Australian Forestry

ISSN: 0004-9158 (Print) 2325-6087 (Online) Journal homepage: www.tandfonline.com/journals/tfor20

Assessing genetic structure and diversity in natural populations of *Eucalyptus bosistoana*: insights from SNP genotyping for New Zealand breeding trials

S.-J. Kim, T. E. Steeves, C. M. Altaner, L. A. Apiolaza & P. B. Pelser

To cite this article: S.-J. Kim, T. E. Steeves, C. M. Altaner, L. A. Apiolaza & P. B. Pelser (25 Jun 2026): Assessing genetic structure and diversity in natural populations of *Eucalyptus bosistoana*: insights from SNP genotyping for New Zealand breeding trials, Australian Forestry, DOI: 10.1080/00049158.2026.2687719

To link to this article: https://doi.org/10.1080/00049158.2026.2687719

AUSTRALIAN FORESTRY
https://doi.org/10.1080/00049158.2026.2687719

Assessing genetic structure and diversity in natural populations of *Eucalyptus bosistoana*: insights from SNP genotyping for New Zealand breeding trials

S.-J. Kim [a,b], T. E. Steeves [b], C. M. Altaner [a], L. A. Apiolaza [a] and P. B. Pelser [b]

[a]School of Forestry, University of Canterbury, Christchurch, New Zealand; [b]School of Biological Sciences, University of Canterbury, Christchurch, New Zealand

ABSTRACT

Eucalyptus bosistoana (Myrtaceae) is a breeding species of NZDFI (New Zealand Dryland Forests Innovation), which aims to establish forestry plantations of ground-durable, high-value hardwoods in New Zealand. For this purpose, plants from *E. bosistoana* seeds collected from natural source locations in Australia between 2008–2012 have been used since 2009 to establish breeding trials in New Zealand. To inform the NZDFI breeding programme, leaf samples of 177 *E. bosistoana* breeding families were genotyped using a *Eucalyptus* 72K single nucleotide polymorphism (SNP) Axiom array. This was done to identify patterns of genetic structure among *E. bosistoana* breeding families and to assess the genetic structure and diversity of the natural source populations from which their seeds were obtained. PCoA and STRUCTURE analyses indicated that the *E. bosistoana* seeds were obtained from three natural source populations with similar levels of genetic diversity. Weak genetic structure among most natural source locations of *E. bosistoana* within these populations was observed, and families from nearby source locations were generally genetically more similar to each other than to those established from more distant locations. We also found that some families in the *E. bosistoana* breeding programme represent *E. melliodora*. These combined genetic patterns provide the genetic basis for establishing, maintaining, or improving NZDFI's *E. bosistoana* breeding populations.

ARTICLE HISTORY
Received 13 May 2025
Accepted 4 June 2026

KEYWORDS
Eucalyptus bosistoana; *E. melliodora*; *Eucalyptus*; forestry, population genomics; tree improvement

Introduction

Understanding patterns of genetic structure and diversity of natural source populations is useful for tree breeders. For example, knowledge of the genetic differences among populations aids the selection of suitable source populations for establishing, maintaining or improving a breeding population (Allier et al. 2020). Information about genetic diversity is also essential for supporting a breeding programme with a broad genetic base, necessary for identifying and selecting commercially valuable genotypes or creating them through crossbreeding, eventually enhancing the genetic gain of the traits (Johnson et al. 2001; Jing et al. 2023). A broad genetic base can allow for selection against pests and diseases, developing new forestry products and preventing inbreeding depression in breeding populations (Eldridge 1990). In addition, genetic structure and diversity data can support forestry programmes with efforts to maintain genetic diversity in forestry plots for the purpose of increasing resilience to climate change (Eldridge 1990; Lefèvre et al. 2014). Ensuring that breeding populations have sufficient adaptive potential is also critical for the successful establishment of plantations across areas with different environmental conditions (Hamilton and Miller 2016). Further, because the genetic structure among natural populations and breeding families established from them influences the accuracy of predicting genetic potential (Werner et al. 2020), understanding patterns of genetic structure is important for the effective application of genomic selection in breeding programmes.

Eucalyptus bosistoana F.Muell. (Myrtaceae) is a species native to New South Wales and Victoria in Australia, where it is mostly found in the south and south-eastern parts of these states (Slee et al. 2006; Flores-Renteria et al. 2017). *Eucalyptus bosistoana* is a breeding species of NZDFI (New Zealand Dryland Forests Innovation). It produces Class 1 naturally ground-durable timber (Australian Standard, AS5604-2005) suitable to supply domestic and global markets for agricultural posts, outdoor joinery, and engineered wood products (Bootle 2005; Sharma et al. 2024). NZDFI has been establishing *E. bosistoana* breeding trials in New Zealand since 2009. To date, seeds of 254 mother trees (i.e. breeding families) identified in the field as naturally occurring *E. bosistoana* have been collected from various locations in southeast Australia (Figure 1) to establish a genetically diverse breeding population of the species in New Zealand.

CONTACT S.-J. Kim seoljong.kim@gmail.com; P. B. Pelser pieter.pelser@canterbury.ac.nz School of Biological Sciences, Private Bag 4800, Christchurch 8140, New Zealand

Supplemental data for this article can be accessed online at https://doi.org/10.1080/00049158.2026.2687719

