used to estimate genetic variance components, calculate heritability and their standard errors using a first-order Taylor series expansion to approximate the variance of a ratio of variances. Heritability of each trait was estimated using univariate analyses and bivariate analyses estimated correlations between traits. Families were modelled as nested within provenances. Incomplete blocks and plots were modelled without covariance, family effects were modelled with covariance and heterogeneous errors for each trait were included to determine genetic correlations. Narrow-sense heritability (family-within-provenance) was estimated for each trait as:

$$\hat{h}^2 = \frac{2.5\hat{\sigma}_f^2}{\hat{\sigma}_f^2 + \hat{\sigma}_e^2} \quad (2.2)$$

where $\hat{\sigma}_f^2$ is the variance of half-sib families and $\hat{\sigma}_e^2$ is the error variance and 2.5 represents a coefficient of relationship (ρ) of 0.4 which appears to be suitable for this species given its relatively high levels of inbreeding (Bush et al. 2011a). For the DBH trait, the plot variance component was also added to the denominator of Eq. 2.2.

Results

Diameter and heartwood proportion

The mean DBHOB of whole- and core-sample-population stems (n=1440, n=222 respectively) was 12.8 and 13.2 cm respectively. There were significant differences for this trait (p<0.001) between ROP (regions-of-provenance), provenances and families-within-provenance for both sample populations (Table 2.2). The coefficient of phenotypic variation (CV_P) of the whole-trial population was relatively lower for the selected trees (18 %) than for the whole population (29 %). The heritability estimate for the whole population was 0.26 (s.e. 0.08) but was significantly higher for the (biased) cored sample 0.59 (s.e. 0.08).

There were significant differences (p<0.001) at the ROP and family levels for bark thickness; South Flinders trees had significantly thicker bark than either Kangaroo Island or Eyre Peninsula (1.4 > 1.1 ≈ 0.9 cm radial thickness respectively) corresponding to 11-18 % of DBHOB. There was no significant difference in heartwood and sapwood symmetry along the axis in the north-south direction in which the cores were taken.

There were significant differences (p≤0.007) in heartwood diameter (expressed as proportions of both DBHOB and DBHUB) among ROP and families (nested within provenances within ROP), but not among provenances (Table 2.2). South Flinders Ranges provenances have significantly more heartwood as a percentage of DBH than the Kangaroo Island material, and slightly more than the Eyre Peninsula provenances. Heritability (family-within-provenance) was estimated at 0.38 and 0.30 for heartwood percentage of DBHOB and DBHUB respectively, with CV_P 28 and 29 % respectively. Mean heartwood DBH for the sampled trees was 72 mm, with Kangaroo Island and South Flinders Ranges ROPs having 58 and 82 mm respectively.

Basic density

Significant differences among ROP exist for heartwood, sapwood and whole-of-core BD (p<0.001), though the significance of the effect at the family level is higher for heartwood than sapwood. South Flinders Ranges material tended to be denser than Kangaroo Island and Eyre Peninsula. Sapwood was less dense (639 kg m^{-3}) than heartwood (673 kg m^{-3}) for all provenances with mean wood BD 661 kg m^{-3}. Heartwood BD was moderately heritable $\hat{h}^2$ = 0.41 (0.10) whereas sapwood BD was less so $\hat{h}^2$= 0.23 (0.12). Phenotypic variation in this trait was low: CV_P = 5-6 %.

Methanol extractives

Mean methanol extractive content for the sample set was 12 %, with strong region-of-provenance and provenance differences (F Pr. ≤0.003), but non-significant family-within-provenance effects (F pr.=0.06). The South Flinders Ranges and Eyre Peninsula provenances had a much higher percentage of heartwood extractives than the Kangaroo Island provenances. Heritability was estimated at 0.25 and CV_P was 23 %. All 11 of the Kangaroo Island families are ranked among the bottom 14 of 49 families for this trait, the lowest family mean extractive content being 6.3 %.

Resistance to biodeterioration

The mass-loss results indicate differential responses of provenances to the three different fungi, though overall, mass loss was low to moderate, with better families from the South Flinders Ranges having less than 5 % mass loss for all three fungi. The most aggressive fungus, *F. lilacino-gilva [1109],* was a brown rot. The Kangaroo Island provenances were the least decay resistant. Heritability estimates for *P. tephropora* [7904] and *G. abietinum* [13851] were high (0.52 and 0.58 respectively) while the estimate for *F. lilacino-gilva* was low at 0.13, with a high standard error (0.14). While there was a large amount of variation among provenances for decay resistance to *F. lilacino-gilva* (appearing as a highly significant F probability at the ROP stratum in Table 2.2), there is little family-within-provenance variation.

Phenotypic and genetic correlations

Phenotypic (r_P) and genetic (r_G) correlations (Table 2.3) among the three fungal mass-loss estimates were all positive, with *P. tephropora* [7904] having strong r_G with *F. lilacino-gilva* [1109] and *G. abietinum* [13851]. Though r_P between 1109 and 13851 was significant and moderate, r_G was lower than between the other two fungi and had a high standard error.

Table 2.2 Trait means, significance of effects (F probabilities), heritability estimates ($\hat{h}^2$) and coefficient of phenotypic (CV_P) variation of an *E. cladocalyx* trial at 8 years of age at Benalla, Victoria

Provenance	DBHOB (whole trial) (cm)	DBHOB (core samples) (cm)	Heart % of DBHUB	Heart % of DBHOB	Bark DBH [2 x radial thickness] (cm)	Basic density heart (kgm^{-3})	Basic density sap (kgm^{-3})	Basic density wood (kgm^{-3})	% heart mass-loss *(F. lilacino-gilva)* [1109]	% heart mass-loss *(P. teph.)* [7904]	% heart mass loss *(G. abietinum)* [13851]	Methanol extractives (% of heart weight)
South Flinders Ranges												
S Wilmington	12.5	16.4	65	54	2.9	718	687	707	6.3	3.9	2.7	11
Wilmington	12.1	16.2	63	53	2.7	701	679	693	8.3	3.5	2.9	11
Wirrabara	14.4	15.7	64	54	2.5	716	680	702	9.7	4.8	3.9	12
Eyre Peninsula												
Marble Range	11.0	14.8	57	51	1.7	683	653	670	12.0	5.2	2.8	12
Wanilla	12.1	13.3	57	48	2.2	671	644	661	11.8	4.9	3.4	11
Kangaroo Island												
Flinders Chase NP 1	14.5	15.8	42	36	2.0	681	684	683	14.5	13.0	9.3	7
Flinders Chase NP 2	12.5	14.6	43	37	2.2	710	695	701	13.7	7.2	5.0	9
Grand mean	*12.8*	*13.2*	*62*	*51*	*2.5*	*673*	*639*	*661*	*11.9*	*6.13*	*2.4*	*12*
SED	0.70	0.81	3.9	3.6	0.37	13	13	12	2.9	2.4	1.8	0.72
ROP	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Prov. within ROP	<0.001	<0.001	0.272	0.221	0.213	0.004	0.605	0.064	0.500	0.020	0.020	0.003
Family within prov.	<0.001	<0.001	0.007	<0.001	<0.001	0.002	0.04	<0.001	0.092	<0.001	<0.001	0.057
$\hat{h}^2$ (SE)	0.26 (0.08)	0.59 (0.08)	0.30 (0.12)	0.38 (0.11)	0.43 (0.11)	0.41 (0.10)	0.23 (0.12)	0.39 (0.11)	0.13 (0.14)	0.52 (0.11)	0.58 (0.10)	0.25 (0.13)
$\%CV_P$	29	18	28	29	49	6	6	5	75	115	128	23

SED standard error of difference, *ROP* region of provenance

Table 2.3 Phenotypic r_P and genetic correlations r_G (standard error) between pairs of traits in 8-year-old *E. cladocalyx*

	Decay 13851	Decay 7904	Decay 1109	Basic density of heartwood	Extractive	Heartwood % of DBHUB	heartwood % of DBHOB	DBHOB
Phenotypic correlation r_P								
Decay 7904	0.27**							
Decay 1109	0.38**	0.45**						
Basic density heartwood	-0.30**	-0.21*	-0.29**					
Extractive	-0.50*	-0.31**	-0.23*	0.19*				
Heartwood % of DBHUB	-0.36*	-0.15	-0.34**	0.25*	0.41**			
Heartwood % of DBHOB	-0.32*	-0.06	-0.28**	0.23*	0.31**	0.91**		
DBHOB	0.12	0.00	-0.21*	0.08	-0.34**	0.10	0.07	
Bark thickness	0.00	-0.15	-0.12	0.03	0.07	0.09	-0.30**	0.40**
Genetic correlation r_G								
Decay 7904	0.95 (0.12)							
Decay 1109	0.33 (0.44)	1.04 (0.28)						
Basic density heartwood	-0.67 (0.20)	-0.54 (0.14)	0.09 (0.46)					
Extractive	-0.61 (0.27)	-0.44 (0.28)	-0.14 (0.55)	0.20 (0.31)				
Heartwood % of DBHUB	-0.61 (0.14)	-0.78 (0.10)	*ND*	-0.19 (0.33)	-0.00 (0.20)			
Heartwood % of DBHOB	-0.53 (0.16)	-0.70 (0.12)	*ND*	-0.11 (0.29)	-0.01 (0.36)	0.98 (0.02)		
DBHOB	0.15 (0.24)	-0.40 (0.20)	-1.02 (0.27)	-0.16 (0.25)	-0.42 (0.26)	0.50 (0.21)	0.44 (0.21)	
Bark thickness	-0.20 (0.27)	-0.57 (0.23)	-0.21 (0.46)	-0.08 (0.28)	-0.19 (0.33)	-0.34 (0.33)	-0.54 (0.24)	0.39 (0.21)

ND Not determined due to non-convergence of REML algorithm

** $p<0.001$, * $p<0.05$ (phenotypic correlations significantly different from zero)

Trait-trait correlations involving 1109 could not be determined in some cases, and had high associated standard errors in others – variance was mainly partitioned at the provenance level for this trait as discussed in the preceding section. However, there was a very strong negative r_G between 1109 and DBHOB (i.e. larger trees are less susceptible to decay). Decay and heartwood extractive percentage were moderately strongly and negatively correlated to heartwood basic density for both of 13851 and 7904 (i.e. denser wood and wood with a higher extractive content were less susceptible to decay). Heartwood basic density was weakly negatively correlated with heartwood proportion, DBHOB and bark thickness. Heartwood basic density was moderately genetically correlated with both mass loss (negatively) and extractive content (positively) for 13851 and 7904.

NIR Calibration

Calibration models for the prediction of physico-chemical properties from the NIR spectra of the woodmeal are presented in Table 2.4. The extractives content of the cores can be predicted from NIR spectra of the ground wood. The absolute error of prediction is ± 0.009 % which corresponds to a relative error of 8.8 % of the mean extractives value. While both first and second derivative spectra provided similar results, the second derivative spectra were used as they provided convergence of the model with fewer principal components.

The calibration of basic density for the heartwood was able to be determined at a modest level of accuracy (RMSEP 31 %).

Table 2.4 Calibration statistics for NIR Partial Least Squares (PLS-1) calibration models giving PCs (principal components), R^2 (coefficient of determination of calibration and validation), RMSECV and RMSEP

	Treatment	No. PCs / outliers	R^2 (calib.)	RMSECV (%)	R^2 (valid.)	RMSEP (%)
Heartwood methanol extractives (% w/w)	2nd derivative	4/0	0.933	0.007	0.88	0.009
Density (kgm^{-3})	2nd derivative	3/3	0.48	24	0.15	31
Decay (Individual)						
- 1109	1st derivative	5/4	0.43	3.6	0.27	4.1
- 1109	2nd derivative	3/3	0.46	3.6	0.28	4.2
- 7904	2nd derivative	2/1	0.24	3.5	0.07	3.9
- 13851	2nd derivative	2/4	0.44	3.9	0.32	4.3
Decay (Combined)						
- 1109	2nd derivative	4/2	0.47	3.6	0.27	4.2
- 7904	2nd derivative	4/2	0.19	3.9	0.08	4.1
- 13851	2nd derivative	4/2	0.35	4.1	0.22	4.6

RMSECV root mean square error of cross-validation, *RMSEP* root mean square error of prediction

Correlation of NIR spectra with the decay data as percent mass loss was only slightly successful for one of the decay fungi [1109] (*F. lilacino-gilva*), based on the coefficients of determination (R^2) and the root mean square error of prediction (RMSEP) (see Appendix 1 for calibration plot). Two approaches were modelled, one using each individual decay fungus as the response variable and one using the three decay fungi as a multivariate response. In both instances the error in prediction of the mass loss for 1109 was ± 4.2 % (or 38 % of the mean). While the coefficient of determination and error in prediction for *G. abietinum* was similar to that of 1109 for the individual calibration (R^2 = 0.44 and RMSEP 4.3 %), the error as a percentage of the mean decay is 109 %.The calibration statistics for *P. tephropora* [7904] were similarly poor with an R^2 of 0.34 and a RMSEP of ±3.9 % (68 % of the mean).

Discussion

Growth and heartwood diameter analyses show that significant heartwood was present at breast height at eight years (72 and 88 mm in the whole and core-sample populations respectively), to indicate that a round post of 75 mm large-end heartwood diameter could probably be produced within 8-10 years from this site. Moreover, it would appear that heartwood proportion is a moderately heritable trait with a high CV and there is also significant scope for provenance-level selection. Similar findings have been made for *Pinus sylvestris* ($\hat{h}^2$ = 0.3-0.5) (Ericsson and Fries 1999) and *E. grandis* ($\hat{h}^2$ = 0.39) (dos Santos et al. 2004) heartwood. We did not find that DBHOB and heartwood proportion were strongly genetically correlated (r_G = 0.44), a finding at odds with Paques (2001) who reported r_G = 0.87-0.92 between various *Larix* spp. radial growth and heartwood. The South Flinders provenances (particularly Wirrabara) would be the target of selection at the provenance level as they have both good diameter growth and a higher proportion of heartwood on average. While selection for heartwood proportion would be an important trait for production of small poles, it would be less so for later-rotation sawlog production, as the proportion of sapwood to heartwood would be expected to decline significantly, and it is likely that the CV of heartwood proportion would decrease over time.

Significant genetic variation in basic density was found, the heartwood being denser than the sapwood, and the Kangaroo Island and South Flinders Ranges ROP being significantly denser than the Eyre Peninsula. However, the range of difference is only circa 20-30 kg m^{-3} around a grand mean of 660 kg m^{-3} (CV_P=6 %), making all of the young-aged wood quite dense: the least-dense Kangaroo Island material would still be too dense for pulp production for example. The prediction of this trait using NIR was only moderately successful (+/- 31 kgm^{-3}), a result similar to that of Schimleck et al. (2001). NIR would be insufficiently precise for resolving the small differences between genotypes. Genetic correlations between basic density and two of the fungal decay results (13851 and 7904) were negative and moderate. Yu et al. (2003) found a similar negative correlation for basic density and *Gloeophyllum trabeum* decay in *Picea glauca*, but in reviewing the literature on several other softwood species, noted that both positive and negative phenotypic and genetic correlations between these traits have been reported.

The very good NIR prediction (R^2 = 0.88) for methanol extractives makes it a quick and cost-effective method of assay, a result similar to that previously determined for *E. globulus* (R^2 = 0.89) (Poke et al. 2005). The overall extractive content (12 %) in *E. cladocalyx* is high compared with less-durable species such as *E. globulus* (typically circa 5 %, e.g. Poke et al. 2006). It is often stated that resistance to biodeterioration is as a direct result of extractive content (e.g. Standards Australia 2005). This is

supported by the moderately strong (and negative) genetic correlations between extractive content and biodeterioration from 13851 and 7904 and the correlation between predicted extractives and mass loss for 1109 (Table 2.3). Estimated heritability of methanol extractives was moderate and imprecisely estimated at 0.25 ± 0.13 (c.f. 0.37 ± 0.23 in *E. globulus*, Poke et al. 2006), and CV_P was also moderate. However the variation among ROP and provenances was marked, and selection at this level would be an important first step in breeding programs.

The heritability of resistance to fungal decay, based on the three chosen fungi, was estimated to be high, at least for two of the fungi (7904, 0.52 ± 0.11 and 13851, 0.58 ± 0.10), and there is also significant scope for improving this trait by provenance or ROP selection in all three. By comparison, Yu et al. (2003) estimated moderate $\hat{h}^2 = 0.21$ and 0.27 for *P. glauca* decay resistance to brown and white rots respectively. An interesting finding was the very strong negative genetic correlation between DBHOB and decay from 1109 and moderately strong correlation for 7904. Bearing in mind that genetic parameter estimates for 1109 were generally of poor precision, with most of the variance at the provenance level, this phenomenon might be explained by very aggressive attack of this fungus on core samples of small diameter. It is well known that the pith region of the heartwood is less resistant to decay than the outer heartwood (Hillis 1987). In preparing samples for decay testing from the 12 mm cores, outer heartwood was taken preferentially, but in small-sized trees, increasing proportions of inner heartwood were necessarily included. This may have contributed to the more rapid decay of the Kangaroo Island material, as it tended to have a smaller heartwood core, though extractive content within that core, of which there was less on average, is also likely to have a bearing on decay rates.

NIR was not a successful tool for predicting decay, with R^2 in the validation populations only 0.07-0.32 for the three fungi. Natural durability is a complex trait, resulting from wood physical as well as chemical properties (Hillis 1987). It is likely that NIR-predicted durability would be more precise in cases where more of the durability was attributable to clearly-NIR-predictable chemical substances, such as the extractives. However, as it stands, the NIR has best-predicted decay from 1109 which had r_G of only -0.14±0.55 between decay and methanol extractives, and performed worst for 7904, which had r_G = -0.61±0.27 for this pair of traits. It is disappointing that NIR was a poor predictor of decay, because the soil jar technique is expensive, labour and space intensive, presenting a practical limit on the scale of studies that can be performed. All of the reviewed studies involving accelerated biodeterioration traits including our own have been on relatively small numbers of trees (100-400) from very small samples of families or clones (20-50) nearly always sampled from a single site. While this is adequate for identifying genetic variation and coarse estimation of variance components, larger studies would be needed in a later-stage breeding program to provide precise and unbiased estimates

of genetic parameters. Clearly a low-cost predictive tool like NIR would be very useful if accurate. In contrast to our finding, Gierlinger et al. (2003) achieved R^2 of 0.84 and 0.94 with RMSEP = 0.08 (both) for *Coniophora puteana* and *Poria placenta* decay of *Larix* spp., indicating that the technique can be highly useful in at least some conifers.

A very encouraging finding is that the South Flinders Ranges provenances and families within this ROP tended to have the highest heartwood proportion, the strongest resistance to decay and higher extractive content than the Kangaroo Island provenance material tested. This has important implications for breeding programs for this species. If naturally durable and/or increased young-age heartwood production is a priority, it would be advantageous to screen the populations for some of the traits studied here, or as a lower cost alternative, reduce the proportion of Kangaroo Island-origin individuals in breeding populations.

Conclusion

This study has shown that there is considerable genetic variation in natural durability-related wood properties, including heartwood/sapwood ratio, extractive content and resistance to biodeterioration (soil jar test) in *E. cladocalyx*. The use of NIR for determination of heartwood extractives was highly effective, but the technique was not sufficiently precise for accurate determination of fungal mass-loss. Strong variation among populations indicates that selection of the correct provenances, particularly those from the South Flinders Ranges region would result in significant genetic improvement. With the caveat that heritability estimates here are based on a relatively small sample of families within provenances from a single site, and that there was bias towards sampling of larger trees and families, there would also appear to be good prospects for genetic improvement through recurrent selection and breeding within populations. Though precision of estimation was in many cases low, unfavourable genetic correlations between growth and durability traits are not indicated. It has previously been observed that there is considerable within-species variation in natural durability in *Eucalyptus*, and this study indicated that a significant proportion of the variation may be genetic.

Acknowledgements

We would like to acknowledge the useful comments of two anonymous reviewers on a draft of the paper. Paul Macdonell, Luke Bulkeley, Nick Ebdon, Brioni Brammall and Jenny Carr assisted with field and laboratory work. The study was jointly funded by Rural Industries Research and Development Corporation and CSIRO.

Chapter 3: Marker-based adjustment of the additive relationship matrix for estimation of genetic parameters – an example using *Eucalyptus cladocalyx*

Tree Genetics & Genomes (2011) 7:23–35

David Bush[1,2,*] Dominic Kain[1] Colin Matheson[1,2] Peter Kanowski[1]

1. Australian National University, Fenner School of Environment and Society, Canberra ACT, Australia
2. CSIRO Division of Plant Industry
* Corresponding author: CSIRO Black Mountain Laboratories, Clunies Ross Street Black Mountain Canberra ACT, Phone +61 2 6246 4829 david.bush@csiro.au

Communicated by R. Burdon

Received: 19 November 2009 / Revised: 9 May 2010 / Accepted: 21 May 2010 / Published online: 22 June 2010

Abstract

The effects of adjusting additive (numerator) relationship matrices (**A**) for inbreeding estimates taken from molecular markers were investigated using a small, model population of *Eucalyptus cladocalyx*. A number of individual-tree, mixed-models were compared, incorporating estimates of population- and family-level selfing and ancestral inbreeding applied either as average values to the entire population or as variable estimates for subpopulation and family groups. The consequences of ignoring inbreeding were inflated additive genetic variance estimates and underestimation of residual variance, with resulting inflation of heritability. We found models that correct for differential inbreeding at the subpopulation level give similar results to more complex ones including family-level estimates. Our analysis indicates that the commonly applied coefficient of relationship for first-generation eucalypt progeny of $\rho = 1/2.5$ appears to be quite suitable for correcting variance component and heritability estimates. However, if inbreeding is not specifically corrected for by adjustment of **A**, some minor rank changes of individual breeding values can occur, especially where levels of inbreeding vary among families, and some suboptimal selections and loss of genetic gain may ensue.

Keywords

Eucalyptus cladocalyx, Inbreeding, Numerator relationship matrix, Selfing

Introduction

Estimating variance and covariance components is the basis of estimation of quantitative genetic parameters such as heritability and genetic correlations between traits. In populations of open-pollinated forest trees, a coefficient of relationship (ρ) of 0.25 (see for example Wright 1922) may be assumed, implying that relatedness among individuals within families is equal to the covariance among half-siblings, or one quarter of the additive genetic variance. However, this assumption is predicated on parents being unrelated and maternal parents being pollinated by a large effective number of paternal parents (Askew and El-Kassaby 1994; Squillace 1974). In fact, this is only rarely likely to be valid in open-pollinated families, due to a number of factors that lead to average relatedness of $\rho > 0.25$. In forest trees with mixed-mating systems, self-fertilisation ('selfing') and neighbourhood inbreeding are common, especially where pollen and/or seed dispersal are spatially limited, as is generally the case in *Eucalyptus* (Eldridge et al. 1993). Progeny resulting from neighbourhood inbreeding and selfing will be more closely related than are half-sibs, and if there are significant numbers present in populations that are wrongly assumed to represent half-sibs, the resulting heritability estimates will be inflated (Namkoong 1966). This is due to factors including within-population heterogeneity of additive genetic variance and dominance effects such as inbreeding depression (Griffin and Cotterill 1988; Hodge et al. 1996). However, while the frequency of embryos with significant levels of inbreeding may be high, selection against inbred germinants, seedlings and young trees, causing their mortality during early development, might result in considerably lower levels of observed inbreeding in mature, wild trees, a result often observed in eucalypts (Potts and Wiltshire 1997).

Eucalyptus cladocalyx F. Muell. is an open forest and woodland species endemic to South Australia. It is the subject of a breeding program in Australia because of its suitability for low rainfall sites where it might be planted for production of timber and poles as well as environmental remediation (Harwood et al. 2007). It occurs in three geographically disjunct regions with a total range of less than 500 km; two of these are situated in mainland Australia and one on Kangaroo Island off the southern Australian coast. McDonald et al. (2003) found evidence for clustering of three subpopulations within the species, corresponding to these regions, using eight allozyme markers. They also showed that there is a relatively high overall level of inbreeding in *E. cladocalyx,* with an estimated multilocus outcrossing rate ($\hat{t}_m$) = 0.57. Most of the inbreeding would appear to be due to selfing rather than neighbourhood inbreeding as gauged by the small difference between single- and multilocus outcrossing estimates (see Brown 1990) at the population level ($\hat{t}_m - \hat{t}_s = 0.05$). Their estimate of

multilocus outcrossing is towards the lower end of published estimates in eucalypts ($\hat{t}_m = 0.44 - 0.96$) (Potts and Wiltshire 1997). However, McDonald et al. (2003) also showed that there was substantial variation in outcrossing between and within the subpopulations ($\hat{t}_m = 0.44 - 0.66$), with subpopulations having proportions of completely selfed families ($\hat{t} = 0$) ranging from 14 % to 50 %. Their estimate of Wright's inbreeding coefficient ($\hat{f}$) for the species was 0.56, but this parameter also varied significantly between subpopulations, ranging from 0.10 to 0.60.

Historically, family models have been used to estimate genetic parameters in forest tree populations (White et al. 2007). The usual approach where inbreeding is likely or known to have occurred is to assume a coefficient of relationship (ρ) greater than 1/4 - typically between 1/3 and 1/2.5 for eucalypts (Eldridge et al. 1993), though some authors have used values as high as 1/1.8 (Hodge et al. 1996). This is then applied post hoc to estimate the additive genetic variance component for the population as a whole, which is the numerator of the narrow-sense heritability ($\hat{h}^2$) ratio. This implies that levels of inbreeding are uniform within the entire population. The first drawback to this approach is that the application of a single value of ρ to a number of subpopulations may not give accurate global estimates of variance components where the level of inbreeding within and among the subpopulations varies significantly. Secondly, best linear unbiased predictor (BLUP) breeding values will be improperly shrunken (regressed towards the mean – see Chapter 15, pp. 419-423 of White et al. 2007 for a general treatment of this subject) in all cases where the levels of inbreeding are different from the chosen value of ρ. This has implications for the formation of selection indices and the prediction of genetic gain.

During the last decade, the flexibility of the mixed-model equations (Henderson 1975) has been further realised due to increased computing power and associated software. The use of individual-tree models, the forestry equivalent of the animal model, has become increasingly common in recent years (e.g. Costa e Silva et al. 2006; Gapare et al. 2008; Jarvis et al. 1995). The individual-tree model is advantageous because it allows the simultaneous prediction of BLUP breeding values for both parents and progeny. It is also possible to specify the additive covariance between every pair of trees via its progenitors using the *numerator relationship matrix* (NRM), even when the relationship is several generations deep and otherwise complex.

Where pedigree is uncertain, such as for the male parentage of open-pollinated trees, it should be possible to improve the reliability of the genetic relationships in the NRM using estimates of subpopulation, family or individual-specific rates of inbreeding obtained from molecular marker data. This should afford several improvements on the traditional approach of applying a single coefficient of relationship to the entire population. Firstly, estimates of variance components, functions and ratios

related to additive genetic effects will also be automatically adjusted for estimated levels of differences in inbreeding. Secondly, the rates at which family- and within-family deviations are regressed towards their means will be adjusted for differences in inbreeding, possibly leading to rank changes in both family- and individual-tree BLUPs. Information about inbreeding accumulates as the pedigree deepens through recurrent selection. Inbreeding has not been accounted for in first-generation analyses of forest tree species, though estimates of outcrossing rates are available from molecular genetic studies of numerous forest species; for example, Potts and Wiltshire (1997) summarise studies in eucalypts.

This study examines the effect of variable rates of inbreeding in open-pollinated families of *E. cladocalyx* on genetic analysis of a small dataset of 4-year growth from a trial established in 2001 in western Victoria, Australia. Family- and subpopulation selfing rates were estimated by re-analysing an isozyme dataset used to estimate population parameters for the species (McDonald et al. 2003). These were applied to mixed models to estimate variance components and thence heritabilities and breeding values for the dataset. The impact of differential selfing and inbreeding on heritability estimates and breeding-value prediction was assessed.

Materials and methods

Phenotypic data

The phenotypic data are drawn from a subset of individuals from 4-year measurements of a progeny trial at Lismore, western Victoria. The trial is one of eleven sites established in 2001 in Australia's southern States by the Australian Low Rainfall Tree Improvement Group (ALRTIG). The trial was designed as five-tree family plots in four complete block replicates with nested incomplete blocks as rows and columns using CycDesign (Whitaker et al. 2002).

The trial contained a total of 119 families from 14 subpopulations. There were only 40 families for which information on outcrossing and subpopulation inbreeding rates could be calculated from assays carried out by McDonald et al. (2003). Two datasets, referred to hereafter as the large and small dataset, were delineated as follows. The large dataset comprised all 40 families (727 trees) from three subpopulations, Cygnet River (seven families), Flinders Chase (12 families) and Wirrabara (21 families), for which estimates of t_m were available for subpopulations from 12, 7 and 7 families respectively (see Table 3.2). The small dataset incorporated the 19 families (318 trees) from Cygnet River and Flinders Chase for which discrete estimates of t_m were available for each family: the seven Wirrabara families used to make the outcrossing estimates listed in Table 3.2 could not be included, as they comprise a different subpopulation sample to that used in the field trial. Data for two growth

traits were analysed: tree height and diameter at breast height (DBH). These traits were both measured in October 2005. Survival was 90 % in the study data subset, with between 14 and 20 trees per family remaining at 4 years.

Calculation of inbreeding rates and coefficients of relationship

Population parameter estimates were drawn from McDonald et al. (2003). Their study used some family seedlots common to those used in the field trials from which our phenotypic data are drawn, but grown separately for 8-12 days in laboratory growth cabinets, after which time they were destructively harvested for tissue extraction and isozyme assay. Family-level outcrossing rates were estimated by re-analysing their scored assays. The data used in this study included those from 12 Flinders Chase families, 7 from Cygnet River and 7 from Wirrabara. Between 10 and 50 progeny per family were assayed, totalling approximately 200 per subpopulation.

The analyses were performed using the methods described in Ritland (2002) implemented using the MLTR computer program (Ritland 2008). Firstly *subpopulation* single-locus $(\hat{t}_s)$ and multilocus $(\hat{t}_m)$ outcrossing estimates were made on the basis of the segregation of alleles at eight polymorphic loci in seedling populations of Cygnet River, Flinders Chase and Wirrabara. The estimation is by likelihood maximisation using the Newton–Raphson method, and standard errors for the estimates were based on 500 bootstraps, re-sampling progeny within families. The difference between multi- and single-locus estimates $(\hat{t}_m - \hat{t}_s)$ was calculated for each subpopulation to give the lower bound of inbreeding that could be attributed to biparental inbreeding, i.e. mating between related individuals (Brown 1990). The rate of selfing $(\hat{s})$ was estimated as $\hat{s} = 1 - \hat{t}_m$. Single-locus $(\hat{t}_s)$ and multilocus $(\hat{t}_m)$ outcrossing estimates were then made for each family within subpopulation. These estimates were made using the method of moments, applying gene-frequency estimates from the subpopulation-level analysis.

Several assumptions about inbreeding were investigated, including that each of the three subpopulations is in inbreeding equilibrium. Inbreeding equilibrium occurs when there is a balance between the creation (through outcrossing) and loss of heterozygotes (through mating between relatives) in each generation. The expected inbreeding equilibrium coefficient, $f_{\text{eq.}}$, is given by

$$f_{\text{eq.}} = \frac{s}{2-s} \qquad (3.1)$$

assuming that the inbreeding is due to self-fertilisation (e.g. Hedrick and Cockerham 1986).

The coefficient of relationship (Fisher 1918; Wright 1922) between individuals, ρ, is used when estimating additive genetic variance under the assumption that inbreeding levels are homogeneous throughout a population. For individuals i and j:

$$\rho_{ij} = \frac{2\theta_{ij}}{\sqrt{(1+f_i)(1+f_j)}} \qquad (3.2)$$

where f_i and f_j are their inbreeding coefficients and θ_{ij} is the coefficient of co-ancestry. The coefficient of co-ancestry of two individuals reflects the probability that two gametes taken at random, one from each, carry alleles that are identical by descent and is equivalent to the inbreeding coefficient of their progeny should they be mated together (Lynch and Walsh 1998). For selfs $f_i = f_j = 1/2$, $\theta_{ij} = 1/2$ and $\rho_{ij} = 2/3$; for half-sibs $f_i = f_j = 0$ and $\theta_{ij} = 1/8$, $\rho_{ij} = 1/4$. The coefficient of relationship is effectively a correlation between breeding values. Though ρ was not directly used in the models used here, equivalent values were calculated for comparative purposes.

Construction of the additive or numerator relationship matrix

The mixed-model equations (Henderson 1953; Searle 1968) can be implemented for prediction of breeding values and genetic parameters:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \qquad (3.3)$$

where **y** is the data matrix and **b** and **u** are the vectors of solutions for the fixed and random effects, respectively. **X** and **Z** are the design matrices for the fixed and random effects, respectively. Also included are terms **R** for the variance/covariance matrix of the residuals and, if the random-effect factors are assumed to be independent, **G** is the direct sum of the variance/covariance matrices of each of the random-effect factors.

It is also possible to assume that the levels of the random effects are not independent. For breeding values, the additive (or numerator) relationship matrix (**A** or NRM) is then used. The elements are equivalent to twice the co-ancestry (θ) of the individuals concerned and therefore equivalent to the numerator of ρ [Eq. 3.2]: **A** is then a covariance while ρ is a correlation structure. Manipulation of **A** then allows the calculation of breeding values in populations with complex inter-generational

pedigrees for both parents and progeny (Henderson 1976). The additive genetic covariance term **G** in Eq. 3.3 therefore takes the general form shown in Eq. 3.4.

$$\mathbf{G} = \sigma_a^2 \mathbf{A} = \sigma_a^2 \begin{bmatrix} 1+f_i & 2\theta & 2\theta & 2\theta \\ 2\theta & 1+f_i & 2\theta & 2\theta \\ 2\theta & 2\theta & 1+f_i & 2\theta \\ 2\theta & 2\theta & 2\theta & 1+f_i \end{bmatrix} \qquad (3.4)$$

Structurally, **A** is sparse, as individuals usually only have relationships to a few specific others, with much of the information contained on the diagonal. The diagonal elements of **A** are formed as the direct sum of a submatrix of parent auto-relationships and a submatrix of progeny auto-relationships. For base parents that are usually assumed to be non-inbred and unrelated, the parent auto-relationship submatrix is an identity matrix since f_i=0. The rest of **A** consists of parent-offspring relationships and offspring inter-relationships. However, in this case, we have access to information on both extant levels of inbreeding in the parental population and evidence that a proportion of the progeny are the result of selfing. This will affect all non-zero elements of **A**.

We have extended the Dutkowski and Gilmour (2001) rules for construction of elements of **A** in populations that are subject to partial selfing. They developed simple formulae for calculation of elements of the **A** matrix assuming proportions of selfed individuals. These have been extended to incorporate the assumption that each subpopulation is in inbreeding equilibrium. Squillace (1974) developed similar rules to calculate coefficients of relationship (i.e. correlations) for partially selfed populations incorporating ancestral inbreeding. Squillace used path analysis (developed by Wright) to derive his rules. He makes the important point that, between partially selfed offspring, there are potentially four types of relationship (and coefficients of co-ancestry): (1) self-full-sibs (SFS), (2) half-sibs (HS), (3) self-half-sibs and (4) full-sibs. For partial selfing, relationships 1-3 will certainly exist. Figure 3.1 shows path diagrams which illustrate the nature of the relationships between parents and offspring in these three cases (see Lynch and Walsh 1998 for further information on path analysis). The parents of selfed offspring (SFS) are shown as two separate parents with a coefficient of co-ancestry (θ) equal to one half. The proportion of outcrossed full-sibs (i.e. cases where a pollen parent other than the tree that is also the female parent) is independent of the rate of selfing. We have assumed random mating in cases other than selfs, after the mixed-mating model (Clegg 1980; Fyfe and Bailey 1951), though this is unlikely to be a completely realistic assumption (see later discussion). Assuming no biparental full-sib offspring, the proportions (p) of the other three relationships are as follows:

$$(p_{SFS} + p_{HS} + p_{SHS}) = [s^2] + [(1-s)^2] + [2s(1-s)] = 1 \qquad (3.5)$$

where s is the proportion of selfed progeny. These proportions can then be used to weight Eqs. 3.6-8 (Lynch and Walsh 1998) for calculation of the required matrix elements:

$$\theta_{ii} = \frac{1}{2}(1 + f_P) \qquad (3.6)$$

$$\theta_{PO} = \frac{1}{4}(1 + f_P + 2f_O) \qquad (3.7)$$

$$\theta_{ij} = \frac{1}{8}(2 + f_k + f_m + 4\theta_{km}) \qquad (3.8)$$

Where $P = \{k, m\}$ denotes the set of k (female) and m (pollen) parents and $O = \{i, j, ...\}$ denotes the set of offspring including individuals i and j. The inbreeding coefficients of the female and male parents are f_k and f_m respectively, and those of the offspring are denoted f_i and f_j. θ_{PP} will take the same form as Eq. 3.6. The difference between θ_{ii} and θ_{PP} is that f in θ_{ii} is the sum of both selfing and ancestral inbreeding, whereas θ_{PP} only includes ancestral inbreeding. The solutions for selfing without ancestral inbreeding (Dutkowski and Gilmour 2001), and with both selfing and ancestral inbreeding, simplify to the values given in Table 3.1.

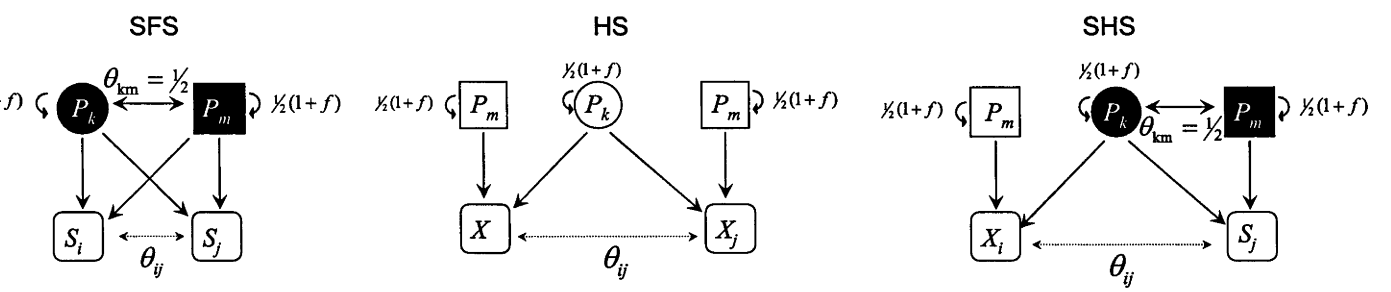


Fig. 3.1 Path diagrams for deriving coefficients of co-ancestry (θ_{ij}) for partially selfed progeny (after Squillace 1974 and Lynch and Walsh 1998). The parents (female, P_k and male, P_m) are shown in the *top row*, progeny, selfed (*S*) and outcrossed (*X*) *below*. In the self-full-sib (*SFS*) and self-half-sib (*SHS*) relationships, male and female components of a single, self-fertilising parent are shown as *solid black shapes* with θ_{km}=1/2. In the half-sib (*HS*) relationship, two unrelated pollen parents are shown. The *solid-arrowed lines* indicate paths by which progeny might share common alleles by descent. Inbreeding in the parents further increases θ_{ij} with a contribution of ½(1+*f*) per parent where *f* is Wright's inbreeding coefficient.

Modification of the matrix

The rules defined in Table 3.1 were used to modify the elements of **A** to assess several different models of ancestral inbreeding and selfing. Values of f were calculated using Eq. 3.1, with $f_{eq.}$

calculated at the subpopulation level, with the subpopulation value of s taken as the average of raw, MLTR-derived estimates. Individual-family estimates of s were truncated to fall between the theoretical limits (0-1), recognising the bias this causes (Ritland 2008), to avoid later matrix inversion problems. The values so-calculated were used to form a sparse, row-column indexed listing of the lower triangle elements of **A**. These data can be read into ASREML version 2 (VSN International, Hemel Hempstead, UK), which inverts **A** and substitutes it for the $\mathbf{A}^{-1}$ matrix that would otherwise be formed directly from a user-defined pedigree file relating to the phenotypic data (Gilmour et al. 2006, ASREML User Guide, Version 2, Chapter 9.6).

Several versions of **A** were applied to the large and small datasets, incorporating different assumptions about ancestral inbreeding and selfing. Firstly, a series of **A** matrices were constructed assuming population-wide rates of selfing in increments of 0.1 for both the datasets. This was then repeated to form a second set of **A** matrices assuming $f_{eq.}$ equal to the average equilibrium inbreeding coefficient for the entire population in addition to the selfing. Secondly, matrices incorporating *variable* levels of selfing and inbreeding were constructed. For the small dataset, **A** matrices were constructed incorporating family estimates of s with (a) no ancestral inbreeding and (b) estimates of $f_{eq.}$ corresponding to the two subpopulation estimates. For the large dataset, **A** matrices were constructed incorporating subpopulation estimates of s with (a) no ancestral inbreeding and (b) estimates of $f_{eq.}$ corresponding to the three subpopulation estimates.

