

Characterising a *Eucalyptus cladocalyx* breeding population using SNP markers

David Bush · Bala Thumma

Received: 25 July 2012 / Revised: 14 November 2012 / Accepted: 27 November 2012 / Published online: 16 January 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract Population structure, family relatedness and inbreeding within a first-generation *Eucalyptus 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 derived from SFR subpopulations 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.

Keywords 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

Communicated by R. Burdon

Electronic supplementary material The online version of this article (doi:10.1007/s11295-012-0589-1) contains supplementary material, which is available to authorized users.

D. Bush (✉) · B. Thumma
CSIRO Plant Industry, GPO Box 1600, Canberra,
ACT 2601, Australia
e-mail: david.bush@csiro.au

D. Bush
Australian National University, Fenner School of Environment and
Society, Canberra, ACT, Australia

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 Online Resource 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 KI 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

(ALRTIG) (see Bush et al. 2009). Samples were drawn from 119 families within 13 subpopulations (Table 1; for subpopulation location data, see Online Resource 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, the 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 2,185 trees) of the 20 trees per family originally planted were measured at 5 and 7.8 years.

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 seed lots 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 Online Resource 2 which gives the GenBank accession numbers and places the sequences 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

Table 1 MAF, average observed (\hat{H}_O) and expected (\hat{H}_E) heterozygosity, percentage of polymorphic loci (%P), relatedness ($\bar{\theta}$) and inbreeding coefficient (\bar{f}) calculated by the Wang triadic method

Subpopulation and location of trial	Description	MAF	\hat{H}_E (SE)	\hat{H}_O (SE)	%P	N (families/trees)	Average $\bar{\theta}$ (relatedness) not corrected for bias		\bar{f}
							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 (0.02)	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
Mount 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
Cygnnet 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
Mount 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-provenance 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	119/1,199	0.42	0.12	0.17

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

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 (Online Resource 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 :

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

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-estimation 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 biallelic data were 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):

$$\overline{2\theta}_{ij} = \frac{1}{4} (1 + \overline{f} + 2s + s^2 + 2\overline{f}s + \overline{f}s^2) \quad (2)$$

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) (Online Resource 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' untranslated regions, promoter regions and intron and exon regions of *E. grandis* genes (Table 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 EcxCBS gene was identified as

Table 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

an outlier. A plot of subpopulation differentiation (F_{ST}) versus heterozygosity is included in Online Resource 3.

Figure 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.

SNP was variable among subpopulations. For the KI group of subpopulations, only one marker from the EcxHPA 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 EcxMYB83 and EcxNOD genes, which had an MAF <0.03 in the KI subpopulations, were monomorphic in all SFR subpopulations, though

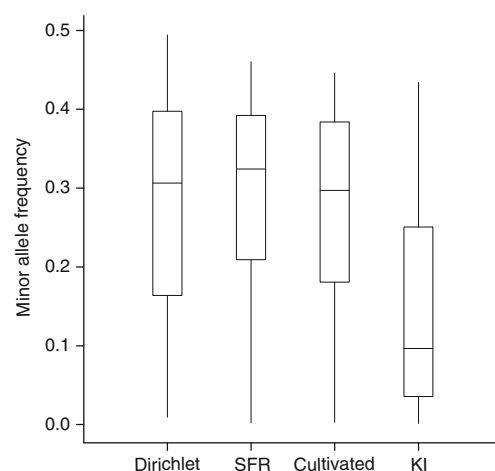
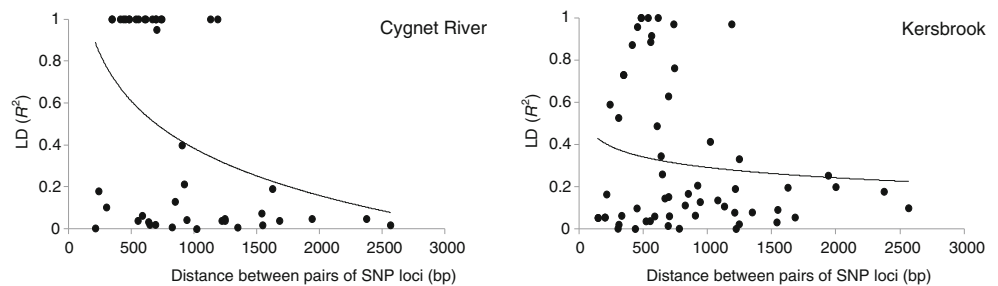


Fig. 1 Box and whisker diagram showing MAF from a simulated Dirichlet distribution, the South Flinders Ranges (SFR), Kangaroo Island (KI) and cultivated groups of *E. cladocalyx* subpopulations. Whiskers indicate the full range of the data, boxes indicate the interquartile range and bars within boxes indicate the median

Fig. 2 Intra-gene LD (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



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 EcxMYB83 and EcxNOD 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. 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 2,500 bp, LD had decayed significantly ($R^2 < 0.1$) in all subpopulations.