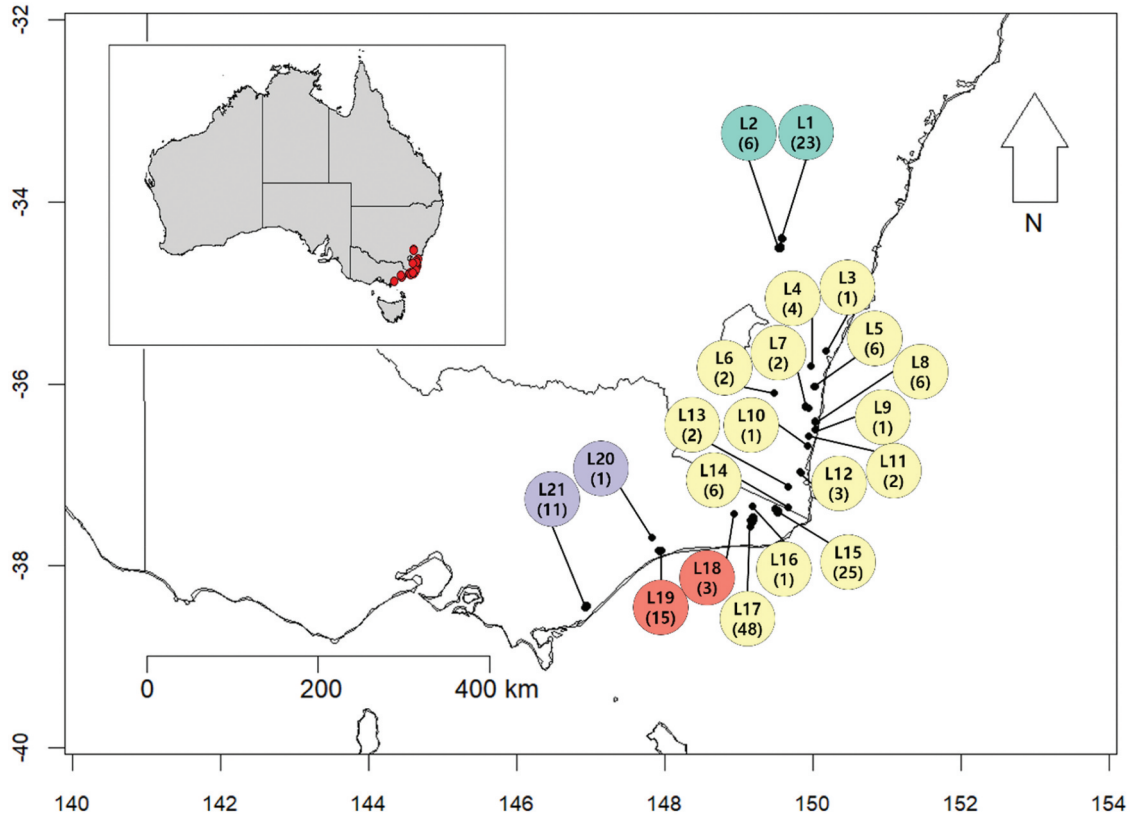

Figure 1. Geographical distribution in Australia of 21 natural source locations of families in the NZDFI *E. bosistoana* breeding programme. Each circle represents a sampling location, labelled L1-21. The number of families from each location that were included in our study is presented in brackets. One individual per family was genotyped. The colours of the circles indicate the population to which plants from that location belong, as indicated by the results of the STRUCTURE analyses and PCoA of SNP data (Figures 2 and 3). L1-2 (green) represent *E. melliodora* and L3-21 (yellow, red and purple) are *E. bosistoana* families.

Progeny trees in the NZDFI trials have been phenotyped for traits that are of commercial interest, including growth and form, heartwood quantity and quality, growth strain, grain and essential oils (Davies et al. 2017; Li et al. 2018; Rajapaksha et al. 2023b; Sharma et al. 2024). This allowed the identification of *E. bosistoana* families with superior characteristics for propagation and deployment in plantations (Davies et al. 2017; Li et al. 2018; Rajapaksha et al. 2023a). Information about the genetic relationships among families and their natural source populations, as well as the genetic diversity of these source populations would allow more accurate estimation of breeding values and enable better informed management of the genetic diversity in the breeding programme (Werner et al. 2020). However, this information is currently not available.

In this study, we used a *Eucalyptus* 72K SNP Axiom array (de Oliveira et al. 2023) to genotype 177 plants (one per breeding family) of the 254 families in the *E. bosistoana* NZDFI breeding programme. These plants were grown from seeds that were sourced from 21 natural source locations in Australia (Figure 1). Although the families were collected for commercial purposes, the 21 source locations span the known natural geographic range of *E. bosistoana* in southeastern Australia (Slee et al. 2006). The specific objectives of this study were to identify patterns of genetic structure among natural *E. bosistoana* source locations and breeding families established from them, to delineate the number of natural source populations in the breeding programme, and to assess the genetic diversity of these source populations.

Materials and methods

Sampling

Seed collection from natural *E. bosistoana* locations in Australia (Figure 1) was organised by Proseed NZ Ltd, a tree seed company and NZDFI partner, in 2009, 2010 and 2012 (Nicholas & Millen 2012). Mother trees used for establishing *E. bosistoana* breeding families were selected based on traits of potential commercial interest (e.g. form and health), seed availability and accessibility. This seed collecting project was part of a broader NZDFI initiative that involved collecting various *Eucalyptus* species considered suitable for establishing plantations in the New Zealand environment. Seeds of mother trees were germinated and grown at various breeding trial sites in New Zealand to establish breeding families (Supplementary Figure S1).

Leaf samples of 177 breeding families (one tree per family) were collected for this study. These were

obtained from trees grown in New Zealand from seed lots from 21 Australian natural source locations (L1–21; Figure 1). L1 and L2 were originally collected and labelled as *E. bosistoana*, and subsequently included in the NZDFI breeding population, but they were identified as *E. melliodora* as a result of the present study (see Results and Discussion). We also included samples from eight families for which some uncertainty regarding the provenance of their mother trees existed (U1–8). These trees were recorded as grown from seeds from western Victoria, but other metadata instead indicated that they were most likely collected in eastern Victoria. Following collection, the leaf samples were frozen at −17°C and subsequently stored at −80°C. All samples were completely desiccated at 35°C in a drier at the University of Canterbury Herbarium (CANU) prior to DNA extraction. Voucher herbarium specimens for each location were deposited at CANU.