Table 3.1 Additive relationships for elements of **A** incorporating partial selfing (s) and partial selfing and inbreeding in previous generations (f)

Element	Selfing	Selfing and ancestral inbreeding
$2\theta_{PP}$	1	$1+f$
$2\theta_{PO}$	$\frac{1}{2}(1+s)$	$\frac{1}{2}(1+f+s+fs)$
$2\theta_{ii}$	$1+s/2$	$(1-s)+ s((3+f)/2)$
$2\theta_{ij}$	$\frac{1}{4}(1+s)^2$	$\frac{1}{4}(1+f+2s+s^2+2fs+fs^2)$

Variance-component estimation and BLUP breeding value prediction

Mixed-models were solved by restricted maximum likelihood implemented in ASREML to estimate variance components and BLUP breeding values for height and diameter. This software is flexible and can accommodate user-defined **A** and $\mathbf{A}^{-1}$ matrices as described in the previous section. The analyses were based on general linear mixed-models of the form given in Eq. 3.3, which can be simplified to the form of Eq. 3.9 below:

$$\mathbf{y} = \mathbf{X}\hat{\mathbf{b}} + \mathbf{Z}\hat{\mathbf{u}} + \mathbf{e} \qquad (3.9)$$

where **y** is the vector of observations on the trait, **b** and **u** are vectors of fixed- and random-effect estimates respectively, **X** and **Z** are incidence matrices for fixed and random model terms and **e** is a vector of random residual effects. The vector **b** contained sub-vector estimates for fixed effects of replicate and subpopulation effects, and **u** contained sub-vectors for the random effects of incomplete blocks, plots the individual trees and the relationship matrix for the pedigree as described. The subpopulations specified in **b** are effectively genetic groups (see for example Westell et al. 1988), since no individual has a relationship to more than one genetic group. The BLUP breeding values corrected for genetic group effects are obtained by adding the family or individual effects to the subpopulation best linear unbiased estimator.

Individual-tree narrow-sense heritability was estimated as:

$$\hat{h}^2 = \frac{\hat{\sigma}_A^2}{\hat{\sigma}_A^2 + \hat{\sigma}_P^2 + \hat{\sigma}_e^2} \qquad (3.10)$$

where σ_A^2 is the additive genetic variance, σ_P^2 is the plot variance and σ_e^2 is the residual variance. 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. 2006; Lynch and Walsh 1998).

Correlations

Nonparametric Spearman rank correlations (Spearman 1904) between (a) families and (b) single tree BLUPs as estimated for the various models were calculated with the SPEARMAN procedure in Genstat 12 (VSN International, Hemel Hempstead, UK). Spearman's rank correlation coefficient is a measure of association between the rankings of two variables, and so is useful for examining the extent to which families and individual-tree BLUPs are effectively re-ordered by the various models. The same software was also used to calculate product-moment correlations between (a) the estimated outcrossing rate $(\hat{t}_m)$ and phenotypic growth measures and (b) $\hat{t}_m$ and number of surviving progeny per family. These correlations are analogous to heterozygosity-fitness correlations used to identify inbreeding depression (Mitton 1993), with rate of selfing assumed to be negatively correlated to heterozygosity.

Genetic gain

A common practical use of BLUP ranking indices is for the selection of elite subpopulations such as clonal seed orchards. Typically a selection of the best individuals is made, and their breeding values can be compared to the population average as a measure of predicted genetic gain (PGG). To study the modelled effects of selfing, a small PGG example was calculated. The 15 top-ranking individuals from models assuming various levels of selfing and inbreeding were selected for a clonal seed orchard (putting aside family selection considerations in this hypothetical example). PGG for these selections was taken by averaging the BLUPs of the same individuals, as predicted by the maximal model (viz. that including the more detailed assumption of family-level selfing plus subpopulation-level equilibrium inbreeding). Rank changes - i.e. selections other than the top 15 from the maximal model, resulting from varying assumptions among the alternative models - will therefore result in some loss of genetic gain. Genetic gain calculated from the BLUPs of the top 15 individuals from other models were also calculated to gauge the effects of the alternative assumptions, which assumed less or no inbreeding, on the overall magnitude of genetic gain.

Results

Estimates of selfing rates

As expected from McDonald et al. (2003), estimates of single- and multilocus outcrossing at the family level were variable among subpopulations and high on average (Table 3.2). Estimates of t_s carry high standard errors at the family level; the sampling was designed originally for subpopulation estimates. Flinders Chase families ranged from almost completely outcrossing to completely selfed, with an average rate of selfing of 36 %. Average selfing was 46 % in Cygnet River, with two of the seven families having low levels of selfing, and one nearly entirely selfed. At Wirrabara, no families appeared to be predominantly selfed, with the population selfing estimate at 34 %.

Biparental inbreeding appears to be low on average, though the high standard errors on t_s estimates make the results inconclusive at the family level. There is some suggestion that some neighbourhood inbreeding has occurred in some families, but estimates of $t_m - t_s$ are approximately zero at the subpopulation level.

Estimates of inbreeding equilibrium coefficients ($\hat{f}_{\text{eq.}}$) (Eq. 3.1) for the subpopulations calculated from the subpopulation mean selfing rates were 0.27 (Flinders Chase), 0.20 (Cygnet River) and 0.19 (Wirrabara). The population average was 0.25. The population average rate of selfing, based on the estimates in Table 3.2, and weighted by numbers of progeny, in both the small and large datasets was 0.36. This corresponds to $\rho = 1/2.7$

Table 3.2 Estimates of family single- $(\hat{t}_s)$ and multilocus $(\hat{t}_m)$ outcrossing rates with standard errors computed from arrays of *n* progeny and eight isozyme loci per family with MLTR

Family	*n*	$\hat{t}_s$	$s.e.(\hat{t}_s)$	$\hat{t}_m$	$s.e.(\hat{t}_m)$	$(\hat{t}_m - \hat{t}_s)$	$s.e.(\hat{t}_m - \hat{t}_s)$	$\hat{s} = 1 - \hat{t}_m$
Flinders Chase								
1	15	0.50	0.20	0.93	0.05	0.42	0.16	0.07
2	10	0.88	0.15	1.02	0.01	0.15	0.15	-0.02
3	10	1.01	0.05	1.06	0.02	0.05	0.04	-0.06
4	20	-0.27	0.22	-0.31	0.34	-0.04	0.15	1.31
5	25	0.07	0.16	-0.01	0.25	-0.08	0.10	1.01
6	20	-0.14	0.22	-0.42	0.42	-0.27	0.24	1.42
7	15	0.60	0.27	0.82	0.34	0.22	0.21	0.18
8	20	0.72	0.24	0.79	0.20	0.07	0.06	0.21
9	15	0.48	0.35	0.47	0.40	-0.01	0.15	0.53
10	10	0.17	0.40	0.66	0.35	0.48	0.31	0.34
11	20	1.14	0.03	1.15	0.03	0.00	0.05	-0.15
12	20	0.49	0.24	0.73	0.21	-0.24	0.15	0.27
Population	*200*	*0.60*	*0.08*	*0.64*	*0.08*	*0.04*	*0.02*	*0.36*
Cygnet River								
13	45	0.26	0.30	0.02	0.38	-0.24	0.16	0.98
14	40	0.57	0.12	0.65	0.13	0.09	0.06	0.35
15	15	0.34	0.23	0.72	0.12	0.38	0.23	0.28
16	25	0.12	0.31	-0.28	0.51	-0.40	0.40	1.28
17	40	1.20	0.30	2.16	0.20	0.96	0.29	-0.16
19	35	0.50	0.17	0.37	0.22	-0.13	0.11	0.63
20	30	0.89	0.09	1.00	0.01	0.11	0.09	0.00
Population	*230*	*0.49*	*0.14*	*0.54*	*0.15*	*0.05*	*0.02*	*0.46*
Wirrabara								
21	20	0.91	0.07	0.96	0.02	0.06	0.06	0.04
22	30	0.36	0.10	0.50	0.10	0.14	0.05	0.50
23	30	0.31	0.11	0.68	0.09	0.37	0.06	0.32
24	10	0.25	0.13	0.48	0.15	0.23	0.07	0.52
25	50	0.79	0.12	0.57	0.31	-0.22	0.19	0.43
26	30	0.57	0.12	0.84	0.08	0.27	0.09	0.16
27	30	0.53	0.12	0.72	0.14	0.19	0.09	0.28
Population	*200*	*0.58*	*0.09*	*0.66*	*0.09*	*0.08*	*0.03*	*0.34*

Variance components and heritability

Variance component and heritability estimates for a selection of the models are given in Table 3.3. Variance component estimates for incomplete block and plot structures were all very low or zero. The plot term was therefore dropped from Eq. 3.10 in estimating heritability.

The heritability estimates and standard errors were relatively higher from the small dataset than the large for both traits. The effect of uniformly applying successively higher levels of selfing to each whole population in the small and large phenotypic datasets is a reduction in the estimate of heritability as the estimated additive genetic variance component decreases and the residual increases. Heritability estimates, modelled using the small dataset, for uniform, population-wide rates of selfing from 0 % to 100 %, with and without ancestral inbreeding (population assumed to be in inbreeding equilibrium with $f_{eq.}$ =0.25), are shown in Fig.3.2. There is a marked decrease in heritability with increased selfing, with ancestral inbreeding further deflating the estimates. If the average population selfing rate of 0.36 is applied to construct elements of **A**, the heritability estimate for height decreases from 0.76 to 0.44 (57 % of the estimate assuming no inbreeding). If ancestral inbreeding is also assumed, the estimate is further reduced to 0.35 (46 %). Standard errors as proportions of the heritability estimates were quite stable (<3 % change) for the various models applied to both traits in both datasets.

Assumptions of variable selfing (at the family level in the small dataset and the subpopulation level in the large) and variable inbreeding (at the subpopulation level in both datasets) gave heritability estimates very close to the simpler models that assumed average values of selfing and inbreeding applied population-wide.

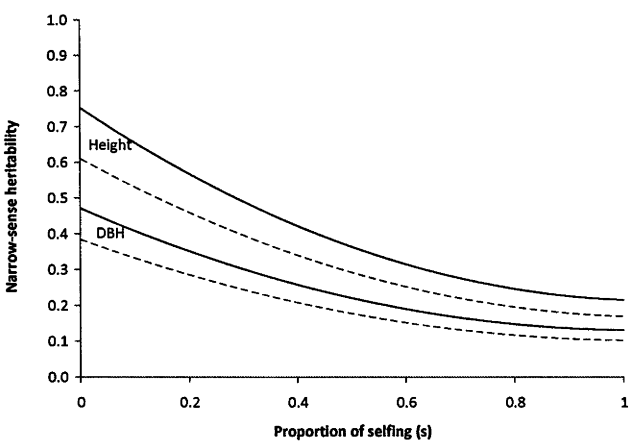


Fig. 3.2 Narrow-sense heritability estimates ($\hat{h}^2$) modeled using the small dataset as a function of selfing proportion assuming no ancestral inbreeding (*solid lines*) and population-wide inbreeding equilibrium with $f_{eq.}$=0.25 (*dashed lines*)

Table 3.3 Genetic parameter estimates ($\hat{h}^2$ narrow-sense heritability, $\hat{\sigma}^2_A$ additive genetic variance and $\hat{\sigma}^2_e$ residual variance) and standard errors for DBH and height assuming varying levels of ancestral inbreeding and selfing

Dataset	$f_{eq.}$	s	$\hat{h}^2$	$\hat{h}^2$ s.e.	$\hat{\sigma}^2_A$	$\hat{\sigma}^2_A$ s.e.	$\hat{\sigma}^2_e$	$\hat{\sigma}^2_e$ s.e.
DBH: family variable (small)	0	0	0.48	0.22	247	2.0	266	2.6
	0	0.36	0.27	0.13	132	2.0	357	5.7
	0	Var.	0.27	0.13	134	1.9	365	5.9
	0.25	0.36	0.22	0.10	105	2.0	384	7.3
	Variable	Variable	0.21	0.10	106	1.9	393	7.7
Height: family variable (small)	0	0	0.76	0.29	7409	2.2	2287	0.9
	0	0.36	0.44	0.18	3947	2.2	5018	3.5
	0	Var.	0.42	0.16	3844	2.3	5354	4.1
	0.25	0.36	0.35	0.14	3158	2.2	5807	5.1
	Variable	Variable	0.33	0.13	3026	2.3	6156	6.0
DBH: provenance variable (large)	0	0	0.32	0.13	156	2.24	336	5.6
	0	Var.	0.18	0.08	85	2.24	393	9.9
	0	0.36	0.18	0.08	84	2.24	393	10.0
	0.25	0.36	0.14	0.06	67	2.24	410	12.1
	Variable	Variable	0.15	0.06	69	2.24	408	11.8
Height: provenance variable	0	0	0.64	0.18	5613	3.0	3137	2.1
	0	0.36	0.37	0.11	3035	3.0	5169	6.0
	0	Var.	0.37	0.11	3050	3.0	5163	6.0
	0.25	0.36	0.29	0.09	2393	3.0	5805	8.4
	Variable	Variable	0.30	0.09	2502	3.0	5716	8.0

BLUP shrinkage and rankings

Application of the models that assumed variable rates of selfing and inbreeding resulted in marked shrinkage (regression towards the mean) when compared with the no-inbreeding model. The magnitude of the effect can be visualised for height from the small dataset using point histograms (Fig. 3.3). The pattern of the uniform, population-wide selfing histogram with s corresponding to the estimate of the population mean appears to be broadly similar to that of the variable selfing model, while the assumption of ancestral inbreeding in addition to selfing has resulted in slightly more shrinkage. The precision of these estimates should be considered in the context of the small sample sizes, particularly at the family level, used to estimate s. Small sample size has likely led to over-dispersal of the selfing estimates, and bias caused by the truncation of some estimates that fell outside the theoretically possible range of 0 and 1 will have consequent effects on the BLUP shrinkage estimates.

The comparison of BLUP rank Spearman correlations (Table 3.4) for each of the models revealed that generally, the subpopulation and family rankings differed little. However, in the small dataset which includes two subpopulations, the rankings for the subpopulations reversed for the DBH trait. In the large dataset, the Spearman rank correlations for families were all unity for height, and the lowest

correlation between any pair of models was 0.996 between the variable selfing plus inbreeding and no-inbreeding models in the small dataset.

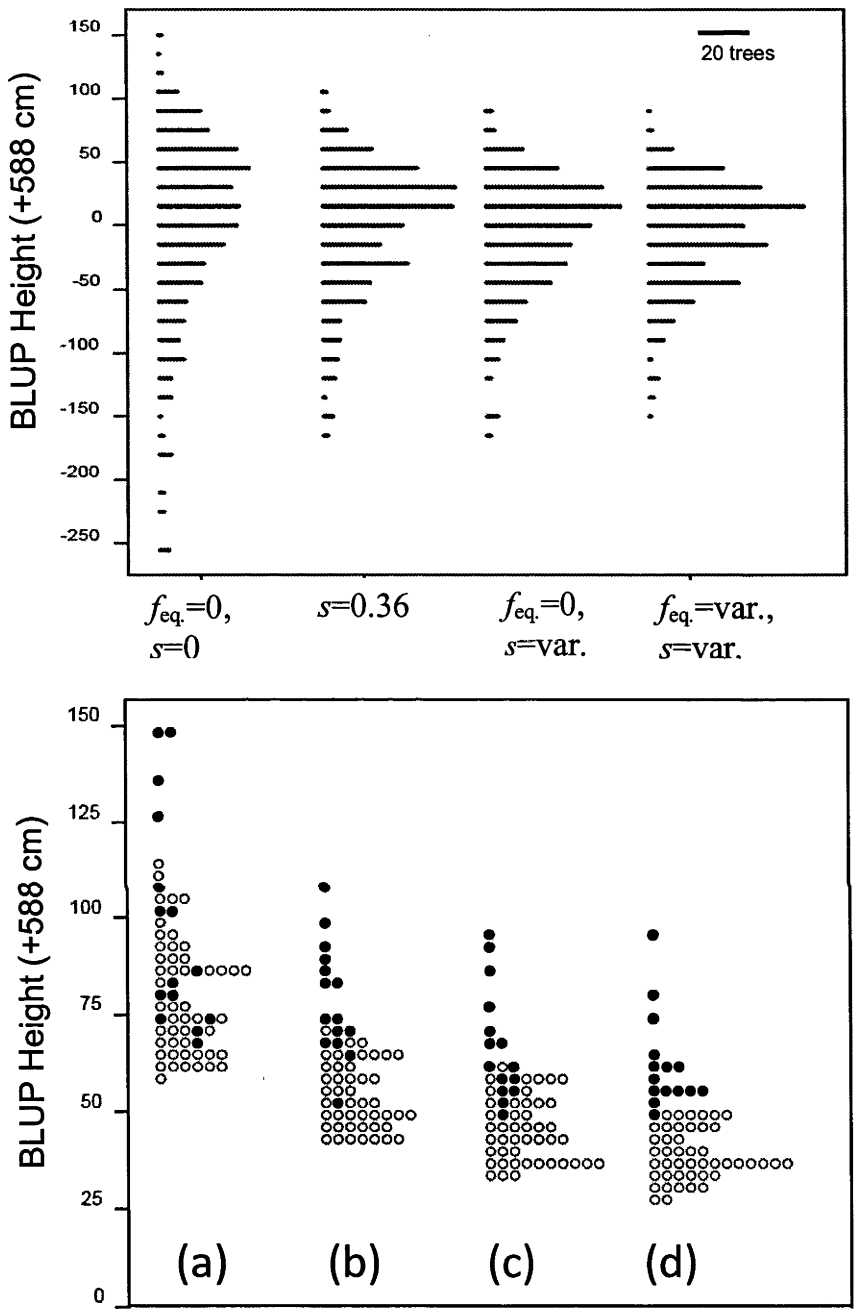


Fig. 3.3 BLUP point (individual-tree) histograms showing the degree of shrinkage for models using individual-family selfing for height. Each *pane* compares models assuming (a) no inbreeding, (b) selfing equivalent to the population average applied to all parents, (c) variable selfing (s = var.) for each family and (d) variable selfing plus variable subpopulation equilibrium inbreeding ($f_{eq.}$ = var.). The *top pane* shows the entire population (318 trees). The *bottom pane* shows the top 60-ranked trees in each model. The *filled circles* indicate the top 15 trees selected using the maximal model (variable equilibrium inbreeding and variable selfing).

The Spearman rank correlation coefficients for individual-tree BLUPs (Table 3.4) were generally high for all models that assumed uniform rates of selfing. However, the correlations involving the variable models (particularly those assuming ancestral inbreeding as well as selfing) and the no-inbreeding model were lower, indicating that some rank changes are occurring.

Table 3.4 Pairwise Spearman rank correlations between BLUP height and DBH (individual trees) for models assuming (a) no inbreeding ($f = 0$, $s = 0$), (b) selfing equivalent to population average s=0.36) applied to all trees, (c) variable selfing (Var. s) and (d) variable selfing plus subpopulation equilibrium inbreeding

	Height			DBH		
Small dataset	$f_{eq.} = 0$, $s = 0$	$s = 0.36$	Var. s	$f_{eq.} = 0$, $s = 0$	$s = 0.36$	Var. s
$s = 0.36$	0.946			0.947		
Var. s (family level)	0.933	0.995		0.890	0.897	
Var. $f_{eq.}$+s	0.869	0.977	0.987	0.940	0.947	0.991
Large dataset						
$s = 0.36$	0.947			0.957		
Var. s (subpop. level)	0.948	1.000		0.957	1.000	
Var. $f_{eq.}$+s	0.901	0.992	0.991	0.922	0.994	0.994

The variable models were individual family selfing models

The practical implication of rank changes and differential shrinkage for the various models was gauged by carrying out predicted genetic gain calculations for a hypothetical, 15-clone seed orchard (Table 3.5). For each model evaluated, the 15 highest-ranked individuals were identified, and the mean of their breeding values predicted by the maximal model was calculated. These 'corrected' means of selections from each model (i.e. means of breeding values taken from the maximal model) reflect the impact of rank changes relative to the selections taken from the maximal (variable selfing plus inbreeding) model. The difference in genetic gain is minimal (both traits and both datasets) for the variable selfing and fixed population average selfing models, but is markedly lower if no inbreeding is assumed, with six to nine trees that are not amongst the top 15 selections in the maximal model being 'incorrectly' selected. This indicates that there are appreciable rank changes, due to differential shrinkage, at the top of the index. The more detailed histogram of the tallest 60 trees from the small dataset (Fig. 3.3, lower pane) from each model shows the relative ranks of the trees selected using the maximal model according to the alternative models and shows the minor changes in the three models assuming some degree of inbreeding. The 'inflated' means are calculated as averages of the BLUPs from each of the models used to make the selections. All of the alternative models shrink BLUPs relatively less than the maximal model, and therefore overestimate the magnitude of genetic gain, to a large degree in the case of those assuming neither selfing nor inbreeding.

Table 3.5 Mean of selections and genetic gain from the top 15-ranked individuals from each model, expressed in terms of their breeding values from either the maximal model (MAX) or predicted using the other models (OM)

	Large dataset				Small dataset			
Model assumption: †	$f_{eq.}$, s variable (maximal)	s variable	s =0.36	$f_{eq.}$=0, s=0	$f_{eq.}$, s variable (maximal)	s variable	s =0.36	$f_{eq.}$=0, s=0
Height (cm)								
Mean of selections (MAX)	78.2	76.7	76.7	69.9	59.4	58.1	58.9	51.7
Predicted gain (MAX)	13.3 %	13.0 %	13.0%	11.9 %	9.8 %	9.6 %	9.8 %	8.6 %
Alternative selections (of 15) (SM)	0	5	5	9	0	5	3	9
Predicted mean of selections (OM)	-	83.1	83.0	108.6	-	66.3	78.7	110.8
DBH (mm)								
Mean of selections (MAX)	9.62	9.48	9.54	8.81	8.68	8.58	8.64	8.17
Predicted genetic gain (MAX)	11.3%	11.2%	11.2%	10.4%	10.6%	10.4%	10.5%	9.9%
Alternative selections (of 15) (SM)	0	3	2	8	0	3	2	6
Predicted mean of selections (OM)	-	10.05	10.13	12.85	-	10.07	9.38	16.35

Note: †$f_{eq.}$=equilibrium population inbreeding rate; s=rate of selfing; variable indicates family(s) or population($f_{eq.}$)-specific values were used; 0.36 is mean population selfing rate.

Relationship between rate of selfing, survival and phenotypic measurements

There was no evidence of a heterozygosity-fitness correlation. Product-moment correlations of family rates of outcrossing (both $\hat{t}_m$ and $\hat{t}_m$ truncated at 0 and 1) with phenotypic values of the two growth traits were all in the range 0.09 to -0.15 and statistically non-significant. The correlations between number of surviving progeny per family were 0.00 and 0.05 with $\hat{t}_m$ and $\hat{t}_m$ truncated, respectively.

Discussion

The study of a small, model population of *E. cladocalyx* has indicated the likely importance of accounting for inbreeding effects when estimating additive genetic parameters and effects in this species, though the effects of adjusting for differential inbreeding were relatively small. The consequence of ignoring selfing

in the case of the datasets used here is a marked inflation of the heritability estimates as additive genetic variance components are overestimated and residuals are underestimated. Proportionate standard errors associated with $\hat{h}^2$ are quite stable for all trait and model combinations (maximum ± 3%), regardless of the amount of selfing assumed. This general result agrees with the simulation studies undertaken by Dutkowski and Gilmour (2001), which indicated inflated additive and deflated residual variance with increasing selfing and that heritability is inflated by about 20% per 10% of selfing.

It is common practice to apply an estimate of the likely population average coefficient of relationship (Eq. 3.2) to the additive genetic variance component in the estimation of narrow-sense heritability. Application of a coefficient of relationship of $\rho = 1/2.7$ (corresponding to population average s=0.36 and $2\theta_{ij}$=0.46) would be sufficient to correct for the average amount of selfing estimated for this population, while $\rho =$ 1/2.3 would be needed if the population inbreeding equilibrium assumption of ancestral inbreeding is also to be accounted for. The commonly suggested value of $\rho = 1/2.5$ for eucalypts would therefore appear to be well suited to these data. It is probable that studies of first-generation eucalypt breeding populations assuming ρ=1/4 will have significantly overestimated heritability. Examples are *Eucalyptus longirostrata,* $\hat{h}^2$ = 0.33-0.49 for height and 0.38-0.45 for stem straightness at three sites at age 31 months (Henson et al. 2007), and for *E. cladocalyx,* $\hat{h}^2$ = 0.28 for height at age 5 years and 0.52 for flowering precocity at age 3 years (Mora et al. 2009). By comparison, a comprehensive review of heritability in forest tree species (Cornelius 1994) gives a usual range of $\hat{h}^2$ = 0.1-0.3 for growth traits, and for wood specific gravity, the most heritable of commonly reported traits, a median value of 0.48. This places the *E. longirostrata* estimates well into the upper range, while the *E. cladocalyx* estimate for height in the upper part of the usual range and the precocity trait, though high, is comparable to estimates for traits less directly related to fitness, such as specific gravity.

In this study we have employed marker-based estimates of subpopulation- and family-level selfing to obtain average values of θ for application to **A**. An alternative approach is to estimate θ directly from the marker data, as, for example, Kumar and Richardson (2005) have done for a *Pinus radiata* breeding population. This approach, resulting in a marker-informed *genomic relationship matrix*, should theoretically have the advantage of being able to identify within-family variability in identity-by-descent, an occurrence that can have significant consequences for selection and breeding programs (Franklin 1977). Also, individuals assumed to be unrelated in the pedigree-based **A**, which would take 0 for off-diagonal elements and 1 for diagonal elements, may actually have more or fewer alleles in common and more or less heterozygosity than average when genotypes are examined (Van Raden 2007). However, the marker-based measures of heterozygosity and inbreeding have not always been highly correlated with those accurately determined from pedigrees. Kumar and Richardson (2005) found a poor correlation between marker and pedigree-based estimates of co-ancestry using eight SSR markers. Slate et al. (2004) found that in a study of sheep involving 138 SSR loci spanning all autosomes, the relationship between

heterozygosity and pedigree-based $\hat{f}$ was also weak, and that while heterozygosity-fitness correlations for the seven-generation deep pedigree were strong they were not significant for the marker data.

The analyses in the current study, based on the mixed-mating model, have assumed that seed collected from wild mothers represents progeny that are the product of matings that are either selfs or outcrosses to an infinite number of male parents. Neighbourhood inbreeding, giving rise to relationships other than half-sibs and self-full-sibs and increasing the likelihood of biparental full-sibs, has been assumed absent, an assumption supported by the difference between single and multi-locus outcrossing rates at the subpopulation level. Due to the modest size of the progeny assays, the data were not sufficiently precise to gauge whether neighbourhood inbreeding may differentially occur in some families, though it would seem likely, especially given the open forest and woodland forest types where the species occurs. Neighbourhood inbreeding is more difficult to model with precision because the exact relationship among progeny resulting from mating between relatives is unknown. It would be possible to estimate the upper bound of the impact of the neighbourhood-inbred individuals by substituting t_s for t_m in the calculation of elements of **A**.

Two further assumptions of the mixed-mating model are that (a) the probability of an outcross will not be affected by the maternal genotype, and (b) that selection does not occur between mating and determination of genotype distributions (Shaw and Allard 1981). Both of these assumptions are likely to be flawed. Ellis and Sedgley (1992) found that individual *E. cladocalyx* trees range from completely self-incompatible to self-compatible, a fact that complicates the modelling of ancestral inbreeding, especially if the trait is itself heritable. Though we have not found evidence of selection against selfed progeny, we do not know to what extent such selection occurs in the wild populations. The most detailed models considered here assume selfing and ancestral inbreeding in wild populations, both of which result in a deficit of heterozygotes. The estimates of selfing are derived from sibling progeny of the mothers that are represented in the phenotypic dataset. However, the DNA was extracted at the early stages of seedling growth in laboratory growth cabinets, whereas the field trials were raised in a commercial nursery. There may have been relatively heavy selection against selfed individuals in the nursery, either through natural mortality or anthropogenic. However, for some families, the estimated rate of outcrossing was virtually nil, and there is no evidence from the nursery records (from two separate nurseries) that these seedlots were difficult to germinate. Nor is there any correlation between rate of selfing and survival in the field trials, with survival being generally high.

The validity of the assumption that the wild populations are in inbreeding equilibrium, with an associated increase in relatedness between individuals, is more difficult to assess. It has been shown in a number of studies of eucalypts that, while f in laboratory-raised seedling populations is often significantly positive, there is usually a slightly negative f value in the wild maternal population (Potts and Wiltshire 1997), indicating that selection against homozygotes in the wild is very strong. If this proves to be the case in *E.*

cladocalyx, the assumption of inbreeding equilibrium should be dropped, and inclusion of this assumption in the analysis of actual breeding program data would be questionable, especially given the additional complexity in constructing the **A** matrix that it entails.

The study has not indicated that the family rate of selfing is correlated with phenotypic vigour or survival, a correlation that, if present, would suggest inbreeding depression. This phenomenon, caused by accumulation of deleterious recessive alleles, has been documented in many plant species including *Eucalyptus globulus* (Hardner and Potts 1995; Hodge et al. 1996), *Eucalyptus nitens* (Hodge et al. 1996) and *Eucalyptus regnans* (Griffin and Cotterill 1988; Hardner and Potts 1997). There are two factors that might help explain our result. The first, already mentioned, is the lack of selection in the assay population and potentially strong selection and mortality of inbred progeny in the nursery. A second possibility is that the effects of selfing have not yet fully manifested. Hardner and Potts (1995) found that the impact of inbreeding depression increased quite markedly over time in *E. globulus*, although it was already significant by 2 years of age. In *E. regnans,* Hardner and Potts (1997) found that inbreeding depression evidenced by mortality increased markedly after 4 years, once crown closure had occurred. The alternative is that inbreeding depression is absent or not as strong in this species. Hodge et al. (1996) identified neither evidence of inbreeding depression nor dominance effects in *E. nitens,* despite identifying both in *E. globulus*. A larger study incorporating more families and over an extended time-period would help elucidate the situation in *E. cladocalyx*. Simulation studies by Borralho (1994) have indicated that, if inbreeding depression and selfing are present and variable among families, overestimation of heritability may result. The models presented here would need to be augmented to deal with non-additive genetic effects, for example the dominance genetic relationship matrix $(\sigma_D^2\mathbf{D})$ would need to be incorporated, with additional data from control-pollinated matings being needed to support the more complex model.

While application of an appropriate average coefficient of relationship will appropriately reduce the heritability estimate, this case study indicates that incorrect BLUP ranks will result, especially if differential inbreeding is present within the population. Using an individual-tree model incorporating appropriate levels of selfing applied to subpopulations and/or families results in increased shrinkage of breeding values, deflation of PGG of elite selections and rank changes, resulting in different individuals being selected. However, application of the much simpler uniform-selfing models gave good approximations to the results obtained using the variable models with these datasets. A further refinement to the approach taken here would be to assign appropriate values in **A** for each individual on the basis of their being outcrossed or selfed. This would result in differential shrinkage and re-ranking within families as well as among families and genetic groups.

This study has opportunistically employed data that apply to a small subset of families and populations from a pre-existing population genetics study. While the sample size employed in this study is too small for reliable species-level parameter estimates, the general findings are likely to apply to the larger

ALRTIG breeding population, which is comprised of these and similar wild, open-pollinated families as well as some from planted stands; each of these may well have quite different patterns of inbreeding. In the past, the cost and effort to obtain such estimates would have been prohibitive. However, the costs of genotyping large numbers of individuals are rapidly falling (e.g., through use of molecular markers such as single nucleotide polymorphisms). The cost of DNA extraction is also likely to fall. This may result in it being quite possible to assay breeding populations of several thousands of trees at a cost similar to that of acquiring phenotypic data. If this eventuates, it may be quite possible to routinely incorporate more detailed assumptions on the co-ancestry of individuals to attain more accurate parameter estimates and make more effective selections for subsequent generations.

Acknowledgements

Maurice McDonald is gratefully acknowledged for making available the scored isozyme assays on which this study was based. Philippa Noble, Tim Jackson and Tom Baker are thanked for collaboration on the measurement of the Australian Low Rainfall Tree Improvement Group trial at Lismore, Victoria – the source of our phenotypic data. Charlie Bell and Tony Brown are thanked for useful discussions during the course of the work. We would also like to thank Rowland Burdon and three anonymous reviewers for their very helpful suggestions for the improvement of our draft manuscript.

Chapter 4: Characterising a *Eucalyptus cladocalyx* breeding population using SNP markers

Tree Genetics & Genomes (2013) 9:741–752

David Bush[1,2] and Bala Thumma[1]

1. CSIRO Plant Industry, GPO Box 1600, Canberra ACT 2601, Australia
2. Australian National University, Acton, Australia

Received: 25 July 2012 / Revised: 14 November 2012 / Accepted: 27 November 2012

Communicated by R. Burdon

Abstract

Population structure, family relatedness and inbreeding within a first generation *E. cladocalyx* breeding population were analysed with single-nucleotide polymorphism markers to underpin quantitative trait analysis and breeding program management. The breeding population, comprising families selected from wild and cultivated stands, was found to be strongly structured ($\hat{F}_{ST} = 18\%$), with two geographically-defined groups of South Australian wild subpopulations: Kangaroo Island (KI) and South Flinders Ranges (SFR). The selections from cultivated stands were shown to be derivative of SFR and had similar levels of diversity, suggesting that they were established from a broad genetic base. Relatedness and inbreeding among families was heterogeneous, ranging from completely outcrossed and predominantly half-sib (HS) to completely selfed. Families from the cultivated stands had minimal inbreeding and were close to HS on average. Among SFR subpopulations, family-average inbreeding was negatively correlated with growth, suggesting inbreeding depression (ID). Inbreeding was high, on average, in the KI subpopulations; however evidence of ID was absent, with highly inbred families amongst the most vigorous, perhaps indicative of purging of deleterious recessive alleles in a bottleneck event. The marker-based information suggested that modification of the usual assumptions of relatedness made in undertaking quantitative analysis of the first generation populations would be desirable.

Key words: co-ancestry, inbreeding, linkage disequilibrium

Introduction

Domestication and tree breeding of eucalypt species typically commences with first-generation testing of families from a range of subpopulations (or provenances) in provenance-progeny trials established across a range of sites in the target planting region. From quantitative genetic analysis of phenotypic data, subpopulations, families and individuals with superior properties for inclusion in the second generation are identified. In some cases, first-generation families will have been selected only from provenances identified as better-performing in previous provenance tests, though inclusion of untested provenances and selections from land races is also common. At the time of breeding program commencement, questions of subpopulation structure, such as whether 'provenances' from which seed was collected are genetically discrete entities, genetic origins of land-race selections and breeding system variables, e.g. outcrossing rates, are unresolved. Ideally, these questions should be answered before quantitative analysis commences, as assumptions concerning population structure and relatedness within and among families will affect the accuracy of selection and genetic gain in future generations. Molecular markers are a potentially powerful tool, complementary to quantitative analysis of phenotypic data, for managing breeding populations. Markers can be used to quantify within-population genetic diversity and manage the genetic base, identify population structure and hierarchy and elucidate pedigree to help manage inbreeding (Burdon and Wilcox 2007).

As eucalypts have a mixed mating system, there is potentially a mix of half-sib (HS), full-sib (FS), self-full-sib and self-half-sib relationships within families. In estimating genetic parameters related to future genetic gain, such as narrow-sense heritability (h^2), it is usual to accommodate the effects of mixed mating by dividing the additive variance component by a coefficient of relationship (ρ): in eucalypts, a value of 1/2.5 rather than 1/4, the value for true HS is commonly used (e.g. Eldridge et al. 1993). However, as the necessary paternal pedigree information does not exist, adjustments for heterogeneity of relatedness among subpopulations, families and individuals within families are not made, and the ranking of individual trees and families is equivalent to that made under the assumption of HS families.

Marker-based assays of young eucalypt germinant cohorts usually demonstrate a deficiency of heterozygous alleles relative to Hardy-Weinberg expectation (HWE) and positive values of the coefficient of inbreeding (f) (Potts and Wiltshire 1997). This is indicative of the effects of selfing and/or other forms of inbreeding associated with mixed mating and is more marked in some eucalypt species than others. For example McDonald et al. (2003) showed that wild *Eucalyptus cladocalyx* is less outcrossed and more inbred than most widely planted eucalypt species. They also showed marked heterogeneity in the degree of heterozygosity and inbreeding among populations and families. However for a number of eucalypts, inbreeding has resulted in marked inbreeding depression (ID) of

fitness-related traits (e.g. vigour, survival), resulting in strong selection against inbred individuals with nearly complete mortality of selfed individuals before sexual maturity (e.g. Costa e Silva et al. 2010b; Griffin and Cotterill 1988; Hardner and Tibbits 1998; Hardner and Potts 1997). Observed heterozygosity (H_O) in adult populations of wild eucalypts is, therefore, usually slightly higher than expected heterozygosity (H_E) (Potts and Wiltshire 1997). Low vigour and/or rapid mortality in progeny trials involving control-pollinated selfs and open-pollinated progeny has been demonstrated in *Eucalyptus regnans* (Griffin and Cotterill 1988; Hardner and Potts 1997) and *Eucalyptus globulus* (Costa e Silva et al. 2010b; Hardner et al. 1996), though ID of vigour and survival was not evident in a small number of *E. cladocalyx* families estimated to be completely selfed using isozyme markers at age 5 years (Bush et al. 2011a).

It is usually assumed that the mothers of progeny within a first-generation breeding population are unrelated. Wild eucalypt mothers should ideally be sampled at sufficient spacing, twice the height of the trees (Gunn 2001) or approximately 100 m apart, to minimise the selection of related individuals within the family clusters typical of eucalypts (e.g. Skabo et al. 1998) , which disperse most of their seed over a short distance (Potts and Wiltshire 1997). However mating and relatedness between quite distant trees is still possible, especially if stand flowering is sparse, some individuals flower asynchronously and/or pollinators are foraging over relatively long distances. Another common practice in first generation breeding populations is the integration of land-race selections, the genetic base of which is typically unknown. If established from a very narrow genetic base, trees sampled from distant parts of a plantation or even region may well be close relatives (e.g. Eldridge 1978).

In this study, we investigate the use of single-nucleotide polymorphism (SNP) markers to test assumptions that underpin orchard design and breeding program management of an *E. cladocalyx* F. Muell breeding population situated in southeast Australia. *E. cladocalyx* occurs in three disjunct regions of South Australia: Kangaroo Island (KI), Eyre Peninsula and the South Flinders Ranges (SFR) between latitudes 32-36°S (for grid references of specific subpopulations, see Appendix 2.1). In the SFR, *E. cladocalyx* typically forms extensive open forests, as either pure or mixed eucalypt species stands. Though the total extent of the forest there has been reduced by clearing in the last 140 years, reserves now protect large populations. Stands in the Flinders Chase National Park and at Cygnet River on KI are generally small and isolated within creek lines, where they often form a dominant overstorey. Another Kangaroo Island population at American River forms a small coastal strip on the foreshore of the town, the existing trees probably a remnant of a more extensive stand that has been removed in the last two centuries. On the Eyre Peninsula, *E. cladocalyx* occurs as low woodlands, the trees often having a stunted appearance. The populations on Eyre Peninsula are very fragmented due to clearing for agriculture, occurring in two disjunct, upland localities. Mean annual rainfall in the natural range is between 350 and 650 mm and mean annual temperature is 16-17.5°C

(Commonwealth of Australia Bureau of Meteorology records). The species is predominantly insect-pollinated and has a mixed-mating system, with individual trees ranging from self-compatible to self-incompatible (Ellis and Sedgley 1992). It has been widely planted since the 1870s in southern Australia on farms for naturally durable utility timber, fuelwood and windbreaks. It is the subject of tree breeding programs for low-rainfall planting environments in Australia (Harwood et al. 2007) and Chile (Mora et al. 2009). Population genetic diversity, structure, relatedness and inbreeding parameters within and among subpopulations and families in the breeding population are estimated. The implications of the findings in the context of domestication and breeding program management are discussed.

Materials and methods

Breeding populations and sampling

The *E. cladocalyx* breeding population was established at Lismore, Victoria (37°51'S 143°51'E), as 1 of 11 sites established throughout southeast Australia in year 2001 by the Australian Low Rainfall Tree Improvement Group (see Bush et al. 2009). 119 families from 13 subpopulations were sampled (Table 4.1; for subpopulation location data, see Appendix 2.1). Studies of range-wide selections by McDonald et al. (2003), using isozyme markers, and Bush et al. (2011b), who examined quantitative growth and wood trait data, have demonstrated that wild *E. cladocalyx* subpopulations form three groups, of which two are represented in the breeding population sampled in this study (KI and SFR). Families from the third, Eyre Peninsula group, were not included in the ALRTIG breeding population as prior provenance tests had demonstrated the group's markedly inferior growth and form (Harwood and Bulman 2001). The breeding population also contains families from cultivated stands including land-race selections from western Victoria and a planted stand and multi-provenance seed production area (SPA) in South Australia. Genomic DNA from an average of 10 random selections per family was sampled at age 5 years from the 20 available trees per family in the trial. Height and diameter at breast height (DBH) from all survivors (total of 2185 trees) of the 20 trees per family originally planted were measured at 5 and 7.8 years.