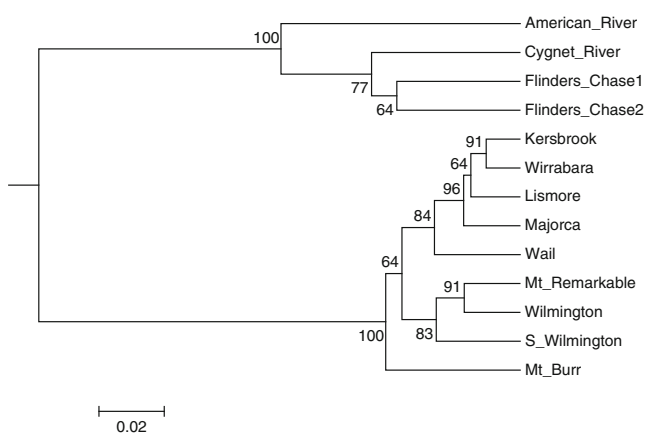


Fig. 3 UPGMA phenogram for the *E. cladocalyx* breeding population at Lismore, Victoria based on \hat{F}_{ST} uncorrected genetic distances (Takezaki et al. 2010); 10,000 bootstraps over loci were performed and the percentage of replications in which a branch bifurcated is given at nodal bases

Measures of heterozygosity and diversity

Heterozygosity

Heterozygosity (\hat{H}_E) averaged over subpopulations was 0.28 (standard error [SE], 0.01) (Table 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 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. 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.

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

Table 3 Bias estimates from triadic estimation of relatedness using 75 simulated SNP loci with allele frequencies drawn from the *E. cladocalyx* empirical dataset

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		

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 1,000 bootstraps over loci

on allele frequencies from the empirical dataset (Table 3) and Dirichlet-distributed allele frequencies (full results not shown, though compared with the 75-loci entry in Table 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 0.06 for unrelated and HS dyads, respectively.

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 (500–1,000) were employed (Table 4). Indications are that 75 SNP markers give only modest pairwise precision of

Table 4 From 500 simulated dyads (allele frequency 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 ($\bar{2\theta}_f$) and inbreeding coefficient (\bar{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	
1,000	0.017	0.00–0.07	0.011	0.20–0.33	0.014	0.47–0.59	–0.008	0.91–1.08	
$\bar{2\theta}_f$ (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	
1,000	0.015	0.00–0.07	0.017	0.00–0.08	0.017	0.00–0.08	–0.005	0.43–0.56	
\bar{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	

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 HS dyads). With 500 loci, the 95 % CI narrow substantially to 0.00–0.10 and 0.18–0.37 for unrelated and half-sib (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.

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 Online Resource 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 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, \overline{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 $\overline{2\theta}_f = 0.6$ and $\overline{f}_f = 0.28$. A whole-population positive, linear relationship between f and 2θ was found, with $R^2=0.93$.

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 ($\overline{2\theta}_{f-f}$) was 0.17, with some pairs of families having $\overline{2\theta}_{f-f} > 0.5$, 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 ($\overline{2\theta}_{f-f} = 0.15$) 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.