DNA extraction and SNP genotyping

An Oscillating Mill MM400 (Retsch GmbH, Haan, Germany) was used to grind c. 20 mg of leaf tissue per sample using three metal beads. DNA extraction was subsequently conducted using either the DNeasy Plant Mini Kit DNA (Qiagen GmbH, Hilden, Germany) or the Genomic DNA Mini Kit (Plant) (DNAture, Gisborne, New Zealand). DNA extraction followed each manufacturer's protocol, but we extended the incubation time to one hour to improve yield. Following DNA extraction, DNA quality and concentration were measured using a Nanodrop ND-1000 Spectrophotometer and Qubit 2.0 fluorometer (ThermoFisher Scientific (TF), Waltham, MA, U.S.A.). Only genomic samples with an OD260/OD280 ratio between 1.8 and 2.0 and an OD260/OD230 ratio greater than 1.5 were submitted for genotyping. The DNA concentration of genomic samples selected for genotyping ranged between 3.59 and 161 ng/µl, as estimated with a Qubit fluorometer.

SNP data were generated with the *Eucalyptus* 72K SNP Axiom array (TF). TF conducted Quality Control using PicoGreen (TF) analysis at the Applied Biosystems Microarray Research Services Laboratory to verify the DNA concentration of the samples. To produce better quality genotyping data for the samples and to mitigate potential ascertainment bias of the array, custom genotyping probeset lists were created. SNP data were stored as CEL files and converted to Variant Call Format (VCF) files for data analysis.

To assess the reproducibility of the genotyping results, 83 repeat samples were included alongside the study samples. The average cluster call rate across all samples was 99.51%, and sample reproducibility was 99.82%, indicating a high level of genotyping consistency. Repeat samples were removed prior to downstream analyses.

Data filtering

Only polymorphic SNPs with a Call Rate over 97.32% were used in our analyses. SNP loci with a Minor Allele Count of 1 and 2 were removed from the data set following the recommendation of Schmidt et al. (2021), as including singletons and doubletons in genetic structure analyses can confound model-based inferences of genetic structure (Linck & Battey 2019). Linkage Disequilibrium pruning was carried out using BCFtools (Danecek et al. 2021). We tested different values of r^2 (0.2, 0.4, 0.6 and 0.8) and window sizes (w) (500, 1k and 10k) to determine if the results of the genetic structure analyses (i.e. PCoA) would change if different LD filters were used. Genetic structure analyses using various combinations of r^2 and window size values resulted in very similar genetic structure patterns. Thus, a window size of 10k was chosen as it is the most conservative value tested and an r^2 value of 0.8 was selected to minimise the number of SNPs excluded in the analyses (i.e. 1,110 SNPs with the selected LD pruning settings). The final data set after filtering included 6,898 SNPs.

Data analysis for genetic structure and diversity analyses

To infer genetic structure among natural source locations and breeding families of *E. bosistoana*, STRUCTURE analyses were conducted using Structure_threader (Pritchard et al. 2000; Pina-Martins et al. 2017). The filtered VCF data files were converted into STRUCTURE format using PGDSpider version 2.1.1.5 (Lischer & Excoffier 2012). Location information was provided for each of the 21 source locations using the 'indfile' option in Structure_threader. Each STRUCTURE analysis was run 20 times with 200K generations each. A total of 100K generations per run were discarded as burn-in. The method of Evanno et al. (2005) was used to determine the optimal K value which is defined as the highest ΔK value in the STRUCTURE analysis results. The estimated optimal K value was further compared with the second-order rate of change in the likelihood of the data (LnP(K)) with respect to K to identify the K value where there is an 'elbow' in the curve (Rosenberg et al. 2002). Graphs of both ΔK and LnP(K) estimates were plotted using Microsoft Excel (Supplementary Figure S2). Finally, membership probability plots for each K value of interest were generated using the 'plot' mode in Structure_threader (Pina-Martins et al. 2017).

The ΔK method for selecting the K value in STRUCTURE analyses commonly indicates K = 2 as the most likely number of clusters, even when more than

two populations or subpopulations are present (Meirmans 2015; Janes et al. 2017; Lawson et al. 2018). We therefore also used Principal Coordinate Analyses (PCoA) in GenAlEx 6.5 (Peakall & Smouse 2006, 2012) to study genetic structure. To convert the VCF file to a GenAlEx format, the *vcfR2genind* function in the vcfR 1.15.0 package was used to load the VCF file as a 'genind' object in RStudio and R version 4.2.3 (Knaus & Grunwald 2017; R Core Team 2021; RStudio Team 2021). The genind object was then reformatted for GenAlEx using the *genind2genalex* function in the poppr 2.9.6 package (Kamvar et al. 2014, 2015).

The STRUCTURE and PCoA analyses revealed patterns of genetic structure at different hierarchical levels. We further explored these by exposing three data sets to an Analysis of Molecular Variance Analysis (AMOVA): a data set containing samples from all source locations (i.e. L1–21), a dataset composed of samples from L3–21, and data only from L3–17. For each of these data sets, only source locations composed of at least four individuals were used. AMOVA were performed in GenAlEx 6.5 using 999 permutations and a global F'_{ST} (fixation index) value was calculated for each of the three analyses to quantify genetic differentiation.

Isolation by distance (IBD) analysis was conducted to detect a possible correlation between geographic and genetic distances among *E. bosistoana* individuals from different locations. A Nei genetic distance matrix and a geographic distance matrix were generated in GenAlEx 6.5. Mantel tests (Mantel 1967) with 999 permutations were performed with these matrices to test for IBD.

To quantify the genetic diversity of each population identified in the genetic structure analyses, observed heterozygosity (H_o), expected heterozygosity (H_e) and inbreeding coefficients (F_{IS}) were calculated in GenAlEx 6.5. To reduce the impact of varying sampling sizes among populations, we determined allelic richness (Ar) and richness of private alleles (ap) using the rarefaction method available in HP-Rare v.1.1 (Kalinowski 2005).