Table 4.1 MAF (minor allele frequency), average observed ($\hat{H}_O$) and expected ($\hat{H}_E$) heterozygosity, percentage of polymorphic loci, (%P) relatedness ($\overline{2\theta}$) and inbreeding coefficient ($\overline{f}$) calculated by the Wang (2007) triadic method

Subpopulation and location of trial	Description	MAF	$\hat{H}_E$ (SE)	$\hat{H}_O$ (SE)	%P	n (families/ trees)	Average $\overline{2\theta}$ (relatedness) not corrected for bias		$\overline{f}_t$
							Trees within families	Between trees in different families	
Wirrabara	Wild – SFR	0.29	0.37 (0.02)	0.34 (0.02)	97	22/216	0.38	0.14	0.14
Wilmington	Wild – SFR	0.29	0.35 (0.02)	0.31 90.02)	95	7/60	0.46	0.12	0.18
S. Wilmington	Wild – SFR	0.30	0.37 (0.02)	0.34 (0.02)	95	9/71	0.34	0.12	0.14
Mt Remarkable	Wild – SFR	0.30	0.37 (0.02)	0.33 (0.02)	96	7/166	0.42	0.13	0.15
Region	*SFR*	*0.30*	*0.38 (0.02)*	*0.33(0.02)*	*96*	*55/513*	*0.40*	*0.13*	*0.15*
Flinders Chase 1	Wild - KI	0.24	0.20 (0.02)	0.12 (0.02)	60	7/32	0.72	0.13	0.28
Flinders Chase 2	Wild - KI	0.17	0.23 (0.02)	0.17 (0.02)	99	21/208	0.56	0.17	0.24
American R.	Wild - KI	0.15	0.16 (0.02)	0.11(0.01)	72	6/72	0.70	0.13	0.28
Cygnet R.	Wild - KI	0.15	0.17 (0.02)	0.14 (0.02)	81	5/32	0.55	0.11	0.28
Region	*KI*	*0.18*	*0.22 (0.02)*	*0.14 (0.01)*	*78*	*44/344*	*0.63*	*0.14*	*0.27*
Mt Burr	Cultivated - SA	0.27	0.33 (0.02)	0.35 (0.02)	93	5/68	0.25	0.12	0.06
Majorca	Cultivated - VIC	0.29	0.36 (0.02)	0.36 (0.02)	95	5/60	0.29	0.12	0.10
Wail	Cultivated - VIC	0.29	0.34 (0.02)	0.33 (0.02)	92	3/28	0.27	0.06	0.10
Lismore	Cultivated - VIC	0.30	0.37 (0.02)	0.38 (0.02)	95	3/30	0.22	0.08	0.09
Kersbrook	Multi-prov. SPA	0.29	0.37 (0.02)	0.36 (0.02)	99	15/156	0.28	0.15	0.11
Group	*Cultivated stands*	*0.29*	*0.37 (0.02)*	*0.36(0.02)*	*95*	*32/342*	*0.26*	*0.11*	*0.09*
Population		*0.26*	*0.33 (0.01)*	*0.28(0.01)*	*90*	*119/1199*	*0.42*	*0.12*	*0.17*

SA South Australia, *VIC* Victoria, *KI* Kangaroo Island, *SFR* South Flinders Ranges, *SPA* seed production area

Molecular markers

SNP markers, though individually less informative than highly multi-allelic markers such as microsatellites (simple sequence repeats [SSR]), can provide useful information if employed in sufficient numbers. Wang and Santure (2009) estimate that for the purposes of co-ancestry estimation and pedigree reconstruction, 10 SNPs give similar information to 1 SSR. SNP discovery was carried out by creating DNA bulks from seedling tissue of three individuals from each of 149 open-pollinated family seedlots drawn from the SFR, KI and Eyre Peninsula subpopulation groups selected to be representative of the *E. cladocalyx* natural range. Amplicons of 39 genes (predominantly known and putative genes related to aspects of wood, vascular tissue development and some other traits identified in *Eucalyptus nitens*) that had been previously sequenced in *E. nitens* and other eucalypts, and for which existing primers were available, were selected for this study (see Appendix 2.2 which lists and places the amplicons within the *Eucalyptus grandis* genome reference assembly; Myburg et al. 2011). Sequencing was performed using the Roche/454 pyrosequence technology, producing ~200 bp reads. Sequencing reads were aligned using CLC Genomics Workbench Version 4 (CLC bio, Aarhus, Denmark) to identify SNPs. SNPs were selected primarily on the basis of read depth and sequence

quality score and absence of proximal SNPs (within approximately 10 bp of the target SNP) that can interfere with primer design. This resulted in a minimum minor allele frequency (MAF) >0.1. Target spacing of SNPs within genes was >500 bp. A total of 75 selected SNPs were genotyped using the Sequenom platform at the Australian Genome Research Facility. In many cases, multiple SNPs were selected from single genes (Appendix 2.2), allowing exploration of intra-gene linkage disequilibrium (LD) breakdown which is normally rapid in eucalypts (Grattapaglia and Kirst 2008). SNP identity was established by aligning SNP flanking sequence (200-300 bp) to the *E. grandis* genome reference assembly using the Phytozome Version 8 (Goodstein et al. 2012) online *basic local alignment search tool* (BLAST).

Allele frequency-dependent measures

Parameter estimates of observed and expected heterozygosity and tests for HWE and F statistics (F_{ST}, allelic diversity with subpopulations relative to the entire population; F_{IS}, inbreeding coefficient of an individual relative to the subpopulation) were calculated using Arlequin (Excoffier and Lischer 2010). These allele frequency parameters were used to study genetic diversity, population sub-structure and inbreeding. Genetic distance matrices were constructed and used as the basis for analysis of molecular variance (AMOVA; Arlequin) and exact tests of population differentiation (Arlequin). Unweighted pair group method with arithmetic mean (UPGMA) cluster analysis (POPTREE, Takezaki et al. 2010) was undertaken to examine population differentiation.

Marker neutrality

Marker neutrality is an important consideration for population genetics studies because markers that are under selection are typically outliers with respect to at least some of their properties, are then likely to cause biased estimates of certain parameters and may produce misleading results. Their removal is therefore desirable (Luikart et al. 2003). Beaumont and Nichols (1996) have shown that loci likely to be under selection will show unusually high levels of genetic differentiation (F_{ST}), and they and others have modelled selective neutrality by comparing F_{ST} and H_O against confidence limits simulated under various assumptions about population structure. Excoffier et al. (2009), in a refinement of the Beaumont and Nichols method, showed that false-positive identification of SNPs under selection may result where regional structuring of subpopulations exists. As *E. cladocalyx* is known to form three groups of subpopulations, confidence limits were simulated for 3 regional groups and 14 subpopulations. Calculations were implemented in Arlequin.

Linkage disequilibrium

SNPs from all genes were combined to estimate LD. Within-gene SNP positions obtained from BLAST-searching SNP flanking sequences against the preliminary *E. grandis* reference genome were used for plotting LD decay. LD measure R^2 was calculated using the Haploview program (Barrett et al. 2005). LD decay was investigated by plotting R^2 values against the relative positions of SNPs.

Relatedness and inbreeding

Measures of relatedness between individual trees, within and among families, subpopulations and whole populations are important but often unknown in first generation eucalypt breeding populations derived from mixed-mating in the wild and land-race selections. We used *Coancestry* Version 1.0 (Wang 2010) to estimate relatedness of pairs of trees (*dyads*) within each of the breeding populations. This software estimates the coefficient of relatedness (2θ), equivalent to twice the coefficient of co-ancestry (θ), for individual dyads. The coefficient of co-ancestry of two individuals reflects the probability that two gametes taken at random, one from each, carry alleles that are identical by descent (IBD) and is equivalent to the inbreeding coefficient (f) of their progeny should they be mated together (Lynch and Walsh 1998). The coefficient of relatedness and the coefficient of relationship (ρ), the latter often applied uniformly to whole populations to scale variance components in quantitative analysis, are related as follows:

For individuals i and j:

$$\rho_{ij} = \frac{2\theta_{ij}}{\sqrt{(1+f_i)(1+f_j)}} \qquad (4.1)$$

where f_i and f_j are their coefficients of inbreeding and θ_{ij} is the coefficient of co-ancestry. Assuming no ancestral inbreeding: for selfs, $f_i = f_j = 1/2$, $2\theta_{ij} = 1$ and $\rho_{ij} = 2/3$; and for outcrossed relationships, 2θ and ρ will be equivalent, since f is nil, e.g. for HS $f_i = f_j = 0$ and $2\theta_{ij} = \rho_{ij} = 1/4$.

The package *Coancestry* estimates 2θ by various methods. We selected the Wang triadic expectation maximisation strategy which has been shown to perform well when family structure is present and which also accounts for inbreeding (Wang 2007), an important consideration given previous findings of elevated selfing in *E. cladocalyx* (Bush et al. 2011a; McDonald et al. 2003).

Wang (2010) recommends that a critical first step in undertaking estimation of relatedness is determination of bias and precision of the selected method using data simulated to emulate the characteristics of the empirical marker system and data. Depending on the specific characteristics of the markers used, systematic under – or over-estimation of relatedness parameters such as f and 2θ are likely to occur. The opportunity is to quantify the likely direction and magnitude of these biases using marker data, simulated for specific, known relationship types, and that emulate the allele frequency characteristics of those used in the empirical study. This then allows an appropriate correction to be applied. Simulated data for varying numbers of biallelic markers was generated and analysed using *Coancestry's* in-built functionality. Allele frequencies from the empirical marker set, as well as those

drawn from a Dirichlet distribution, were used to simulate datasets with specified relationships among sets of dyads of varying sizes. *Coancestry* estimates confidence intervals (CI) by bootstrapping over loci. The effect on bias and precision of averaging over 45 dyads, representing the average *E. cladocalyx* family size, giving $\overline{2\theta}_f$ and $\overline{f}_f$, was investigated, with CI determined by bootstrapping over sets of within-family dyad estimates.

Other measures commonly used to describe mixed-mating systems are the proportion of outcrossed progeny, t and its complement of selfed progeny, s. The relationship between s, f and 2θ for families resulting from mixed mating, where FS relationships are absent, is given by Bush et al. (2011a):

$$\overline{2\theta}_{ij} = \tfrac{1}{4}(1 + \overline{f} + 2s + s^2 + 2\overline{f}s + \overline{f}s^2) \qquad (4.2)$$

Relationship between fitness and inbreeding

The relationship between family-average relatedness estimates ($\overline{2\theta}_f$ and $\overline{f}_f$) and fitness-related traits (tree height, DBH and survival) was explored by multiple linear regression (with and without subpopulation and subpopulation group terms) implemented in GenStat 14 (VSN International, Hemel Hempstead, UK).

Results

SNP marker properties

Genomic location of SNPs

BLAST searches were used to position SNPs within 10 of the 11 *E. grandis* genome large chromosome scaffolds (Myburg et al. 2011) (Appendix 2.2). SNPs were located on the basis of highly homologous (typically> 90 %) alignment of SNP primer sequence (~200-300 bp). SNPs, including several of those from the novel genes, were further annotated as 5' and 3' un-translated regions, promoter regions and intron and exon regions of *E. grandis* genes (Table 4.2). Of all SNPs, only four were non-synonymous (i.e. result in amino acid changes).

Average spacing of SNPs within genes was 818 bp. Using the method of Excoffier et al. (2009), one non-synonymous SNP from the PE2 gene was identified as an outlier. A plot of subpopulation differentiation (F_{ST}) versus heterozygosity is included in Appendix 2.3.

Table 4.2 Classification of SNP markers used to characterise the *E. cladocalyx* breeding population at Lismore, Victoria. Genomic locations were determined by BLAST of SNP primer sequence against the preliminary *E. grandis* genome sequence scaffolds (Myburg et al. 2011)

SNP genomic location	Number
Unannotated region	7
Promoter region	3
Intronic	33
Exonic	32
Synonymous	*27*
Non-synonymous	*4*
Untranslated	*1*
Total SNPs	**75**
From genes	39
From chromosome scaffolds	10

Figure 4.1 shows the distributions of MAF of the marker sets in each of the subpopulation groups compared with those of a simulated biallelic marker set with frequencies drawn from a Dirichlet distribution. The cultivated and SFR groups of subpopulations had similar MAF (median MAF=0.3 and 0.32, respectively). Most notably divergent from the Dirichlet allele frequency distribution was the KI group of subpopulations, which had a markedly lower MAF distribution.

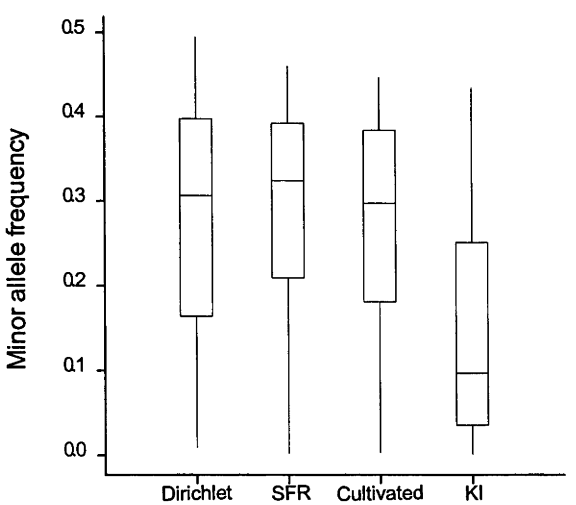


Fig. 4.1 Box and whisker diagram showing MAF from a simulated Dirichlet distribution, the South Flinders Ranges (SFR), Kangaroo Island (KI) and planted stand groups of *E. cladocalyx* subpopulations. *Whiskers* indicate the full range of the data, *boxes* indicate the inter-quartile range and *bars within boxes* indicate the median

Single-nucleotide polymorphism was variable among subpopulations. For the KI group of subpopulations, only one marker from the HPA gene was monomorphic in all subpopulations, though the percentage of polymorphism (%*P*) ranged between 60 % for the small Cygnet River subpopulation and 99 % for the larger Flinders Chase 2 population. SNPs from the MYB83 and PIP2 genes, which had MAF <0.03 in the KI subpopulations, were monomorphic in all SFR subpopulations, though generally the SFR %*P* was high (95-97 %). All markers were polymorphic in the cultivated stand group of subpopulations taken as a whole, though the aforementioned SNPs from MYB83 and PIP2 were present only in the Kersbrook SPA families. All except one Kersbrook family is of SFR maternal origin, but KI pollen parents were present in this mixed-provenance stand.

Linkage disequilibrium

LD decayed rapidly (Fig. 4.2). No evidence of LD (R^2>0.3) between inter-gene (for gene-within-chromosome scaffold) SNP pairs was found involving SNPs with MAF >0.05. On average, inter-gene LD (R^2) was 0.01. Where pairs of loci were near-monomorphic, as is the case in some of the subpopulations, some inter-gene SNP estimates were 0<R^2<0.3, though this is caused by correlation of numerous pairs of monomorphic individuals, probably not indicative of LD. Patterns of LD decay differed among the subpopulations. High levels of LD (R^2 values of 1) were more frequent, and LD persisted for longer distances in the KI subpopulations, whereas LD decayed more rapidly in the SFR and cultivated subpopulation groups. However, by 2500 bp, LD had decayed significantly (R^2<0.1) in all subpopulations.

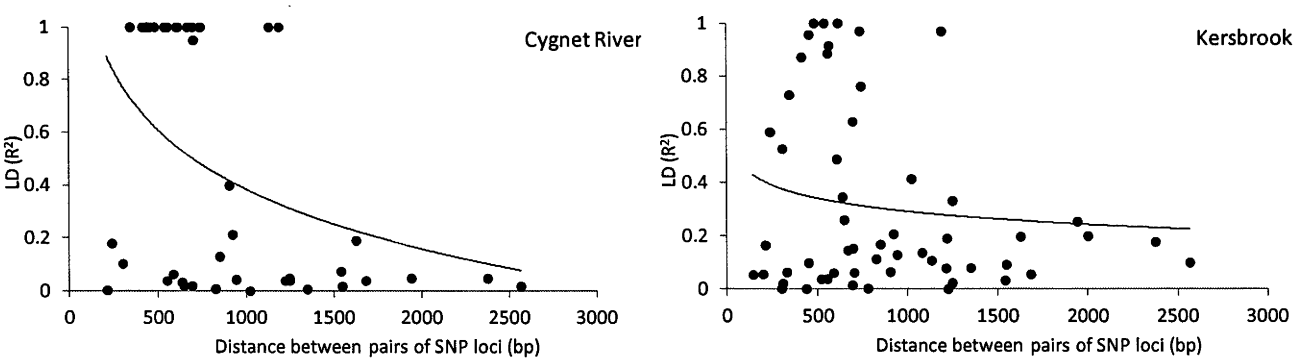


Fig. 4.2 Intra-gene (from within 39 genes) LD (R^2) for wild (Cygnet River) and cultivated (Kersbrook) *E. cladocalyx* subpopulations. Logarithmic trend lines indicate rapid decay of LD

Measures of heterozygosity and diversity

Heterozygosity

Heterozygosity ($\hat{H}_E$) averaged over subpopulations was 0.28 (standard error [SE], 0.01) (Table 4.1), though in the KI population it was 0.19, while in the SFR population and planted-stand groups

$\hat{H}_E$ averaged 0.37 and 0.35 respectively (Table 4.1). Observed heterozygosity ($\hat{H}_O$) was much lower than expected in trees from wild subpopulations especially KI (only 52 % of loci met HWE, $p<0.05$), but was very similar to expected in the cultivated stands (92 % of loci met HWE).

Partitioning of genetic diversity

Strong differentiation among subpopulations was indicated, with $\hat{F}_{ST}$ (taken as an average across individual-locus AMOVAs) ranging from 13.50 % among the three groups of subpopulations (SFR, KI and the planted stands) and 5.34 % among subpopulations within the groups. Significant inbreeding ($p<0.001$) was indicated for all wild subpopulations, with $\hat{F}_{IS}$ averaging 0.26 in KI and 0.07 in the SFR subpopulations. $\hat{F}_{IS}$ was not significantly different from zero for any of the planted-stand subpopulations.

The UPGMA phenogram (Fig. 4.3) shows wild subpopulations segregating strongly into two regions-of-provenance (KI and SFR), and exact tests of population differentiation showed this division to be significant ($p<0.001$). Most cultivated stands clustered together with Wirrabara, while Wilmington, South Wilmington and Mount Remarkable formed another group with Mount Burr (cultivated) being an outlier to the other SFR populations.

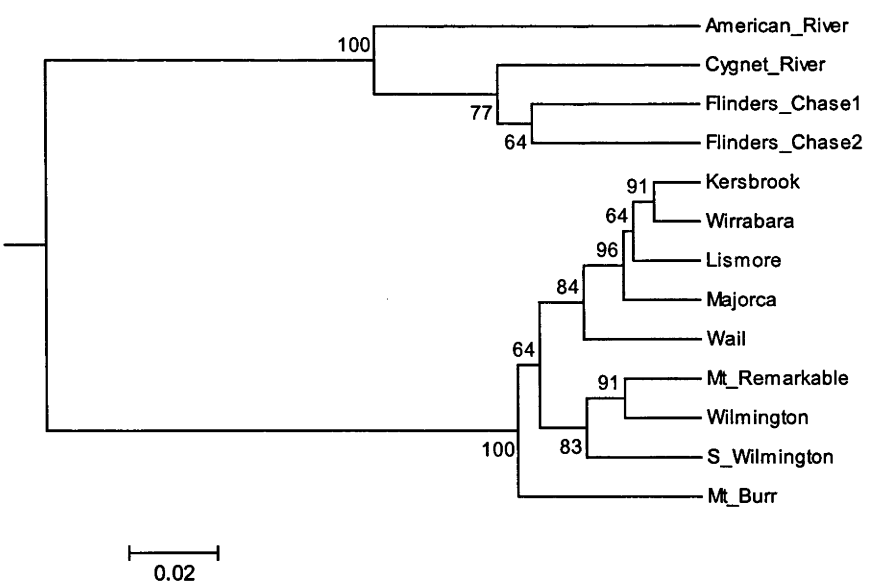


Fig. 4.3 UPGMA phenogram for the *E. cladocalyx* breeding population at Lismore, Victoria based on $\hat{F}_{ST}$ uncorrected genetic distances (Takezaki et al. 2010); 10 000 bootstraps over loci were performed and the percentage of replications in which a branch bifurcated is given at nodal bases

Relatedness

Bias and precision of estimates of relatedness and inbreeding– simulation study

Simulations based on sets of 45 and 500 dyads showed that there are consistent biases in estimated relatedness (2θ) and inbreeding (f) with very similar results for simulations based on allele frequencies from the empirical dataset (Table 4.3) and Dirichlet-distributed allele frequencies (full results not shown, though compare with 75-locus entry in Table 4.4). The magnitude and direction of bias for each relationship type is similar, decreasing only slightly as dyad-count increases. The magnitude of bias in 2θ decreased as relatedness increased and was positive for non-inbred relationships and negative for inbred ones. The results and their general stability across small and large sets of dyads indicate that correction of bias in estimation from empirical data would be possible and desirable; for example, estimates of 2θ should be reduced by around 0.07 and 0.05 for unrelated and HS dyads respectively.

Table 4.3 Bias from triadic estimation of relatedness using 75 simulated SNP loci with allele frequencies drawn from the *E. cladocalyx* empirical dataset. Simulated sets of 500 dyads (corresponding to the smallest *E. cladocalyx* subpopulation) and 45 dyads (corresponding to a family of 10 siblings) are presented. The 95 % CI (which apply to single-dyad estimates of 2θ and single-tree estimates of f) are calculated from 1000 bootstraps over loci

Relationship	Unrelated (CI)	Half-sib (CI)	Full-sib (CI)	Self-full-sib (CI)
Expected values				
2θ	0	0.25	0.5	1
f	0	0	0	0.5
500 dyads				
2θ	0.074 (0.01-0.33)	0.307 (0.12-0.59)	0.549 (0.40-0.84)	0.980 (0.73-1.30)
f	0.052 (0.00-0.26)	0.070 (0.00-0.28)	0.056 (0.00-0.30)	0.478 (0.30-0.70)
Bias 2θ	*0.07*	*0.06*	*0.05*	*-0.02*
Bias f	*0.05*	*0.07*	*0.06*	*-0.02*
45 dyads				
2θ	0.088 (0.01-0.33)	0.282 (0.09-0.57)	0.543 (0.37-0.82)	0.965 (0.73-1.32)
f	0.055 (0.00-0.25)	0.065 (0.00-0.26)	0.050 (0.01-0.30)	0.486 (0.27-0.73)
Bias 2θ	*0.09*	*0.05*	*0.043*	*-0.035*
Bias f	*0.06*	*0.07*	*0.05*	*0.01*

Dirichlet frequency-based simulations indicate that the magnitude of bias and CI for 2θ would decrease significantly if a considerably greater number of biallelic markers (500-1000) were employed (Table 4.4). Indications are that 75 SNP markers give only modest pairwise precision of estimation of 2θ, especially for unrelated or relatively distantly related dyads (95 % CI of 0.01-0.31 for unrelated dyads and 0.11-0.57 for half-sib dyads). With 500 loci, the 95 % CI narrow substantially to 0.01-0.12 and 0.18-0.37 for unrelated and HS relationships, respectively. Averaging across a number of dyads will also give satisfactory estimates from 75 loci. For example, with a 10-individual,

45-dyad HS family (the average size in the breeding population sample), $\overline{2\theta}_f$ is 0.29 (95 % CI 0.26-0.33) which is close to expectation once a bias of ~0.04 is subtracted from the estimate.

Table 4.4 From 500 simulated dyads (allele frequencies distribution drawn from Dirichlet distribution): (1) individual-dyad bias of 2θ and single-tree bias of f (deviation from expected values, $E[2\theta]$ and $E[f]$) and 95 % CI obtained from 100 bootstraps (over varying numbers of loci) for different relationships and (2) bias and CI of family-average relatedness ($\overline{2\theta}_f$) and inbreeding coefficient $\overline{f}_f$ obtained from bootstraps over a family of 10 siblings 45 dyads

	Unrelated		Half-sib		Full-sib		Self-full-sib	
Expected values								
2θ	0		0.25		0.5		1	
f	0		0		0		0.5	
Loci	Bias	CI	Bias	CI	Bias	CI	Bias	CI
2θ (single dyads)								
50	0.094	0.01-0.39	0.060	0.11-0.67	0.044	0.37-0.92	-0.029	0.68-1.39
75	0.067	0.01-0.31	0.041	0.11-0.57	0.040	0.39-0.82	-0.032	0.73-1.30
150	0.051	0.01-0.20	0.026	0.13-0.49	0.034	0.41-0.74	-0.017	0.79-1.23
250	0.036	0.00-0.15	0.025	0.15-0.43	0.026	0.42-0.68	-0.015	0.83-1.17
500	0.026	0.00-0.10	0.017	0.18-0.37	0.015	0.45-0.63	-0.011	0.87-1.11
1000	0.017	0.00-0.07	0.011	0.20-0.33	0.014	0.47-0.59	-0.008	0.91-1.08
$\overline{2\theta}_f$ (10 sibling, 45 dyad family average)								
75	0.068	0.00-0.09	0.042	0.26-0.33	0.040	0.51-0.58	-0.033	0.93-1.02
f (single trees)								
50	0.068	0.01-0.34	0.073	0.01-0.35	0.070	0.01-0.38	-0.030	0.26-0.77
75	0.053	0.00-0.25	0.052	0.00-0.26	0.053	0.01-0.29	-0.030	0.29-0.71
150	0.041	0.00-0.22	0.042	0.00-0.20	0.037	0.00-0.21	-0.030	0.33-0.66
250	0.031	0.00-0.14	0.034	0.00-0.15	0.029	0.00-0.16	-0.016	0.27-0.62
500	0.028	0.00-0.10	0.024	0.00-0.11	0.023	0.02-0.11	-0.010	0.40-0.58
1000	0.015	0.00-0.07	0.017	0.00-0.08	0.017	0.00-0.08	-0.005	0.43-0.56
$\overline{f}_f$ (10 sibling family average)								
75	0.043	0.00-0.25	0.051	0.00-0.25	0.054	0.00-0.28	-0.030	0.30-0.70

Relatedness within subpopulations

Trees from KI were closely related within families, with high proportions of selfing and/or ancestral inbreeding at the last mating required to explain subpopulation average $\overline{2\theta}_p$ values above 0.5 and up to 0.7, well above those expected for FS. Only four KI families had estimated $\overline{2\theta}_f < 0.4$ (see Appendix 2.4). Trees-within-families of the SFR population were also quite closely related on average, with $\overline{2\theta}_p$ of 0.4; however, the range of values was greater with several families appearing to be predominantly HS ($\overline{2\theta}_f < 0.3$ allowing for bias). Estimates of subpopulation inbreeding ($\overline{f}_p$) ranged between 0.14 and 0.28 for the wild populations and between 0.06 and 0.11 (close to zero once likely bias is accounted for) in the planted stands. This represents subpopulation average outcrossing

rates from $t = 0.5$ to $t = 1$ assuming selfing is the main form of inbreeding. For several wild families, $\overline{2\theta}_f$ estimates approached 1 (and, therefore, $\bar{f}_f$ approaching 0.5 and $t = 0$), indicating nearly complete selfing. Three individuals from KI (two from the same family) were homozygous at all 75 loci, indicating several generations of ancestral selfing. For most families from cultivated stands, $\overline{2\theta}_f < 0.3$, which, allowing for bias, is close to HS expectation, and only one family (from Kersbrook) showed signs of high levels of inbreeding with $\overline{2\theta}_f = 0.6, \bar{f}_f = 0.28$. A whole-population positive, linear relationship between f and 2θ was found, with coefficient of correlation R^2=0.93.

The average co-ancestry estimate for dyads-within-subpopulations that are expected to be unrelated (i.e. between trees from different families) ranged between 0.06 and 0.17 and was, on average, 0.12, which is slightly > 0 on average, allowing for bias ≈0.07. The cause of this class of inter-relatedness can be attributed to inter-relatedness between progeny of some specific families rather than generally elevated relatedness between all dyads within subpopulation. This was particularly strong in the Flinders Chase 2 subpopulation, where average inter-family relatedness ($\overline{2\theta}_{f\text{-}f}$) was 0.17, with some pairs of families having $\overline{2\theta}_{f\text{-}f}$ >0.5, indicating pollen flow between mothers, a common pollinator of the two mothers, relatedness of mothers and/or elevated ancestral inbreeding. Some evidence of relatedness among progeny of particular pairs of families was found within all of the wild provenances of *E. cladocalyx*, and in at least one pair of families per subpopulation in the planted stands, except for Majorca. Inter-family relatedness was uniformly elevated ($\overline{2\theta}_{f\text{-}f}$ = 0.15) in the Kersbrook SPA subpopulation.

Average 2θ between dyads in different subpopulations was 0.08: this value is very close to zero once bias has been accounted for.

Relatedness, inbreeding and fitness

Differences in response of the fitness-related trait, DBH at age 7.8 years, to family-average relatedness ($\overline{2\theta}_f$) and inbreeding ($\bar{f}$) were observed among subpopulation groups (Fig. 4.4). Multiple linear regression analyses including subpopulation group and $\bar{f}$ then $\overline{2\theta}_f$ explained 25 and 31 % of the variance in DBH at 7.8 years, respectively, with the interaction between the SFR group and the inbreeding and relatedness parameters being the only significant terms in the models (p<0.001) (Table 4.5). The significant negative relationship was similarly observed between inbreeding and the DBH and height traits in the SFR group of subpopulations, but not others, at the

age of 5 years. Wild families from KI subpopulations, many of which were highly related and inbred (Table 4.1), showed no significant regression response in height and DBH traits to either elevated $\overline{2\theta_f}$ or $\overline{f}$. The cultivated subpopulations had a narrow range of $\overline{2\theta_f}$ and $\overline{f}$ was generally low. Though weakly significant relationships between survival and $\overline{2\theta_f}$ and $\overline{f}$ ($p = 0.04$ and $p = 0.05$, respectively) were observed in the SFR subpopulation group, removal of a single family with high leverage resulted in non-significant relationships. Survival was generally very high in all of the progeny trials at measurement age (trial populations averaged 88-92 %).

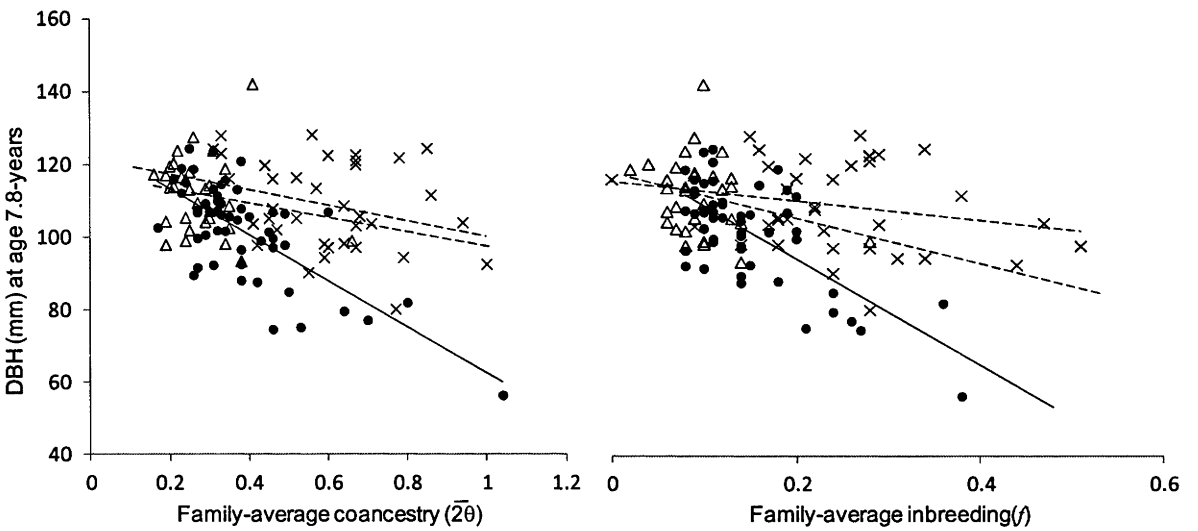


Fig. 4.4 Relationship between family-average relatedness ($\overline{2\theta_f}$), inbreeding ($\overline{f}$) and fitness trait, DBH, at age 7.8-years. Separate regression lines are fitted for three groups of subpopulations: SFR (*circles*, significant at $p<0.001$, R^2=0.38 and 0.48 for $\overline{2\theta_f}$ and $\overline{f}$, respectively), KI (*crosses*, not significant) and cultivated stands (*triangles*, not significant)

Table 4.5 Summary of regression analysis of family-mean relatedness ($\overline{2\theta_f}$) and inbreeding coefficient ($\bar{f}_f$) against phenotypic trait means (DBH, height and percent survival at various ages) with percent variance accounted for by regression and probability that the fitted line is different to zero for groups of subpopulations in *E. cladocalyx* (where significant, the regression coefficients were all negative)

Trait	Trait mean	$\overline{2\theta_f}$		$\bar{f}_f$	
		Percent variance explained	Significance (p)	Percent variance explained	Significance (p)
E. cladocalyx - Kangaroo Island					
DBH 7.8 years	10.9 cm	7	NS	2	NS
DBH 5 years	7.9 cm	<1	NS	<1	NS
Height 5 years	5.8 m	5	NS	3	NS
Survival 7.8 years	93 %	<1	NS	<1	NS
Survival 5 years	94 %	2	NS	<1	NS
E. cladocalyx - South Flinders Ranges					
DBH 7.8 years	10.2 cm	50	<0.001	47	<0.001
DBH 5 years	8.0 cm	45	<0.001	46	<0.001
Height 5 years	5.5 m	41	<0.001	38	<0.001
Survival 7.8 years	91 %	7	0.04	6	0.052
Survival 5 years	90 %	6	0.05	2	NS
E. cladocalyx - cultivated stands					
DBH 7.8 years	11.2 cm	<1	NS	4	NS
DBH 5 years	8.7 cm	<1	NS	4	NS
Height 5 years	5.9 m	<1	NS	<1	NS
Survival 7.8 years	92 %	<1	NS	3	NS
Survival 5 years	92 %	1	NS	2	NS

Discussion

SNP marker-based assays of the breeding population yielded a variety of information that can be used to manage the breeding program. The panel of 75 loci proved suitable for making allele frequency-based population-level estimates and also provided information on short-distance LD decay. In accordance with other eucalypts (Grattapaglia 2007; Grattapaglia and Kirst 2008; Thavamanikumar et al. 2011; Thumma et al. 2005), LD was found to decay rapidly, though it was more extensive in the KI subpopulations – a result consistent with elevated levels of inbreeding observed in this subpopulation group. Only one SNP from the PE2 gene was identified as likely to be under selection on the basis of its heterozygosity. This may indicate that this locus is under selection, though it may also be as a result of allelic differentiation caused by founder or bottleneck effects. Dillon et al. (2013) identified 8 % of SNPs likely to be under selection in a broadly similar set of *P. radiata* genes. We found that *E. cladocalyx,* from the subgenus *Symphyomyrtus* and monotypic section *Sejunctae,* appears to be highly homologous with *E. grandis* (subgenus *Symphyomyrtus* section *Latoangulatae*), with successful location of all SNPs within the currently mapped extent of the *E. grandis* genome.

The breeding population was strongly structured, with $\hat{F}_{ST}$ totalling 18 % among subpopulations, the majority of which could be apportioned to the two groups of subpopulations corresponding to the wild SFR and KI regions identified by McDonald et al. (2003) using allozymes and by Steane et al. (2011) using Diversity Arrays Technology (DArT) markers. The cultivated stands clustered with the SFR subpopulations, and were similarly diverse, indicative of a broad, original genetic base in the SFR, though they were less inbred, on average, probably reflecting the lack of neighbourhood structure (i.e. fewer closely situated near-relatives) in planted stands. Strong segregation in quantitative trait variation of growth, form and wood property traits corresponding to the wild groups of subpopulations has previously been identified (Bush et al. 2011b), with the phenotypically selected families from cultivated stands performing particularly well (Callister et al. 2008). Further selections from these cultivated sources could be made as infusions in future generations. Our estimate of H_E was 0.28 on average, higher than 0.15, the isozyme-based estimate of McDonald et al. (2003) from a young seedling population of *E. cladocalyx*. However the patterns of diversity among subpopulations in this study were very similar to those found by McDonald et al. (2003), with strong geographical regional differentiation of subpopulations, SFR having considerably greater diversity than KI, which supports their finding of SFR origins for the Wail cultivated stand. Wild subpopulations, particularly those from KI, showed strong signs of inbreeding, as evidenced by low MAF, marked deficit of heterozygotes, low genetic diversity (H_E and $\%P$) and greater persistence of LD.

Estimates of relatedness between pairs of individual trees would be beneficial in first generation breeding populations, which are typically derived from sources where only the mother's identity is known – even the original subpopulation identity can be obscure in the case of land-race selections. Pairwise estimates can be used to construct genomic relationship matrices, which substitute for pedigree-based additive relationship matrices, where pedigree does not exist, as is the case in testing and domestication of wild selections, to identify pedigree errors or to better calculate variance-covariance in accounting for IBD (Van Raden 2007). However our simulation study indicated that, though bias in estimation of 2θ and f was of minor practical consequence and in any case could be consistently identified and corrected for using a panel of 75 SNP markers, single-dyad estimates are not sufficiently precise to use reliably in a relationship matrix, and that precision would need to be improved by employing more markers: 500-1000 SNPs would be required. However, averaging across groups of dyads, families (~10 individuals or more) or subpopulation groups, for example, can provide useful, cost-effective estimates. The cost-effectiveness of family-level estimates arises not only because fewer markers are used, but far fewer individuals are genotyped – the family-level estimates can be applied to numerous family members across sites. For first-generation breeding populations, it would then be possible to construct a marker-based genomic relationship matrix, using

a relatively modest genotyping effort, which will allow quantitative analysis to proceed while accounting for heterogeneity of family and subpopulation relatedness.

The *E. cladocalyx* breeding population demonstrated a large degree of heterogeneity in relatedness among families, a result presaged by the study of McDonald et al. (2003) on germinant seedlings using eight isozyme loci. Our study indicates that f ranges between 0 and 0.5 and, accordingly, family outcrossing (t) ranges between 0 and 1. Elevated f might be largely attributed to selfing in the parental generation; however in some families, evidence of very high levels of inbreeding were detected, with three individuals from KI, two of which were from the same family, found to be homozygous at all loci, indicating several generations of selfed matings or other ancestral inbreeding leading to allele fixation. Though the SNP loci used in this study are widely distributed throughout the genome, the individuals are being examined, as a follow-up, at a greater number of loci to determine the extent of fixation: it is possible that heterogeneous variable regions of the genome still exist. The strong correlation between 2θ and f indicates that much of the higher relatedness in *E. cladocalyx* is attributable to inbreeding. However, in several families, high 2θ is not accompanied by elevated f, and this is indicative of a significant proportion of FS relationships, a result not indicated by the isozyme study of McDonald et al. (2003).

Contrasting patterns of relationship between fitness-related trait (DBH) and measures of relatedness and inbreeding were observed among the breeding populations. Among the cultivated stands, significant relationships between neither DBH nor survival and inbreeding were indicated. Relatedness was generally quite close to HS on average, with low levels of inbreeding. The wild subpopulation groups (SFR and KI) had a wide range of family-average coefficients of relatedness and inbreeding. SFR family-average growth declined significantly in response to increased relatedness and inbreeding. Controlled pollination experiments have demonstrated a similar marked decline in fitness associated with selfing in a number of eucalypt species including *E. grandis* (Hodgson 1976), *E. nitens* (Hardner and Tibbits 1998), *E. globulus* (Hardner and Potts 1995) and *E. regnans* (Griffin and Cotterill 1988). In *E. regnans* and *E. globulus* this has been accompanied by quite rapid mortality in progeny trials. The relationship between inbreeding and survival is only weakly significant for the SFR subpopulation group in the progeny trial at Lismore, though in the long run, suppression of the smaller, inbred families by the more vigorous ones would probably result in increased mortality. Costa e Silva et al. (2011) have shown that ID in *E. globulus* survival is differentially expressed across environments, and it may be that selection against inbred *E. cladocalyx* families is stronger in the wild and on more demanding plantation sites. However, KI families, which were on average more inbred than SFR, appeared to be showing no significant signs of ID of survival or growth traits at this site. This is an unusual finding in eucalypts, though Costa e Silva et al. (2010a)

have recently shown that ID is variable among families and subpopulations of *E. globulus*. This result may indicate a lower genetic load of deleterious recessive alleles in the KI subpopulations. Stands on KI are generally small and isolated within creek lines in the Flinders Chase National Park and at Cygnet River, where anthropogenic disturbance has probably been relatively minimal. The American River stand is a small coastal strip on the foreshore of the town, the existing trees probably a remnant of a more extensive stand that has been removed in the last two centuries. Mimura et al. (2009) have similarly found that levels of outcrossing are lower in small fragmented stands of *E. globulus* than in stands occurring as more extensive forest. It is possible that the KI group of subpopulations, which are markedly less genetically diverse than those of the SFR, may have gone through a genetic bottleneck which has resulted in purging of recessive alleles. McDonald et al. (2003) postulated that this may have resulted from changes in climate and sea levels during the Pleistocene during which time the species retreated to small refugia on KI and more extensive open forests in the uplands of the SFR.

An often-implicit assumption in the analysis of first generation eucalypt populations is that levels of inbreeding and relatedness among populations and families are homogenous. This assumption appears to be critically flawed for *E. cladocalyx*, where inbreeding levels varied from low to very high within and among subpopulations. Taking the traditional approach of using a single value of ρ to scale variance parameters, a value of $\rho = 1/2.8$ is indicated. However the respective values of ρ for the SFR, KI and cultivated subpopulations would be approximately 1/3, 1/2 and 1/4. Quantitative analysis could be improved by using the marker-based estimates of the coefficient of co-ancestry to adjust the additive relationship matrix at the family level, a relatively complex step not normally carried out for first generation domestication breeding parameter estimation. This procedure was carried out for a small selection of *E. cladocalyx* wild families by Bush et al. (2011a), who demonstrated that some re-ranking of breeding values may result.

While correction of the additive relationship matrix will result in a more realistic additive variance model, the problem of non-additive variation, particularly dominance, is not addressed. The effects of dominance in this breeding population may be considerable, given the clearly different responses to inbreeding among subpopulations. The relative magnitude of the additive and non-additive genetic effects will impact on selection efficiency in the early phase of domestication. Substantial non-additive genetic variation associated with inbreeding was found in *E. globulus* (Costa e Silva et al. 2010a), and because of the persistence of highly inbred individuals in the *E. cladocalyx* breeding population, this is likely to be an important consideration. Inbreeding can be theoretically expected to give rise to greater between- family and within-family variance relative to outcrossed populations (Kelly 2004), and Costa e Silva et al. (2010b) have shown that this can amount to very large within- and between-family differences in experimentally selfed and outcrossed populations of *E. globulus*.