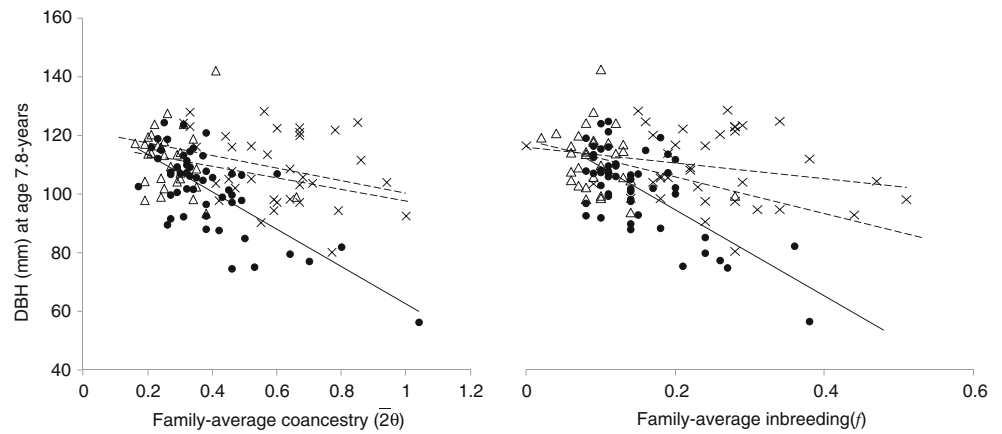
Relatedness, inbreeding and fitness

Differences in response of the fitness-related trait, DBH at age 7.8 years, to family-average relatedness ($\overline{2\theta}_f$) and inbreeding (\overline{f}) were observed among subpopulation groups (Fig. 4). Multiple linear regression analyses including subpopulation group and \overline{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 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 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 moment resulted in non-significant relationships. Survival was generally very high in all of the progeny trials at measurement age (trial populations averaged 88–92 %).

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 *ExcCBS* 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

Fig. 4 Relationship between family-average relatedness ($2\theta_f$), inbreeding (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 $2\theta_f$ and f , respectively), KI (crosses, not significant) and cultivated stands (triangles, not significant)



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 *Pinus 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

Table 5 Summary of regression analysis of family-mean relatedness ($2\theta_f$) and inbreeding coefficient (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	$2\theta_f$		f	
		Percent variance explained	Significance (p)	Percent variance explained	Significance (p)
<i>E. cladocalyx</i> —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
<i>E. cladocalyx</i> —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
<i>E. cladocalyx</i> —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

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: 500–1,000 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, where stands are generally small and isolated, either by natural and/or anthropogenic causes. 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 homogeneous. 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 coefficient of relationship to scale variance parameters, a value of $\hat{\rho} = 1/2.8$ is indicated. However, the respective values of $\hat{\rho}$ 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 among-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-family and among-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 $\hat{\rho} = 1/2.5$, 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 ID 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.

Acknowledgments We would like to acknowledge Bronwyn Matheson and Charlie Bell for the assistance with DNA extraction and preparation of samples for sequencing. Tim Jackson, Des Stackpole, Paul Macdonell and David Spencer are thanked for the 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.

References

- Barrett JC, Fry B, Maller J, Daly M (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265
- Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proc: Biol Sci* 263:1619–1626
- Burdon RD, Wilcox PL (2007) Population management: potential impacts of advances in genomics. *New For* 34:187–206
- Bush D, Kain D, Matheson C, Kanowski P (2011a) Marker-based adjustment of the additive relationship matrix for estimation of genetic parameters—an example using *Eucalyptus cladocalyx*. *Tree Genet Genomes* 7:23–35
- Bush D, McCarthy K, Meder R (2011b) Genetic variation of natural durability traits in *Eucalyptus cladocalyx* (sugar gum). *Ann For Sci* 68:1057–1066
- Bush DJ, Jackson TJ, Driscoll J, Harwood CE (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, 143 pp
- Callister A, Bush DJ, Collins S, Davis W (2008) Prospects for genetic improvement of *Eucalyptus cladocalyx* in Western Australia. *N Z J For Sci* 38:211–226
- Costa e Silva J, Hardner C, Potts BM (2010a) Genetic variation and parental performance under inbreeding for growth in *Eucalyptus globulus*. *Ann For Sci* 67:606p1–606p8
- Costa e Silva J, Hardner C, Tilyard P, Pires AM, Potts BM (2010b) Effects of inbreeding on population mean performance and observational variances in *Eucalyptus globulus*. *Ann For Sci* 67:605