Results

Genetic structure

STRUCTURE analyses were carried out to understand the patterns of genetic structure among natural source locations and breeding families established in New Zealand from 21 locations in Australia (L1–21; Figure 1) for the NZDFI *E. bosistoana* breeding programme. We also included eight samples of uncertain provenance (U1–8). The ΔK and LnP (K) plots suggested that K = 2 was the optimal value (Supplementary Figure S2A), indicating that the genetic variation of the breeding families included in this study is best structured into two genetic clusters (Figure 2A). Individuals from locations L1 and L2 were placed with high probability in one cluster, whereas those from L3–17 and U1–8 were assigned to the other cluster (Figure 2A). The results indicated admixture between these two clusters in L18–21, which was strongest for L20–21 (Figure 2A). However, this pattern was less pronounced at higher levels of K (i.e. K = 3 and K = 4, Figure 2B, C).

A subsequent examination of voucher herbarium specimens collected from each location indicated that *Eucalyptus* plants from locations L1–2 (Figure 1) represent *E. melliodora* A.Cunn. ex S.Schauer instead of

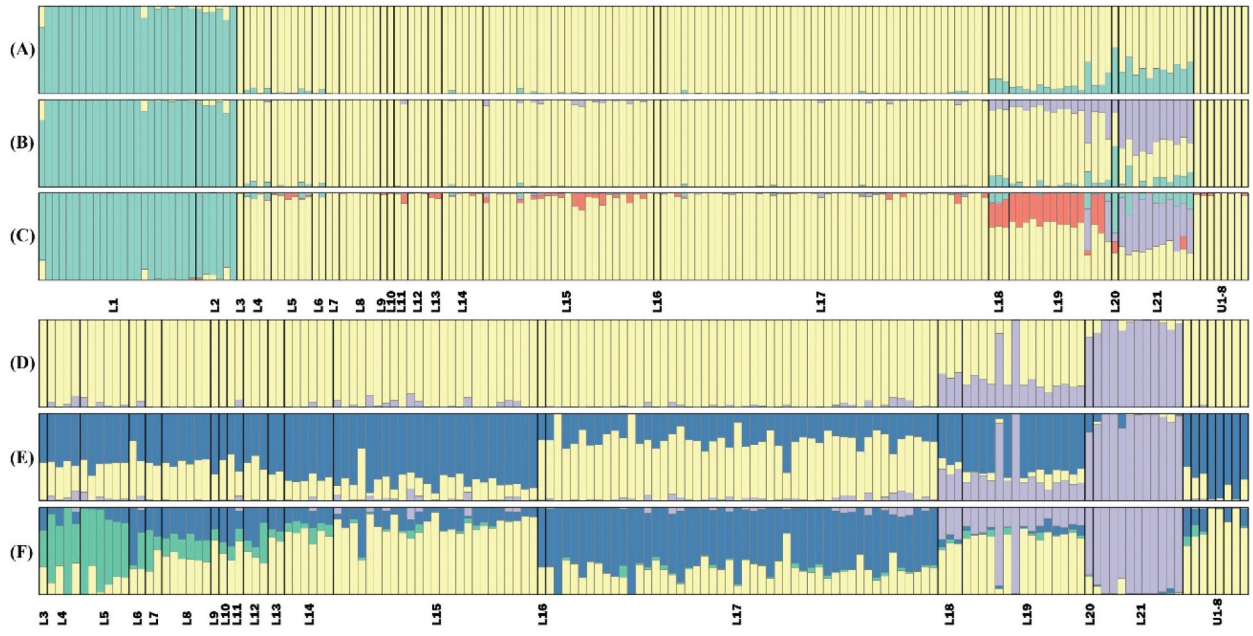

Figure 2. STRUCTURE plots of SNP data from plants grown from seeds from 21 natural source locations (L1–21) and eight families of uncertain provenance (U1–8) in the NZDFI *E. bosistoana* breeding programme. The dataset with *E. melliodora* (L1–2) included (A) K = 2, (B) K = 3 and (C) K = 4. The dataset without *E. melliodora* samples (D) K = 2, (E) K = 3 and (F) K = 4.

E. bosistoana. Because *E. bosistoana* is the targeted species of the NZDFI breeding programme, we therefore also conducted genetic structure analyses of a data set from which *E. melliodora* (L1–2) was removed (Figure 2D–F). The optimal K value for the *E. bosistoana* locations (i.e. L3–21 and U1–8) was K = 2 (Supplementary Figure S2B). At this value, L20–21 were shown as genetically distinct from L3–17 and U1–8 (Figure 2D). Plants from seeds collected from locations L18–19 presented evidence of admixture between the cluster formed by L3–17 and U1–8 and that composed of L20–21. This admixture pattern between the two main *E. bosistoana* clusters (L3–17/U1–8 and L20–21) is also visible at higher values of K (K = 3 and K = 4), with most sampling locations of L18–19 showing some degree of admixture (Figure 2E, F). Notably, families from L18 and L19 exhibited a very similar admixture pattern, even though the two locations are geographically well separated and each lies closer to a different main cluster (L18 nearer to L3–17, L19 nearer to L20–21; Figure 1).

The patterns of genetic structure obtained with PCoA (Figure 3) were congruent with those generated by STRUCTURE (Figure 2). Plants grown from seeds from locations L1–2 were identified as genetically distinct from those of all other locations (Figure 3A). Plants from L20–21 were distinct from individuals grown from L3–17 and U1–8, with L18–19 plants taking a position between these two groups in the PCoA plots (Figure 3A, B). A PCoA with only samples from L3–17 and U1–8 performed to reveal more detailed genetic structure patterns among plants from these locations (Figure 3C) showed some, albeit lower levels of genetic differentiation among locations.

Based on the genetic patterns observed in the results of the STRUCTURE analyses at K = 2–4 (Figure 2) and the PCoA results (Figure 3), four putative source populations for the breeding programme can be recognised: L1–2, L3–17, L18–19, and L20–21. Additionally, some structure was observed within population L3–17, with samples from nearby locations showing genetic similarity (Figures 2E,F and 3C).