Heritability estimates for growth traits from three *E. cladocalyx* sites in Western Australia, which include a subset of the families described here, were unusually high (0.41-0.85) (Callister et al. 2008) despite application of $\rho=1/2.5$, a value that, as it applies to additive variation, is too high on average given the large proportion of material from cultivated stands in the Western Australian breeding population plantings. The result may well indicate significant underlying non-additive components of variation that have led to inflated variance associated with the inbred families.

Concluding remarks

The study has shown that molecular markers can be used to characterise eucalypt breeding population structure, clarifying genetic origins of land-race selections, revealing subpopulation structure and providing information of relatedness within and between families. We have shown that the breeding population is strongly structured and that the levels of inbreeding in some wild subpopulations are unusually high, while selections from cultivated stands are generally outcrossed. Modification of the usual assumption of homogeneity of relatedness within and between first-generation families, based on the marker data, may, therefore improve quantitative analysis. We also observed differential responses to inbreeding among the subpopulations, with indications of inbreeding depression in the SFR group of subpopulations. Accounting for the likely effects of non-additive variation on within-family and between-family variance is likely to be another important consideration for this species.

Data Archiving

Nucleotide sequences described in Appendix 2.2 have been deposited in the National Center for Biotechnology Information (NCBI) Genbank repository

Acknowledgements

We would like to acknowledge Bronwyn Matheson and Charlie Bell for their assistance with DNA extraction and preparation of samples for sequencing. Tim Jackson, Des Stackpole, Paul Macdonell and David Spencer are thanked for their assistance with management, sampling and measurement of the field trial. Chris Harwood and Shannon Dillon made useful suggestions on a draft version of the paper.

Chapter 5: Genetic parameter estimates informed by a marker-based pedigree: a case study with *Eucalyptus cladocalyx* in southern Australia

Tree Genetics & Genomes (2015) 11:798

David Bush[1, 2], Dominic Kain[2,3], Peter Kanowski[1], Colin Matheson[1,2]
1. Commonwealth Scientific Industrial Research Organisation (CSIRO) Australia
2. Australian National University
3. HQPlantations, Queensland

Communicated by R. Burdon

Received: 10 June 2014 / Revised 14 September 2014 / Accepted 24 October 2014

Abstract

Analysis of stem diameter, height and axis persistence (AP) in a first generation *Eucalyptus cladocalyx* breeding population comprising 137 wild and land-race families planted at 11 sites in southern Australia revealed significant genetic variation among subpopulations and among families within subpopulations. Alternative analyses were carried out using individual-tree mixed models that (i) assumed the trees within families were half siblings (HS) and (ii) used molecular marker-based information to account for highly heterogeneous relatedness and inbreeding depression among families resulting from mixed mating (MM). For certain site and trait combinations, the HS models would not converge, as estimates of additive variance exceeded the total phenotypic variance, demonstrating the fundamental unsuitability of the HS assumption for this breeding population. Where HS models converged, moderate to very high heritability estimates resulted for growth traits. The MM assumption resulted in re-ranking of individual-tree breeding values and markedly lower estimates of narrow-sense heritability for all trait-site combinations. In some cases, however, heritability remained moderate to high, probably reflecting unquantified dominance variance in some highly-inbred subpopulations. Genotype-by-environment interaction was significant overall, due to reactivity of genotypes on a few sites, with type-B correlations between pairs of sites ranging from 0.06-0.99. Generally, families from the Australian land race were found to perform particularly well for both growth and AP traits. Some wild families were found to be vigorous, despite significant inbreeding. The study has demonstrated that traditional models assuming non-relatedness and/or homogeneous inbreeding in first generation eucalypt breeding populations can be significantly improved upon by flexible mixed models that integrate marker-based data.

Key words: *Eucalyptus*, inbreeding depression, mixed mating

Introduction

Eucalyptus cladocalyx F. Muell occurs naturally in a restricted natural range in South Australia. There are three disjunct regions of provenance (groups of subpopulations): Kangaroo Island (KI), the Eyre Peninsula and the South Flinders Ranges (SFR). The species is well suited to planting on dryland sites in temperate and Mediterranean environments. Significant plantations exist in western Victoria, the Western Cape Province of South Africa (de Lange et al. 2013) , northern Africa and Mediterranean Europe (Jacobs 1979) and it is being genetically improved for planting in Chile (Mora et al. 2009). The species' uses include shelter belts, sawn timber, naturally-durable poles and posts (Bush et al. 2011b) and apiary (de Lange et al. 2013; Mora et al. 2009).

Open-pollinated families of forest trees are commonly assumed to comprise half-sibs, which would result in the event of panmictic mating, each offspring being the result of an outcross to a different male parent. In fact this is unlikely to be true for progeny of wild eucalypts, as most species studied have a mixed-mating system (Eldridge et al. 1993), producing a mix of outcrossed half-sib (HS) and full-sib progeny as well as inbred progeny resulting from selfing and mating with close relatives. Bush and Thumma (2013) used 75 single-nucleotide polymorphism (SNP) markers to estimate the relatedness and inbreeding within and among five-year-old open-pollinated families of *E. cladocalyx*. This study showed that, overall, estimated relatedness ($2\hat{\theta}$) was highly heterogeneous (Fig. 5.1); in the wild populations it was much higher than ¼ which is the expected value for half-sibs; and was associated with elevated coefficients of inbreeding (*f*). This was particularly marked for many of the KI families (subpopulation average $2\hat{\theta}$=0.63; $\hat{f}$=0.27). In contrast, the families from cultivated stands were predominantly outcrossed, and were close to half-sib on average with $2\hat{\theta}$ = 0.26. Bush and Thumma (2013) also found that the subpopulation groups differed in their inbreeding response. Phenotypic growth of KI families, which were most inbred on average, appeared not to decline significantly in response to elevated *f*, while those from the SFR showed a decline in growth in response to increasing *f* indicating inbreeding depression (ID). Though ID is known to be variable within eucalypt populations (Costa e Silva et al. 2010a) the response of the KI material is unusual. Typically, high *f* leads to poor growth (as demonstrated in the SFR group) and high mortality (e.g. Costa e Silva et al. 2010b; Griffin and Cotterill 1988; Hardner and Potts 1995). Inbreeding depression can also be a source of inflated heritability estimates, particularly if it is variable among families (Borralho 1994). One way of accounting for inbreeding depression in linear mixed models is to include *f* as a fixed-effect covariate (White and Hodge 1989). This approach has recently been taken by Doerksen et al. (2014) who used marker-based estimates of *f* to model ID in a population of white spruce.

The classical approach to analysis of first generation open-pollinated progeny trials of tree species with mixed-mating systems is to assume that all trees within-families are uniformly related, i.e. they are either half-sib, or somewhat closer than half-sib, because of inbreeding including selfing as well as the presence of some full-sibs. In calculating narrow-sense heritability, family variance component (σ^2_f) estimates are usually multiplied by a coefficient of four for half-sib families, or less than four to correct for bias attributable to closer relatedness (Squillace 1974). A value of 2.5 is commonly used for eucalypts (Eldridge et al. 1993). With advances in computing power and software that enable flexible mixed-modelling approaches, individual-tree models, which are a forestry analogue of the animal model (Mrode 1996), can now be employed. These models estimate the additive genetic variance among individual trees (σ^2_t) rather than family variance, and account for expected relatedness among individuals via an additive relationship matrix (**A**) that is usually constructed from known pedigree information (e.g. Baltunis et al. 2008; Costa e Silva et al. 2006; Gapare et al. 2008; Jarvis et al. 1995). In the first generation, **A** is normally populated with values corresponding to half-sib relationships. If closer relatedness is likely, then the variance component estimates are simply scaled back *post hoc*, or global adjustment of the matrix resulting in homogeneous within-family relatedness is possible in some software (Dutkowski and Gilmour 2001; Gilmour et al. 2009). However, such an approach would be flawed in the ALRTIG *E. cladocalyx* breeding population given the indications of highly heterogeneous relatedness among families, a situation that is known to extend, at least to some extent, to other species (e.g. Doerksen et al. 2014; Surles et al. 1990).

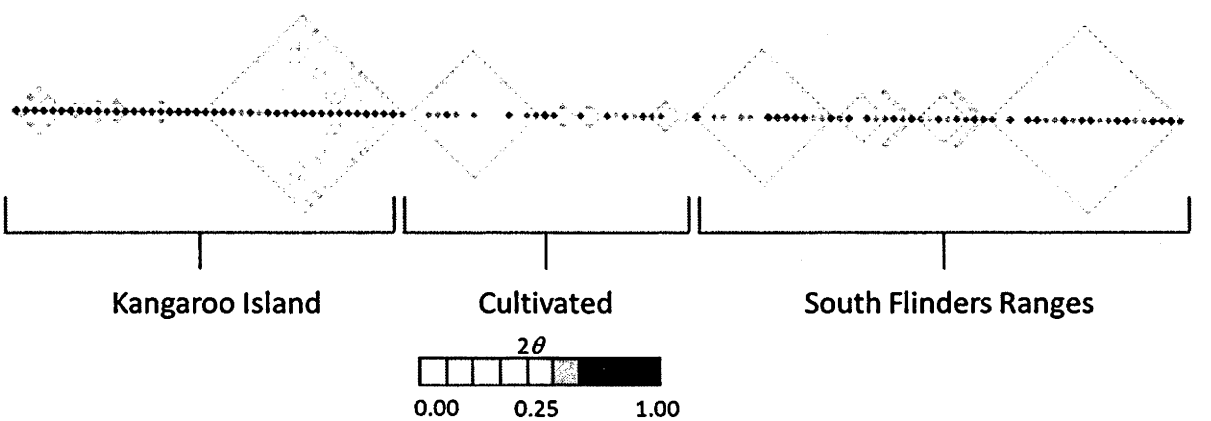


Fig. 5.1 Graphical representation of family-average relatedness ($2\hat{\theta}$) within and between families based on data from Bush and Thumma (2013). Each coloured block on the central "spine" indicates within-family relatedness for each of the 137 families. Flanking blocks depict average inter-relatedness of families within provenances. These data were used to construct **A** under the mixed mating assumption. The half-sib model assumption would be graphically represented by a yellow spine with no flanking blocks.

An alternative approach is to substitute a *genomic relationship matrix* (Van Raden 2007) for the classical pedigree-based one; i.e. pairwise relatedness of all trees is estimated using molecular markers. The use of marker data to infer relatedness of progeny can be usefully applied to base populations where no pedigree information, other than the maternal and subpopulation identities of individuals, is known. If sufficient markers are available, Mendelian sampling - the variation of average relatedness among members of a family due to random sampling of the two possible alleles at each locus from each parent during gamete formation - can also be accounted for. This cannot be determined by traditional pedigree methods or records (Klápště et al. 2013) and is often referred to as *realised relatedness* (Visscher et al. 2006).

Though gathering dense marker data for moderately large numbers of individuals has become cheaper, the present cost of genotyping entire, across-sites breeding populations and/or large numbers of trees at sufficient density to allow accurate individual tree (i.e. pairwise) estimates of realised relatedness appears to have limited its use with larger tree breeding populations. Existing studies in conifers have been restricted to relatively modest numbers of trees and/or markers (Doerksen and Herbinger 2010; Gaspar et al. 2009; Klápště et al. 2013; Kumar and Richardson 2005; Zapata-Valenzuela et al. 2013), but have nevertheless shown that it is possible to considerably improve information on relatedness and heritability estimates and even attempt to estimate non-additive effects using information from inferred full-sibs in open-pollinated progeny (e.g. Klápště et al. 2013).

A compromise solution for larger, first-generation populations for which there is no established pedigree is to estimate the family-average relatedness from a relatively small number of trees per family inferred from a modest-sized panel of markers (Bush and Thumma 2013). This average result could then be applied to the entire breeding population by adjusting the assumed half-sib family pedigree such that trees within families are at least half-sib, or more closely related as indicated. This is effectively a hybrid between the fully-fledged genomic realised relatedness matrix and the pedigree-based approaches. It does not account for Mendelian sampling, but it does allow a correction for the heterogeneity of relatedness to be made at the family level and for adjustment of the default assumption that each family comprises half-siblings.

Strong differentiation in performance has been demonstrated in *E. cladocalyx* at the region of provenance, provenance and family levels. This extends to growth and stem-form (Callister et al. 2008; Harwood et al. 2007), wood-property (Bush et al. 2011b) and flowering traits (Mora et al. 2009). Genotype-by-environment interaction (GxE) of *E. cladocalyx* growth and form traits has previously been found to be minimal, based on a study of a subset of the data included here, comprising three sites in Western Australia (Callister et al. 2008). A further finding of that study was

that narrow-sense heritability estimates were unusually high for growth ($\hat{h}^2$ = 0.40-0.85) and stem straightness ($\hat{h}^2$ = 0.29) traits. Heritability estimates for growth, form and flowering precocity from Chile (Mora et al. 2009; Vargas-Reeve et al. 2013) and for wood-property traits (Bush et al. 2011b) have also been moderate to high.

As previous traditional analyses of first-generation *E. cladocalyx* trials have given higher than normal estimates of heritability, and heterogeneous family-level relatedness and inbreeding have been demonstrated, we have sought to integrate available marker-based data to improve quantitative genetic analysis. We compare single- and across-sites additive genetic parameter estimates from a first-generation breeding population established in 2001 by the Australian Low Rainfall Tree Improvement Group (ALRTIG) across 11 sites in southern Australia (Harwood et al. 2007). Analyses were conducted using an assumed half-sib pedigree and an alternative model where the degree of inbreeding and coancestry within each family has been estimated using molecular markers. The marker-based model accounts for ID, and relatedness is adjusted via the additive relationship matrix, **A**. We discuss the impact of heterogeneous relatedness and inbreeding at the subpopulation and family level and the importance of accounting for it in the context of a first-generation breeding program.

Methods

Breeding population

The breeding population comprised 137 open-pollinated families from two wild regions of provenance (groups of provenances), South Flinders Ranges (SFR) and Kangaroo Island (KI), both in South Australia, and open-pollinated families from phenotypically-selected candidate plus trees from cultivated stands in South Australia and Victoria (Table 5.1). Fifteen families were sourced from a multi-provenance seed production area. The families were established across 11 sites in southern Australia: near Moora (MOR), Kojonup (KOJ), Wellstead (WEL) and two sites near Esperance (ESP1, ESP2), in Western Australia; in the Avenue Range (AVE) and near Bordertown (BTN) in South Australia; near Hamilton, (HAM) and Lismore (LIS) in Victoria; and, near Corowa (COR) and Wagga Wagga (WGW) in New South Wales (Table 5.2). These sites were selected as typical of temperate/Mediterranean plantation sites for the species, with mean annual rainfall between about 450 and 700 mm.

Six of the 11 sites include a relatively large number of families (84-133) and these were established as primary progeny trial sites (Table 5.2). The remaining five have between 42 and 64 families, and in these, many of the provenances are not represented at all or by fewer than five families (see Appendix

3.1 for distribution of families across the sites). These sites were established as secondary sites that could be selectively thinned for seed production based on information from across the whole trial series.

Table 5.1 Summary of seedlots included in the ALRTIG breeding population

ATSC seedlot	ATSC seedlot name	No. families	Lat. (°S)	Long. (°E)	Altitude (m)	MAR (mm)	Forest type
Flinders Ranges							
19348	Wilmington	9	32.70	138.11	550	600	Open forest
20388	Mt Remarkable	16	32.73	138.10	580	600	Open forest
16089	9 km S. of Wilmington	10	32.72	138.10	580	600	Open forest
20414*	Wirrabara SF	6	33.08	138.18	480	580	Open forest
20268	Wirrabara SF	10	33.08	138.18	480	580	Open forest
Kangaroo Island							
20265	American River	6	35.77	137.77	176	520	Open forest
20266*	Cygnet River	7	35.72	137.48	20	520	Open forest
16022	Flinders Chase NP	8	35.75	136.63	80	660	Open forest
20267*	Flinders Chase NP	12	35.95	136.70	60	660	Open forest
19717	Flinders Chase NP	12	35.95	136.73	70	660	Open forest
Cultivated stands							
N/A	Kersbrook SPA	15	34.73	138.83	360	750	Seed production area
N/A	Mt Burr	5	37.43	140.35	50	710	Block planting
21044	Wail	5	36.50	142.07	110	410	Block planting
21043	Majorca	5	37.12	143.80	230	520	Block planting
21042	Lismore	4	38.40	142.57	20	600	Shelter belt

N/A not applicable, *SF* State Forest, *NP* National Park, *SPA* seed production area, *MAR* mean annual rainfall

The trials were established with 4-5 complete replicates of 4- or 5-tree family plots giving 20 or 25 trees per family at each trial. Planting rows and/or columns formed additional incomplete blocks. The trials were planted between July and September 2001. Establishment details of the KOJ, WEL and ESP2 trials were previously reported by Callister et al. (2008) and by Bush et al. (2009) for the others.

Table 5.2 Summary site data for the ALRTIG breeding population

	Site										
	MOR	KOJ	WEL	ESP1	ESP2	AVE	BTN	HAM	LIS	COR	WGW
State	WA	WA	WA	WA	WA	SA	SA	VIC	VIC	NSW	NSW
Latitude (°S)	30.87	34.05	34.65	33.48	33.7	36.77	36.25	36.36	37.85	35.98	34.99
Longitude (°E)	115.97	117.15	118.3	122.02	122.15	140.20	140.57	142.12	143.50	146.38	147.48
Altitude (m)	200	290	160	150	100	40	30	205	200	175	210
Soil description	Sand over sandy-clay-loam	Gritty sand /loam duplex	Deep grey sand	Deep yellow sand	Deep yellow sand	Sandy loam	Sand over clay	Duplex clay-loam over clay	Buck-shot gravel, clay-loam	Alluvia l loam	Red clay-loam to sandy - clay-loam
Rainfall (mm/year)	460	495	539	620	570	600	460	685	600	540	570
Evaporation (mm/y)	1640	1325	1426	1679	1694	1460	1600	1298	1300	1570	1860
Measure age (y)	4.3	5.5	5.5	4.7	5.5	5.5	4.2	3.8	5.2	4.4	5.3
Families	117	42	42	61	42	64	130	96	119	84	109

Site and phenotypic data

The trials were measured at ages between 3.8 and 5.5 years (Table 5.2). Diameter at breast height (DBH) was measured at every site. Height (HT) was measured at all sites except MOR and ESP1. Axis persistence (AP) was measured at all sites except MOR. AP was rated on a scale of 1-6 with 1 corresponding to the lowest occurring fork (a fork being defined as an apically-dominant second leader with diameter greater than half that of the main stem at the point of divergence) at the base; 2, a fork in the lowest quarter of the tree; 3, a fork in the second quarter; 4, a fork in the third quarter; 5 a fork in the uppermost quarter and; 6 no fork. Two of the trials, BTN and AVE, were selectively thinned from 5 to 3 trees per plot prior to the measurement reported here. Mean annual rainfall and pan evaporation records were taken from the Australian Bureau of Meteorology weather station nearest to each site. Soils were characterised by excavation of soil pits at each site.

Family phenotypic trait means, based on all trees at each site, were regressed on estimated family inbreeding coefficients ($\hat{f}$) to investigate ID. To provide trait measures adjusted for differences in overall growth rates across sites, the data from each site were standardised so that for each tree $z=(y-\mu)/\sigma$ where y is the individual's phenotypic value, μ is the site mean, σ is the site standard deviation and z is the individual's standardised value. The standardised phenotypic values for each family were then averaged across sites to give an overall measure of family phenotypic values for each trait. Multiple linear regression analyses including subpopulation group terms were implemented in GenStat 16 (VSN International, Hemel Hempstead, UK).

Marker data

Marker data and the analysis briefly described in this section are drawn from an earlier study (Bush and Thumma 2013). A panel of 75 single-nucleotide polymorphism (SNP) markers was used to genotype an average of 10 trees per family from the majority (119 of 137) of families in the breeding population. The data were used to estimate family-average relatedness ($\overline{2\theta}$) and the inbreeding coefficient ($\overline{f}$) for each family using the method of Wang (2007) which is well suited to data with significant inbreeding levels. Simulations based on biallelic allele frequency equivalent to that of the SNP panel were undertaken to estimate the likely bias of 2θ and f. For use in this study, the raw estimates were adjusted to account for the estimated biases. Both $2\hat{\theta}$ (Fig. 5.1) and $\hat{f}$ (see later) were found to be highly heterogeneous among families. Provenance-mean values of $\hat{f}$ and $2\hat{\theta}$ were applied to families for which marker-based estimates were not available.

Genetic analysis

Individual-tree linear mixed models of the general form

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Zs} + \mathbf{e} \qquad (5.1)$$

were fitted where **y** is the vector of individual (non-standardised) tree observations on each trait, **b** is a vector of fixed- effect estimates, **a** is a vector of random additive genetic effects, **s** is a vector of random within-site environmental effects including incomplete blocks and plots and **e** is a vector of random residual effects. **X** and **Z** are incidence matrices for fixed and random model terms. The additive, within-site environmental and error effects were assumed uncorrelated with zero means and normally distributed as $\mathbf{a} \sim N(0, \mathbf{A}\sigma_a^2)$, $\mathbf{s} \sim N(0, \mathbf{I}\sigma_s^2)$ and $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$ where **I** is an identity matrix and **A** is the additive relationship matrix. Matrix **A** comprises off-diagonal elements of 2θ, describing the relatedness between pairs of individuals that is equivalent to twice the probability that two gametes taken at random, one from each, carry alleles that are identical by descent (IBD) and diagonal elements, calculated as $1+f$.

We present model variants based on two different assumptions with respect to **A**. Firstly the 'default' assumption, that would be made in the absence of a known pedigree, is that individual trees within families are uniformly related to each other with $2\theta = ¼$ and are not inbred so that $f=0$ and that individuals from different families are unrelated with $2\theta = 0$. This is termed the half-sib (HS) model. Secondly, we assume a mixed mating (MM) model that incorporates the SNP marker estimates of relatedness and inbreeding. Individuals within families have heterogeneous values of 2θ and f taken as the family-average values $2\hat{\theta}$ and $\hat{f}$. Estimates of average relatedness between different families within subpopulations of $2\hat{\theta} > 0$ are also incorporated. Values of $2\hat{\theta}$ within and between families are shown graphically in Fig. 5.1, while values of $\hat{f}$ are shown on the x axis of Fig. 5.2. We have also made some assumptions to include in **A** the female parents of the progeny being tested. For some families, $2\hat{\theta} > 0.25$, but there is no evidence of inbreeding (i.e. $\hat{f} \approx 0$). We have assumed that this indicates a proportion of full sibs are present. For each of these families, we have included phantom male parents (Westell et al. 1988) in the pedigree and adjusted **A** to account for the contribution of both male and female parents. A benefit of including the parental model is that best linear unbiased predictor (BLUP) estimates for parents as well as progeny are available. Secondly, $\mathbf{A}^{-1}$, which must be formed to solve the mixed model equations, is sparser, which was an important factor for of achieving model convergence within a practical time-frame. The $\mathbf{A}^{-1}$ was initially not positive definite for some individual sites. The problem was investigated by iteratively removing subpopulations and families

until causative families (5 from KI and 1 from SFR with high $2\hat{\theta}$ and/or $\hat{f}$) were located. Constraining maximum values of $2\hat{\theta}$ and $\hat{f}$ to 1 and 0.5, respectively, was an effective remedy.

Inbreeding will not only affect relatedness, but may also lead to ID of fitness-related traits, an effect indicated for HT and DBH in the SFR subpopulation at the LIS site (Bush and Thumma 2013) and see Results section of this chapter 'Phenotypic means and evidence of inbreeding depression'. White and Hodge (1989) advocate inclusion of f as a fixed-effect covariate to account for this in the linear model. We have also incorporated this approach into the MM model for all traits.

Single site analyses

Referring to Eq. 5.1, for the single-site models, vector **b** is the vector of fixed effects (mean, complete-block replicates, provenance group, provenance nested within provenance group) and **X** is the incidence matrix relating the observations in **y** for each trait to the fixed effects in **b.** For the MM models, **b** also included a vector of family inbreeding coefficients ($\hat{f}$). This was modelled with a separate term for each of the three provenance groups, allowing exploration of differential responses to ID among subsections of the breeding population. Vector **a** contained random additive genetic effects for individual trees incorporating $\mathbf{A}^{-1}$ for the HS or MM pedigree assumptions as previously described. The random design factors (plots, incomplete blocks as trial rows and columns and 'standard' non-directional incomplete blocks) variously present at the sites were included in vector **s**.

Narrow-sense heritability (h^2) for each trait at each site was calculated as:

$$\hat{h}^2 = \frac{\hat{\sigma}_t^2}{\hat{\sigma}_t^2 + \hat{\sigma}_p^2 + \hat{\sigma}_e^2}, \qquad (5.2)$$

where σ_t^2 is the additive variance of individual trees, σ_p^2 is the plot variance and σ_e^2 is the residual variance. For comparative purposes $\hat{h}^2$ from the HS individual-tree models were also reduced by multiplying $\hat{\sigma}_t^2$ in the denominator and numerator of Eq. 5.2 by 1/1.6. This reduces the additive variance so that it is equivalent to the generally recommended coefficient of relationship (ρ) of 1/2.5 used to account for mixed mating in eucalypts (Eldridge et al. 1993) and that has been used previously to scale additive variance in *E. cladocalyx* family models (Bush et al. 2011b; Callister et al. 2008).

Bivariate genetic correlation estimates (r_A) between traits x and y were obtained from the estimated single-site additive covariance and variance components as:

$$r_A = \frac{\hat{\sigma}_{t_x t_y}}{\sqrt{(\hat{\sigma}^2_{t_x}\hat{\sigma}^2_{t_y})}} \qquad (5.3)$$

where $\sigma_{t_x t_y}$ is the additive genetic covariance component between traits, and $\sigma^2_{t_x}$ and $\sigma^2_{t_y}$ are the additive variance components for traits x and y respectively.

Breeding value rank correlations

Individual-tree breeding values from the HS and MM models were compared by product-moment and Spearman rank correlation to examine whether the different model assumptions caused significant re-ranking. The product-moment test reflects on differences in both magnitude and rank between pairs of BLUPs, both of which will affect predicted genetic gain, while the Spearman reflects only on rank changes that will affect selections.

Analyses across sites

Various approaches can be taken to assess performance of trials that test common genotypes across a number of sites (see for example a review by Smith et al. 2005). The traditional approach is to model main effects for environment (E) and genotype (G) and a multiplicative interaction effect (GxE) – commonly termed the G+GxE model. This approach has the advantage of relative simplicity, with only three terms to be estimated whatever the number of sites. Drawbacks include assumptions of homogeneous genotypic variance with no covariance among sites and also homogeneous error variance among sites. Advances in theory and computing capacity have resulted in the possibility of fitting more complex models that allow for heterogeneous error variance and more detailed assumptions about genotypic variance, for example individual site estimates of genotypic variance with either homogeneous or heterogeneous estimates of covariance among sites. We attempted to fit (a) the traditional G+GxE model; (b) the most general and complex unstructured (US) model, assuming heterogeneous error variances and heterogeneous genotypic variance and covariance; and (c), due to difficulty fitting the US, a reduced-rank approximation to US provided by Factor Analysis (FA). The models are described within the framework of Eq. 5.1 as follows:

Traditional G+GxE

The vector **b** contained sub-vector estimates for fixed effects of site, replicates within sites and subpopulation effects (provenances nested within provenance groups). For the MM models **b** also

included a sub-vector of f coefficients. The vector **s** contained sub-vectors for the random effects of incomplete blocks and plots. The vector **a** was expanded to include sub-vectors for the genotypes (individual trees-within-subpopulations) and the interaction between genotypes and sites. The inverse relationship matrix ($\mathbf{A}^{-1}$) included either the HS or MM pedigree variants as previously described. To further explore the cause of the relatively high overall heritability estimates obtained from the single-site models, separate model runs were made for each of the three provenance groups. To examine variable ID among subpopulations, MM model variants without the f covariate in **b** were also run.

Unstructured model

The US model provides the most general way of modelling genotypic variance. The fixed effects in vector **b** and random effects in vector **s** are the same as specified for the G+GxE model. Vector **a** was expanded to contain $s(s+1)/2$ sub-vectors comprising the additive variance parameters for s individual sites and a separately-estimated covariance for each pair of sites. Additionally, residual variance at each site was assumed to be heterogeneous and was modelled without covariance, expanding **e** to contain s sub-vectors. This gives a total of 66 additive genetic variance-covariance parameters plus 11 residual parameters for the DBH trait that was measured at 11 sites. Convergence was not achievable due to singularities once greater than eight sites were involved. Exploratory analyses involving various subsets of sites showed that no one particular site was responsible for this problem. Such problems with large, unstructured models involving more than just a few sites are common (Smith et al. 2005) as the variance/covariance matrix tends not to be positive-definite.

Factor analytic model

When applied to the genotype effects at each site, the FA (factor analytic) model postulates dependence on a set of k random hypothetical factors where $k < s$: this results in fewer parameter estimates and a more readily-convergent model than the US model (Smith et al. 2001). Factor analytic models provide a satisfactory approximation to the US alternative in terms of genotype selection (Kelly et al. 2007; Smith et al. 2001), and are more likely to converge when there are large numbers of sites and genotypes (Smith et al. 2005). The FA model also enables exploration of the nature of the GxE interaction. In this respect it is similar to the additive main effects multiplicative interactions (AMMI) approach (Gauch 2006), but has the significant advantage of being implemented in a mixed model environment allowing analysis of unbalanced data.

With reference to Eq. 5.1, vectors **b** and **s** contained fixed and random terms, respectively, as described in the previous models and the residual term **e** is heterogeneous as described for the US model. Briefly, (following Hardner et al. 2011), vector **a** comprises individual-tree additive effects ($\mathbf{a}_t$) that were modelled as a function of the effect of each site (site loadings) for k hypothetical

orthogonal factors and the genetic effect for each factor (genotype scores), plus a site-specific genotype variance in addition to that associated with the loadings:

$$\text{var}(\mathbf{a}_t) = (\Lambda \otimes \mathbf{I}_m)\varphi + \delta \qquad (5.4)$$

where Λ is a s x k matrix of individual-site (s) factor loadings (k), φ is a mk x 1 vector of genotype scores for each factor, and δ is a msx1 vector of deviations of the effect of the m^{th} genotype in the s^{th} site from that predicted by the factors. The genotype scores and the specific site deviations were assumed to be normally distributed with zero mean and variances $\mathbf{I}_{mt}$ and $\Psi \otimes \mathbf{I}_m$ respectively, where Ψ was an s x s diagonal matrix containing the specific variances for each site. The variance-covariance matrix for the individual tree effects at each site is then given by:

$$\text{var}(\mathbf{a}_t) = (\Lambda\Lambda' + \Psi) \otimes \mathbf{I}_m \qquad (5.5)$$

When there is more than one factor fitted (i.e. $k > 1$), constraints must be imposed on the elements of Λ in order to ensure identifiability. Following Smith et al. (2001) the leading *k-1* elements of each loading were constrained to zero for FAk >1 models. Given this constraint when k>1, the number of terms in the FA model is then given by $sk+s-k(k-1)/2$. This equates to 32 terms in **a** plus 11 in **e** for the DBH trait for the k=2 model. Log likelihood ratio tests (Wilks 1938) were performed to gauge the significance of differences in log likelihood between pairs of nested models.
The additive variance at each site i is then given by

$$\text{var}(\mathbf{a}_{t_i}) = \sum_{n=1}^{k}(\Lambda_1^2 + \Psi_1, \ldots, \Lambda_k^2 + \Psi_k) \qquad (5.6)$$

and the covariance between sites i and j as:

$$\text{covar}(\mathbf{a}_{t_{ij}}) = \sum_{n=1}^{k}(\Lambda_{1_i}\Lambda_{1_j}, \ldots, \Lambda_{k_i}\Lambda_{k_j}) \qquad (5.7)$$

Graphical representation of the site factor loadings as vector plots and genotype scores as scatter plots for the FA2 models was carried out following principal component rotation of the loading and score matrices, respectively, (Smith et al. 2001), implemented using the FACROTATE directive in Genstat 16 (VSN International, Hemel Hempstead, UK).

As a fixed effect for site was included in the mixed model, loadings were environmentally centred. In this case, the length of the trial loading vector is the estimated genetic standard deviation described by the first two factors. The cosine of the angle between vectors is an estimate of the genetic correlation due to the first two factors (Smith 1999). Genotype (family- and individual-tree) scores corresponding to each loading were also produced, though because of the large number of individual-tree genotypes, plotting the family scores was the most practically useful graphical representation.

Effectiveness of the FA model in terms of the percent genetic variance explained was calculated for the k^{th} factor (λ_k) as (Smith 1999):

$$\frac{\mathrm{tr}(\lambda_k \lambda_k')}{\mathrm{tr}(\Lambda\Lambda' + \Psi)} \times 100 \qquad (5.8)$$

All mixed models were solved using restricted maximum likelihood with AsReml version 3 (VSN International). This software implements testing of fixed-effect significance using Wald F statistics, while standard errors of functions of variance components are calculated using the Taylor series expansion method (Gilmour et al. 2009).

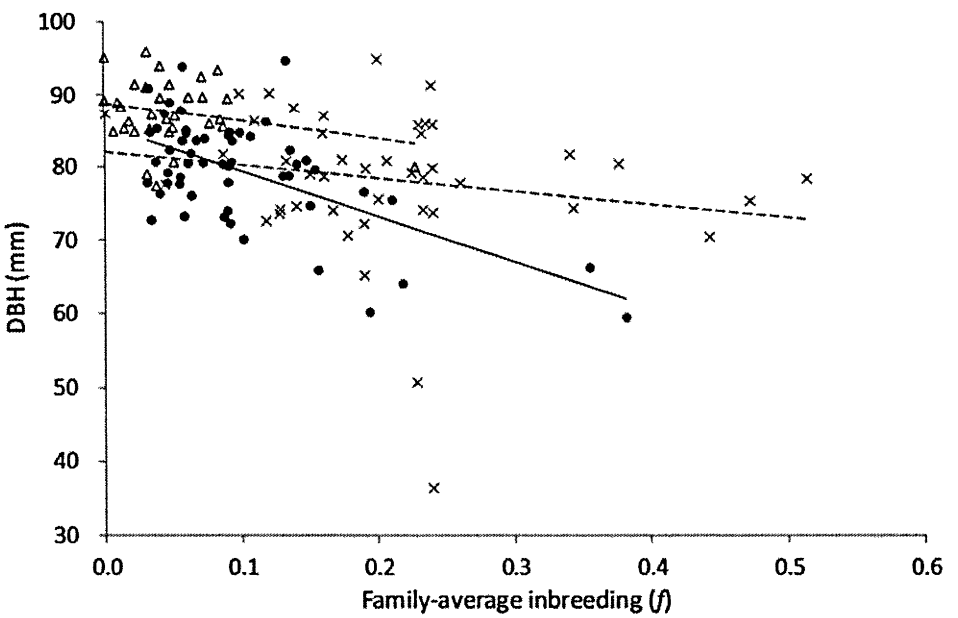


Fig 5.2 Relationship between family-average inbreeding ($\hat{f}$) and fitness trait, DBH (cf. Fig. 4.4). Phenotypic DBH data were converted to standard deviations from the mean of each site and then averaged over sites. The data here have been re-projected into units corresponding to the Lismore, Victoria site. Separate regression lines are fitted for the three groups of subpopulations: SFR (*circles*, significant at $p<0.001$, $R^2=0.36$), KI (*crosses*, not significant) and cultivated stands (*triangles*, not significant).

Results

Phenotypic means and evidence of inbreeding depression

Phenotypic means and coefficients of variation for each of the three traits across the 11 sites are given in Table 5.3. Growth was greatest at the three sites in Western Australia, KOJ, WEL and ESP2. It was lowest at WGW, New South Wales; the most easterly site with the highest pan evaporation. Linear regression on standardised across-sites data for the three provenance groups revealed that the inbreeding coefficient ($\hat{f}$) explained significant (p<0.001) phenotypic variance for the growth traits of KI families but not the other provenance groups or the AP trait (Table 5.4). For SFR families, a 0.1 increase in $\hat{f}$ led to reductions of 8% and 5% for DBH and HT, respectively (see Fig. 5.2 for DBH).

Table 5.3 Single-site estimates of heritability $\hat{h}^2$ (with approximate standard errors) from mixed mating (MM) and half-sib (HS) family assumptions, phenotypic mean and coefficient of variation (CV_p) for height (HT), diameter at breast height (DBH) and axis persistence (AP) traits

Site	MOR	KOJ	WEL	ESP1	ESP2	AVE	BTN	HAM	LIS	COR	WGW
DBH											
$\hat{h}^2$ MM	0.11 (0.07)	0.29 (0.10)	0.39 (0.12)	0.49 (0.13)	0.30 (0.11)	0.26 (0.19)	0.13 (0.04)	0.20 (0.05)	0.16 (0.04)	0.24 (0.07)	0.15 (0.04)
$\hat{h}^2$ HS	0.21 (0.11)	0.53 (0.17)	0.67 (0.19)	0.68 (0.17)	0.55 (0.17	0.68 (0.30)	0.22 (0.06)	0.42 (0.09)	0.41 (0.08)	0.46 (0.11	0.25 (0.07)
Site mean (cm)	5.9	12.7	12.1	6.3	14.1	9.0	8.3	8.8	8.2	6.9	4.8
%CVp	33	19	20	33	22	29	21	21	28	35	33
HT											
$\hat{h}^2$ MM	*	0.49 (0.11)	0.62 (0.15)	*	0.63 (0.17)	0.59 (0.22)	0.28 (0.06)	0.45 (0.09)	0.32 (0.07)	0.35 (0.08)	0.41 (0.07)
$\hat{h}^2$ HS	*	NC	NC	*	*NC*	*NC*	0.45 (0.12)	0.84 (0.14)	0.77 (0.12)	0.76 (0.15)	0.73 (0.12)
Site mean (m)	*	8.6	8.4	*	10.2	7.3	6.1	6.3	5.8	6.0	4.5
$\%CV_p$	*	15	14	*	17	22	16	14	17	22	20
AP											
$\hat{h}^2$ MM	*	0.14 (0.07)	0.18 (0.08)	0.00 (0.06)	0.12 (0.07)	0.09 (0.12)	0.25 (0.04)	0.14 (0.05)	0.27 (0.06)	0.05 (0.05)	0.10 (0.05)
$\hat{h}^2$ HS	*	0.31 (0.12)	0.35 (0.14)	0.00 (0.08)	0.30 (0.14)	0.15 (0.22)	0.50 (0.12)	0.24 (0.08)	0.45 (0.09)	0.17 (0.08)	0.26 (0.08)
Site mean (rating 1-6)	*	4.1	4.3	4.4	4.8	5.2	5.0	3.3	3.0	4.3	4.4
$\%CV_p$	*	48	49	26	33	27	29	44	38	29	27

NC not converged

Table 5.4 Linear regression statistics of growth and form traits on standardised across-sites data on $\hat{f}$

Trait	DBH		HT		AP	
Provenance group	Percent variance explained	Significance (p)	Percent variance explained	Significance (p)	Percent variance explained	Significance (p)
KI	2.5	0.157	3.7	0.114	4.1	0.099
SFR	36.0	<0.001	23.0	<.001	4.4	0.247
Cultivated stands	1.9	0.216	0.1	0.450	0.1	0.949

Comparison of HS and MM models

Single-site analyses with the HS models resulted in very high estimates of additive variance as a proportion of total variance at many of the sites (summarised in Table 5.3 and see also Appendix 3.2 for a full table of random-effect variance estimates and heritability estimates for each site). The HS models for the HT trait would not converge for AVE, ESP2, KOJ or WEL. Inspection of the convergence sequence prior to model abortion indicated that that this was due to absence of fitted residual variation resulting in singular mixed-model equations, with a large proportion of this partitioned to the individual-tree (additive) component at the final iteration. Further diagnostic analysis using a simpler but equivalent family model (i.e. fitting Equation 3 to estimate σ^2_f rather than σ^2_t) revealed that assuming half-sib families with $\rho = 1/4$ results in $4\hat{\sigma}_f^2 > \sigma_p^2$ (total phenotypic variance), which explains why these individual tree models would not converge under the HS assumption. Where they did converge, individual-tree HS models indicated considerable additive variation for each of the traits resulting in narrow-sense heritability estimates that were moderate to very high (0.21-0.84) for HT and DBH and low to moderate (0-0.50) for AP.

Models including the MM assumptions all converged with reductions of $\hat{\sigma}_t^2$ relative to the HS version for all trait and site combinations except AP at ESP1 which had $\hat{h}^2 \approx 0$ under both assumptions. Single-site MM estimates averaged 51% of HS, though some growth estimates were still in the moderate-high range, especially for HT which ranged from 0.28-0.63. Precision of additive variance component estimation was high at the 'major' trial sites (BTN, COR, HAM, LIS, MOR, WGW); typically between 5 and 40% of the component estimate. Approximate standard errors were moderately higher at the 'minor' sites. MM model variants generally had slightly higher standard errors as a proportion of the $\hat{h}^2$ parameter estimate. In most cases design effects were significant with replicate F tests having $p<0.05$ and plots, rows and columns having variance components exceeding their approximate standard errors.

Inclusion of *f* as a fixed-effect covariate in the MM models explained significant variance at all sites and across sites ($p<0.001$). Fitting *f* separately for each of the three provenance groups showed that it was significant across sites for all three provenance groups for DBH ($p<0.001$), for the cultivated and SFR groups for HT ($p<0.001$), and for the KI and cultivated groups for AP ($p<0.006$) (full results at each site and across sites are presented in Appendix 3.2).