- Costa e Silva J, Hardner C, Tilyard P, Potts BM (2011) The effects of age and environment on the expression of inbreeding depression in *Eucalyptus globulus*. *Heredity* 107:50–60
- Dillon SK, Nolan MF, Bragg JG, Matter P, Gapare WJ, Southerton SG (2013) Contrasting patterns of hierarchical structure inferred among the mainland populations of *Pinus radiata* (D. Don) from putatively neutral and adaptive SNP markers. *Tree Genet Genomes* (resubmitted)
- Eldridge K (1978) Genetic improvement of eucalypts. *Silvae Genetica* 27:205–209
- Eldridge K, Davidson J, Harwood CE, van Wyk G (1993) *Eucalypt* domestication and breeding. Clarendon, Oxford, 288 pp
- Ellis MF, Sedgley M (1992) Floral morphology and breeding system of three species of *Eucalyptus*, section *Bisectaria* (Myrtaceae). *Aust J Bot* 40:249–262
- Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity* 103:285–298
- Excoffier L, Lischer H (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10:564–567
- Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res* 40:D1178–D1186
- Grattapaglia D (2007) Marker-assisted selection in *Eucalyptus*. In: Guimarães EP, Ruane J, Scherf BD, Sonnino A, Dargie JD (eds) Marker-assisted selection: current status and future perspectives in crops, livestock, forestry and fish. Food and Agriculture Organization of the United Nations, Rome, pp 251–282
- Grattapaglia D, Kirst M (2008) *Eucalyptus* applied genomics: from gene sequences to breeding tools. *New Phytol* 179:911–929
- Griffin AR, Cotterill PP (1988) Genetic variation in growth of outcrossed, selfed and open-pollinated progenies of *Eucalyptus regnans* and some implications for breeding strategy. *Silvae Genetica* 37:124–131
- Gunn B (2001) Australian tree seed centre operations manual. CSIRO, Canberra, 150 pp
- Hardner C, Tibbits W (1998) Inbreeding depression for growth, wood and fecundity traits in *Eucalyptus nitens*. *For Genet* 5:11–20
- Hardner C, Vaillancourt RE, Potts BM (1996) Stand density influences outcrossing rate and growth of open-pollinated families of *Eucalyptus globulus*. *Silvae Genetica* 45:226–228
- Hardner CM, Potts BM (1995) Inbreeding depression and changes in variation after selfing in *Eucalyptus globulus* ssp. *globulus*. *Silvae Genetica* 44:46–54
- Hardner CM, Potts BM (1997) Postdispersal selection following mixed mating in *Eucalyptus regnans*. *Evolution* 51:103–111
- Harwood C, Bulman P (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 CE, Bush DJ, Butcher T, Bird R, Henson M, Lott R, Shaw S (2007) Achievements in forest tree genetic improvement in Australia and New Zealand. 4. Tree improvement for low-rainfall farm forestry. *Aust For* 70:23–27
- Hodgson LM (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 Afr For J* 98:32–43
- Kelly JK (2004) Family level inbreeding depression and the evolution of plant mating systems. *New Phytol* 165:55–62
- Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nat Rev Genet* 4:981–994
- Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer Associates Inc., Sunderland, 980 pp
- McDonald MW, Rawlings M, Butcher PA, Bell JC (2003) Regional divergence and inbreeding in *Eucalyptus cladocalyx* (Myrtaceae). *Aust J Bot* 51:393–403
- Mimura M, Barbour RC, Potts BM, Vaillancourt RE, Watanabe KN (2009) Comparison of contemporary mating patterns in continuous and fragmented *Eucalyptus globulus* native forests. *Mol Ecol* 18:4180–4192
- Mora F, Gleadow R, Perret S, Scapim CA (2009) Genetic variation for early flowering, survival and growth in sugar gum (*Eucalyptus cladocalyx* F. Muell) in southern Atacama Desert. *Euphytica* 169:335–344
- Myburg A, Grattapaglia D, Tuskan G, Jenkins J, Schmutz J, Mizrachi E, Hefer C, Pappas G, Sterck L, Peer YVD, Hayes R, Rokhsar D (2011) The *Eucalyptus grandis* Genome Project: genome and transcriptome resources for comparative analysis of woody plant biology. *BMC Proc* 5:I20
- Potts B, Wiltshire J (1997) *Eucalypt* genetics and genecology. In: Williams J, Woinarski J (eds) *Eucalypt ecology: individuals to ecosystems*. Cambridge University Press, Cambridge, pp 56–91
- Skabo S, Vaillancourt RE, Potts BM (1998) Fine-scale genetic structure of *Eucalyptus globulus* ssp. *globulus* forest revealed by RAPDs. *Aust J Bot* 46:583–594
- Steane D, Nicolle D, Sansaloni C, Petroli C, Carling J, Kilian A, Myburg A, Grattapaglia D, Vaillancourt R (2011) Population genetic analysis and phylogeny reconstruction in *Eucalyptus* (Myrtaceae) using high-throughput, genome-wide genotyping. *Mol Phylogenet Evol* 59:206–224
- Takezaki N, Nei M, Tamura K (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
- Thavamanikumar S, McManus LJ, Tibbits JFG, Bossinger G (2011) The significance of single nucleotide polymorphisms (SNPs) in *Eucalyptus globulus* breeding programs. *Aust For* 74:23–29
- Thumma BR, Nolan MF, Evans R, Moran GF (2005) Polymorphisms in cinnamoyl CoA reductase (CCR) Are associated with variation in microfibril angle in *Eucalyptus* spp. *Genetics* 171:1257–1265
- Van Raden PM (2007) Genomic measures of relatedness and inbreeding. *Interbull Bull* 37:33–36
- 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. *Mol Ecol Resour* 11:141–145
- Wang J, Santure AW (2009) Parentage and sibship inference from multilocus genotype data under polygamy. *Genetics* 181:1579–1594