Patterns of genetic structure were further explored by calculating global F'_{ST} values and conducting IBD analyses. The estimated global F'_{ST} values were significant at $p = 0.001$ for all three data sets. The global F'_{ST} for the data set in which *E. melliodora* (L1–2) and the three *E. bosistoana* populations (L3–17, L18–19 and L20–21) were included was 0.106. A separate AMOVA from which *E. melliodora* samples were excluded obtained an F'_{ST} value of 0.034. Pairwise F'_{ST} values among the four putative source populations were all statistically significant after Benjamini–Yekutieli correction ($p \leq 0.024$; Supplementary Table S1) (Narum 2006), ranging from 0.020 (L3–17 vs L18–19) to 0.158 (L1–2 vs L3–17). We quantified genetic structure within the largest *E. bosistoana* population (L3–17) by calculating the global F'_{ST} for a data set in which samples from locations from which we sampled four or more individuals were included (i.e. L4, L5, L8, L14, L15 and L17). The resulting F'_{ST} value was 0.023. An IBD test resulted in a statistically significant correlation between genetic and geographic distance for *E. bosistoana* ($R = 0.219$, $p = 0.03$, Supplementary Figure S3).

Genetic diversity

Overall, the four source populations of the breeding programme showed similar levels of genetic diversity (Table 1). Population L1–2 had the highest genetic diversity as measured by allelic richness (1.4), richness of private alleles (0.09) and H_e (0.212). Population L20–21 showed the highest H_o (0.163). The highest F_{IS} was found in population L3–17 (0.254) and the lowest in L20–21 (0.156).

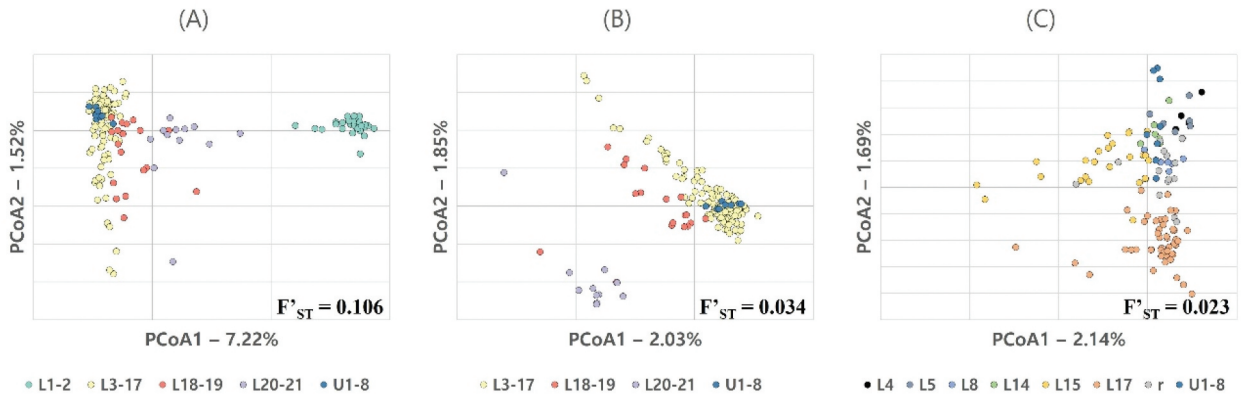

Figure 3. PCoA plots of SNP data from samples from 21 natural source locations (L1–L21) and eight families of uncertain provenance (U1–8) in the *E. bosistoana* breeding programme. L1–2 represent *E. melliodora* and L3–21 are *E. bosistoana* families. F'_{ST} values (excluding samples from U1–8; all statistically significant at $p = 0.001$) are provided for each data set. (A) Sampling locations L1–21 and U1–8; (B) L3–21 and U1–8; (C) L4, L5, L8, L14, L15, L17, U1–8 (coloured dots) and families from the remaining L3–L17 locations (r; grey dots).

Table 1. Genetic diversity indices from four natural source populations represented in the NZDFI *E. bosistoana* breeding programme: L–-2 (*E. melliodora*), L3–17, L18–19, and L20–21 (*E. bosistoana*) (Figure 1). # samples, number of samples included; Ar, allelic richness corrected for sample size; ap, richness of private alleles corrected for sample size; H_o, observed heterozygosity; H_e, expected heterozygosity; and F_{IS}, inbreeding coefficient.

Populations	# Samples	Ar	ap	H_o	H_e	F_{IS}
L1–2 (*E. melliodora*)	29	1.40	0.09	0.156	0.212	0.236
L3–17 (*E. bosistoana*)	110	1.39	0.06	0.151	0.212	0.254
L18–19 (*E. bosistoana*)	18	1.39	0.05	0.156	0.204	0.191
L20–21 (*E. bosistoana*)	12	1.40	0.05	0.163	0.200	0.156

Discussion

Identification of E. melliodora in the E. bosistoana breeding programme

STRUCTURE analyses and PCoA revealed that breeding families grown from seeds from two locations north of Canberra (L1 and L2) were genetically distinct from all other families included in our analyses (Figures 2 and 3). Subsequent investigation of their morphology indicated that these families in the NZDFI *E. bosistoana* breeding programme belong to a different *Eucalyptus* species. All stamens of *E. bosistoana* flowers are fertile (Slee et al. 2006), but flowers of the L1–2 breeding families had both fertile and infertile stamens (i.e. staminodes). Further, the anthers of *E. bosistoana* dehisce through lateral pores, but those of plants from L1–2 had terminal pores (Slee et al. 2006). We also observed that L1–2 trees in the New Zealand breeding trials showed differences in leaf morphology with those obtained from L3–21. They had lanceolate instead of oblong, elliptic or ovate juvenile leaves. These features suggest that population L1–2 represents *E. melliodora* (yellow box, honey box or yellow ironbark). Both species have been classified in *Eucalyptus* sect. *Adnataria* and may be confused because of morphological similarities (Slee et al. 2006). Two other sect. *Adnataria* species in New South Wales sharing similarities with *E. bosistoana* (i.e. *E. sideroxylon* A.Cunn. ex Woolls and *E. tricarpa* L.A.S. Johnson & K.D.Hill) resemble *E. melliodora* in having flowers with staminodes and terminal anther pores. However, both are ironbark species and therefore unlikely to be confused with *E. bosistoana* which, like *E. melliodora*, has a box-type bark. Our finding that the NZDFI *E. bosistoana* breeding programme contains some breeding families of a different species than the targeted taxon highlights the importance of verifying the taxonomic identification of collected families through preserving fertile specimens from mother trees. Since floral characteristics, including stamens, are often required for confident species identification, flowering *Eucalyptus* specimens may be necessary to ensure accurate identification.