Provenance variation

Provenance nested within provenance group was modelled as a fixed effect for each trait. The provenance groups correspond to the KI and SFR wild regions of provenance and the cultivated stand group. Overall, there was a large amount of provenance-level variation for all three traits, with consistent patterns across sites (see Appendix 3.2 for a full table of predicted means and F-test results for all traits). Predicted means were very similar between the HS and MM models for those site-trait combinations where both converged. Differences across sites between MM models were significant ($p<0.04$) at both the provenance-group and provenance levels for HT and DBH and at the provenance-group level only for AP ($p<0.001$).

Mean DBH growth was generally similar for the KI and SFR groups and highest in the cultivated provenance group at individual sites, a result reflected by means across sites (MM FA2 model) of 86, 88 and 92 mm, respectively, and a standard error of difference (SED) of 1.5 mm. Generally, the cultivated and KI provenance groups were significantly taller than the SFR group. Estimates across sites (MM, FA2 model) were 671, 714 and 730 mm (SED 11 mm) for the SFR, cultivated and KI groups, respectively. There were significant HT differences among and within provenance groups ($p<0.036$), for example within the KI group the American River and Flinders Chase provenances had estimates of 716 and 774 mm, respectively, (SED 21 mm).

KI provenances had consistently poorer AP than the SFR and cultivated provenances at all sites. Axis persistence rating estimates across sites (MM, FA2 model) were 3.7, 4.5 and 4.5 (SED 0.1) for the KI, SFR and cultivated and groups, respectively. There were significant differences among provenance groups at all sites and across sites, but no significant differences among provenances within groups at any site arising from either the HS or MM model variants.

Trait-trait correlations (r_A)

There were generally strong correlation estimates between HT and DBH at each site (r_A between 0.71 and 1.00) from MM models (Table 5.5). The correlations between AP and the two growth traits ranged from around zero to moderately positive. The two sites where DBH-AP and HT-AP correlations were of practical significance were Corowa and Lismore, as at the other sites either r_A was close to zero and/or the precision of estimation was low. The implication, taken together with the high $\hat{h}^2$ and coefficient of phenotypic variance (CVp) (Table 5.3) is that selection of more vigorous trees at these sites would also result in improvement of axis persistence. The r_A estimates from the HS models were on average only 0.01 units higher for DBH-HT, 0.06 for DBH-AP and 0.07 for HT-AP with approximate standard errors averaging +/- 0.01 those of the MM model for all traits.

Table 5.5 Estimates of trait-trait genetic correlation (r_A) for single sites from the MM model

Site	DBH-HT	s.e.	DBH-AP	s.e.	HT-AP	s.e.
Avenue	1.00	0.02	-0.12	0.45	-0.02	0.37
Bordertown	0.80	0.08	0.29	0.18	0.32	0.15
Corowa	0.94	0.03	0.66	0.26	0.42	0.24
Esperance 2	0.95	0.05	0.22	0.31	0.27	0.27
Hamilton	0.87	0.05	0.36	0.19	0.43	0.16
Kojonup	0.71	0.11	0.33	0.26	-0.14	0.23
Lismore	0.87	0.04	0.60	0.12	0.71	0.09
Wellstead	0.83	0.08	0.02	0.29	0.04	0.24
Wagga	0.92	0.04	-0.01	0.25	0.35	0.18

Genotype-by-environment interaction

Analyses using the traditional G+GxE model with homogeneous additive variance revealed moderate genotype-by-environment interactions with interaction variance components of just over 50% (HT and DBH) and 100% (AP) of the additive estimates arising from the HS models. The MM models resulted in re-partitioning of variance giving rise to an increased proportion of interaction variance for all three traits, and increased residual variance (Table 5.6). The partitioning of the genotype-by-environment interaction component in the across-sites analyses resulted in narrow-sense heritability estimates that were substantially lower than those indicated by the single site analyses, for example the heritability averaged over the 11 single-site MM DBH estimates is 0.25 while the estimate across sites is 0.14. As expected from the single-site estimates, the across-sites estimates from the MM models were also markedly lower than those from the HS models. Notwithstanding the reductions to $\hat{h}^2$ arising from the application of the MM assumptions across sites, the estimate for HT remained moderate at 0.32 (0.05).

Heritability estimates for each of the three provenance groups, including MM variants both with and without the $\hat{f}$ covariate, are given in Table 5.6. Analyses of the separate provenance groups across sites show that the KI group produced the highest $\hat{h}^2$ for each trait under both HS and MM assumptions. The HS model run on the KI dataset did not converge for HT, for the reason noted in some of the single-site HT analyses. Application of the MM assumption (with f) to the KI data gave a moderately-high HT estimate of 0.53. Inclusion of the $\hat{f}$ covariate in the MM models had the greatest effect on growth traits for the SFR provenance group, reducing $\hat{h}^2$ for HT from 0.18 to 0.10 and for DBH from 0.27 to 0.20. The covariate had lesser effects on the KI provenance group, for which f was generally elevated but ID overall was apparently minimal, and almost no effect on the cultivated group, for which f was low and ID negligible.

Table 5.6 HS and MM variance component estimates for genotype ($\hat{\sigma}_t^2$), site-by-genotype interaction $\hat{\sigma}_{t.s}^2$) and residual ($\hat{\sigma}_e^2$) terms and narrow-sense heritability ($\hat{h}^2$) estimates with standard errors from across-sites analyses based on the traditional G+GxE model. The ratio of genotype-by-environment to genotype variance is also given. Separate results are presented for the whole population (ALL) and the three provenance groups.

Trait	$\hat{\sigma}_t^2$	$\hat{\sigma}_{t.s}^2$	$\hat{\sigma}_e^2$	$\frac{\hat{\sigma}_{t.s}^2}{\hat{\sigma}_t^2}$	$\hat{h}^2$ ALL	s.e.	SFR	s.e.	KI	s.e.	Cult.	s.e.
HS												
DBH	121	68	208	0.56	0.36	0.05	0.35	0.08	0.53	0.13	0.15	0.05
HT	5728	3015	236	0.53	0.82	0.11	0.52	0.11	*NC*	-	0.45	0.13
AP	0.30	0.32	1.04	1.07	0.21	0.04	0.10	0.04	0.30	0.10	0.20	0.07
MM (no *f* covariate)												
DBH	62	42	277	0.68	0.18	0.03	0.18	0.04	0.21	0.05	0.12	0.04
HT	2651	1706	3760	0.64	0.36	0.05	0.27	0.06	0.58	0.12	0.32	0.09
AP	0.15	0.18	1.28	1.20	0.10	0.02	0.06	0.02	0.10	0.04	0.17	0.06
MM (with *f* covariate)												
DBH	47	35	282	0.73	0.14	0.02	0.10	0.03	0.18	0.05	0.11	0.04
HT	2352	1705	3968	0.72	0.32	0.05	0.20	0.05	0.53	0.11	0.32	0.09
AP	0.14	0.18	1.29	1.31	0.09	0.02	0.06	0.02	0.07	0.03	0.17	0.06

NC not converged

The unstructured (US) models would not converge with more than 8 sites, whatever the combination of sites. The factor analytic approach was therefore taken. FA models with up to k= 3 factors were fitted. Log likelihood ratio test results are given in Table 5.7: in all cases the FA2 models explained significantly more than the FA1 while the FA3 models did not provide a significant improvement. Models with $k \geq 4$ factors would not converge.

Table 5.7 Log likelihoods from G+GxE and factor analytic models for half-sib (HS) and mixed-mating (MM) assumptions. The number of terms listed refers to the random genetic effects with s sites and k factors. Asterisks denote models with significantly higher likelihood (likelihood ratio test, p<0.05) than the model listed immediately below. Log Likelihoods are not comparable between MM and HS models.

		DBH			HT			AP	
Model	Terms	MM	HS	Terms	MM	HS	Terms	MM	HS
US	$s(s+1)/2$	*NC*	*NC*	45	*NC*	*NC*	55	*NC*	*NC*
FA3	$sk+s-k(k-1)/2$	41	-46181	-46208	33	-57351	-57379	37	-8899	-8906
FA2		32	-46188*	-46213*	26	-57353*	-57383*	29	-8904*	-8911*
FA1		22	-46198*	-46223*	18	-57363*	-57394*	20	-8913*	-8922*
G+GxE	2	2	-46266	-46278	2	-57746	-58981	2	-8951	-8961

NC not converged

Type-B correlation matrices and variance component estimates for the 11 sites are given in Appendix 3.4. As foreshadowed by the single-site analyses, the HS models gave higher estimates of additive variance and lower residual variance than did the MM models: though the models converged, the HS residual variance was estimated to be approximately zero for DBH at the AVE site and for HT at the KOJ, WEL, ESP2 and AVE, all 'minor' sites and the four most westerly measured for this trait. Type-B correlations from the MM models ranged from 0.06 to 0.95 for DBH, 0.03 to 0.93 for HT and -0.42 to 0.99 for AP. Figure 5.3 illustrates these correlations for DBH and AP with vector plots. These are provided for all models in Appendix 3.3. There is no clearly interpretable pattern of regional clustering based on climate or soils, though the ESP1 and AVE sites cluster separately for all traits. A negative genetic correlation for AP between BTN and ESP1 and BTN and AVE was found, which contrasts with the otherwise positive correlation estimates. Both of these estimates have high standard errors and the BTN and AVE sites were selectively thinned prior to measurement. Three major eastern sites, HAM, LIS and COR have generally high correlations for all three traits. Though genotypes were not explicitly fitted in the selected FA models for each trait, the first loading effectively corresponds to an average environment in terms of genotype performance, and the genotype scores are then a weighted average of genotype effects across environments (Smith 1999). The largest weights went to those sites that have a large covariance with the average site (i.e. large loadings associated with the first factor). The scores are quite closely correlated to the genotype main effect in the G+GxE model (R^2 between first factor scores and the G+GxE BLUPs was 0.87 for DBH, 0.79 for HT and 0.88 for AP).

Figure 5.4 shows the FA1 HT scores for each family with estimates of f for some more outlying values. It is evident that some of the best-performing families (in terms of DBH growth in the average environment) are highly inbred, particularly those of the KI provenance group. The scores on the second loading give an indication of the stability of the genotypes.

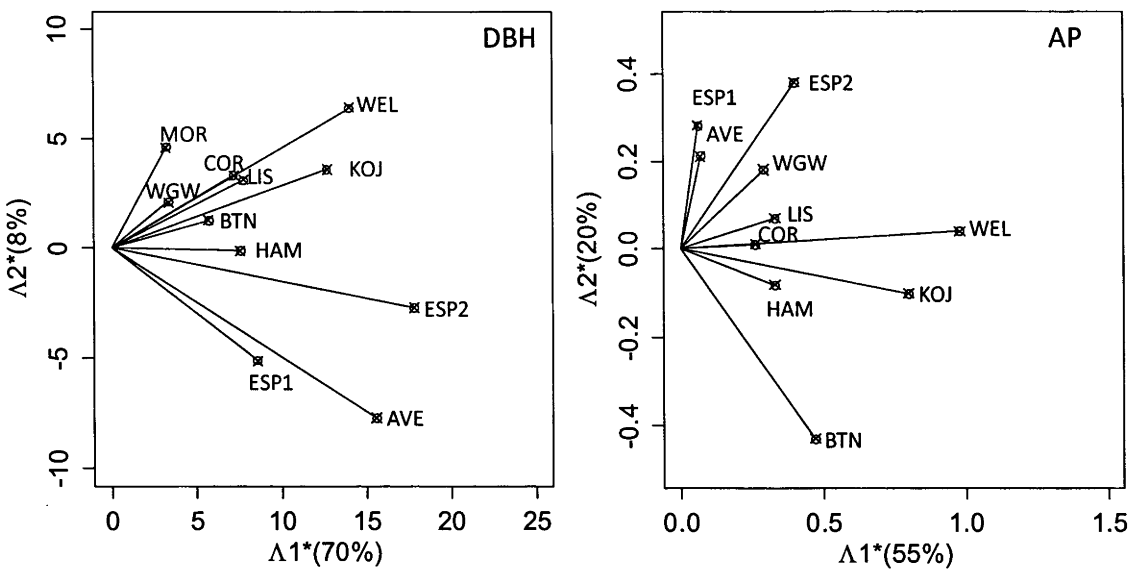


Fig. 5.3 Vector plot showing principal component-transformed loadings ($\Lambda 1^*$ and $\Lambda 2^*$) from the FA2, mixed mating model for DBH and AP. The axes are labelled with the percentage of explained variance. The cosine of the angle between each pair of vectors is indicative of the type-B correlation between the sites. In this case, the length of the trial loading vector is the genetic standard deviation described by the first two factors.

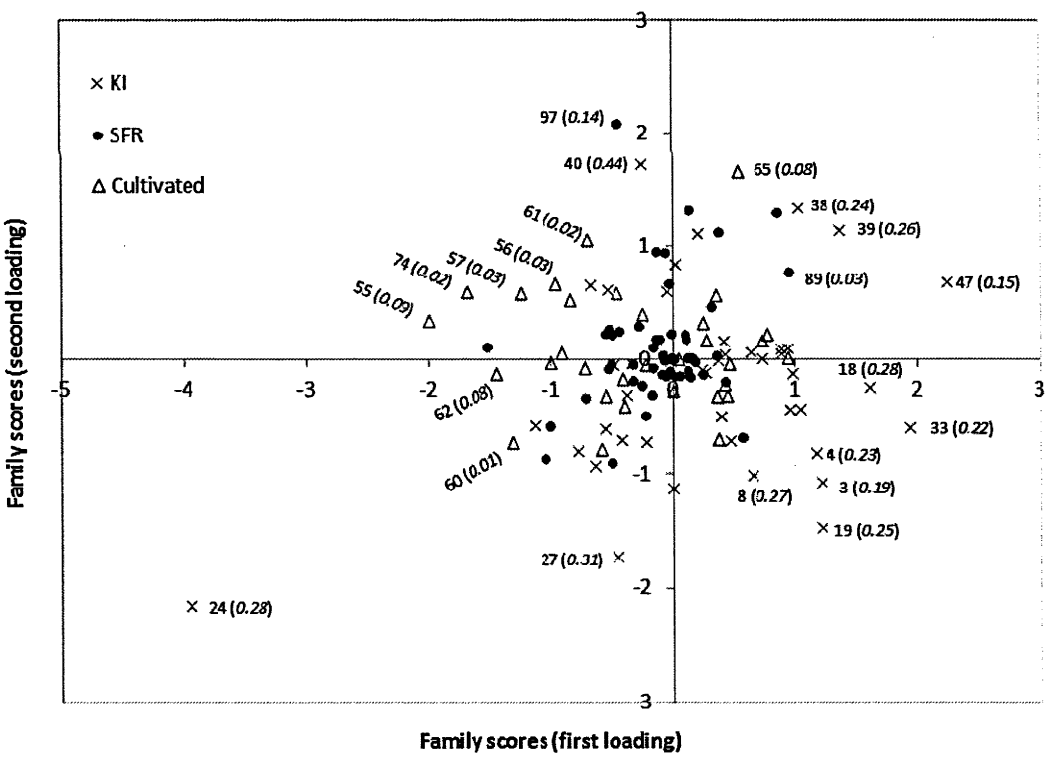


Fig. 5.4 Plot of family scores for the first and second loading of the FA2 MM model for HT. The first axis corresponds to the performance of the families in an average environment, with those to the right being the better overall performers. The second loading scores give an indication of the reactivity of the genotypes, with those towards the top of the graph being more stable. Some families are labelled with their family number followed by their estimated *f*.

Breeding value correlations

Product-moment and Spearman ranks were used to compare the correlations between individual-tree BLUP breeding values resulting from the HS and MM models across sites. Considering the overall individual-tree means (predicted at the mean value for each of the first and second loadings), the product-moment correlations between HS and MM models were 0.91, 0.80 and 0.97 for DBH, HT and AP, respectively. Spearman correlations were within 2% of these values. However, selections and culls are usually made among the highest and lowest breeding values, where re-ranking between the models was far greater. The mean BLUP height of the tallest 200 trees selected using the MM model is 669 cm, while selecting the tallest 200 from the HS model would result in an average of 657 cm. The overall mean BLUP HT was 596 cm, so a loss of 12 cm represents a 16% reduction. Losses of 4% in AP and 9.5% in DBH result from similar comparisons between selection using the HS and MM models. All preceding comparisons were made using units of the MM BLUP scale which is relatively more 'shrunken' or regressed towards the mean than that of the HS.

Comparison with family model assuming a coefficient of relationship

The heritability estimates from the HS models were deflated by a factor of 1/1.6, which approximately corresponds to the result that would arise from a family model where heritability is calculated using a coefficient of relationship of $\rho = 1/2.5$, a value used in other *E. cladocalyx* studies. For the individual-site models, this resulted in heritability estimates that were between -0.06 and +0.05 of the estimates given by the MM assumption, with average deviations of only -0.01 for HT and AP and -0.03 for HT (though HS models for HT converged for only five sites). For the estimates across sites (Table 5.6), the adjustments result in $\hat{h}^2$ of 0.23, 0.51 and 0.13 for DBH, HT and AP, respectively, compared with the MM estimates of 0.18, 0.36 and 0.10.

Discussion

Analysis of the effects of inbreeding on growth traits showed that the SFR subpopulation of *E. cladocalyx* is affected by ID while the KI subpopulation does not appear to be. This corroborated an earlier finding (Bush and Thumma 2013) from the LIS site. A shortcoming of this analysis, that applies relatively few markers, is the expected low precision of estimation of f, especially as the effect of heterogeneous, within-family inbreeding is not accounted for. A simulation study by Bush and Thumma (2013) and comparison of a smaller subset with the full, relatively-dense marker panel (5844 SNPs) used by Doerksen et al. (2014) both conclude that a large number of markers are required to estimate f with precision. A possible consequence of this may be downward bias of the regression coefficients, in which case the tests for their significance may be conservative. Notwithstanding this possible limitation, the response of the KI subpopulation presents an unusual and perhaps extreme

example among tree species, as most do not have the capacity to carry high levels of inbreeding to relatively mature life stages (i.e. to reproductive maturity). Eucalypts are thought to carry a high genetic load that results in severe ID and mortality as demonstrated in many of the commercially important species (Costa e Silva et al. 2010b; Griffin and Cotterill 1988; Hardner and Tibbits 1998). Bush and Thumma (2013) hypothesise that a bottleneck event involving the KI subpopulation may have led to purging of deleterious recessive alleles. Nevertheless, other eucalypt species are likely to possess heterogeneously-related individuals within families caused by inbreeding and/or full-sib relationships, especially in the first generation. Heterogeneous outcrossing or relatedness among young seedling families from tree seed orchards has frequently been observed (Butcher et al. 2004; Lai et al. 2010; Moran et al. 1989; Surles et al. 1990), and in a small number of cases where it has been studied, this has also extended to the parents (Bush et al. 2011a; Doerksen et al. 2014). Wild families infused into later generations of breeding populations may also exhibit heterogeneous levels of relatedness and ID. The approach taken here of using marker-based, family-average estimates of relatedness and inbreeding and applying them to the entire breeding population across a number of sites provides more realistic estimates of relatedness for a modest genotyping cost.

Results from individual-tree models incorporating a marker-based pedigree contrast strongly with those of the classical model that assumes homogeneous half-sib within-family relatedness and no inbreeding. The close within-family relatedness and inbreeding depression in some subpopulations, particularly wild families from Kangaroo Island, results in non-convergence of the individual-tree mixed models for some trait and site combinations if half-sib relationships are assumed. This is because the estimate of additive variance exceeds the total phenotypic variance, due to variable ID, incorrect scaling of relationships that are closer than HS, and probably, dominance variance. Even if convergence was always achievable, perhaps by assuming full-sib rather than half-sib families, the assumption of homogeneous relatedness and no inbreeding is clearly unsuitable as gauged by the marker-based estimates for this breeding population. In fact, the HS assumption applies quite well to the cultivated stand provenance group, while members of some of the wild families are more closely related than are full sibs and are inbred. This leads to re-ranking of breeding values relative to MM assumptions, especially of the highest and lowest. An alternative approach would have been to run family models to estimate σ^2_f and then apply an appropriate coefficient of relationship in calculating heritability. If $\rho = 1/2.5$ was selected, this would be approximately equivalent to down-scaling to 63% of the HS estimate. For across-sites HT this would give $\hat{h}^2 = 0.52$ which is still substantially greater than the MM estimate of $\hat{h}^2 = 0.32$. Nor would this approach appropriately re-rank breeding values, resulting in loss of genetic gain from making sub-optimal selections or culling.

The study has confirmed considerable genetic variation in growth and form traits at both the subpopulation and family levels. This finding is in general accord with other studies of *E. cladocalyx* which have identified considerable variation in all traits studied at the provenance level including flowering- (Mora et al. 2009), wood-property (Bush et al. 2011b) and growth traits (Callister et al. 2008; Mora et al. 2009). The subpopulation differences indicated by the present study were partitioned mainly to the provenance-group level rather than the provenance level, which accords with finding of Bush and Thumma (2013) that the species' considerable F_{ST} (fixation index, a measure of population differentiation) was partitioned 13.50% among the three provenance groups and 5.34% among provenances within the groups. Provenance ranks were generally quite stable across sites. Axis Persistence was found to be significantly superior in the SFR and cultivated stand provenance groups relative to the KI group, whereas growth (HT and DBH) was in most cases better at age 5 years in the KI and cultivated stands. A likely explanation for the good all-round performance of the cultivated stand material is that it came from phenotypically-selected candidate plus trees that Bush and Thumma (2013) showed to be derived from the SFR. It is also significantly more outcrossed than both the KI and SFR population groups.

The differential response to inbreeding of the SFR and KI subpopulations presents a challenge for breeding-population management. Evidence for ID among families in the SFR group suggests that the standard approach of inbreeding avoidance is likely to be advantageous for that material. However, the superior performance of some of the highly inbred families from KI suggests that this might be managed differently. The breeding biology of the KI material should also be more closely examined – it is possible that KI material may be strongly preferentially selfing or have other unusual reproductive processes that may lead to inbreeding, such as asynchronous or very sparse flowering (see review in Potts and Wiltshire 1997), rendering it non-amenable to management in a multi-provenance seed-orchard design.

The traditional G+GxE analysis revealed that GxE is significant for each trait, with interaction components exceeding 50% of the additive component. Heritability estimates were therefore lower from the across-sites analyses. Matheson and Raymond (1986) point out that while GxE interactions are frequently shown to be statistically significant, it is often only a few sites and/or genotypes that are responsible for much of the interaction. Moreover, the practical significance of the result is limited if specific causative links that allow the manager to predict the performance of particular genotypes at specific sites are absent. In this study relatively high reactivity of genotypes for specific traits on particular sites is evident. For example, much of the interaction for AP is caused by genotype response at Bordertown (see Fig. 5.3) and estimated genetic correlations between BTN and some other sites are negative. Considering some subsets of the data (as Callister et al. 2008 did for ESP2, WEL and KOJ) GxE is of negligible practical consequence. Referring to the climatic and basic soil

summaries given in Table 5.2, it is difficult to draw any predictive conclusion about what has caused the GxE interaction and resulting low Type-B correlations among certain pairs of sites. It is apparent from Fig. 5.3 that some sites appear to have clustered together in terms of performance of different traits (e.g. for AP and DBH, the ESP1-AVE cluster, which have similar soils but are geographically distant) and the general tendency for the eastern trials (LIS, COR, WGW) to cluster with high type-B correlations. Many more sites would probably be required to resolve the relationship between genotype response and specific environmental factors.

The analysis with MM assumptions is likely to be more realistic for this breeding population with highly heterogeneous relatedness and ID. The assumption has resulted in considerable reductions in estimates of population-wide additive variance and heritability. Separate analyses of the two wild and cultivated subpopulation groups have demonstrated that it is with the more-inbred, wild material that inflated additive variance estimates arise from the HS assumption. However, even after adjusting the additive relationship matrix using the marker-estimated values, the additive variance, particularly of HT, remained moderately high for the KI provenance region. While fitting f as a covariate accounted for significant growth-trait variance attributable to variable ID among SFR families, the effect on KI families, for which ID was non-significant, was weaker. It would seem unlikely, from examination of a wide range of growth trait heritabilities across a broad range of tree species (Cornelius 1994), that the resulting heritability ($\hat{h}^2 = 0.53$ for HT) is close to correct. A likely explanation for this is that non-additive dominance variance (σ_D^2), which accumulates under inbreeding, is likely confounded with the additive variance. Simulation studies (Borralho 1994) show that σ_D^2 and ID in combination with heterogeneous σ_A^2 can cause either positive or negative bias of $\hat{h}^2$. The apparent magnitude of the upwards bias of $\hat{h}^2$ in the KI subpopulation is quite large for HT, and if caused by heterogeneous dominance variance, is likely to have implications for accurately making selections in the first-generation breeding population. For breeding-population management, it may be necessary to formally estimate the amount of non-additive variation by undertaking controlled pollinations. This may also be warranted to better understand the breeding biology of the species, which from estimates of relatedness indicating more than one generation of close inbreeding, appears to be highly amenable to self-pollination and/or near-relative inbreeding. A controlled-pollination study on three *E. cladocalyx* trees found that they ranged from highly self-compatible to self-incompatible (Ellis and Sedgley 1992).

A further question arising from our findings is the likely bias of existing heritability estimates for this species. This study has included the data of Callister et al. (2008), and provides a direct explanation for the overall high heritability estimates reported in this study, despite *post hoc* application of a

coefficient of relationship of 1/2.5 to their estimates of σ^2_f. Bush et al. (2011b) also applied $\rho = 1/2.5$ in their study of wood properties which included only wild families from the KI, SFR and Eyre Peninsula regions that are also likely to be inbred (McDonald et al. 2003). The present study has indicated that $\rho = 1/2.5$ would be required as a minimum to correct for relatedness, though this neither addresses the problem of ID, that is present in the SFR provenance group, nor additional non-additive variance that may also be present. Other studies of *E. cladocalyx* (Mora et al. 2009; Vargas-Reeve et al. 2013), using a population with many families common to those included here, (Australian Tree Seed Centre unpublished records), appear to have made heritability estimates of growth, form and flowering traits on the assumption of HS families. In this case it is likely that the heritability estimates are upwardly biased.

Concluding remarks

The use of molecular markers to model more realistically the pedigree in this first-generation breeding population of *E. cladocalyx,* a eucalypt species with highly heterogeneous relatedness among and within families resulting from a mixed-mating system, yields genetic parameter estimates that contrast strongly with those generated under the assumption that trees within families are half sibs. The reduction in bias of $\hat{h}^2$ resulting from incorporating assumptions of closer relatedness and variable ID due to MM was expected, and was similar to or greater than the reductions that would be made using a coefficient of relationship of 1/2.5 applied to σ^2_f estimates from family models. However the marker-based information showed that the breeding population was highly structured in terms of relatedness and ID, and this had two implications: (i) individual-tree models would not converge for some trait and site combinations, as $\hat{\sigma}^2_t > \sigma^2_p$, due to closer than half-sib relatedness and ID in the wild subpopulations and (ii) heterogeneous relatedness resulted in re-ranking of breeding value estimates. Despite the reduction in $\hat{h}^2$ made using the marker-based pedigree and inbreeding estimates across sites, estimates remained moderate in some cases, probably indicating unquantified dominance variance. More experimentation is required to quantify the non-additive variance, which given the heterogeneous levels of inbreeding, is also likely to be heterogeneous among families and subpopulations. The apparent propensity for inbreeding to accumulate without ID in some sections of this breeding population also raises further questions about the species' breeding biology and how to best manage the breeding population. The use of markers to better model the pedigree and account for ID has proven to be very informative for this breeding population. Our strategy of using a modest number of markers to give information on the entire breeding population represents a practical compromise between genotyping effort and information that can be used to significantly refine quantitative parameter estimation.

Acknowledgements

The set of field trials reported here were established by members of the Australian Low Rainfall Tree improvement Group. We would like to thank the teams of people who established and managed the trials led by Trevor Butcher and Andrew Callister in WA, Tim Jackson and Des Stackpole in Victoria, Hans Porada in NSW and Mick Underdown and Bob Boardman in South Australia. We would also like to thank Chris Harwood, who was instrumental in establishing the breeding program for *E. cladocalyx* and for helpful comments on a draft of the manuscript.

Chapter 6: Discussion and conclusion

The four preceding chapters have explored important aspects of the population and quantitative genetics of *E. cladocalyx* that will influence and impact on genetic improvement of the species. This chapter discusses the implications of the four studies, taken together, for the species' domestication and breeding. Firstly, in Section 6.1, the drivers for domestication of the species are re-examined, providing background that leads to discussion (Section 6.2) of how the results from the wood properties study of Chapter 2 might be practically applied.

Section 6.3 then turns to the findings of the Callister et al. (2008) study on GxE across three of the ALRTIG sites in WA. Their paper, which reported unusually high heritability for growth traits, had strong implications for this study, suggesting that the traditional assumptions underpinning the quantitative analysis of this species may need to be revisited. As a probable cause of the inflated heritability estimates was thought likely to be heterogeneous relatedness and inbreeding, a two-stage approach was taken to investigate the impact of modifying the traditional first generation half-sib pedigree assumption with marker-based information.

The first stage, or "pilot study" (Chapter 3), used a modest panel of isozyme markers to estimate the relatedness of a small number of families and examine whether the approach could appropriately be scaled up to apply to the ALRTIG breeding population. The findings of this study are discussed in section 6.4. The second stage involved acquisition of SNP marker data from the majority of the families, as reported in Chapter 4. The applications of these data are further discussed in Section 6.5. The findings of Chapter 5 are discussed in two sections. In Section 6.6, results of models based on the traditional assumption that open-pollinated families are half-sibs are compared with models integrating SNP marker data, reflective of mixed mating. Section 6.7 then examines the implications of the GxE study reported in Chapter 5.

Section 6.8 examines possible causes of apparent differences between *E. cladocalyx* and other eucalypts in terms of the relatively high degree of inbreeding identified throughout the study. Section 6.9 draws together the findings of Chapters 2-5, presenting overall implications for the domestication and breeding of *E. cladocalyx*. Section 6.10 offers concluding comments from the study as a whole.

6.1 Drivers for the domestication process

The Australian Low Rainfall Tree Improvement Group (ALRTIG) first generation breeding strategy for *E. cladocalyx* (Harwood and Bulman 2001; Harwood et al. 2005) outlines the breeding objective, which is production of small sawlogs. The strategy employs cyclical recurrent selection for general

combining ability focussed on traits including survival, growth, stem and branch form and apparent resistance to pests and diseases. It is noted that properties related to wood strength are not likely to cause problems as, unlike softwoods such as *P. radiata*, the mature wood is typically very dense and therefore likely to have more than adequate strength. Harwood et al. (2005) and Harwood et al. (2007) also explain that the underlying reason for the development of the species is not only the need for timber, which in fact is only in modest demand. An important driver is improvement of the commercial attractiveness of wide-scale establishment of the species in southern Australia's sheep-wheat belt, a region where past clearing for agriculture has resulted in a variety of environmental problems such as dryland salinity and other forms of soil degradation. To improve short-term financial returns, and thereby encourage planting, development of uses for thinnings would therefore be beneficial. As *E. cladocalyx* is among the most naturally-durable of eucalypts, poles and posts that would not require chemical treatment were identified as a potential product that could fill this need (Bush and McCarthy 2008).

6.2 Wood properties

Chapter two presents a study examining the wood properties relating to natural durability of *E. cladocalyx*. Given that the problems of high wood density and long distances from the target planting zone to the processing centres would make the species unsuitable for pulpwood, the most likely use for thinnings from sawlog plantations would be either fuelwood or posts and poles. Past use of the species for naturally-durable products on farms prompted an investigation into the genetic variation of durability-related traits as well as density of young-aged (pole-sized) logs. The importance of naturally-durable wood products is likely to increase, especially in food and beverage sectors where organic goods that have been produced with minimal exposure to toxic chemicals are in increasing demand (Bush 2011). Either dedicated plantations or thinnings from sawlogs might be used for poles and posts. This end-use and the associated traits were not considered in the original breeding strategy, and Chapter 2 provides insight into whether or not the set of originally-proposed traits and breeding objective could be expanded to encompass natural durability.

As outlined in Chapter 1, making selections for genetic improvement of a particular end-product based on only a subset of important traits can lead to problems: low strength *P. radiata* wood that has resulted from early selection based on growth and form traits, but not wood properties, is a salient example in Australasia. This, together with financial pressure to shorten rotation times, has resulted in a greater proportion of low-density, low-strength, juvenile corewood being produced. It should of course be recognised that the assessment of wood traits was not as readily addressable by early tree breeders, as, firstly, carrying out wood property sampling was not as easy as it is today and, secondly,

the financial circumstances leading to reduced rotation times were not easily foreseeable. With lessons learned from this precedent and the possibilities afforded by current wood property measurement techniques in mind, it is worth re-examining the strategy for *E. cladocalyx*.

The results presented in Chapter 2 reveal that there is a large amount of genetic variation in the durability-related wood property traits of *E. cladocalyx*. While the study showed that there was significant variation in density at both the region of provenance and family levels; it did confirm the assumption made in the original breeding strategy that the wood has a generally high density, an overall average of 660 kgm^{-3} and $CV_P = 6\%$ for 8-year old trees cored at breast height was estimated. The trial assessed was planted earlier than the ALRTIG trial series discussed in later chapters and it also contained families from the Eyre Peninsula region of provenance (ROP). This ROP was not included in the ALRTIG breeding population because its growth and form in provenance trials in South Australia were clearly inferior to the other two ROP (Bulman et al. 2000). Significant differences among the three wild ROP were indicated for all traits (Table 2.2), including DBH which was significantly lower for the Eyre Peninsula material and slightly but not significantly higher for Kangaroo Island (KI) than for the South Flinders Ranges (SFR). An important result was the overall superiority of decay resistance of the SFR. If natural wood durability was considered to be a desirable trait in addition to growth and form, but durability-trait screening was deemed to be too expensive, focussing selection on other traits within this ROP would be a simple but effective strategy. Low-moderate narrow-sense heritability for growth traits and extractives content and moderate-high heritability of basic density and decay resistance to two of the three fungi were indicated. Selection at both the provenance and family levels should result in significant genetic gain.

As explained in Chapter 1, Type-A genetic correlations among growth and wood property traits are an important factor in early domestication, as selection for growth traits in isolation of wood properties has been known to lead to degradation of wood properties in some pine species. Though this study only examined a relatively small number of families, it indicated that density of the heartwood of young-aged trees did not appear to be adversely correlated to diameter. Both heartwood extractives and wood density are known to be factors that confer durability, and both these traits were found to be negatively correlated to decay mass-loss. This is an important preliminary result for the breeding program, especially if durable products are a perceived use for thinnings, as it indicates that genetic improvement for growth will not compromise the value of the thinnings, and that it should be possible to select for both durability and growth traits if desired. Further examination of this relationship would however be desirable, as the relatively small numbers of trees screened gave type-A correlation estimates of only modest precision. One of the main limitations of durability-related assessments of wood is the large amount of space, time and money needed to screen large numbers of individuals, particularly for the decay traits. Studies larger than the one described in Chapter 2 would be difficult

to implement using traditional, laboratory-based techniques, which is why development of NIR-based assessment was seen as an important tool.

The use of NIR to rapidly screen the durability-related wood properties at low cost was only partially successful. While the technique worked well for methanol extractives, predictions were not sufficiently precise to accurately screen individual genotypes for decay mass-loss. This is a disappointing outcome, as wider, low-cost screening of this trait within the ALRTIG breeding population would be highly desirable. The immediate, practical implication of this, given the modest nature of the *E. cladocalyx* genetic improvement program, is to focus on the SFR ROP for production of planting material for this purpose, as screening families from KI with acceptable decay resistance may be too expensive unless the market value of this end product becomes high enough in future to justify undertaking the extra work. Early identification of this decision-point is an important practical outcome of the study: if the strategy of focussing on SFR material without further screening is chosen, it should be done early to avoid introgression of KI genes that may potentially reduce the durability characteristics of the breed.

6.3 High heritability in Western Australian trials

The publication of genetic parameter estimates for three of the ALRTIG sites in Western Australia (Kojonup, Wellstead and Esperance 2) by Callister et al. (2008) revealed strong differences between ROPs, moderate-high heritabilities and little GxE for growth and form traits. Of these results, the moderate-to- high estimates of narrow-sense heritability for DBH and height ($\hat{h}^2$ = 0.44 and 0.85, respectively) were the most surprising. These heritabilities are well outside the usual range for growth traits in forest trees (Cornelius 1994) including eucalypts (Eldridge et al. 1993) which, as fitness-related traits, would normally expected to be low (Falconer and Mackay 1996). The authors selected a coefficient of relationship of 1/2.5, reflecting the likelihood of relatedness closer than half-sib (HS) and variance attributable to ROP and provenance differences were removed by the inclusion of a fixed effect. In fact, heritabilities were calculated from variance components estimated from a family model as:

$$\hat{h}^2 = \frac{2.5\sigma_f^2}{\sigma_f^2 + \sigma_P^2 + \sigma_B^2 + \sigma_e^2}, \qquad (6.1)$$

Where σ_f^2 denotes family nested within ROP, $\sigma_P{}^2$ corresponds to plots, $\sigma_B{}^2$ corresponds to incomplete blocks and $\sigma_e{}^2$ denotes residual variance. The inclusion of the block term in the denominator was relatively unusual, but served to reduce the heritability estimate. Possible sources of

the apparently inflated estimate attributed by Callister et al. (2008) included the highly variable outcrossing rates indicated by McDonald et al. (2003) and differential inbreeding depression, a phenomenon that might logically accompany differential outcrossing and that Hodge et al. (2001) had suggested, in a general context, might be responsible for both inflated heritability estimates and underestimation of the magnitude of GxE.

6.4 Modification of the A matrix using isozyme markers: the "pilot study"

In light of the fact that one the major objectives of the research program was to assess the additive variance and GxE across the 11 ALRTIG sites, it was decided that following up on the apparently inflated variance parameters from the Western Australian sites was necessary. This was conducted in two stages, the first being a "pilot study" that could make use of existing CSIRO isozyme marker data generated by the McDonald et al. (2003) study. The study (Chapter 3) presented both theory and results from analyses undertaken with an additive relationship (**A**) matrix modified by values resulting from heterogeneous selfing among families and subpopulations including ancestral inbreeding. While the marker panel (eight isozyme loci across seedling progeny) and the number of families studied (30 families) were sufficient to develop the theory, they were not necessarily considered to be indicative of the results that might be obtained from a more comprehensive study. Moreover, the plant material was sampled from very young seedlings which would not be expected to give parameter estimates representative of a mature population. The theory presented was an extension of work undertaken by Dutkowski and Gilmour (2001), developed here to account for ancestral inbreeding. The study was successful in demonstrating the potential effect of differential family-level inbreeding on the overall additive variance parameter estimates. It showed that the coefficient of relationship of 1/2.5 suggested by Eldridge et al. (2003) would be quite appropriate for this small population given that the population average for selfing was $s = 0.36$, which would imply $\rho = 1/2.7$ or $\rho = 1/2.3$ if ancestral inbreeding was also to be accounted for.

Chapter 3 also investigated the effect of adjusting **A** for differential selfing, at the provenance and family levels, on BLUP ranks. While the application of a coefficient of relationship will globally adjust heritability, to the correct level if suitably chosen, BLUP breeding values will not be suitably scaled and re-ranking will not be accounted for. Though the ASREML software (Gilmour et al. 2009) that is popularly used for carrying out single-tree analyses in forestry trials now has a facility to globally adjust **A** for prescribed levels of selfing (a function based on the work of Dutkowski and Gilmour [2001]), that will provide overall shrinkage of breeding values and reduction of additive variance estimates, this will still not properly account for differential relatedness among provenances, families and progeny. The finding of the pilot study was that re-ranking due to differential family-level outcrossing would lead to some differences among selections for a hypothetical 30-clone seed

orchard. The impact on genetic gain, however, would be quite low, as even 'incorrect' selections would not generally be amongst the worst.

A further indication from the pilot study was that there was a lack of correlation between fitness (indicated by survival and DBH measured in four-year old families on the Lismore, Victoria site) and outcrossing. This result was potentially exciting because all other eucalypts hitherto studied have been found to suffer from marked inbreeding depression, a phenomenon common in preferentially-outcrossing plants (Potts and Wiltshire 1997). However, because the isozyme assays had been carried out on seedlings, it was thought likely that lack of time for selection against inbred seedlings with deleterious recessive alleles might explain the result. Indeed it is commonly found that eucalypt seedling populations have lower heterozygosity than expected, but that there is typically a small excess of heterozygotes relative to expectation by the time sexual maturity has been reached due to strong selection against inbred trees (Potts and Wiltshire 1997).

The overall implications of the pilot, isozyme-based study were that differential outcrossing could effectively be accounted for using marker data and modification of the additive matrix, and that there were some potentially unusual indications of lack of a heterozygosity-fitness relationship. Though the isozyme assays had been performed on very young material, and hence might not be properly indicative of the families measured in the field trial at Lismore or elsewhere in the ALRTIG breeding population, the levels of selfing indicated would be required to justify selection of the coefficient of relationship (1/2.5) used by Callister et al. (2008) – a value apparently still not sufficient to deflate variance estimates to 'normal' levels for eucalypt growth traits. This was especially the case as a significant proportion of full-sib progeny was not indicated by the isozyme study, a result that in itself was surprising given the apparently high levels of selfing and the wide-spaced woodland setting of many of the parent trees. Overall, the results suggested that a more-detailed molecular analysis of the relatively mature trees in the breeding population itself would be justified.