While access to breeding families that do not belong to the primary target species of a breeding programme allows breeders to introduce or enhance traits of commercial interest through hybridisation (Kopecký et al. 2022), correct taxonomic identification is important to inform this process.

Genetic structure and diversity of E. bosistoana breeding families

The K = 2 STRUCTURE plot for our data set (Figure 2A) suggested possible admixture between population L1–2 and other families included in the analyses. This could indicate that some populations (i.e. L18–19 and L20–21) are composed of hybrids between *E. bosistoana* and *E. melliodora*, or backcrosses of these hybrids with *E. bosistoana*. However, this pattern of admixture was much less pronounced at higher values of K. This may indicate that populations L18–19 and L20–21 might not be of hybrid origin but are genetically distinct *E. bosistoana* populations instead. In addition, plants grown from seed from these populations did not display morphological characters supporting the hypothesis that they were interspecific hybrids (e.g. Barbour et al. (2005)). Nonetheless, closely related *Eucalyptus* species are known to hybridise (Potts & Dungey 2004; Larcombe et al. 2015; Bush 2022) and backcrossed hybrids can be morphologically cryptic (e.g. Sale et al. (1996); de Oliveira et al. (2023)). Therefore, additional research is needed to determine if L18–21 plants are interspecific hybrids between *E. melliodora* and *E. bosistoana*.

Under the assumption that L18–21 plants were not of hybrid origin, our STRUCTURE analysis and PCoA revealed three genetic groups within *E. bosistoana*. These groups are further supported by significant pairwise genetic differentiation among them (Supplementary Table S1). These are here considered as natural populations. All samples from L3–17 compose a population with a mainly New South Wales distribution (Figure 1), which extends into the easternmost part of Victoria. A second population (L20–21) was identified in southeast Victoria. The third population (L18–19) is geographically located between the other two populations and shows admixture between them (K = 2, Figure 2D). STRUCTURE results at higher levels of K (i.e. K = 3 and K = 4; Figure 2) as well as the results of the PCoA analyses (Figure 3C) showed that genetic clustering largely reflected geographic

relationships between locations, with families from geographically proximal sites tending to group together. Overall, our results therefore showed that breeding families from nearby sampling areas are genetically less differentiated than those from areas further away. Evidence of IBD for families from locations L3–21 provided further support for this pattern. Together with the statistically significant but generally low global F'_{ST} values among different locations (Figure 3), this suggested that geographic distance among locations is a driver of genetic structure in *E. bosistoana*, but that this species has relatively high levels of genetic connectivity. This finding is consistent with the results of studies of other *Eucalyptus* species. For instance, low genetic differentiation has been reported in *E. camaldulensis* Dehnh (Dillon et al. 2014), *E. tricarpa* L.A.S.Johnson & K.D.Hill (Andrew et al. 2010), and *E. urophylla* S.T.Blake (Lu et al. 2018), for which limited genetic differentiation among populations was explained by high levels of genetic connectivity.

The NZDFI breeding programme includes eight *E. bosistoana* families with locations in western Victoria originally recorded as their provenance (U1–8). A morphological examination (results not presented) and their genomic similarity with southeast Victorian and New South Wales *E. bosistoana* populations (L3–21; Figures 2 and 3) confirm that these families indeed represent this species. However, *E. bosistoana* has not previously been reported from western Victoria, and its distribution is instead considered to be restricted to areas east of Melbourne (Chippendale 1988; Slee et al. 2006; Flores-Renteria et al. 2017). Other data associated with these eight families, however, suggested that their seed lots might instead have been obtained from eastern Victoria (i.e. East Gippsland). The latter hypothesis is well supported by our genomic data. Our STRUCTURE results (Figure 2F) suggested that U1–8 are genomically most similar to trees from seeds from L14 and L15, which are in East Gippsland (L15) and a nearby area in Southern Tablelands (L14). This finding is relevant for the breeding programme, because knowing their likely natural source location would enable enriching the *E. bosistoana* breeding population by collecting seeds from additional mother trees from these sites in the future.

The three *E. bosistoana* source populations used to establish the NZDFI breeding population had similar levels of genetic diversity (Table 1). This implies that differences in genetic diversity among source populations do not need to be taken into consideration when determining the relative numbers of families from specific source populations that should be added or retained in the breeding population to optimise its genetic diversity. All *E. bosistoana* source populations contained fewer heterozygotes than expected under the assumption that they are in Hardy-Weinberg equilibrium, with similar F_{IS} values among them (Table 1). This heterozygote deficiency is likely attributable to the mixed mating system of the species, which involves both self-pollination and outcrossing (Phillips & Brown 1977; Breed et al. 2015; Griffin et al. 2019). This interpretation is supported by the generally weak genetic differentiation among source locations and the consistent heterozygote deficiency observed in other *Eucalyptus* species (including *E. melliodora* in the present study; Table 1) (Jones et al. 2005; Mora et al. 2017).

This study provided insights into patterns of genetic structure and diversity among Australian natural source populations of a New Zealand *E. bosistoana* breeding programme. The results of this study provide the first detailed genetic baseline for the NZDFI *E. bosistoana* breeding population. They can be used to make informed decisions on which families or individuals to select, retain or cross when developing breeding lines. For example, breeders can preferentially select genotypes from the three identified source populations (L3–17, L18–19 and L20–21) to maximise genetic diversity and minimise inbreeding or prioritise high-performing families within each genetic cluster based on both genomic and phenotypic data. However, maintaining a breeding population that represents the genetic diversity of a species throughout its natural distribution area can also have benefits beyond commercial forestry. Most of the *E. bosistoana* locations included in our study (L3–L14) are within the Lowland Grassy Woodland ecosystem in southeastern Australia, which is experiencing significant fragmentation and a reduction in ecological connectivity (Manning et al. 2020). Although *E. bosistoana* is not currently threatened, the NZDFI breeding populations in New Zealand could potentially serve as an *ex-situ* resource to support future conservation efforts if needed, by providing genetic material for restoration programmes aimed at increasing the adaptive potential of natural populations (Valbuena-Urena et al. 2017). In turn, the conservation of *in situ* genetic diversity in natural source populations can be of benefit to commercial forestry by providing access to more genetic diversity than what can be maintained in a breeding programme, which is needed to ensure its long-term productivity (Eldridge 1990).