6.5 SNP study of breeding population

Chapter 4 presents the results of the SNP-based analysis of co-ancestry and other population parameters relating to the ALRTIG breeding population. This study differed from the isozyme-based pilot study of Chapter 3 in that: (i) DNA from relatively mature trees (age 5-years) that were part of the ALRTIG breeding population at Lismore, Victoria was sampled rather than from very young seedlings, (ii) many more markers were employed, allowing for better estimates of relatedness parameters, and (iii) more families, 119 of the 137 in the ALRTIG breeding population, were included. This allowed a more-detailed analysis of the population structure, the relationship between fitness and heterozygosity (or inbreeding) and estimates of family-average co-ancestry to be made.

The Chapter 4 study generally confirmed many findings of the earlier study by McDonald et al. (2003). The breeding population was found to be highly structured with F_{ST} = 18 %. Most (13.5 %) of this was partitioned among the three subpopulation groups (the two wild ROPs and the planted stand group). All of the land-race selections were found to be derived from the SFR ROP and it was confirmed that, collectively, these have similar genetic diversity to the wild SFR population. As indicated by the isozyme study (Chapter 3), relatedness and inbreeding was highly variable among families and consistent patterns corresponding to the two wild ROPs and the cultivated material were evident.

The land-race material was found to be highly outcrossed, with estimates of co-ancestry close to HS on average, and little inbreeding. Wild families from the SFR ranged between highly outcrossed and highly inbred, as indicated by the McDonald et al. (2003) study. Elevated inbreeding also caused inbreeding depression in this ROP, as evidenced by lower growth in families with higher *f*. While the KI families were found to be very closely related and highly inbred on average, there was no significant evidence of inbreeding depression in this ROP. Some of the best-performing families from KI, in terms of growth, comprised closely related trees likely to have been the product of selfing. In fact three individuals from this ROP were found to be homozygous at all 75 SNP loci, suggesting several generations of inbreeding rather than just a single selfing event, and a number of families were found to have $2\theta \approx 1$. This lends evidence to the McDonald et al. (2003) theory that the species, at least on Kangaroo Island, may have experienced a significant genetic bottleneck, possibly associated with rapid climate and sea-level changes during the Pleistocene. A bottleneck event has the capacity to purge deleterious recessive alleles, allowing survivors to potentially tolerate close inbreeding (Taft and Roff 2012). The propensity of KI trees to undergo self-fertilisation and outcrossing is an interesting area of future research: it is unclear why the inbreeding has occurred, though possible mechanisms could be sparse and/or asynchronous flowering among family groups in the wild, preferential self-fertilisation mechanisms or other unknown causes. Crosses performed by Ellis and Sedgley (1992), indicating that *E. cladocalyx* ranges from self-compatible to self-incompatible, involved only three cultivated trees of unknown origin (Ellis et al. 1991). This aspect should be further investigated.

An important difference between the studies in Chapter 3 and Chapter 4 was in the estimation of diversity using the H_E (expected heterozygosity) measure. Whereas McDonald et al. (2003) estimated H_E of 0.18 and 0.10, the estimates from Chapter 4 were 0.38 and 0.22 for SFR and KI subpopulations, respectively (Table 4.1). The estimates for the cultivated stands were 0.17 and 0.37 from the earlier and later studies, respectively, though the study in Chapter 4 involved a greater range of land-race selections. While the general pattern of diversity among the two studies was similar, the greater

diversity in the latter study was unexpected, as the population samples would be expected to be similar in actual diversity. The difference between studies may be due to the low number of markers employed in the McDonald et al. (2003) study. Though direct comparisons between isozymes and SNPs are few, due to the passage of time between these technologies being considered 'state of the art', differences in estimates of parameters including H_E among isozymes, SNPs and intervening technologies (e.g. RFLPs, SSRs) are common (e.g. Coates et al. 2009; Nybom 2004) . Despite these differences, the practical implications of the two studies (Bush and Thumma 2013; McDonald et al. 2003) with respect to population parameters, origins of the land-race population, patterns of genetic diversity among subpopulations and inbreeding were highly congruent.

6.6 Comparison of mixed-mating and half-sib assumptions on analysis of ALRTIG breeding population

Chapter 5 presents genetic-parameter estimates and an analysis of GxE from across the 11 ALRTIG sites in southern Australia. Two alternative models were compared, one assuming that trees within families are half-siblings (the HS model) and an alternative where the Chapter 4 estimates of relatedness and inbreeding resulting from mixed mating (denoted the MM model) were used to update the additive relationship matrix (**A**). The family inbreeding estimates were also fitted as a covariate to account for inbreeding depression that was found to be significant in the SFR population. As expected from Chapter 3, the impact of adjusting **A** was very significant, with resultant deflation of additive variance and heritability estimates. Fitting $\hat{f}$ as a covariate also significantly reduced the additive variance estimate attributable to ID in the SFR subpopulation. The overall unsuitability of the HS approach for this dataset was demonstrated by non-convergence of the models for the height trait at four of the sites, Kojonup, Esperance 2 and Wellstead (all in Western Australia) and Avenue, South Australia. This was caused by the estimates of additive variance exceeding the total phenotypic variance in the individual-tree models. This problem did not occur in the Callister et al. (2008) analysis of the same data (for the Western Australian sites) because family models were used. Family models allow estimates of additive variance that can ultimately exceed the total phenotypic variance once divided by the coefficient of relationship, as only a proportion of the additive variance is directly estimated. However the result was extremely high heritability estimates for height (HT) (0.60, 0.79 and 0.58 for the three sites) after application of $1/\rho = 2.5$ and removal of plot and incomplete block variance in the denominator of $\hat{h}^2$ (see Eq. 6.1). Inspection of Table 4 in Callister et al. (2008) reveals that if $1/\rho = 4$ had been applied then the additive variance components would have been 126%, 96 % and 94 % of the total phenotypic variance at Wellstead, Kojonup and Esperance, respectively.

From the results of Chapter 4, an obvious potential cause of the very high additive variance components estimated for the breeding population in Chapter 5 might be the elevated inbreeding,

especially in the KI subpopulation. To examine the relative contributions to the inflated additive variance estimates, separate analyses were performed using both the HS and MM assumptions for the KI, SFR and cultivated sections of the breeding population. These analyses (Table 5.6) revealed that the KI subpopulation is the greatest contributor to the inflated estimates, with the across-sites model not converging for HT under the HS assumption and $\hat{h}^2 = 0.53$ under the MM assumption.

Of some concern is the still-moderate $\hat{h}^2$, particularly for height (0.32 across the whole population, Table 5.6), even after the adjustments to the pedigree and adjustment for ID made under the MM assumption. The explanation postulated for this in Chapter 5 is the likelihood of non-additive variance being present and confounded with additive variance estimates. Dominance variance is highly likely to accompany the elevated inbreeding demonstrated in Chapter 4 for the KI and SFR subpopulations and is a probable cause of the ID identified for the SFR population. Estimation of non-additive variance is usually undertaken by analysis of field trials of controlled-cross progeny that give rise to full-sib families. This resource is not presently available for *E. cladocalyx*, though it may be necessary to adequately separate non-additive variance so that the breeding population can be properly modelled and selections can be undertaken. One other possibility is that a dense marker panel might be applied to sections of the subpopulation (probably a similar sized sample to that of Chapter 4) so that precise individual-dyad co-ancestry can be estimated. Klápště et al. (2014) have shown that if this can be done, then non-additive variance could potentially be estimated from full-sibs identified via their co-ancestry. Attempting this with the current level of genotyping is impractical, as the simulation studies undertaken in Chapter 5 show that many more SNP markers (thousands) would be needed to adequately estimate pair-wise co-ancestry between trees, a general conclusion supported by Doerksen et al. (2014). However in practical terms, the relative cost of undertaking control-pollinations, establishing and measuring field trials and waiting for meaningful results should be carefully weighed against the cost of undertaking more genotyping with dense markers, especially as phenotypic data collection and DNA extraction have already been undertaken for this breeding population. This approach might be considered to be a variation on the 'breeding without breeding' concept of El-Kassaby et al. (2011).

6.7. Genotype-by-environment interactions

Chapter 5 also examined GxE across 11 sites in southern Australia. Provenances and ROPs across sites were significantly different and their rankings quite stable. For axis persistence (AP) the Kangaroo Island ROP was consistently inferior to the SFR and cultivated subpopulations (Appendix 3.2). For HT and DBH, KI and the land race performed well. Taking both growth and form traits together, the land-race material performed particularly well, perhaps because trees were phenotypically selected from the planted stands. They were also more-highly outcrossed than the

wild, SFR material from which they were derived. It possible this is the result of phenotypic selection for vigour resulting in selection against inbred material, though it is also the case that the neighbourhood inbreeding that occurs among clusters of related individuals in wild stands is likely to be absent in the planted stands due to wide dispersal of closely related individuals. Family-level GxE was significant overall, but not between all pairs of sites (Appendix 3.4). Generally, there was no interpretable trend relating to geographic location or edaphic parameters at the reactive sites, though r_B for some of the Western Australian sites tended to be lower, albeit with different sites involved for each trait. Estimates of r_B for the growth (DBH and HT) traits were high (all > 0.5, average 0.7) for the set of trials in the eastern states comprising Bordertown, Hamilton, Lismore, Corowa and Wagga Wagga. However in this cluster, r_B was generally low for the axis persistence trait (average 0.5). Table 6.1 presents a simple means of identification of clusters by their type-B correlations averaged across all three measured traits. This method would not be used for selection purposes (where careful consideration of weights applied to each trait would be necessary), but is used here to simplify the discussion of general patterns in GxE across sites. It is evident , for example, that [KOJ, ESP2,WEL] (as identified by Callister et al. 2008) and [ESP1, ESP2,AVE] form clusters as do [KOJ, WEL, LIS] and [KOJ, WEL, HAM], but not [KOJ, WEL, HAM, LIS], with $r_B = 0.66$ being used as an approximate bound beneath which GxE becomes of practical importance (Shelbourne 1972). WEL stands out as being well correlated with all other sites except ESP1 and AVE whereas WGW has significant type-B interaction with all others except WEL.

A practically useful application of GxE, once identified, is relating the reactivity of certain sites to their climatic or edaphic parameters to formulate predictive capacity (Kanzler 2002). In the case of this study, it is difficult to relate site edaphic factors to performance for the three traits. The Western Australian and South Australian sites are all situated on sandy soils with relatively high evapotranspiration, and while some of these sites have high type-B correlations (e.g. the closely situated ESP1 and ESP2) others are very weakly correlated (e.g. ESP1 with each of MOR,KOJ and WEL). Though the analysis has been carried out within a BLUP framework, which deals with missing genotype entries in the family-by-site table by appropriately shrinking (regressing towards the mean) genotypes for which information is absent, this cannot fully compensate for missing information. It should be remembered that only a small subset of families are present on the KOJ, WEL, ESP1, ESP2 and AVE sites (Table 6.2) and these sites are involved in some of the lowest type-B correlations with some of the highest standard errors (Appendix 3.4).

Table 6.1 heat-mapped table of type-B correlations averaged across the DBH, HT and AP traits (separate Tables for each trait are given in Appendix 3.4).

Site	KOJ	WEL	ESP1‡	ESP2	AVE	BTN	HAM	LIS	COR	WGW
Moora†	0.66	0.71	0.09	0.45	0.24	0.64	0.46	0.65	0.56	0.49
Kojonup		0.81	0.33	0.76	0.50	0.77	0.70	0.68	0.65	0.50
Wellstead			0.37	0.80	0.54	0.82	0.79	0.79	0.77	0.75
Esperance1‡				0.81	0.78	0.39	0.51	0.51	0.46	0.47
Esperance2					0.82	0.58	0.65	0.66	0.62	0.62
Avenue						0.30	0.61	0.59	0.54	0.54
Bordertown							0.68	0.67	0.64	0.49
Hamilton								0.58	0.60	0.53
Lismore									0.60	0.60
Corowa										0.60

†Only DBH was measured at MOR ‡ Height was not measured at ESP1

6.8 Why is the species apparently different?

One intriguing question arising from the study is why *E. cladocalyx* appears to exhibit such curious behaviour in respect to its level of inbreeding as revealed by molecular markers, the heterozygosity-fitness response of the KI material and the unusually large amount of 'additive' variance, which is presumed to include a significant proportion on non-additive variance, identified in the quantitative analysis. These are all unusual when compared with other studied eucalypts. The most likely explanation for these results, put forward by McDonald et al. (2003) and discussed in Chapters 4 and 5, is that the species, particularly the subpopulation group found on KI, may have gone through a bottleneck event that significantly reduced the population size. Populations emerging from bottleneck events may be less susceptible to inbreeding depression because of very strong selection pressure ("purging selection") against deleterious, recessive alleles during the event, though the result is often incomplete purging of sub-lethal alleles (Charlesworth and Charlesworth 1987; Frankham et al. 2001). Sea-level changes during the Pleistocene could plausibly have led to the species becoming isolated in upland refugia, bearing in mind that the parts of Kangaroo Island that are now at relatively low altitudes above sea level were once relatively more elevated when sea levels were lower and the island land mass was connected to the mainland. This theory is supported by the disjunct populations that are restricted to quite distinctive upland parts of the landscape.

While the bottleneck theory provides an explanation for the observations made here on *E. cladocalyx*, the question of why similar results have not been found in other eucalypts remains. One possible explanation is that the number of detailed studies on eucalypt breeding populations is still quite limited, despite the fact that the genus contains species that are among the world's most important forest plantation species. As discussed in Chapter 1.1, much of the focus on eucalypt domestication and breeding has centred on species that are well adapted to better quality, high-rainfall sites.

Harwood (2011) estimates that over 90% of the world's eucalypt plantations are of just nine species and hybrids among them, (*E. camaldulensis*, *E. dunnii*, *E. globulus*, *E. grandis*, *E. nitens*, *E. saligna*, *E. urophylla*, *E. pellita* and *E. tereticornis*), and much of the detailed analysis of breeding systems and sophisticated quantitative genetic analysis has focussed on them and a few other high-rainfall species from the ash group such as *E. regnans*. A feature that sets *E. cladocalyx* apart from these species (with the possible exception of *E. camaldulensis* that inhabits riparian environments) is the relatively low-rainfall open-forest and woodland environment that it occupies. It is possible that the species' unusual attributes are present in other eucalypts that have not yet been examined. There is now growing interest in testing eucalypts that are adapted to marginal land (e.g. Gardner et al. 2007; Harwood et al. 2007; Hobbs et al. 2008; Swain and Louw 2011). The assumption that inbred progeny will not be present in progeny tests and breeding populations should be examined more closely for these species.

6.9 Practical implications of the study

The study has examined a number of aspects of the species' genetics that will have practical implications for the domestication process. Chapter 2 investigated the genetic control of traits that relate to natural durability of wood. The finding was that traits including heartwood proportion, extractive content and decay mass loss were all variable at both the region of provenance level and the family level with narrow-sense heritability ranging from low to high. The commercial motivation for this study was the potential to use pole-sized logs in applications such as vine trellis posts where naturally-durable, high-strength posts that do not need chemical preservatives would be a marketable product from thinnings and/or dedicated shorter-rotation plantations. This market is yet-to-be developed, and technical specifications in terms of durability and strength of *E. cladocalyx* poles have not been proven. Though promising, it is therefore not certain that this end-use will be an important objective. The study revealed that the SFR subpopulation generally had significantly higher decay resistance than either the Eyre Peninsula or KI material. Since Eyre Peninsula has already been removed from the ALRTIG breeding population in Australia, due to its poor growth and form, this leaves a decision about how to best manage the mix of SFR and KI material. As discussed in Chapters 1 and 5, the ALRTIG breeding population also contains a considerable proportion of land-race families. These were shown in Chapter 4 to have originated from the SFR, so they could reasonably be expected to have relatively good decay resistance. The options for management include (i) screening the entire breeding population for trees with good durability properties and adding this as a selection trait, (ii) concentrating on SFR, perhaps by segregating the breeding population into separate breeds *sensu* Jayawickrama and Carson (2000), without further screening under the assumption that this will give generally acceptable durability (iii) screening and forming a breed from the SFR material only and (iv) as a short-term measure, identify enough trees with good durability properties (including from the already-screened Benalla population discussed in Chapter 2) and form a clonal seed orchard dedicated to seed production for durable post plantations. That these options are now

available to tree breeders underlines the importance of screening wood properties early in the domestication process as advocated in Chapter 1, as proceeding under the direction of the original strategy would result in the loss of option (ii and iv) and potential loss of gain from delays in commencing option (ii).

Another strategy that could be considered is complete removal of KI material from the breeding population. This would be justifiable in that the ROP has markedly lower resistance to decay and young-age heartwood proportion (Chapter 2) and poorer form, as demonstrated in Chapter 5 by a lower axis persistence rating. However at a number of the sites, height and diameter growth for the KI ROP has been excellent, equalling or outperforming the land race and SFR wild material. To avoid the loss of growth potential it may be best to retain the KI material, possibly as a separate breeding population or subline.

Another reason to restructure the breeding population, separating the KI families, is because the mating system of this subpopulation group, in particular, should be more closely examined. The evidence for more than one generation of selfing or close inbreeding in a number of the KI families (Chapter 4) raises questions about the breeding biology of this subpopulation. One possibility is that some trees preferentially self, or that pollen production or vectors are limited which has led to the close relatedness among many of the progeny of wild mothers. Panmixis is assumed within the ALRTIG progeny trials, many of which have now been thinned for seed production. Studies using molecular markers to determine outcrossing rates and parentage among the progeny of these stands should be undertaken to determine whether this is actually the case, and whether certain families that have already demonstrated ancestral inbreeding need to be specially managed.

An important general finding of the study was that molecular markers can provide important information that complements phenotypic measurement data in quantitative analysis. Though the cost of genotyping individual trees has decreased dramatically in the last decade, the cost of collecting tissue samples and processing DNA for large breeding populations is still considerable: this cost is less likely to decrease in future, as it is intrinsically related to the cost of labour and field sampling. The method employed in Chapter 4 was to sample a small number (an average of ten) trees per family and use this sample to determine average relatedness and inbreeding at the family level. This restricted the genotyping effort to around 1200 trees, the information from which was then applied to the entire breeding population of more than 13 000 trees. Such an approach is presented as a cost-effective strategy for low-input breeding programs. This approach does not provide some of the detailed information such as inference of Mendelian sampling and relatedness between pairs of trees that could be gained from a dense marker pane (i.e. thousands of SNPs) applied to the entire breeding population. However this level of information is probably not required and certainly not financially

justifiable to collect for relatively small-scale breeding programs such as that of *E. cladocalyx* in Australia. While the cost of genotyping more trees is not likely to decrease markedly, the cost of gathering more markers per individual is likely to continue to fall (Grattapaglia et al. 2011). It is possible that high-throughput sequencing of individual tree genomes may become feasible for a relatively low cost and dense marker panels of tens of thousands of SNPs are already a reality for eucalypts (Silva-Junior et al. 2013). In the case of this study, a dense marker panel would have given greater precision of estimation of relatedness between individual dyads (as indicated by the simulation study in Chapter 4). This would have enabled more-precise estimation of relatedness and inbreeding and the possibility of estimating non-additive variance as described by Klápště et al. (2013). The implication for future studies is that inclusion of marker-based information is likely to become increasingly feasible and useful.

6.10 Conclusion

The study has examined aspects of *E. cladocalyx* genetics that underpin its domestication. The objective has been to use current methodology and techniques in wood property screening, molecular and quantitative genetics to optimise the process of selection and to better inform decision making in the tree breeding program. Chapter 2 addresses the question of whether there is significant genetic variation in wood properties related to naturally-durable posts and poles. This study revealed that young-aged, pole-sized material is probably sufficiently durable for the target product, and that there is considerable genetic variation at both the subpopulation and family level with low to high narrow-sense heritability estimates for most traits. This presents practical options including concentrating on the more-durable South Flinders Ranges ROP and/or recurrent selection and breeding focussed on the durability traits. The genetic correlation between the durability traits and diameter growth was also examined. Preliminary indications from this small-scale study are that there are no adverse genetic correlations between growth and durability traits in this species. One disappointing aspect of this section of the study was the poor correlation between NIR predictions of decay mass-loss and laboratory-determined mass loss, despite excellent NIR predictive ability for extractive content. This is a key selection trait that is both time consuming and laborious to measure. It was hoped that more extensive screening of the ALRTIG breeding population could be carried out using the NIR technique.

An important conclusion is that molecular marker-based information can be highly informative and complementary to traditional quantitative analysis of first-generation breeding populations. In this case, SNP markers were used to characterise the breeding population in terms of its subpopulation structure, diversity, origins of land-race accessions and co-ancestry of individual trees and families. This information can be used directly, for example information about the land race shows that families within the breeding population are not closely related and are quite genetically diverse, and

that further selections could be made from the land race itself for infusion into the breeding population. Marker-based estimates of inbreeding within the wild populations, particularly the KI population were very high. This raises questions about the mating system of some families within the KI subpopulation, which should be further examined to ensure trees in multi-provenance seed orchards outcross as expected and to gain further insight into the mechanism of selfing. The estimates of co-ancestry and inbreeding were used to modify the additive relationship matrix used to estimate additive variance components at single sites and across sites for determination of GxE interaction. This was an important step as individual-tree models would not converge at some sites under the traditional assumption of HS families due to the extreme departure from this attributable to inbreeding in some families. This also led to reduction in heritability estimates and re-ranking of BLUP breeding values.

Though the molecular marker-based approach to adjusting the pedigree and accounting for ID in the breeding population was successful in amending estimates of additive variance, heritability estimates, particularly for the height trait, remain moderately high. This is assumed to be due to significant non-additive variance, particularly in the wild, inbred families of Kangaroo Island. Further steps to confirm and quantify this apparent non-additive component of variance should be made. This could be achieved either by the traditional method of establishing and assessing progeny trials based on controlled crosses, or possibly by employing a denser panel of markers so that individual dyad coancestry (including that of full-sibs) can be reliable estimated. The former approach could also be taken as an opportunity to study possible unusually high self-compatibility in the species.

Overall, the prospects for genetic improvement of *E. cladocalyx* are good, due to considerable genetic variation in the traits of interest. However the species also presents some significant challenges and opportunities due to its unusually high levels of inbreeding. Though some of these may be unique to this particular species, the general approach may be more widely applicable to others.

Consolidated References

Adams, W.T. and D.G. Joyce. 1990. Comparison of selection methods for improving volume growth in young coastal Douglas-fir. Silvae Genetica. 39:219-226.

Adams, W.T., D.B. Neale and C.A. Loopstra. 1988. Verifying controlled crosses in conifer tree-improvement programs. Silvae Genetica. 37:147-152.

Annicchiarico, P. 2002 Genotype x environment interaction: challenges and opportunities for plant breeding and cultivar recommendations. *In* FAO Plant Production and Protection Paper (No.174), Rome, p 115.

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

Araújo, J.A., N.M.G. Borralho and G. Dehon. 2012. The importance and type of non-additive genetic effects for growth in *Eucalyptus globulus*. Tree Genetics & Genomes. 8:327-337.

Askew, G.R. and Y.A. El-Kassaby. 1994. Estimation of relationship coefficients among progeny derived from wind-pollinated orchard seeds. Theoretical and Applied Genetics. 88:267-272.

Bailleres, H., F. Davrieux and F. Ham-Pichavant. 2002. Near infrared analysis as a tool for rapid screening of some major wood characteristics in a *Eucalyptus* breeding program. Annals of Forest Science. 59 479-490.

Baltunis, B.S., T.A. Martin, D.A. Huber and J.M. Davis. 2008. Inheritance of foliar stable carbon isotope discrimination and third-year height in *Pinus taeda* clones on contrasting sites in Florida and Georgia. Tree Genetics & Genomes. 4:797-807.

Barrett, J.C., B. Fry, J. Maller and M. Daly. 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 21:263-265.

Beaumont, M.A. and R.A. Nichols. 1996. Evaluating loci for use in the genetic analysis of population structure. Proceedings: Biological Sciences. 263:1619-1626.

Beckman, C.H. 2000. Phenolic-storing cells: keys to programmed cell death and periderm formation in wilt disease resistance and in general defence responses in plants? Physiological and Molecular Plant Pathology. 57:101-110.

Blakemore, P. 2004. Density and shrinkage of four low-rainfall plantation-grown eucalypts. Australian Forestry. 67:152-155.

Blouin, M.S. 2003. DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. Trends in Ecology and Evolution. 18 503-510.

Boomsma, C. 1946. The vegetation of the Southern Flinders Ranges, South Australia. Transactions of the Royal Society of South Australia. 70:259-276.

Boomsma, C.D. 1969. The forest vegetation of South Australia. Woods and Forests Department Bulletin No. 18. Woods and Forests Department:, Adelaide.

Borralho, N.M.G. 1994. Heterogeneous selfing rates and dominance effects in estimating heritabilities from open-pollinated progeny. Canadian Journal of Forest Research. 24:1079-1082.

Borralho, N.M.G. 1995. The impact of individual tree mixed model methods (BLUP) in tree breeding strategies. *In* Eucalypt plantations: Improving Fibre Yield and Quality, Proc. CRCTHF-IUFRO Conference Eds. B. Potts, N. Borralho, J. Reid, R. Cromer, W. Tibbits and C. Raymond. CRC for Temperate Hardwood Forestry, Hobart, Tasmania.

Brawner, J.T., D.J. Lee, C.M. Hardner and M.J. Dieters. 2011. Relationships between early growth and *Quambalaria* shoot blight tolerance in *Corymbia citriodora* progeny trials established in Queensland, Australia. Tree Genetics & Genomes. 7:759-772.

Brooker, M.I.H. and D.A. Kleinig. 2006. Field guide to eucalypts: volume 1 south-eastern Australia, third edition. Bloomings Books, Melbourne.

Brown, A. 1990. Genetic characterisation of plant mating systems. *In* Plant population genetics, breeding and genetic resources Eds. A. Brown, M. Clegg, A. Kahler and B. Weir. Sinauer Associates, Sunderland, Massachusetts, pp 145-162.

Bulman, P. and J. Fairlamb. 1998. Farm Tree Improvement Project (project report). Primary Industries SA, Murray Bridge, SA, p 60.

Bulman, P., C.E. Harwood and J. Fairlamb. 2000. Early growth of *Eucalyptus cladocalyx* in six provenance trials in South Australia, unpublished draft report. ALRTIG, Canberra.

Burdon, R.D. 1977. Genetic correlation as a concept for studying genotype-environment interaction in forest tree breeding. Silvae Genetica. 26:168-175.
Burdon, R.D. 2010. Wood properties and genetic improvement of radiata pine. New Zealand Journal of Forestry. 55:22-27.
Burdon, R.D. and P.L. Wilcox. 2007. Population management: potential impacts of advances in genomics. New Forests. 34:187-206.
Burley, J. and R.H. Kemp. 1972. International tropical provenance trials and genotype-environment interactions. *In* Proceedings of the joint symposia for the advancement of forest tree breeding of Genetics Subject Group, International Union of Forest Research Organizations and Section 5, Forest Trees, Society for the Advancement of Breeding Researches in Asia and Oceania. Government Forest Experiment Station of Japan, Tokyo.
Bush, D. 2011. Selecting and breeding eucalypts for natural durability. *In* Developing a eucalypt resource: learning from Australia and elsewhere Ed. J. Walker. Wood Technology Research Centre, University of Canterbury, Christchurch, New Zealand, pp 125-136.
Bush, D., C. Harwood and D. Stackpole. 2001. Overview of species selection process and tree improvement strategies. *In* Australian Low Rainfall Tree Improvement Group: Compendium of Hardwood Breeding Strategies. Rural Industries Research and Development Corporation, Canberra, pp 1-26.
Bush, D., T.J. Jackson, J. Driscoll and C.E. Harwood. 2009. Australian Low Rainfall Tree Improvement Group: metadata from measures of hardwood tree improvement trials in southern Australia. *In* RIRDC Report 09/078. Rural Industries Research and Development Corporation, Canberra, p 143.
Bush, D., D. Kain, P.J. Kanowski and A.C. Matheson. 2015. Genetic parameter estimates informed by a marker-based pedigree: a case study with *Eucalyptus cladocalyx* in southern Australia. Tree Genetics & Genomes. 11:798.
Bush, D., D. Kain, C. Matheson and P. Kanowski. 2011a. Marker-based adjustment of the additive relationship matrix for estimation of genetic parameters—an example using *Eucalyptus cladocalyx*. Tree Genetics & Genomes. 7:23-35.
Bush, D. and K. McCarthy. 2008. Scope for genetic improvement of natural durability in low rainfall eucalypts for vine trellis posts. Rural Industries Research and Development Corporation, Canberra. 45 p.
Bush, D., K. McCarthy and R. Meder. 2011b. Genetic variation of natural durability traits in *Eucalyptus cladocalyx* (sugar gum). Annals of Forest Science. 68:1057-1066.
Bush, D. and B. Thumma. 2013. Characterising a *Eucalyptus cladocalyx* breeding population using SNP markers. Tree Genetics & Genomes. 9:741-752.
Butcher, P., C. Harwood and T.H. Quang. 2004. Studies of mating systems in seed stands suggest possible causes of variable outcrossing rates in natural populations of *Acacia mangium*. Forest Genetics. 11:303-309.
Callister, A., D.J. Bush, S. Collins and W. Davis. 2008. Prospects for genetic improvement of *Eucalyptus cladocalyx* in Western Australia. New Zealand Journal of Forestry Science. 38:211-226.
Campbell, R.K. 1992. Genotype x environment interaction: a case study for Douglas-fir in western Oregon. Research Paper PNW-RP-455. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, Oregon, U.S.A. 21 p.
Cane-Retamales, C., F. Mora, F. Vargas-Reeve, S. Perret and R. Contreras-Soto. 2011. Bayesian threshold analysis of breeding values, genetic correlation and heritability of flowering intensity in *Eucalyptus cladocalyx* under arid conditions. Euphytica. 178:177–183.
Charlesworth, D. and B. Charlesworth. 1987. Inbreeding depression and its evolutionary consequences. Annual Review of Ecology and Systematics. 18:237-268.
Chen, Z.-Q., M.R.G. Gil, B. Karlsson, S.-O. Lundqvist, L. Olsson and H.X. Wu. 2014. Inheritance of growth and solid wood quality traits in a large Norway spruce population tested at two locations in southern Sweden. Tree Genetics & Genomes. 10:1291-1303.
Clegg, M. 1980. Measuring plant mating systems. Bioscience. 30:814-818.
Coates, B.S., D.V. Sumerford, N.J. Miller, K.S. Kim, T.W. Sappington, B.D. Siegfried and L.C. Lewis. 2009. Comparative performance of single nucleotide polymorphism and microsatellite markers for population genetic analysis. Journal of Heredity. 100:556-564.
Cornelius, J. 1994. Heritabilities and additive genetic coefficients of variation in forest trees. Canadian Journal of Forest Research. 24:372-379.
Costa e Silva, J. and L. Graudal. 2008. Evaluation of an international series of *Pinus kesiya* provenance trials for growth and wood quality traits. Forest Ecology and Management. 255:3477-3488.
Costa e Silva, J., C. Hardner and B.M. Potts. 2010a. Genetic variation and parental performance under inbreeding for growth in *Eucalyptus globulus*. Annals of Forest Science. 67 606p1-606p8.

Costa e Silva, J., C. Hardner, P. Tilyard, A.M. Pires and B.M. Potts. 2010b. Effects of inbreeding on population mean performance and observational variances in *Eucalyptus globulus*. Annals of Forest Science. 67:605

Costa e Silva, J., C. Hardner, P. Tilyard and B.M. Potts. 2011. The effects of age and environment on the expression of inbreeding depression in *Eucalyptus globulus*. Heredity. 107:50-60.

Costa e Silva, J., B.M. Potts and G.W. Dutkowski. 2006. Genotype by environment interaction for growth of *Eucalyptus globulus* in Australia. Tree Genetics & Genomes. 2:61-75.

Crocker, R.L. 1946. An introduction to the soils and vegetation of the Eyre Peninsula, South Australia. Transactions of the Royal Society of South Australia. 70:83-107.

Crocker, R.L. and J.G. Wood. 1947. Some historical influences on the development of the South Australian vegetation communities and their bearing on concepts and classification in ecology. Transactions of the Royal Society of South Australia. 72:91-136.

CSIRO. 2014. State of the Climate 2014. CSIRO, Australian Government Bureau of Meterology, Canberra. 15 p.

Cullis, B., P. Jefferson and R.S. Thompson, AB. 2014 Factor analytic and reduced animal models for the investigation of additive genotype-by-environment interaction in outcrossing plant species with application to a *Pinus radiata* breeding programme. Theoretical and Applied Genetics. 127:2193-2210.

de Lange, W.J., R. Veldtman and M.H. Allsopp. 2013. Valuation of pollinator forage services provided by *Eucalyptus cladocalyx*. Journal of Environmental Management. 125:12-18.

de Resende, M.D.V. and R. Thompson. 2004. Factor analytic multiplicative mixed models in the analysis of multiple experiments. Rev. Mat. Estat. 22:31-52.

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

Digby, P.G.N. 1979. Modified joint regression analysis for incomplete variety x environment tables. Journal of Agricultural Science. 93:81-86.

Dillon, S.K., M.F. Nolan, J.G. Bragg, P. Matter, W.J. Gapare and S.G. Southerton. 2013. Signatures of adaptation and genetic structure among the mainland populations of *Pinus radiata* (D. Don) inferred from SNP loci. Tree Genetics & Genomes. 9:1447-1463.

Doerksen, T.K., J. Bousquet and J. Beaulieu. 2014. Inbreeding depression in intra-provenance crosses driven by founder relatedness in white spruce. Tree Genetics & Genomes. 10:203-212.

Doerksen, T.K. and C.M. Herbinger. 2008. Male reproductive success and pedigree error in red spruce open-pollinated and polycross mating systems. Canadian Journal of Forest Research. 38:1742-1749.

Doerksen, T.K. and C.M. Herbinger. 2010. Impact of reconstructed pedigrees on progeny-test breeding values in red spruce. Tree Genetics & Genomes. 6:591-600.

Doran, J.C. 2000. *E. cladocalyx* species digest. *In* The Forestry Compendium, Module 2. CAB International,, Wallingford, UK.

Doran, J.C. and D.J. Boland. 1978. Collections of *Eucalyptus camaldulensis* Dehnh. and *Eucalyptus tereticornis* Sm. seed from northern Australia in 1977 for *ex situ* conservation/selection stands. Forest Genetic Resources Information. No 8:12-21.

dos Santos, P.E.T., I.O. Geraldi and J.N. Garcia. 2004. Estimates of genetic parameters of wood traits for sawn timber production in *Eucalyptus grandis*. Genetics and Molecular Biology. 27:567-573.

Downes, G.M., I.L. Hudson, C.A. Raymond, G.H. Dean, A.J. Michell, L.R. Schimleck, R. Evans and A. Muneri. 1997. Sampling plantation eucalypts for wood and fibre properties. CSIRO Publishing, Melbourne. 132 p.

Dutkowski, G.W. and A.R. Gilmour. 2001. Modification of the additive relationship matrix for open pollinated trials. *In* Developing the Eucalypt of the Future. Instituto Forestal: Chile, Valdivia, Chile, p 71.

Eberhart, S.A. and W.A. Russell. 1966. Stability parameters for comparing varieties. Crop Science. 6:36-40.

Ednie Brown, J. 1882. Annual progress report upon State Forest Administration in South Australia for the year 1881-2. Woods and Forests Department, Adelaide, p 24.

El-Kassaby, Y.A., E.P. Cappa, C. Liewlaksaneeyanawin, J. Klápště and M. Lstibůrek. 2011. Breeding without breeding: is a complete pedigree necessary for efficient breeding? PLoS ONE. 6:e25737.

Eldridge, K. 1978. Genetic improvement of eucalypts. Silvae Genetica. 27:205-209.

Eldridge, K., J. Davidson, C.E. Harwood and G. van Wyk. 1993. Eucalypt domestication and breeding. Clarendon Press, Oxford. 288 p.

Ellis, M.F. and M. Sedgley. 1992. Floral morphology and breeding system of three species of *Eucalyptus*, section *Bisectaria* (Myrtaceae). Australian Journal of Botany. 40:249-262.

Ellis, M.F., M. Sedgley and J.A.Gardner. 1991. Interspecific pollen-pistil interaction in *Eucalyptus* L'Her. (Myrtaceae):the effect of taxonomic distance. Annals of Botany. 68:185-194.

Ericsson, T. and A. Fries. 1999. High heritability for heartwood in north Swedish Scots pine. Theoretical and Applied Genetics. 98:732-735.
Excoffier, L., T. Hofer and M. Foll. 2009. Detecting loci under selection in a hierarchically structured population. Heredity. 103:285-298.
Excoffier, L. and H. Lischer. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources. 10:564-567.
Falconer, D.S. 1952. The problem of environment and selection. American Naturalist. 86:193-198.
Falconer, D.S. 1981. Introduction to Quantitative Genetics. Longman, London and New York.
Falconer, D.S. and T.F.C. Mackay. 1996. Introduction to Quantitative Genetics, 4th edition. Longman Scientific and Technical, London.
Falkenhagen, E.R. 1996. A comparison of the AMMI method with some classical statistical methods in provenance research: A case of the South African *Pinus radiata* trials. Forest Genetics. 3:81 - 87.
Finlay, K.W. and G.N. Wilkinson. 1963. The analysis of adaptation in a plant-breeding programme. Australian Journal of Agricultural Research. 14:742-754.
Fischer, J., D.J. Abson, V. Butsic, M.J. Chappell, J. Ekroos, J. Hanspach, T. Kuemmerle, H.G. Smith and H.v. Wehrden. 2014. Land sparing versus land sharing: moving forward. Conservation Letters. 7:149-157.
Fisher, R.A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. Transactions of the Royal Society of Edinburgh. 52:399-433.
Florence, R.G. 1996. Ecology and silviculture of eucalypt forests. Melbourne. 432 p.
Frankham, R., D.M. Gilligan, D. Morris and D.A. Briscoe. 2001. Inbreeding and extinction: effects of purging. Conservation Genetics. 2:279-285.
Franklin, I.R. 1977. The distribution of the proportion of the genome which is homozygous by descent in inbred individuals. Theoretical Population Biology. 11:60-80.
Fyfe, J.L. and N.T.J. Bailey. 1951. Plant breeding studies in leguminous forage crops I. Natural cross-breeding in winter beans. Journal of Agricultural Science. 41:371-378.
Galton, F. 1874. English men of science, their nature and their nurture. Macmillan, London. 270 p.
Gapare, W., M. Ivkovic, M.B. Powell, T.A. McRae and H.X. Wu. 2008. Genetics of shrinkage in juvenile trees of *Pinus radiata* D. Don from two test sites in Australia. Silvae Genetica. 57:145.
Gapare, W.J., B.S. Baltunis, M. Ivković and H.X. Wu. 2009. Genetic correlations among juvenile wood quality and growth traits and implications for selection strategy in *Pinus radiata* D. Don. Annals of Forest Science. 66 606-615.
Gapare, W.J., M. Ivković, S.K. Dillon, F. Chen, R. Evans and H.X. Wu. 2012. Genetic parameters and provenance variation of *Pinus radiata* D. Don. 'Eldridge collection' in Australia 2: wood properties. Tree Genetics & Genomes. 8:895–910.
Gardner, R.A.W., K.M. Little and A. Arbuthnot. 2007. Wood and fibre productivity potential of promising new eucalypt species for coastal Zululand, South Africa. Australian Forestry. 70:37-47.
Gaspar, M.J., A.I. de-Lucas, R. Alía, J.A.P. Paiva, E. Hidalgo, J. Louzada, H. Almeida and S.C. González-Martínez. 2009. Use of molecular markers for estimating breeding parameters: a case study in a *Pinus pinaster* Ait. progeny trial. Tree Genetics & Genomes. 5:609-616.
Gauch, H.G. 1992. Statistical analysis of regional yield trials : AMMI analysis of factorial designs. Elsevier, Amsterdam. 278 p.
Gauch, H.G. 2006. Statistical analysis of yield trials by AMMI and GGE. Crop Science. 46:1488-1500.
Gea, L.D., R. McConnochie and S. Wynyard. 2007. Parental reconstruction for breeding, deployment, and seed-orchard management of *Eucalyptus nitens*. New Zealand Journal of Forestry Science. 37:23-36.
Gierlinger, N., D. Jacques, M. Schwanninger, R. Wimmer, B. Hinterstoisser and L.E. Paques. 2003. Rapid prediction of natural durability of larch heartwood using Fourier transform near-infrared spectroscopy. Canadian Journal of Forest Research. 33:1727-1736.
Gierlinger, N., D. Jacques, M. Schwanninger, R. Wimmer and L.E. Paques. 2004. Heartwood extractives and lignin content of different larch species (*Larix* spp.) and relationships to brown-rot decay-resistance. Trees. 18:230-236.
Gilmour, A.R., B.J. Gogel, B.R. Cullis and R. Thompson. 2006. ASReml User Guide Release 2. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK, p 320.
Gilmour, A.R., B.J. Gogel, B.R. Cullis and R. Thompson. 2009. ASReml User Guide: Release 3. VSN International Ltd, Hemel Hempstead. 372 p.
Goodstein, D.M., S. Shu, R. Howson, R. Neupane, R.D. Hayes, J. Fazo, T. Mitros, W. Dirks, U. Hellsten, N. Putnam and D.S. Rokhsar. 2012 Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research. 40:D1178-D1186.
Grattapaglia, D. 2007. Marker-assisted selection in *Eucalyptus*. *In* Marker-assisted selection: current status and future perspectives in crops, livestock, forestry and fish Eds. E.P. Guimarães, J. Ruane, B.D. Scherf, A.

Sonnino and J.D. Dargie. Food and Agriculture Organization of the United Nations, Rome, pp 251-282.
Grattapaglia, D. and M. Kirst. 2008. *Eucalyptus* applied genomics: from gene sequences to breeding tools. New Phytologist. 179:911-929.
Grattapaglia, D., O.B. Silva-Junior, M. Kirst, B.M.d. Lima, D.A. Faria and G.J. Pappas. 2011. High-throughput SNP genotyping in the highly heterozygous genome of Eucalyptus: assay success, polymorphism and transferability across species. BMC Plant Biology. 11:65.
Greaves, A. and R.H. Kemp. 1977. International provenance trials - *Pinus oocarpa* Morelet. *In* IUFRO Symposium: Progress and Problems of Genetic Improvement of Tropical Forest Trees Eds. D.G. Nikles, J. Burley and R.D. Barnes. Commonwealth Forestry Institute: Oxford, Brisbane, Australia, pp 886-90.
Griffin, A.R. and P.P. Cotterill. 1988. Genetic variation in growth of outcrossed, selfed and open-pollinated progenies of *Eucalyptus regnans* and some implications for breeding strategy. Silvae Genetica. 37:124-131.
Gunn, B. 2001. Australian Tree Seed Centre operations manual. CSIRO, Canberra. 150 p.
Hardner, C., M. Dieters, I. DeLacy, J. Neal, S. Fletcher, G. Dale and K. Basford. 2011. Identifying deployment zones for *Eucalyptus camaldulensis* x *E. globulus* and x *E. grandis* hybrids using factor analytic modelling of genotype by environment interaction. Australian Forestry. 74:30-35.
Hardner, C. and W. Tibbits. 1998. Inbreeding depression for growth, wood and fecundity traits in *Eucalyptus nitens*. Forest Genetics. 5:11-20.
Hardner, C., R.E. Vaillancourt and B.M. Potts. 1996. Stand density influences outcrossing rate and growth of open-pollinated families of *Eucalyptus globulus*. Silvae Genetica. 45:226-228.
Hardner, C.M. and B.M. Potts. 1995. Inbreeding depression and changes in variation after selfing in *Eucalyptus globulus* ssp. *globulus*. Silvae Genetica. 44:46-54.
Hardner, C.M. and B.M. Potts. 1997. Postdispersal selection following mixed mating in *Eucalyptus regnans*. Evolution. 51:103-111.
Harris, S. 2000. Sugar gum provenance trials results. Agroforestry News. 9:18-19.
Harwood, C. 2011. New introductions - doing it right. *In* Developing a eucalypt resource: learning from Australia and elsewhere Ed. J. Walker. Wood Technology Research Centre, University of Canterbury, Christchurch, New Zealand.
Harwood, C. and P. Bulman. 2001. Genetic improvement strategy for *Eucalyptus cladocalyx* in southern Australia. *In* Australian Low Rainfall Tree Improvement Group: compendium of hardwood breeding strategies. Rural Industries Research and Development Corporation, Canberra, pp 45-66.
Harwood, C.E., R. Bird, T. Butcher, D.J. Bush, T. Jackson, I. Johnson, D. Stackpole and M. Underdown. 2005. Update of Australian Low Rainfall Tree Improvement Group hardwood breeding strategies. RIRDC Publication No. 05/023. Canberra. 17 p.
Harwood, C.E., D.J. Bush, T. Butcher, R. Bird, M. Henson, R. Lott and S. Shaw. 2007. Achievements in forest tree genetic improvement in Australia and New Zealand. 4. Tree improvement for low-rainfall farm forestry. Australian Forestry. 70:23-27.
Hedrick, P.W. and C.C. Cockerham. 1986. Partial inbreeding: equilibrium heterozygosity and the heterozygosity paradox. Evolution. 40:856-861.
Henderson, C.R. 1949. Estimates of changes in herd environment. Journal of Dairy Science. 32:706.
Henderson, C.R. 1953. Estimation of variance and covariance components. Biometrics. 9:226-252.
Henderson, C.R. 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics. 31:423-447.
Henderson, C.R. 1976. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. Biometrics. 32:69-83.
Henson, M., H.J. Smith and S. Boyton. 2007. *Eucalyptus longirostrata*: a potential species for Australia's tougher sites? *In* Australasian Forest Genetics Conference: breeding for wood quality, Hobart, Tasmania.
Hill, W.G. and B.S. Weir. 2011. Variation in actual relationship as a consequence of Mendelian sampling and linkage. Genetics Research. 93:47-64.
Hillis, W.E. 1987. Heartwood and tree exudates. Springer series in wood science. Springer-Verlag, Berlin. 268 p.
Hobbs, T.J., M. Bennell, D. Huxtable, J. Bartle, C. Nuemann, N. George, W. O'Sullivan and D. McKenna. 2008. Potential agroforestry species and regional industries for lower rainfall southern Australia: Florasearch 2: Report to the Joint Venture Agroforestry Program (JVAP) and the Future Farm Industries CRC. RIRDC, Canberra.

Hodge, G.R., P.W. Volker, B.M. Potts and J.V. Owen. 1996. A comparison of genetic information from open-pollinated and control-pollinated progeny tests in two eucalypt species. Theoretical and Applied Genetics. 92:53-63.

Hodgson, L.M. 1976. Some aspects of flowering and reproductive behaviour in *Eucalyptus grandis* (Hill) Maiden at J.D.M. Keet Forest Research Station. 2. The fruit, seed, seedlings, self fertility, selfing and inbreeding effects. South African Forestry Journal. 98:32-43.

Hope, G.S. 1994. Quaternary vegetation. *In* History of the Australian vegetation: Cretaceous to Recent. Cambridge University Press, Cambridge, pp 368-389.

Isik, F., S. Keskin and S.E. Mckeand. 2000. Provenance variation and provenance-site interaction in *Pinus brutia* TEN.: Consequences of defining breeding zones. Silvae Genetica. 49:213-223

Jacobs, M.R. 1979. Eucalypts for planting. FAO, Rome. 677 p.

Jarvis, S., N. Borralho and B. Potts. 1995. Implementation of a multivariate BLUP model for genetic evaluation of *Eucalyptus globulus* in Australia. *In* Eucalypt plantations: Improving Fibre Yield and Quality, Proc. CRCTHF-IUFRO Conference Eds. B. Potts, N. Borralho, J. Reid, R. Cromer, W. Tibbits and C. Raymond. CRC for Temperate Hardwood Forestry, Hobart, Tasmania, pp 212-216.

Jayawickrama, K.J.S. 2001. Genetic parameter estimates for radiata pine in New Zealand and New South Wales: a synthesis of results. Silvae Genetica. 50:45-53.

Jayawickrama, K.J.S. and M.J. Carson. 2000. A breeding strategy for the New Zealand Radiata Pine Breeding Cooperative. Silvae Genetica. 49:82-90.

Johnson, I.G. 1989. The breeding strategy for radiata pine in New South Wales Part II - proposed operations, Technical Paper No 47. . Forestry Commission of New South Wales.

Kanzler, A. 2002. Genotype x environment interaction in *Pinus patula* and its implications in South Africa. *In* Department of Forestry. North Carolina State University, Raleigh, NC, p 229.

Kelly, A.M., A.B. Smith, J.A. Eccleston and B.R. Cullis. 2007. The accuracy of varietal selection using factor analytic models for multi-environment plant breeding trials. Crop Science. 47:1063-1070.

Kelly, J.K. 2004. Family level inbreeding depression and the evolution of plant mating systems. New Phytologist. 165:55-62.

Kempton, R.A. 1984. The use of biplots in interpreting variety by environment interactions. Journal of Agricultural Science. 103:123-135.

King, J.P. 1965. Seed source x environment interactions in Scotch pine. Silvae Genetica. 14:141-8.

Kjaer, E.D., S. Kajornsrichon and E.B. Lauridsen. 1999. Heartwood, calcium and silica content in five provenances of teak (*Tectona grandis* L.). Silvae Genetica. 48:1-3.

Klápště, J., M. Lstibůrek and Y.A. El-Kassaby. 2013. Estimates of genetic parameters and breeding values from western larch open-pollinated families using marker-based relationship. Tree Genetics & Genomes. DOI 10.1007/s11295-013-0673-1

Kumar, S. and T.E. Richardson. 2005. Inferring relatedness and heritability using molecular markers in radiata pine. Molecular Breeding. 15:55-64.

Lai, B.S., T. Funda, C. Liewlaksaneeyanawin, J. Klápště, A.V. Niejenhuis, C. Cook, M.U. Stoehr, Jack Woods and Y.A. El-Kassaby. 2010. Pollination dynamics in a Douglas-fir seed orchard as revealed by pedigree reconstruction. Annals of Forest Science. 67 808

Lambeth, C., B.-C. Lee, D. O'Malley and N. Wheeler. 2001. Polymix breeding with parental analysis of progeny: an alternative to full-sib breeding and testing. Theoretical and Applied Genetics. 103:930-943.

Lorenzo, D., M. Troya, M. Prieto, C. Baso and M. Touza. 2007. Study of the natural durability of Spanish *Eucalyptus globulus* wood. The International Research Group on Wood Preservation. IRG Document No. IRG/WP 07-10617

Luikart, G., P.R. England, D. Tallmon, S. Jordan and P. Taberlet. 2003. The power and promise of population genomics: from genotyping to genome typing. Nature Reviews Genetics. 4:981-994.

Lynch, M. and B. Walsh. 1998. Genetics and analysis of quantitative traits. Sinauer Associates Inc., Sunderland, Massachusetts, USA. 980 p.

Matheson, A.C. and P.P. Cotterill. 1990. Utility of genotype x environment interactions. Forest Ecology and Management. 30:159-174.

Matheson, A.C. and C.A. Raymond. 1984. The impact of genotype x environment interactions on Australian *Pinus radiata* breeding programs. Australian Forestry Research. 14:11-25.

Matheson, A.C. and C.A. Raymond. 1986. A review of provenance x environment interaction: its practical importance and use with particular reference to the tropics. Commonwealth Forestry Review. 65:283-302

McDonald, M.W., M. Rawlings, P.A. Butcher and J.C. Bell. 2003. Regional divergence and inbreeding in *Eucalyptus cladocalyx* (Myrtaceae). Australian Journal of Botany. 51:393-403.

Meyer, K. 2009. Factor-analytic models for genotype x environment type problems and structured covariance matrices. Genetics Selection Evolution. 41:21 doi:10.1186/1297-9686-41-21.
Míguez-Soto, B. and J. Fernández-López. 2012. Genetic parameters and predicted selection responses for timber production traits in a *Castanea sativa* progeny trial: developing a breeding program. Tree Genetics & Genomes. 8:409-423.
Mimura, M., R.C. Barbour, B.M. Potts, R.E. Vaillancourt and K.N. Watanabe. 2009. Comparison of contemporary mating patterns in continuous and fragmented *Eucalyptus globulus* native forests. Molecular Ecology. 18:4180-4192.
Mitton, J. 1993. Theory and data pertinent to the relationship between heterozygosity and fitness. *In* The Natural History of Inbreeding and Outbreeding Ed. N. Thornhill. University of Chicago Press, Chicago.
Mohareb, A., P. Sirmah, L. Desharnais, S. Dumarçay, M. Pétrissans and P. Gérardin. 2010. Effect of extractives on conferred and natural durability of *Cupressus lusitanica* heartwood. Annals of Forest Science. 67:504-504.
Mora, F., R. Gleadow, S. Perret and C.A. Scapim. 2009. Genetic variation for early flowering, survival and growth in sugar gum (*Eucalyptus cladocalyx* F. Muell) in southern Atacama Desert. Euphytica. 169:335-344.
Moran, G., J. Bell and A. Griffin. 1989. Reduction in levels of inbreeding in a seed orchard of *Eucalyptus regnans* F. Muell. compared with natural populations. Silvae Genetica. 38:32-36.
Moran, G.F. and J.C. Bell. 1987. The origin and genetic diversity of *Pinus radiata* in Australia. Theoretical and Applied Genetics. 73:616-622.
Morgenstern, E.K. and A.H. Teich. 1969. Phenotypic stability of height growth of jack pine provenances. Canadian Journal of Genetical Cytology. 11:110 - 117.
Mosedale, J.R., B. Charrier and G. Janin. 1996. Genetic control of wood colour, density and heartwood ellagitannin concentration in European oak (*Quercus petraea* and *Q. robur*). Forestry. 69:111-124.
Mrode, R.A. 1996. Linear models for the prediction of animal breeding values. CAB International, Oxford, p 187.
Munoz, P.R., M.F.R. Resende, D.A. Huber, T. Quesada, M.D.V. Resende, D.B. Neale, J.L. Wegrzyn, M. Kirst and G.F. Peter. 2014. Genomic Relationship Matrix for Correcting Pedigree Errors in Breeding Populations: Impact on Genetic Parameters and Genomic Selection Accuracy. Crop Science. 54:1115-1123.
Murray Darling Basin Commission. 2007. River Murray System Drought Update No. 11, December 2007. Murray Darling Basin Commission, Canberra, p 6.
Myburg, A., D. Grattapaglia, G. Tuskan, J. Jenkins, J. Schmutz, E. Mizrachi, C. Hefer, G. Pappas, L. Sterck, Y.V.D. Peer, R. Hayes and D. Rokhsar. 2011. The *Eucalyptus grandis* Genome Project: genome and transcriptome resources for comparative analysis of woody plant biology. BMC Proceedings. 5:I20
Namkoong, G. 1966. Inbreeding effects on estimation of genetic additive variance. Forest Science. 12:8-13.
Namkoong, G., R.D. Barnes and J. Burley. 1980. Screening for yield in forest tree breeding. Commonwealth Forestry Review. 59:61-68.
Namkoong, G., H. Kang and J. Brouard. 1988. Tree breeding: principles and strategies. Springer-Verlag, NY.
National Land & Water Resources Audit. 2008. Land salinity - status of information for reporting against indicators under the National Natural Resource Management Monitoring and Evaluation Framework. NLWRA, Canberra. 45 p.
Ng, M.P. and E.R. Williams. 2001. Joint-regression analysis for incomplete two-way tables. Australian and New Zealand Journal of Statistics. 43:201-206.
Nicholls, J.W.P. and A.C. Matheson. 1980. Variation in wood characteristics in thinnings from a field trial of *Eucalyptus obliqua*. Australian Forest Research. 10:239-247.
Nicolle, D. 2013. Native eucalypts of South Australia. Dean Nicolle, Adleaide. 237 p.
Nybom, H. 2004. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Molecular Ecology. 13:1143-1155.
Ogut, F., C. Maltecca, R. Whetten, S. McKeand and F. Isik. 2014. Genetic analysis of diallel progeny test data using factor analytic linear mixed models. Forest Science. 60:119-127.
Owino, F., R.C. Kellison and B.J. Zobel. 1977. Genotype x environment interaction and genotypic stability in loblolly pine. V. Effects of genotype x environment interaction on genetic variance component estimates and gain predictions. Silvae Genetica. 26:131-134.
Palanti, S., D. Susco and E. Feci. 2010. Natural durability of eucalypt from Italian plantations against fungi and cerambicid *Trichoferus holosericeus* Rossi. European Journal of Wood and Wood Products. 68:59-62.
Paques, L.E. 2001. Genetic control of heartwood content in larch. Silvae Genetica. 50:69-75.
Patterson, H. and R. Thompson. 1971. Recovery of interblock information when blocks sizes are unequal. Biometrika. 31:100-109.

Pelletier, M., M. Henson, S. Boyton, D. Thomas and J. Vanclay. 2008. Genetic variation in shrinkage properties of *Eucalyptus pilularis* assessed using increment cores and test blocks. New Zealand Journal of Forestry Science. 38:194-210.

Pemberton, J.M. 2008. Wild pedigrees: the way forward. Proceedings of the Royal Society of London B: Biological Sciences. 275:613-621.

Piepho, H.-P. 1998. Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. Theoretical and Applied Genetics. 97:195-201.

Poke, F.S., B.M. Potts, R.E. Vaillancourt and C.A. Raymond. 2006. Genetic parameters for lignin, extractives and decay in *Eucalyptus globulus*. Annals of Forest Science. 63:813-821.

Poke, F.S., J.K. Wright and C.A. Raymond. 2005. Predicting extractives and lignin contents in *Eucalyptus globulus* using near infrared reflectance analysis Journal of Wood Chemistry and Technology. 24:55 - 67.

Potts, B. and J. Wiltshire. 1997. Eucalypt genetics and genecology. *In* Eucalypt ecology: individuals to ecosystems Eds. J. Williams and J. Woinarski. Cambridge University Press, Cambridge, pp 56-91.

Powell, M.B., T.A. McRae, H.X. Wu, G.W. Dutkowski and D.J. Pilbeam. 2004. Breeding Strategy for *Pinus radiata* in Australia. *In* IUFRO Joint Conference of Division 2: Forest Genetics and Tree Breeding in the Age of Genomics: Progress and Future, Charleston, South Carolina, USA.

Pryor, L.D. 1976. The Biology of the eucalypts. Institute of Biology's Studies in Biology, No. 16. Arnold, London. 82 p.

Raymond, C.A., M. Henson, M.-C. Pelletier, S. Boyton, B. Joe, D. Thomas, H. Smith and J. Vanclay. 2008. Improving dimensional stabilty in plantation-grown *Eucalyptus pilularis* and *E. dunnii*. Forest and Wood Products Australia.

Ritland, K. 1984. The effective proportion of self-fertilization with consanguineous matings in inbred populations. Genetics. 106:139-152.

Ritland, K. 2002. Extensions of models for the estimation of mating systems using *n* independent loci. Heredity. 88:221-228.

Ritland, K. 2008. Multilocus mating system program: MLTR (reference manual - computer program MLTR version 2.2 available from author at http://genetics.forestry.ubc.ca/ritland/programs.html University of British Colombia.

Rudman, P. 1966. Durability in the genus *Eucalyptus*. Australian Forestry. 28:242-257.

Schimleck, L.R., R. Evans and J. Ilic. 2001. Estimation of *Eucalyptus delegatensis* wood properties by near infrared spectroscopy. Canadian Journal of Forest Research. 31:1671-1675.

Schmidtling, R.C. and T.L. Amburgey. 1982. Genetic variation in decay susceptibility and its relationship to growth and specific gravity in loblolly pine. Holzforschung. 36:159-161.

Schwerdtfeger, P. 1985. Chapter 6: Climate. *In* Natural history of the Eyre Peninsula Eds. C.R. Twidale, M.J. Tyler and M. Davies. Royal Society of South Australia, Adelaide, pp 89-100.

Schwerdtfeger, P. and E. Curran. 1996. Chapter 6: Climate of the Flinders Ranges. *In* Natural history of the Flinders Ranges Eds. M. Davies, C.R. Twidale and M.J. Tyler. Royal Society of South Australia, Inc., Adelaide, pp 63-75.

Searle, S.R. 1968. Another look at Henderson's methods of estimating variance components. Biometrics. 24:749-778.

Shaw, D.V. and R.W. Allard. 1981. Analysis of mating system parameters and population structure in Douglas-fir using single-locus and multilocus methods. *In* General Technical Report PSW-48: Proceedings of the Symposium on Isozymes of North American Forest Trees and Forest Insects Ed. M.T. Conkle. United States Department of Agriculture Forest Service, Pacific Southwest Forest and Range Experiment Station, Berkeley, California, pp 18-22.

Shelbourne, C.J.A. 1972. Genotype-environment interaction: its study and its implications in forest tree improvement. *In* Proceedings of the joint symposia for the advancement of forest tree breeding of Genetics Subject Group, International Union of Forest Research Organizations and Section 5, Forest Trees, Society for the Advancement of Breeding Researches in Asia and Oceania. Government Forest Experiment Station of Japan, Tokyo, p 28.

Silva-Junior, O.B., D.A. Faria, R.C. Togawa and D. Grattapaglia. 2013. *Eucalyptus* genotyping taken to the next level: development of the " EucHIP60k.br" based on large scale multi-species SNP discovery and ascertainment. *In* IUFRO Tree Biotechnology Conference, Asheville, NC.

Skabo, S., R.E. Vaillancourt and B.M. Potts. 1998. Fine-scale genetic structure of *Eucalyptus globulus* ssp. *globulus* forest revealed by RAPDs. Australian Journal of Botany. 46:583-594.

Slate, J., P. David, K. Dodds, B. Veenvliet, B. Glass, T. Broad and J. McEwan. 2004. Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. Heredity. 93:255–265.

Smith, A., B. Cullis and R. Thompson. 2001. Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. Biometrics. 57:1138-1147.
Smith, A., B. Cullis and R. Thompson. 2005. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. Journal of Agricultural Science. 143:449-462.
Smith, A.B. 1999. Multiplicative mixed models for the analysis of multi-environment trial data. *In* Department of Statistics. The University of Adelaide, Adelaide.
Spearman, C. 1904. The proof and measurement of association between two things. American Journal of Psychology. 15:72-101
Specht, R.L. 1972. The vegetation of South Australia: handbook of the flora and fauna of South Australia. Government Printer, Adelaide. 328 p.
Squillace, A.E. 1974. Average genetic correlations among offspring from open-pollinated forest trees. Silvae Genetica. 23:149-156.
Standards Australia. 2003. Methods of test for pulp and paper: Method 012s - organic solvent extractives in wood, pulp and paper. Standards Australia, Homebush, NSW, p 12.
Standards Australia. 2005. Timber-Natural durability ratings. AS 5604-2005. Standards Australia, Homebush, New South Wales.
Steane, D., D. Nicolle, C. Sansaloni, C. Petroli, J. Carling, A. Kilian, A. Myburg, D. Grattapaglia and R. Vaillancourt. 2011. Population genetic analysis and phylogeny reconstruction in *Eucalyptus* (Myrtaceae) using high-throughput, genome-wide genotyping. Molecular Phylogenetics and Evolution. 59 206-224.
Stern, N. 2007. Stern review: the economics of climate change. Cambridge University Press, Cambridge, UK.
Surles, S.E., J. Arnold, A. Schnabel, J.L. Hamrick and B.C. Bongarten. 1990. Genetic relatedness in open-pollinated families of two leguminous tree species, *Robinia pseudoacacia* L. and *Gleditsia tricanthos* L. Theoretical and Applied Genetics. 80:49.56.
Swain, T.-L. and A.K. Louw. 2011. Genetic parameter estimates and parental selection in *Eucalyptus longirostrata* and *Corymbia henryi* populations tested in South Africa Silvae Genetica. 60:105-113.
Taft, H.R. and D.A. Roff. 2012. Do bottlenecks increase additive genetic variance? Conservation Genetics. 13:333-342.
Takezaki, N., M. Nei and K. Tamura. 2010. POPTREE2: software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. Mol Biol Evol. 27:747-752.
TAPPI. 1989. Basic density and moisture content of pulpwood. TAPPI no. T258 om-98 (1989)
Thavamanikumar, S., L.J. McManus, J.F.G. Tibbits and G. Bossinger. 2011. The significance of single nucleotide polymorphisms (SNPs) in *Eucalyptus globulus* breeding programs. Australian Forestry. 74:23-29.
Thumma, B.R., M.F. Nolan, R. Evans and G.F. Moran. 2005. Polymorphisms in cinnamoyl CoA reductase (CCR) are associated with variation in microfibril angle in *Eucalyptus* spp. Genetics. 171:1257-1265.
United Nations. 2013. World population prospects: the 2012 revision (highlights and advance tables). Working Paper No. ESA/P/WP.228. United Nations, Department of Economic and Social Affairs, Population Division, New York, p 94.
Van Raden, P.M. 2007. Genomic measures of relatedness and inbreeding. Interbull Bulletin. 37:33-36.
Vargas-Reeve, F., F. Mora, S. Perret and C.A. Scapim. 2013. Heritability of stem straightness and genetic correlations in *Eucalyptus cladocalyx* in the semi-arid region of Chile. Crop Breeding and Applied Biotechnology. 13:107-112.
Visscher, P.M., S.E. Medland, M.A.R. Ferreira, K.I. Morley, G. Zhu, B.K. Cornes, G.W. Montgomery and N.G. Martin. 2006. Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. PLoS Genetics. 2:e41.
Volker, P.W., C.A. Dean, W.N. Tibbits and I.C. Ravenwood. 1990. Genetic parameters and gains expected from selection in *Eucalyptus globulus* in Tasmania. Silvae Genetica. 39:18-21.
Walker, J. 2011. Developing a eucalypt resource: learning from Australia and elsewhere. University of Canterbury, Canterbury, New Zealand. 158 p.
Wang, J. 2007. Triadic IBD coefficients and applications to estimating pairwise relatedness. Genet. Res., Camb. 89:135-153.
Wang, J. 2010. Coancestry: a program for simulating, estimating and analysing relatedness and inbreeding coefficients. Molecular Ecology Resources. 11:141-145.
Wang, J. and A.W. Santure. 2009. Parentage and sibship inference from multilocus genotype data under polygamy. Genetics. 181:1579-1594.
Washusen, R., P. Ades, R. Evans, J. Ilic and P. Vinden. 2001. Relationships between density, shrinkage, extractives content and microfibril angle in tension wood from three provenances of 10-year-old Eucalyptus globulus Labill. Holzforschung. 55 176-182.

Washusen, R. and G. Waugh. 2000. Opportunities for solid wood products from low-rainfall plantation-grown eucalypts. *In* The future of eucalypts for wood products, 19-24 March 2000. IUFRO, Launceston, Tasmania, p 381.

Westell, R.A., R.L. Quaas and L.D. Van Vleck. 1988. Genetic groups in an animal model. Journal of Dairy Science. 71:1310-1318.

Whitaker, D., E. Williams and J. John. 2002. Cycdesign: a package for the computer generation of experimental designs. 2.0 edn. CSIRO and University of Waikato Canberra, Australia and Hamilton, New Zealand.

White, T.L., W.T. Adams and D.B. Neale. 2007. Forest genetics. CABI Publishing, Wallingford, England. 682 p.

White, T.L. and G.R. Hodge. 1989. Predicting breeding values with applications in forest tree improvement. Springer, Dordrecht. 368 p.

Wilks, S.S. 1938. The large-sample distribution of the likelihood ratio for testing composite hypotheses. The Annals of Mathematical Statistics. 9:60-62.

Williams, E.R., A.C. Matheson and C.E. Harwood. 2002. Experimental design and analysis for tree improvement. CSIRO Publishing, Collingwood. 214 p.

Wong, A.H.H., J. Wilkes and W.A. Heather. 1983. Influence of wood density and extractive content on the decay of the heartwood of *Eucalyptus delegatensis* R.T. Baker Journal of the Institute of Wood Science. 54:261-263.

Wood, J.G. 1930. An analysis of the vegetation of Kangaroo Island and the adjacent peninsulas. Transactions of the Royal Society of South Australia. 54:105-139.

Wricke, G. 1962. Über eine methode zur erfassung der ökologischen streubreite in feldversuchen Zeitung für Pflanzenzüchtung. 47:92-96.

Wright, S. 1922. Coefficients of inbreeding and relationship. American Naturalist. 56:330-338.

Wu, H.X., K.G. Eldridge, A.C. Matheson, M.P. Powell, T.A. McRae, T.B. Butcher and I.G. Johnson. 2007. Achievements in forest tree improvement in Australia and New Zealand: 8. Successful importation and breeding of radiata pine to Australia. Australian Forestry. 70: 215-225.

Wu, H.X. and C.C. Ying. 2001. Variation on reaction norm in lodgepole pine natural populations. Theoretical and Applied Genetics. 103:331-345.

Yates, F. and W.G. Cochrane. 1938. The analysis of groups of experiments. Journal of Agricultural Science. 28:556-580.

Yu, Q., D.Q. Yang, S.Y. Zhang, J. Beaulieu and I. Duchesne. 2003. Genetic variation in decay resistance and its correlation to wood density and growth in white spruce. Canadian Journal of Forest Research. 33:2177-2183.

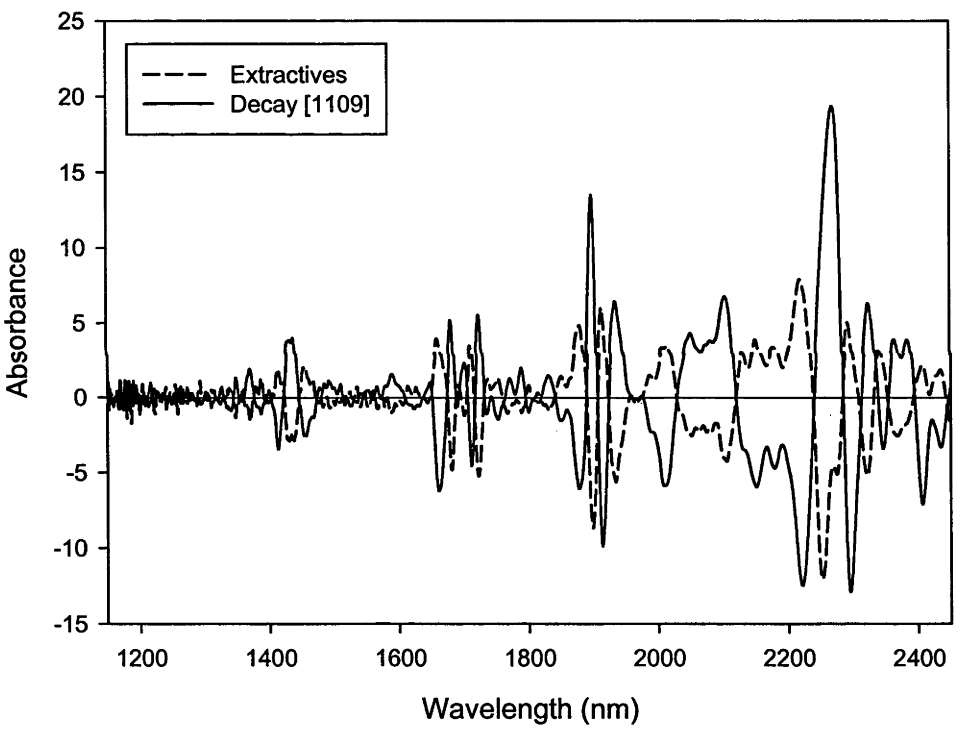
Zapata-Valenzuela, J., R.W. Whetten, D. Neale, S. McKeand and F. Isik. 2013. Genomic estimated breeding values using genomic relationship matrices in a cloned population of loblolly pine. G3: Genes|Genomes|Genetics. 3:909-916.

Appendix 1 Calibration plot for prediction of decay and extractives

Matter presented as Online resource 1 in Bush et al.(2011b)

The loadings plots for all the decay calibrations show two key wavelengths as being responsible for the correlation: 1910 nm (carboxylic acids, esters) and 2250 nm (possibly aldehydes). These wavelengths are consistent with those attributable to the correlation with extractives content (Plot. 1) confirming the role that extractives play in decay resistance.

Plot 1. Loadings plot for PC-1 for NIR prediction of extractives and decay [1109].



Appendix 2.1 Subpopulation/spatial data for the first-generation *E. cladocalyx* breeding population. Summary information for seedlots included in the ALRTIG breeding population at Lismore, Victoria

Material presented as Online Resource 1 in Bush and Thumma (2013)

ATSC seedlot	ATSC seedlot name	Locality or grouping	Latitude (deg.)	Longitude (deg.)	Elevation (m)	Forest stand type
South Flinders Ranges						
19348	Wilmington	Mt Remarkable	-32.70	138.11	550	Open forest
20388	Mt. Remarkable 9 km S. of Wilmington	Mt Remarkable	-32.73	138.10	580	Open forest
16089	Post Office	Mt Remarkable	-32.72	138.10	580	Open forest
20414	Wirrabara SF	Wirrabara	-33.08	138.18	480	Open forest
20268	Wirrabara SF	Wirrabara	-33.08	138.18	480	Open forest
Kangaroo Island						
20265	American River	E. Kangaroo Is.	-35.77	137.77	176	Open forest
20266	Cygnet River	E. Kangaroo Is.	-35.72	137.48	20	Open forest
16022	Flinders Chase NP 1	NW. Kangaroo Is.	-35.75	136.63	80	Open forest
20267	Flinders Chase NP 2	SW. Kangaroo Is.	-35.95	136.70	60	Open forest
19717	Flinders Chase NP 2	SW. Kangaroo Is.	-35.95	136.73	70	Open forest
Cultivated stands						
N/A	Kersbrook SPA	Adelaide Hills	-34.73	138.83	360	Seed production area
N/A	Mt. Burr	S.E. S. Australia	-35.83	143.07	100	Block planting
21044	Wail	W. Victoria	-36.50	142.07	110	Block planting
21043	Majorca	W. Victoria	-37.12	143.80	230	Block planting
21042	Lismore	W. Victoria	-38.40	142.57	20	Shelter belt

N/A not applicable, *SF* State Forest, *NP* National Park, *SPA* seed production area

Appendix 2.2 Description of genes (*n*=39) and SNPs (*n*=75) located within chromosome scaffolds (1-10) and a minor scaffold, as yet unallocated to a chromosome (scaffold 12), with annotation from the *E. grandis* genome (Phytozome Version 7). Gene names are either from known *E. nitens* sequences and/or based on similarity to *Arabidopsis thaliana* gene sequences. SNP descriptors are as follows: U=unannotated *E. grandis* region, I=intron, E=unannotated exon, S=synonymous, NS=non-synonymous, P=probable promoter (from within 1kb of the 5'region of a gene).

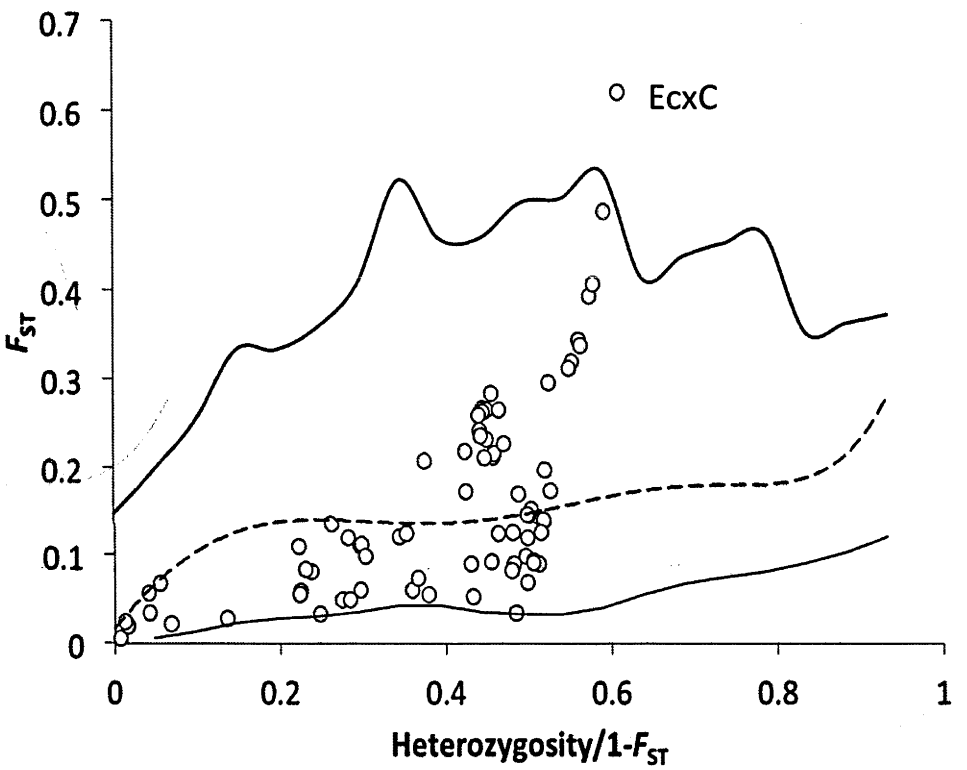
GenBank accession	Gene name	Similar *Arabidopsis thaliana* gene	Function of similar gene	*E. grandis* scaffold	*E. grandis* gene annotation	Function	SNP count & descriptor
KC160180	EcxSAM2	AT4G01850.1	S-adenosylmethionine synthetase 2	1	Eucgr.A01021	1	3(S)
KC160143	EcxCSLA9	AT3G56000.1	Cellulose synthase-like	1	Eucgr.A01558	2	2(I)
KC160180	EcxZNF1	AT1G73760.1	RING/U-box superfamily protein	2	Eucgr.B00829	6	1(I)
KC160176	EcxMLP28	AT1G70830.1	Major latex protein-like	2	Eucgr.B02925	1	1(I)
KC160159	EcxSPL2	AT5G43270.2	Squamosa promoter binding protein	2	Eucgr.B03500	4	1(I)
KC160158	EcxCNX1	AT5G07340.2	Calreticulin family protein	2	Eucgr.B03570	8	1(S)
KC160151	EcxLRRK2	AT2G36570.1	Leucine-rich receptor-like, kinase family protein	3	Eucgr.C00351	1	1(S)
KC160150	EcxHB8	AT2G34710.1	Homeobox-leucine zipper family protein / lipid-binding START domain-containing protein	3	Eucgr.C00605	6	1(I)
KC160168	EcxLRRK	AT1G67510.1	Leucine-rich receptor-like protein kinase family protein	3	Eucgr.C00960	4	1(NS),2(S)
KC160180	EcxVAMP7	AT5G22360.1	Vesicle-associated membrane protein	3	Eucgr.C02758	3	2(I)
KC160158	EcxCESA6	AT5G64740.1	Cellulose synthase 6	3	No annotation	2	2(U)
KC160165	EcxRABA	AT5G45750.1	RAB GTPase homolog	4	Eucgr.D00474	8	1(I)
KC160154	EcxCESA	AT4G18780.1	Cellulose synthase-like	4	Eucgr.D00476	2	4(S)
KC160174	EcxNovel		No close similiarity identified	4	Eucgr.D00708	8	1(I)
KC160180	EcxUPF	AT2G35760.1	Uncharacterised protein family (UPF0497)	4	Eucgr.D01440	6	3(I),2(P)
KC160174	EcxNAC	AT1G25580.1	NAC domain-containing protein	4	Eucgr.D01671	6	1(P)
KC160158	EcxACT	AT5G09810.1	Actin subclass	4	Eucgr.D02537	3	1(S),1(I)
KC160159	EcxNOD	AT4G10380.1	NOD26-like intrinsic protein	4	Eucgr.D02548	5	1(S)
KC160160	Ecx cytochrome	AT3G61750.1	Cytochrome b561/ferric reductase transmembrane with	5	Eucgr.E00054	8	1(S),1(I)

GenBank accession	Gene name	Similar *Arabidopsis thaliana* gene	Function of similar gene	*E. grandis* scaffold	*E. grandis* gene annotation	Function	SNP count & descriptor
	B561		DOMON related domain				
KC160171	EcxLIM2	AT1G10200.1	LIM domain/GATA type zinc finger transcription	6	Eucgr.F00087	1	1(S),1(I)
KC160171	EcxCIP	AT4G27430.1	COP1-interacting protein	6	Eucgr.F01094	6	1(NS),1(E),1(U)
KC160180	EcxUP3	AT5G42710.1	Unknown protein	6	Eucgr.F01564	8	1(S),1(I)
KC160180	EcxWAT1	AT1G75500.1	Walls Are Thin 1/nodulin	6	Eucgr.F01847	3	1(S)
KC160165	EcxC3HC4Z1	AT4G32600.1	Ring/U-box superfamily protein	6	Eucgr.F02028	6	1(S),1(I)
KC160172	EcxTUA1	AT1G04820.1	Tubulin alpha-4 chain	6	Eucgr.F02183	2	2(S),1(I)
KC160167	EcxLIM1	AT1G10200.1	LIM domain protein	6	Eucgr.F02243	1	2(I)
KC160174	EcxHB1	AT1G52150.1	Alpha/beta-hydrolases superfamily protein	6	Eucgr.F03066	7	1(I)
KC160169	EcxCBS	AT4G27460.1	Cystathionine beta-synthase (CBS) family protein	6	Eucgr.F03335	8	1(NS)
KC160170	EcxK1B	AT3G44730.1	Kinesin motor protein/NHL domain-containing protein	6	Eucgr.F03668	2	1(S)
KC160171	EcxCCoAOMT2	AT4G32150.1	Methyltransferase superfamily protein	7	Eucgr.G01417	1	1(I),3(U)
KC160176	EcxMYB83	AT4G38620.1	MYB domain protein	7	Eucgr.G01774	1	1(U)
KC160180	EcxUP15	None	Unknown protein	7	Eucgr.G03122	8	1(S),1(S)
KC160174	EcxCel2	AT4G02290.1	Cellulase	8	Eucgr.H01095	2	1(I)
KC160175	EcxSPDSY	AT5G19530.2	Spermidine synthase (ACAULIS 5)	8	Eucgr.H03531	8	1(NS),1(S)
KC160176	EcxDehydRP	AT3G10200.1	Methyltransferases superfamily protein	9	Eucgr.I00186	5	2(S)
KC160180	EcxUP7	AT4G24530.1	O-fucosyltransferase family protein	9	Eucgr.I00557	6	1(I)
KC160180	EcxSKU5	AT2G23630.1	SKU5 similar	9	Eucgr.I00988	4	1(S),4(I)
KC160179	EcxFAH1	AT4G36220.1	Ferulic acid 5-hydroxylase 1	10	Eucgr.J02393	1	3(I)
KC160180	EcxHPA	AT5G10330.3	Histidinol phosphate aminotransferase	12	Eucgr.L00028	8	1(I)
Total							75 SNPs from 39 genes

1) lignin biosynthesis and deposition; 2) cellulose biosynthesis and deposition; 3) cell wall structure; 4) cell expansion other; 5) plant water relations; 6) transcription factor; 7) vascular development and; 8) other function

Appendix 2.3 Detection of SNPs under selection in *E. cladocalyx* first-generation breeding population using the method of Excoffier et al. (2009) implemented in Arlequin; a hierarchical island model with 3 groups and 14 subpopulations simulated. Solid blue lines delineate the 99% confidence interval envelope. The SNP above the upper bound (from the EcxCBS gene) is an outlier likely to be under selection, or possibly the result of other processes such as a genetic bottleneck or founder effect.