The families used in this study were originally collected for commercial forestry purposes rather than as part of a systematic range-wide genetic survey. Consequently, sample sizes varied considerably among source locations, ranging from 1 to 37 individuals per location (Figure 1) and from 11 to 95 individuals among the four identified source populations (Table 1). Such unbalanced sampling can potentially affect estimates of genetic diversity and population structure. To mitigate the impact of uneven sample sizes on genetic diversity estimates, we used rarefaction-based allelic richness (Ar) and private allelic

richness (ap) measures, which standardise comparisons across populations with different sample sizes (Kalinowski 2005). For AMOVA analyses, we only included source locations with at least four individuals to ensure more reliable estimates of genetic differentiation. STRUCTURE is generally robust to uneven sampling when population structure is strong (Puechmaille 2016), and our main findings (i.e. the distinction between *E. melliodora* [L1–2] and *E. bosistoana* [L3–21] and the differentiation between L3–17 and L20–21) were consistently supported by multiple independent analyses including PCoA and F'_{ST} estimates. Nevertheless, we acknowledge that the smaller sample sizes in some populations (e.g. L20–21 with 12 individuals) may limit the precision of diversity estimates for those populations, and that more balanced sampling across the distribution range of the species would strengthen future genetic assessments.

Conclusion

Understanding patterns of genetic structure and diversity in natural tree populations is important for tree breeders as it helps in selecting suitable source populations, maintaining genetic diversity, and enhancing resilience to pests, diseases, and climate change. This knowledge supports breeding programmes by ensuring a broad genetic base, which is essential for improving commercially valuable traits and preventing inbreeding depression. This study examined patterns of genetic structure and diversity of *E. bosistoana* families in the New Zealand Dryland Forests Innovation (NZDFI) breeding programme. Using a high-density SNP array, we determined that the breeding population was established from seed lots originating from three natural populations in Australia. Our data show relatively little genetic structure within these populations among the different locations from with seeds were collected, but families from nearby locations were generally more genetically similar to each other than to families from more remote locations. All three populations have similar levels of genetic diversity. The results of our genomic analyses also revealed that some families in the *E. bosistoana* breeding programme represent *E. melliodora* instead, and helped clarify the provenance of several *E. bosistoana* families. The genetic structure and diversity patterns reported here provide a baseline against which future changes in genetic diversity, inbreeding levels and the effects of selection can be measured in subsequent generations of the NZDFI *E. bosistoana* breeding programme.

Acknowledgements

Ben Boxshall, Gary Featherston, Paul Millen and Dean Nicolle are acknowledged for providing information about the presence of *E. bosistoana* in Victoria. Dean Nicolle also confirmed some of our *E. bosistoana* and *E. melliodora* identifications. Matthew Alexandra assisted with information about the origin of some locations. Rebecca Jones and Kyu-Suk Kang provided feedback on the PhD thesis that contained an earlier draft of this publication. We thank Chamira Rajapaksha and Ruth McConnochie for their assistance with collecting leaf tissue samples, and Julie Barcelona for her contributions to both sample collection and DNA extraction. Seol-Jong Kim acknowledges the Korea-New Zealand Agricultural Cooperation Scholarships (KNZACS), funded by the governments of the Republic of Korea and New Zealand, for supporting his doctoral studies.

Disclosure statement

The authors have no relevant financial interests to disclose. Clemens Altaner is on the Board of Directors of New Zealand Dryland Forests IP Limited.

Funding

This work was financially supported by the New Zealand's Ministry of Business, Innovation and Employment (MBIE) Specialty Wood Products (SWP) Partnership [FFRX1501].

ORCID

S.-J. Kim http://orcid.org/0000-0001-7924-4072
T. E. Steeves http://orcid.org/0000-0003-2112-5761
C. M. Altaner http://orcid.org/0000-0003-4542-4432
L. A. Apiolaza http://orcid.org/0000-0003-0958-3540
P. B. Pelser http://orcid.org/0000-0002-6990-1419

Data availability statement

The data that support the findings of this study are available from the corresponding authors, Seol-Jong Kim and Pieter B. Pelser, upon reasonable request.

References

Allier A, Teyssedre S, Lehermeier C, Moreau L, Charcosset A. 2020. Optimized breeding strategies to harness genetic resources with different performance levels. BMC Genomics. 21(1):349. doi: 10.1186/s12864-020-6756-0.

Andrew RL, Wallis IR, Harwood CE, Foley WJ. 2010. Genetic and environmental contributions to variation and population divergence in a broad-spectrum foliar defence of *Eucalyptus tricarpa*. Ann Bot. 105(5):707–717. doi: 10.1093/aob/mcq034.

Barbour RC, Potts BM, Vaillancourt RE. 2005. Pollen dispersal from exotic eucalypt plantations. Conserv Genet. 6 (2):253–257. doi: 10.1007/s10592-004-7849-z.

Bootle K. 2005. Wood in Australia. Types, properties and uses. 2nd ed. Sydney, Australia: McGraw-Hill Australia.

Breed MF, Ottewell KM, Gardner MG, Marklund MH, Stead MG, Harris JB, Lowe AJ. 2015. Mating system and early viability resistance to habitat fragmentation in a bird-pollinated eucalypt. Heredity (Edinb). 115(2):100–107. doi: 10.1038/hdy.2012.72.

Bush D. 2022. Long-term research reveals potential role of hybrids in climate-change adaptation. A commentary on 'expansion of the rare *Eucalyptus risdonii* under climate

change through hybridisation with a closely related species despite hybrid inferiority'. Ann Bot. 129(1):i–iii. doi: 10.1093/aob/mcab085.