Material presented as Online Resource 3 in Bush and Thumma (2013)



Appendix 2.4 Family-average relatedness estimates between pairs of trees within families of *E. cladocalyx*. $E(2\theta|\text{HS})=0.25$, $E(f|\text{HS})=0$. KI=Kangaroo Island, SFR=South Flinders Ranges, PS=planted stand. Values given are not corrected for bias (see Table 4.3).

Material presented as Online Resource 4 in Bush and Thumma (2013)

Group	Subpopulation	Family	*n* (trees)	$\overline{2\theta}_f$	$\overline{f}_f$
KI	American River	1	9	0.67	0.24
KI	American River	2	20	0.94	0.47
KI	American River	3	10	0.52	0.19
KI	American River	4	11	0.47	0.23
KI	American River	5	10	0.64	0.22
KI	American River	6	12	0.41	0.17
KI	Cygnet River	7	11	0.67	0.09
KI	Cygnet River	8	7	0.56	0.27
KI	Cygnet River	9	8	0.42	0.51
KI	Cygnet River	10	10	0.35	0.00
KI	Cygnet River	11	11	0.64	0.14
KI	Flinders Chase 1	12	7	0.86	0.38
KI	Flinders Chase 1	13	8	0.67	0.28
KI	Flinders Chase 1	14	8	0.60	0.28
KI	Flinders Chase 1	15	9	0.78	0.21
KI	Flinders Chase 2	16	10	0.60	0.28
KI	Flinders Chase 2	17	11	0.52	0.20
KI	Flinders Chase 2	18	13	0.59	0.31
KI	Flinders Chase 2	19	18	0.31	0.16
KI	Flinders Chase 2	20	9	0.68	0.18
KI	Flinders Chase 2	21	4	0.33	0.29
KI	Flinders Chase 2	22	9	0.46	0.22
KI	Flinders Chase 2	23	11	0.77	0.28
KI	Flinders Chase 2	24	10	0.45	0.18
KI	Flinders Chase 2	25	8	0.85	0.34
KI	Flinders Chase 2	26	7	0.67	0.28
KI	Flinders Chase 2	27	9	0.55	0.24
KI	Flinders Chase 2	28	9	0.67	0.26
KI	Flinders Chase 2	29	10	1.00	0.44
KI	Flinders Chase 2	30	10	0.79	0.34
KI	Flinders Chase 2	31	10	0.57	0.19
KI	Flinders Chase 2	32	10	0.46	0.24
KI	Flinders Chase 2	33	9	0.59	0.18
KI	Flinders Chase 2	34	9	0.71	0.29
KI	Flinders Chase 2	35	10	0.44	0.17
KI	Flinders Chase 2	36	10	0.33	0.15
PS	Kersbrook	37	10	0.20	0.07
PS	Kersbrook	38	10	0.19	0.08
PS	Kersbrook	39	10	0.34	0.10
PS	Kersbrook	40	10	0.27	0.10
PS	Kersbrook (KI mum)	41	5	0.41	0.10
PS	Kersbrook	42	9	0.26	0.09
PS	Kersbrook	43	13	0.23	0.13
PS	Kersbrook	44	10	0.29	0.14
PS	Kersbrook	45	12	0.22	0.08
PS	Kersbrook	46	15	0.25	0.08
PS	Kersbrook	47	11	0.24	0.10
PS	Kersbrook	48	10	0.66	0.28
PS	Kersbrook	49	10	0.19	0.06
PS	Kersbrook	50	11	0.35	0.07
PS	Kersbrook	51	10	0.29	0.12
PS	Lismore	52	10	0.30	0.06
PS	Lismore	53	10	0.19	0.11
PS	Lismore	54	10	0.16	0.09
PS	Majorca	55	12	0.35	0.07
PS	Majorca	56	10	0.25	0.09
PS	Majorca	57	12	0.3	0.13
PS	Majorca	58	15	0.25	0.08

Group	Subpopulation	Family	n (trees)	$\overline{2\theta_f}$	$\overline{f_f}$
PS	Majorca	59	11	0.3	0.13
PS	Mt. Burr	60	8	0.31	0.12
PS	Mt. Burr	61	10	0.34	0.02
PS	Mt. Burr	62	13	0.24	0.06
PS	Mt. Burr	63	9	0.2	0.06
PS	Mt. Burr	64	12	0.21	0.04
	Wail	65	9	0.38	0.14
PS	Wail	66	10	0.24	0.09
PS	Wail	67	9	0.21	0.08
SFR	Mt Remarkable	68	5	0.43	0.11
SFR	Mt Remarkable	69	13	0.23	0.09
SFR	Mt Remarkable	70	10	0.33	0.11
SFR	Mt Remarkable	71	11	0.31	0.08
SFR	Mt Remarkable	72	11	0.24	0.1
SFR	Mt Remarkable	73	10	0.27	0.1
SFR	Mt Remarkable	74	9	0.27	0.11
SFR	Mt Remarkable	75	10	0.25	0.11
SFR	Mt Remarkable	76	14	1.04	0.38
SFR	Mt Remarkable	77	10	0.46	0.27
SFR	Mt Remarkable	78	10	0.46	0.2
SFR	Mt Remarkable	79	10	0.53	0.21
SFR	Mt Remarkable	80	10	0.38	0.08
SFR	Mt Remarkable	81	14	0.34	0.11
SFR	Mt Remarkable	82	10	0.29	0.14
SFR	Mt Remarkable	83	9	0.60	0.19
SFR	S. Wilmington	84	8	0.26	0.14
SFR	S. Wilmington	85	10	0.40	0.12
SFR	S. Wilmington	86	9	0.17	0.10
SFR	S. Wilmington	87	17	0.38	0.15
SFR	S. Wilmington	88	9	0.33	0.14
SFR	S. Wilmington	89	9	0.32	0.2
SFR	S. Wilmington	90	7	0.46	0.11
SFR	Wilmington	91	10	0.29	0.12
SFR	Wilmington	92	10	0.30	0.10
SFR	Wilmington	93	9	0.70	0.26
SFR	Wilmington	94	10	0.27	0.09
SFR	Wilmington	95	10	0.64	0.24
SFR	Wilmington	96	10	0.31	0.10
SFR	Wilmington	97	9	0.80	0.36
SFR	Wirrabara	98	10	0.31	0.19
SFR	Wirrabara	99	7	0.23	0.18
SFR	Wirrabara	100	10	0.50	0.24
SFR	Wirrabara	101	10	0.21	0.09
SFR	Wirrabara	102	10	0.38	0.18
SFR	Wirrabara	103	9	0.42	0.14
SFR	Wirrabara	104	9	0.34	0.17
SFR	Wirrabara	105	7	0.27	0.08
SFR	Wirrabara	106	10	0.49	0.14
SFR	Wirrabara	107	8	0.26	0.08
SFR	Wirrabara	108	10	0.38	0.10
SFR	Wirrabara	109	10	0.37	0.14
SFR	Wirrabara	110	9	0.32	0.2
SFR	Wirrabara	111	10	0.32	0.11
SFR	Wirrabara	112	9	0.38	0.11
SFR	Wirrabara	113	14	0.45	0.14
SFR	Wirrabara	114	9	0.32	0.12
SFR	Wirrabara	115	12	0.33	0.16
SFR	Wirrabara	116	10	0.37	0.09
SFR	Wirrabara	117	11	0.49	0.15
SFR	Wirrabara	118	12	0.46	0.14
SFR	Wirrabara	119	10	0.35	0.11
	Count	**119**	**1202**	**119**	

Appendix 3.1 Distribution of families among sites from two wild regions of provenance (Kangaroo Island and South Flinders Ranges, South Australia) and cultivated stands in Victoria and South Australia. Sites are *MOR* Moora, *KOJ* Kojonup, *WEL* Wellstead, *ESP* Esperance (Western Australia); *AVE* Avenue, *BTN* Bordertown (South Australia); *HAM* Hamilton, *LIS* Lismore (Victoria); *COR* Corowa, *WGW* Wagga Wagga (New South Wales).

Material presented as Electronic Supplementary Matter 1 in Bush et al. (2015)

		Site											
Provenance	ATSC seedlot	MOR	KOJ	WEL	ESP1	ESP2	AVE	BTN	HAM	LIS	COR	WGW	Total
Kangaroo Island													
American River, SA	20265	6	2	2	4	2	4	6	5	6	5	6	6
Cygnet River, SA	20266	5	3	3	1	3	3	7	3	5	2	4	7
Flinders Chase, SA	16022	3	-	-	-	-	2	8	2	4	3	3	8
Flinders Chase, SA	19717	9	3	3	7	3	5	11	8	9	9	9	12
Flinders Chase , SA	20267	12	4	4	7	4	5	12	11	12	9	11	12
South Flinders Ranges													
Mt. Remarkable, SA	20388	16	2	2	15	2	12	16	16	16	16	14	16
Wilmington, SA	19348	7	1	1	6	1	1	8	4	7	-	7	9
Wirrabara (1), SA	20268	15	3	3	10	3	5	16	10	16	4	13	17
Wirrabara (2), SA	20414	6	2	2	5	2	5	6	5	6	6	6	6
S. Wilmington, SA	16089	7	1	1	-	1	2	9	5	7	2	6	10
Cultivated stands													
Wail, VIC		3	-	-	-	-	-	5	3	3	3	2	5
Kersbrook, SA	-	15	10	10	-	10	10	15	13	15	15	15	15
Lismore, VIC	21042	3	2	2	3	2	1	3	1	3	1	3	4
Majorca, VIC	21043	5	5	5	3	5	4	5	5	5	5	5	5
Mt Burr, SA	-	5	4	4		4	5	5	5	5	4	5	5
Total seed sources		15	13	13	10	13	14	15	15	15	14	15	15
Total families		118	42	42	61	42	64	133	96	119	84	109	137

Appendix 3.2a Table of predicted DBH means and fixed effect significance from single-site and across-sites (FA2) analyses

	HS model predicted DBH means (mm)											MM models predicted DBH means (mm)											
Group/ provenance	MOR	KOJ	WEL	ESP1	ESP2	AVE	BTN	HAM	LIS	COR	WGW	MOR	KOJ	WEL	ESP1	ESP2	AVE	BTN	HAM	LIS	COR	WGW	Across sites
American River	47	107	106	58	131	88	76	85	76	61	35	47	107	106	59	130	88	76	85	76	61	35	83
Cygnet River	54	133	119	68	151	101	83	88	79	66	43	55	134	120	68	151	101	83	89	79	66	43	89
Flinders Chase	55					93	85	92	85	77	46	55					94	86	92	85	78	47	94
Flinders Chase NP	55	118	106	62	132	92	81	90	78	62	41	56	117	104	65	130	92	82	95	83	66	41	88
Kersbrook SO	65	134	136		145	92	88	92	86	76	53	66	134	135		143	91	88	91	85	75	52	92
Lismore	65	130	120	71	145	92	88	88	87	75	50	65	130	121	71	145	91	88	87	88	75	50	91
Majorca	58	135	130	72	146	106	91	95	87	75	51	58	135	131	72	146	105	91	95	87	75	51	93
Mt Burr	61	132	129		151	100	93	95	91	79	57	61	132	130		151	100	94	95	91	79	57	93
Wail	56						89	92	78	66	58	56						89	92	79	66	58	89
Mt Remarkable	57	109	111	57	115	73	78	82	78	65	48	58	112	113	57	119	73	78	81	79	65	48	83
S Wilmington PO	56	104	103		119	79	81	83	80	69	44	56	104	103		119	79	81	84	80	68	44	84
Wilmington	54	132	120	54	139	97	77	79	75		44	54	132	120	56	139	96	78	81	78		45	83
Wirrabara SF	59	124	121	67	140	91	87	88	82	74	50	60	124	121	65	141	91	87	89	83	74	52	89
Provenance SED	**4.6**	**9.2**	**9.6**	**7.4**	**11.9**	**12.1**	**3.5**	**4.9**	**5.0**	**7**	**3.2**	**4.9**	**9.9**	**10.8**	**11.7**	**13.6**	**13.6**	**4.3**	**6.1**	**6.6**	**8.8**	**4.2**	**2.9**
Kangaroo Is	53.7	120	109.3	61.5	136.7	93	81.0	88.9	78.4	64	40.4	54	122	111	63	139	93	82	90	80	67	41	88
Cultivated	62.8	133	131.6	71.4	146.3	97	89.3	92.8	86.6	75	53.2	63	133	130	72	147	98	90	93	86	74	53	92
S Flinders	57.6	119	116.4	61.1	131.9	81	81.8	83.9	79.7	68	47.7	58	119	116	60	131	82	81	84	80	69	47	86
Group SED	**1.8**	**4.4**	**5.4**	**4.2**	**5.5**	**4.5**	**1.6**	**1.7**	**2.0**	**3.0**	**1.4**	**2.0**	**5.2**	**4.3**	**6.0**	**6.8**	**5.8**	**2.1**	**2.5**	**3.0**	**4.1**	**2.1**	**1.5**
F probability (*f* KI)												0.001	0.003	<.001	0.468	0.046	0.991	0.003	0.001	0.008	<.001	<.001	<.001
F probability (*f* cult.)												0.038	0.015	<.001	0.016	0.198	0.013	0.022	<.001	0.061	0.072	0.002	<.001
F probability (*f* SFR)												<.001	0.001	0.034	0.299	0.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
F probability (prov. group)	<.001	<.001	<.001	0.067	0.017	0.002	0.001	<.001	<.001	<.001	<.001	0.039	0.987	0.362	0.588	0.928	0.298	0.004	0.234	0.013	0.722	<.001	0.044
F probability (provenance)	0.189	0.071	0.255	0.071	0.185	0.172	<.001	0.201	0.312	0.077	0.002	0.161	0.216	0.390	0.061	0.403	0.168	<.001	0.158	0.374	0.031	0.003	<.001

Appendix 3.2b Table of predicted HT means and fixed effect significance from single-site and across-sites (FA2) analyses

	HS model predicted HT means (cm)											MM models predicted HT means (cm)											
Group/provenance	**MOR**	**KOJ**	**WEL**	**ESP1**	**ESP2**	**AVE**	**BTN**	**HAM**	**LIS**	**COR**	**WGW**	**MOR**	**KOJ**	**WEL**	**ESP1**	**ESP2**	**AVE**	**BTN**	**HAM**	**LIS**	**COR**	**WGW**	**Across sites**
American River	N/A	NC	NC	N/A	1011	NC	623	647	536	574	393	N/A	861	821	N/A	1011	745	625	648	535	573	393	716
Cygnet River					1087		651	653	535	628	444		971	854		1090	826	653	652	536	630	448	743
Flinders Chase							672	640	563	641	475						796	673	642	563	648	479	774
Flinders Chase NP					1008		624	642	533	584	424		889	787		990	748	629	679	566	620	443	719
Kersbrook SO					1049		636	663	560	643	485		872	895		1034	758	632	651	550	635	478	722
Lismore					1033		625	617	572	629	458		838	822		1034	743	625	617	572	631	458	702
Majorca					1021		631	663	548	618	462		824	884		1021	832	631	663	547	617	463	708
Mt Burr					1031		654	652	564	652	495		865	869		1028	769	660	653	561	654	493	716
Wail					886		554	597	502	549	437		787	772		906		552	592	512	549	434	690
Mt Remarkable					851		556	580	492	564	390		704	720		851	617	551	581	492	561	389	656
S Wilmington PO							618	631	508	579	490						627	569	633	513	580	494	636
Wilmington					1001		563	578	487		413		856	843		1001	711	618	592	497		423	663
Wirrabara SF					1009		611	626	533	614	461		812	830		1009	722	610	632	535	613	482	697
Provenance SED					**67**		**24**	**29**	**26**	**46**	**24**		**54**	**62**		**83**	**81**	**30**	**38**	**33**	**54**	**34**	**21**
Kangaroo Is					1028		636	644	537	591	425		917	821		1036	766	646	658	549	613	437	730
Cultivated stand					1038		635	656	555	632	481		845	873		1027	775	634	650	549	622	475	714
S Flinders					961		578	604	512	573	437		796	799		953	661	570	601	510	576	432	671
Group SED					**29**		**10**	**10**	**11**	**16**	**10**		**28**	**31**		**41**	**30**	**15**	**16**	**15**	**25**	**17**	**11**
F probability (*f* KI)													0.039	0.041		0.686	0.484	0.002	0.358	0.985	0.013	<.001	0.923
F probability (*f* cult.)													0.180	0.004		0.035	0.002	0.008	<.001	0.007	0.066	<.001	<.001
F probability (*f* SFR)													<.001	0.034		0.022	<.001	<.001	<.001	<.001	<.001	<.001	<.001
F probability (group)					0.029		<.001	<.001	<.001	0.001	<.001		0.004	0.999		0.289	0.143	<.001	0.059	0.987	<.001	0.102	0.036
F probability (prov.)					0.260		0.002	0.311	0.169	0.113	0.002		0.353	0.557		0.575	0.336	0.001	0.214	0.076	0.020	<.001	<.001

N/A HT was not measured at a particular site. *NC* REML model did not converge

Appendix 3.2c Table of predicted AP (axis persistence) means from single-site and across-sites (FA2) models. The trait was rated 1 – 6 with 1 corresponding to the lowest fork being present at the base of the tree and 6 being free of forks

	HS model predicted AP means (1-6 score)											MM models predicted AP means (1-6 score)											
Group/provenance	**MOR**	**KOJ**	**WEL**	**ESP1**	**ESP2**	**AVE**	**BTN**	**HAM**	**LIS**	**COR**	**WGW**	**MOR**	**KOJ**	**WEL**	**ESP1**	**ESP2**	**AVE**	**BTN**	**HAM**	**LIS**	**COR**	**WGW**	**Across site**
American River	N/A	2.8	2.6	3.9	3.6	5.2	3.9	3.1	2.8	3.6	3.6	N/A	2.8	2.6	3.9	3.7	5.3	3.9	3.1	2.8	3.6	3.6	3.8
Cygnet River		3.5	3.6	4.5	4.8	5.1	4.3	2.8	2.7	3.8	3.9		3.5	3.6	4.5	4.7	5.1	4.3	2.8	2.6	3.8	3.9	3.7
Flinders Chase						4.6	4.3	2.7	2.6	3.4	3.9						4.6	4.2	2.7	2.5	3.5	4.0	3.8
Flinders Chase NP		3.0	3.2	4.1	4.2	4.6	3.9	2.4	2.5	3.5	3.8		3.0	3.2	4.1	4.2	4.6	4.1	2.6	2.7	3.6	3.9	3.7
Kersbrook SO		4.7	4.8		5.0	5.4	5.5	3.8	3.4	4.6	4.6		4.6	4.7		5.0	5.4	5.5	3.7	3.3	4.6	4.6	4.6
Lismore		4.1	3.8	4.6	4.5	5.8	5.7	3.1	3.1	4.1	4.4		4.1	3.8	4.6	4.5	5.8	5.7	3.1	3.2	4.1	4.4	4.5
Majorca		4.4	4.5	4.5	5.2	6.0	5.7	3.6	3.3	4.5	4.5		4.4	4.5	4.5	5.2	6.0	5.7	3.6	3.3	4.5	4.5	4.5
Mt Burr		4.1	4.9		5.3	5.5	5.4	3.3	3.4	4.7	4.7		4.2	4.9		5.3	5.5	5.4	3.2	3.3	4.7	4.7	4.5
Wail		4.5	4.4	4.6	5.0	4.9	5.5	3.6	3.0	4.4	4.5		4.6	4.5	4.6	5.2	4.9	5.5	3.5	3.0	4.4	4.5	4.6
Mt Remarkable		4.8	5.5		4.9	5.6	5.5	3.2	2.8	4.7	4.4		4.8	5.5		4.8	5.6	5.5	3.2	2.7	4.7	4.4	4.4
S Wilmington PO							5.7	3.7	3.4	4.6	4.6							5.6	3.7	3.4	4.6	4.6	4.3
Wilmington		3.8	4.1	4.4	4.3	5.7	5.6	3.3	3.1		4.5		3.8	4.1	4.4	4.3	5.7	5.7	3.3	3.1		4.5	4.5
Wirrabara SF		4.7	5.0	4.7	5.1	5.4	5.5	3.5	3.2	4.7	4.6		4.7	5.1	4.7	5.2	5.4	5.5	3.5	3.3	4.7	4.7	4.5
Provenance SED		**0.6**	**0.67**	**0.2**	**0.5**	**0.5**	**0.331**	**0.3**	**0.3**	**0.3**	**0.2**		**0.6**	**0.7**	**0.2**	**0.5**	**0.6**	**0.4**	**0.4**	**0.4**	**0.3**	**0.3**	**0.2**
Kangaroo Is		3.1	3.2	4.1	4.3	4.8	4.0	2.6	2.6	3.6	3.8		3.1	3.3	4.1	4.3	4.9	4.1	2.8	2.7	3.6	3.8	3.7
Cultivated stand		4.5	4.6	4.6	5.1	5.5	5.5	3.6	3.3	4.6	4.6		4.4	4.5	4.6	5.1	5.6	5.6	3.5	3.3	4.6	4.6	4.5
S Flinders		4.6	4.9	4.6	5.0	5.2	5.5	3.5	3.1	4.5	4.6		4.5	4.8	4.6	5.0	5.2	5.5	3.4	3.0	4.6	4.5	4.5
Group SED		**0.3**	**0.3**	**0.1**	**0.2**	**0.2**	**0.1**	**0.1**	**0.1**	**0.1**	**0.1**		**0.3**	**0.4**	**0.1**	**0.3**	**0.2**	**0.2**	**0.2**	**0.2**	**0.1**	**0.1**	**0.1**
F probability (*f* KI)													<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
F probability (*f* cult.)													0.232	0.524	0.718	0.397	0.240	0.220	0.037	0.027	0.359	0.901	0.006
F probability (*f* SFR)													0.064	0.036	0.329	0.782	0.784	0.096	0.706	0.011	0.211	0.598	0.825
F probability (group)		<.001	0.002	<.001	0.001	0.002	<.001	<.001	<.001	<.001	<.001		0.633	0.702	0.881	0.870	0.238	<.001	0.009	0.046	0.052	0.322	<.001
F probability (prov.)		0.610	0.481	0.300	0.400	0.262	0.838	0.08	0.654	0.543	0.838		0.799	0.349	0.477	0.451	0.227	0.785	0.095	0.677	0.382	0.468	0.437

N/A AP was not measured at a particular site

Appendix 3.3 Variance component and heritability estimates with standard errors for each trait resulting from single-site analyses calculated using models assuming either (i) progeny with common mothers are half-sibs or (ii) progeny have inter-relationships (taken as family averages from SNP marker-estimates) resulting from mixed mating. N/A denotes a parameter that was not included in a model. NC denotes an estimate that could not be made because the REML model would not converge. B denotes a boundary estimate, i.e. an estimate of close to zero variance.

Material presented as Electronic Supplementary Matter 3 in Bush et al. *(2015)*

Trait	(i) Half sib						(ii) Mixed mating					
DBH	$\hat{h}^2$ (se)	$\hat{\sigma}_t^2$	$\hat{\sigma}_e^2$	$\hat{\sigma}_{plot}^2$	$\hat{\sigma}_{row}^2$	$\sigma_{col.}^2$	$\hat{h}^2$ (se)	$\hat{\sigma}_t^2$	$\hat{\sigma}_e^2$	$\hat{\sigma}_{plot}^2$	$\hat{\sigma}_{row}^2$	$\sigma_{col.}^2$
MOR	0.21 (0.11)	69	230**	34	4	13	0.11 (0.07)	36*	257**	32*	5	13
KOJ	0.53 (0.17)	283*	236*	13	B	12	0.29 (0.09)	145*	345**	11	B	11
WEL	0.67 (0.19)	337*	163	B	8	15	0.39 (0.12)	184*	287**	B	8	14
ESP 1	0.68 (0.17)	271*	111	15	N/A	11	0.49 (0.13)	189*	179*	16	N/A	10
ESP 2	0.55 (0.17)	485*	361*	34*	B	B	0.30 (0.11)	246*	553**	29	2	B
AVE	0.68 (0.30)	349*	20	148*	23	108*	0.26 (0.19)	126*	194*	163*	14	111*
BTN	0.22 (0.06)	63*	161**	41*	B	B	0.13 (0.04)	31*	187**	41*	0.55	B
HAM	0.42 (0.09)	123*	168**	0.00	0	10*	0.20 (0.05)	56*	221**	B	B	10*
LIS	0.41 (0.08)	198*	278**	0.72	11	12*	0.16 (0.04)	69*	378**	B	11	12*
COR	0.46 (0.11)	214*	254**	B	2	96*	0.24 (0.07)	106*	346**	B	2	96*
WGW	0.25 (0.07)	52*	160**	B	2	8*	0.15 (0.04)	30*	177**	B	1	8*
HT												
MOR	Not measured						Not measured					
KOJ	NC						0.49 (0.11)	5122*	4576	777*	521	622
WEL	NC						0.62 (0.15)	6531*	2528*	1394*	B	357
ESP 1	Not measured						Not measured					
ESP 2	NC						0.63 (0.17)	10853*	4314	2196*	B	1340*
AVE	NC						0.59 (0.22)	10728*	1940	5554*	B	908*
BTN	0.45 (0.12)	3598*	1848*	2638**	13	1	0.28 (0.06)	2002*	3137**	2640**	18	B
HAM	0.84 (0.14)	5010**	604	324*	3	492	0.45 (0.09)	2450**	2627**	319*	B	484*
LIS	0.77 (0.12)	6274**	1469	449*	5591	521	0.32 (0.07)	2383**	4528**	487*	5703*	499*
COR	0.76 (0.15)	10009*	2729	462*	99	2874	0.35 (0.08)	4333*	7483**	531	102	2853*
WGW	0.73 (0.12)	4660**	1694*	48.6	169	404	0.41 (0.07)	2438**	3458**	51	162	409*
AP												
MOR	Not measured						Not measured					
KOJ	0.31 (0.12)	1.09*	2.39**	B	0.03	B	0.14 (0.07)	0.48*	2.87**	0	0.03	0.01
WEL	0.35 (0.14)	1.35*	2.36	0.15	B	B	0.18 (0.08)	0.67*	2.91**	0.16	B	B
ESP 1	0.00 (0.08)	0.00	1.14**	0.13	N/A	B	0.00 (0.06)	0.00	1.14**	0.13*	N/A	B
ESP 2	0.30 (0.14)	0.70*	1.45**	0.17	B	0.04	0.12 (0.07)	0.28	1.80**	0.18*	B	0.04
AVE	0.15 (0.22)	0.24	1.17*	0.17	0.02	0.23	0.09 (0.12)	0.13	1.27**	0.15	0.02	0.22*
BTN	0.50 (0.12)	0.80*	0.45*	0.35	0.01	B	0.25 (0.04)	0.40*	0.78**	0.34**	0.01	B
HAM	0.24 (0.08)	0.42*	1.33**	0.07	0.05	0.00	0.14 (0.05)	0.24*	1.48**	0.07	0.05	0
LIS	0.45 (0.09)	0.52*	0.54**	0.10	0.01	0.011	0.27 (0.06)	0.31*	0.72**	0.10*	0.01	0.01
COR	0.17 (0.08)	0.20*	0.93**	0.0	B	0.18	0.05 (0.05)	0.06*	1.04**	0.05	B	0.17*
WGW	0.26 (0.08)	0.30*	0.84	0.03	0.05	0.01	0.10 (0.05)	0.11*	0.98**	0.04	0.05*	0.01

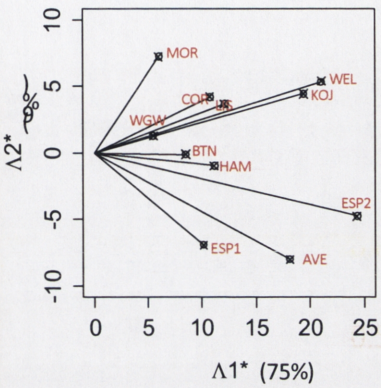
* Indicates that the component estimate is > twice its standard error

** Indicates that the component estimate is > 5 times its standard error

Appendix 3.4 Tables of correlation, variance and parameter estimates from *k*=2 factor analytic (FA2) models. Upper triangle gives correlations (heat mapped); approximate standard errors of additive (individual-tree) variance components ($\hat{\sigma}_t^2$) are given on the diagonal; and, approximate standard errors of correlations in the lower triangle. Estimates of site-specific variance (ψ), first and second loadings ($\Lambda 1$, $\Lambda 2$) and the residual variance $\hat{\sigma}_e^2$ are given below. The accompanying vector plots show the principal-component-rotated loadings ($\Lambda 1^*$ and $\Lambda 2^*$). The cosine of the angle between pairs of vectors is indicative of the type-B correlation between the sites. The length of the vector for each site corresponds to the additive variance explained by the two factors.

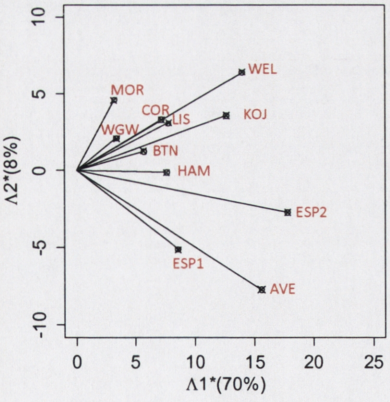
A1) DBH: half-sib, FA2 model

Site	MOR	KOJ	WEL	ESP1	ESP2	AVE	BTN	HAM	LIS	COR	WGW
Moora	34	0.78	0.76	0.08	0.48	0.27	0.63	0.52	0.78	0.69	0.56
Kojonup	0.8	104	0.93	0.53	0.90	0.79	0.96	0.85	0.92	0.77	0.69
Wellstead	0.8	0.93	125	0.49	0.85	0.74	0.91	0.81	0.88	0.74	0.66
Esperance1	0.1	0.53	0.49	70	0.72	0.77	0.66	0.62	0.46	0.35	0.37
Esperance2	0.5	0.90	0.85	0.72	161	0.98	0.98	0.90	0.83	0.67	0.64
Avenue	0.3	0.79	0.74	0.77	0.98	39	0.92	0.86	0.71	0.56	0.56
Bordertown	0.6	0.96	0.91	0.66	0.98	0.92	20	0.90	0.89	0.73	0.68
Hamilton	0.5	0.85	0.81	0.62	0.90	0.86	0.90	30	0.79	0.64	0.60
Lismore	0.8	0.92	0.88	0.46	0.83	0.71	0.89	0.79	38	0.74	0.65
Corowa	0.7	0.77	0.74	0.35	0.67	0.56	0.73	0.64	0.74	47	0.55
Wagga	0.6	0.69	0.66	0.37	0.64	0.56	0.68	0.60	0.65	0.55	13
$\hat{\Psi}$	0	14	56	93	0	0	0	28	22	79	33
$\hat{\Lambda}1$	6	19.4	21.1	10.2	24.3	18.1	8.5	11.2	12.1	10.8	5.5
$\hat{\Lambda}2$	3.7	4.5	5.4	-6.9	-4.7	-8	-0.1	-0.9	3.7	4.2	1.3
$\hat{\sigma}_t^2$	89	410	528	243	614	391	72	153	182	213	65
$\hat{\sigma}_e^2$	217	141	19	133	266	0	155	145	290	255	150



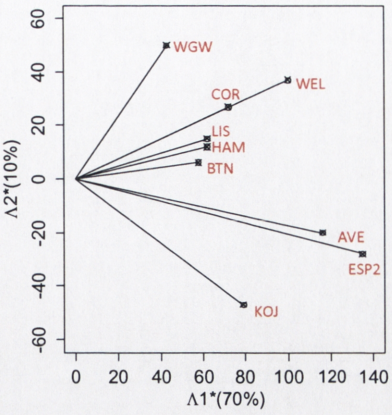
A2) DBH: mixed-mating, FA2 model

Site	MOR	KOJ	WEL	ESP1	ESP2	AVE	BTN	HAM	LIS	COR	WGW
Moora	22	0.57	0.66	0.06	0.35	0.12	0.59	0.38	0.58	0.51	0.47
Kojonup	0.19	53	0.90	0.52	0.86	0.70	0.94	0.78	0.83	0.69	0.58
Wellstead	0.19	0.08	60	0.43	0.81	0.60	0.95	0.75	0.86	0.72	0.62
Esperance1	0.23	0.15	0.16	45	0.73	0.78	0.57	0.59	0.44	0.34	0.23
Esperance2	0.22	0.09	0.10	0.12	88	0.95	0.93	0.85	0.77	0.62	0.48
Avenue	0.27	0.16	0.17	0.14	0.07	98	0.77	0.78	0.60	0.47	0.33
Bordertown	0.21	0.06	0.07	0.17	0.10	0.18	11	0.84	0.88	0.73	0.61
Hamilton	0.19	0.09	0.10	0.12	0.08	0.12	0.08	17	0.71	0.58	0.46
Lismore	0.18	0.10	0.09	0.15	0.11	0.16	0.08	0.10	20	0.66	0.56
Corowa	0.17	0.11	0.11	0.16	0.13	0.18	0.11	0.11	0.11	30	0.48
Wagga	0.17	0.12	0.13	0.16	0.16	0.19	0.13	0.13	0.12	0.12	11
$\hat{\Psi}$	18	22	17	62	0	0	0	20	19	50	23
$\hat{\Lambda}1$	3.2	12.7	14.0	8.6	17.8	15.6	5.7	7.6	7.8	7.2	3.4
$\hat{\Lambda}2$	4.6	3.6	6.4	-5.1	-2.7	-7.7	1.3	-0.1	3.1	3.3	2.1
$\hat{\sigma}_t^2$	49	196	253	162	322	303	34	77	90	113	-7.7
$\hat{\sigma}_e^2$	246	296	228	196	473	55	182	199	358	326	171



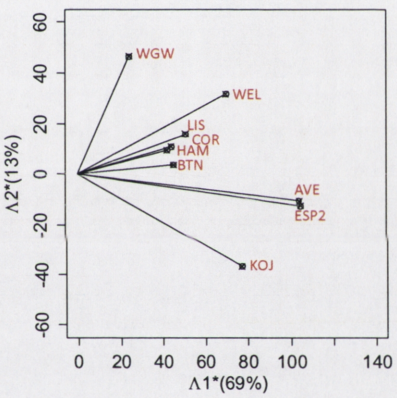
B1) Height: half-sib, FA2 model

Site	KOJ	WEL	ESP2	AVE	BTN	HAM	LIS	COR	WGW
Kojonup	2182	0.50	0.75	0.58	0.61	0.58	0.49	0.41	0.12
Wellstead	0.06	1047	0.83	0.66	0.92	0.95	0.85	0.82	0.77
Esperance2	0.09	0.04	2188	0.76	0.89	0.88	0.76	0.69	0.42
Avenue	0.14	0.09	0.08	3821	0.71	0.70	0.60	0.55	0.35
Bordertown	0.04	0.07	0.09	0.14	831	0.92	0.81	0.76	0.62
Hamilton	0.05	0.10	0.12	0.16	0.07	672	0.82	0.79	0.67
Lismore	0.07	0.03	0.04	0.09	0.07	0.11	856	0.70	0.61
Corowa	0.06	0.02	0.06	0.12	0.06	0.11	0.03	1552	0.64
Wagga	0.07	0.02	0.03	0.07	0.09	0.12	0.03	0.04	1305
Ψ	4727	43	547	9438	347	270	1511	2745	1186
$\hat{\Lambda}1$	79	100	135	116	58	62	62	72	43
$\hat{\Lambda}2$	-47	37	-28	-20	6	12	15	27	50
$\hat{\sigma}_t^2$	13139	11368	19581	23296	3757	4210	5548	8623	5524
$\hat{\sigma}_e^2$	0	0	0	0	1735	1206	2016	3760	1036



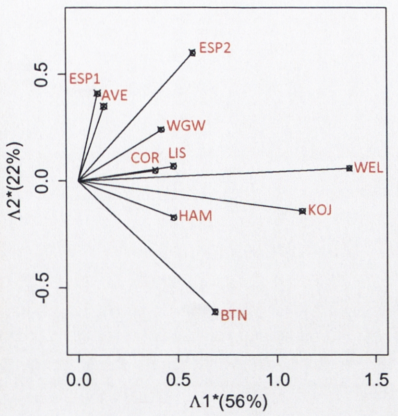
B2) Height: mixed mating, FA2 model

Site	KOJ	WEL	ESP2	AVE	BTN	HAM	LIS	COR	WGW
Kojonup	2501	0.53	0.74	0.58	0.70	0.59	0.54	0.47	0.03
Wellstead	0.12	1140	0.82	0.66	0.93	0.92	0.82	0.79	0.79
Esperance2	0.09	0.07	2598	0.73	0.93	0.84	0.76	0.69	0.34
Avenue	0.11	0.11	0.10	4569	0.73	0.67	0.60	0.55	0.28
Bordertown	0.13	0.10	0.10	0.12	605	0.91	0.82	0.76	0.53
Hamilton	0.11	0.05	0.07	0.10	0.10	454	0.79	0.74	0.62
Lismore	0.12	0.07	0.09	0.10	0.10	0.08	518	0.66	0.54
Corowa	0.12	0.08	0.10	0.10	0.11	0.08	0.09	1039	0.56
Wagga	0.16	0.08	0.14	0.16	0.14	0.10	0.11	0.12	541
Ψ	3616	0	969	8138	66	293	786	1697	0
$\hat{\Lambda}1$	-36.8	31.7	-12.3	-10.6	3.7	11.2	9.5	16.0	46.6
$\hat{\Lambda}2$	77.0	70.1	105.1	104.3	45.3	44.2	42.2	50.9	24.3
$\hat{\sigma}_t^2$	10895	5912	12160	19118	2130	2372	2654	4541	2760
$\hat{\sigma}_e^2$	1999	4454	6121	3236	3054	2681	4347	7086	3228



C1) Axis persistence: (AP) half-sib, FA2 model

Site	KOJ	WEL	ESP1	ESP2	AVE	BTN	HAM	LIS	COR	WGW
Kojonup	0.42	0.99	0.12	0.60	0.15	0.67	0.73	0.64	0.84	0.76
Wellstead	0.04	0.56	0.28	0.72	0.26	0.58	0.69	0.66	0.86	0.84
Esperance1	0.35	0.33	0.11	0.87	0.71	-0.37	-0.07	0.25	0.33	0.65
Esperance2	0.21	0.18	0.18	0.28	0.64	0.04	0.31	0.52	0.69	0.90
Avenue	0.38	0.36	0.22	0.24	0.16	-0.22	0.00	0.22	0.28	0.50
Bordertown	0.12	0.14	0.24	0.23	0.31	0.35	0.55	0.35	0.45	0.25
Hamilton	0.08	0.10	0.28	0.22	0.30	0.10	0.12	0.44	0.57	0.46
Lismore	0.09	0.08	0.23	0.15	0.24	0.12	0.10	0.10	0.58	0.58
Corowa	0.09	0.06	0.31	0.18	0.32	0.17	0.12	0.08	0.07	0.76
Wagga	0.15	0.11	0.24	0.08	0.27	0.18	0.16	0.10	0.11	0.09
Ψ	0	0	0	0	0.14	0.45	0.21	0.29	0.05	0.03
$\hat{\Lambda}1$	1.13	1.37	0.1	0.58	0.13	0.69	0.48	0.48	0.39	0.42
$\hat{\Lambda}2$	-0.14	0.06	0.41	0.6	0.35	-0.61	-0.17	0.07	0.05	0.24
$\hat{\sigma}_t^2$	1.30	1.89	0.18	0.69	0.28	1.29	0.47	0.53	0.20	0.27
$\hat{\sigma}_e^2$	2.20	1.96	1.01	1.46	1.14	0.08	1.29	0.53	0.93	0.86



C2) Axis persistence: mixed mating, FA2 model

Site	KOJ	WEL	ESP1	ESP2	AVE	BTN	HAM	LIS	COR	WGW
Kojonup	0.22	0.99	0.09	0.63	0.12	0.68	0.66	0.57	0.74	0.76
Wellstead	0.04	0.30	0.26	0.75	0.22	0.59	0.64	0.60	0.75	0.85
Esperance1	0.37	0.36	0.06	0.83	0.64	-0.42	-0.01	0.26	0.19	0.68
Esperance2	0.23	0.19	0.21	0.14	0.56	0.05	0.36	0.52	0.56	0.96
Avenue	0.44	0.45	0.60	0.56	0.18	-0.23	0.03	0.20	0.17	0.48
Bordertown	0.16	0.16	0.28	0.25	0.41	0.12	0.49	0.29	0.44	0.22
Hamilton	0.15	0.15	0.29	0.22	0.31	0.15	0.08	0.37	0.48	0.46
Lismore	0.13	0.12	0.24	0.17	0.30	0.14	0.12	0.07	0.45	0.56
Corowa	0.19	0.18	0.33	0.23	0.36	0.20	0.17	0.15	0.05	0.64
Wagga	0.20	0.18	0.31	0.20	0.52	0.20	0.20	0.16	0.23	0.12
Ψ	0.00	0.00	0.00	0.00	0.07	0.18	0.14	0.20	0.05	0.01
$\hat{\Lambda}1$	0.80	0.98	0.06	0.40	0.07	0.47	0.33	0.33	0.26	0.29
$\hat{\Lambda}2$	-0.10	0.04	0.28	0.38	0.21	-0.43	-0.08	0.07	0.01	0.18
$\hat{\sigma}_t^2$	0.65	0.96	0.08	0.31	0.12	0.58	0.26	0.31	0.12	0.12
$\hat{\sigma}_e^2$	2.73	2.72	1.09	1.78	1.28	0.66	1.47	0.71	1.00	0.97