Chippendale GM. 1988. Flora of Australia, Myrtaceae, *Eucalyptus*, *Angophora* (Myrtaceae) Vol. 19. Canberra (Australia): Australian Government Publishing Service Canberra.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of Samtools and Bcftools. GigaScience. 10(2). doi: 10.1093/gigascience/giab008.

Davies NT, Apiolaza LA, Sharma M. 2017. Heritability of growth strain in *Eucalyptus bosistoana*: a Bayesian approach with left-censored data. N Z J For Sci. 47 (1):1–5. doi: 10.1186/s40490-017-0086-2.

de Oliveira DA, da Silva PHM, Novaes E, Grattapaglia D. 2023. Genome-wide analysis highlights genetic admixture in exotic germplasm resources of *Eucalyptus* and unexpected ancestral genomic composition of interspecific hybrids. PLoS ONE. 18(8):e0289536. doi: 10.1371/journal.pone.0289536 .

Dillon S, McEvoy R, Baldwin DS, Rees GN, Parsons Y, Southerton S. 2014. Characterisation of adaptive genetic diversity in environmentally contrasted populations of *Eucalyptus camaldulensis* dehnh. (river red gum). PLoS ONE. 9(8):e103515. doi: 10.1371/journal.pone.0103515.

Eldridge KG. 1990. Conservation of forest genetic resources with particular reference to *eucalyptus* species. Commonw For Rev. 69(1 (218)):45–53.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 14(8):2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x.

Flores-Renteria L, Rymer PD, Riegler M. 2017. Unpacking boxes: integration of molecular, morphological and ecological approaches reveals extensive patterns of reticulate evolution in box eucalypts. Mol Phylogenetics Evol. 108:70–87. doi: 10.1016/j.ympev.2017.01.019.

Griffin AR, Potts BM, Vaillancourt RE, Bell JC. 2019. Life cycle expression of inbreeding depression in *Eucalyptus regnans* and inter-generational stability of its mixed mating system. Ann Bot. 124(1):179–187. doi: 10.1093/aob/mcz059.

Hamilton JA, Miller JM. 2016. Adaptive introgression as a resource for management and genetic conservation in a changing climate. Conserv Biol. 30(1):33–41. doi: 10.1111/cobi.12574.

Janes JK, Miller JM, Dupuis JR, Malenfant RM, Gorrell JC, Cullingham CI, Andrew RL. 2017. The k = 2 conundrum. Mol Ecol. 26(14):3594–3602. doi: 10.1111/mec.14187.

Jing Y, Bian L, Zhang X, Zhao B, Zheng R, Su S, Ye D, Zheng X, El-Kassaby YA, Shi J. 2023. Genetic diversity and structure of the 4th cycle breeding population of Chinese fir (*Cunninghamia lanceolata* (Lamb.) Hook). Front Plant Sci. 14:1106615. doi: 10.3389/fpls.2023.1106615.

Johnson R, St. Clair B, Lipow S. 2001. Genetic conservation in applied tree breeding programs. Proceedings ITTO conference on in situ and ex situ conservation of commercial tropical trees; Yokohama, Japan. p. 215–230.

Jones R, McKinnon G, Potts B, Vaillancourt R. 2005. Genetic diversity and mating system of an endangered tree *Eucalyptus morrisbyi*. Aust J Bot. 53(4):367–377. doi: 10.1071/BT04182.

Kalinowski ST. 2005. Hp-rare 1.0: a computer program for performing rarefaction on measures of allelic richness. Mol Ecol Notes. 5(1):187–189. doi: 10.1111/j.1471-8286.2004.00845.x.

Kamvar ZN, Brooks JC, Grunwald NJ. 2015. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Front Genet. 6:208. doi: 10.3389/fgene.2015.00208.

Kamvar ZN, Tabima JF, Grunwald NJ. 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ. 2:e281. doi: 10.7717/peerj.281.

Knaus BJ, Grunwald NJ. 2017. Vcfr: a package to manipulate and visualize variant call format data in R. Mol Ecol Resour. 17(1):44–53. doi: 10.1111/1755-0998.12549.

Kopecký D, Martín A, Smýkal P. 2022. Interspecific hybridization and plant breeding: from historical retrospective through work of Mendel to current crops. Czech J Genet Plant Breed. 58(3):113–126. doi: 10.17221/19/2022-CJGPB .

Larcombe MJ, Holland B, Steane DA, Jones RC, Nicolle D, Vaillancourt RE, Potts BM. 2015. Patterns of reproductive isolation in *eucalyptus*— a phylogenetic perspective. Mol Biol Evol. 32(7):1833–1846. doi: 10.1093/molbev/msv063.

Lawson DJ, van Dorp L, Falush D. 2018. A tutorial on how not to over-interpret structure and admixture bar plots. Nat Commun. 9(1):3258. doi: 10.1038/s41467-018-05257-7.

Lefèvre F, Boivin T, Bontemps A, Courbet F, Davi H, Durand-Gillmann M, Fady B, Gauzere J, Gidoin C, Karam M-J, et al. 2014. Considering evolutionary processes in adaptive forestry. Ann For Sci. 71(7):723–739. doi: 10.1007/s13595-013-0272-1.

Li Y, Apiolaza LA, Altaner C. 2018. Genetic variation in heartwood properties and growth traits of *Eucalyptus bosistoana*. Eur J For Res. 137(4):565–572. doi: 10.1007/s10342-018-1125-0.

Linck E, Battey CJ. 2019. Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. Mol Ecol Resour. 19(3):639–647. doi: 10.1111/1755-0998.12995.

Lischer HE, Excoffier L. 2012. Pgdspider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics. 28(2):298–299. doi: 10.1093/bioinformatics/btr642.

Lu W, Arnold RJ, Zhang L, Luo J. 2018. Genetic diversity and structure through three cycles of a *Eucalyptus urophylla* S. T.Blake breeding program. Forests. 9(7):372. doi: 10.3390/f9070372.

Manning AD, Cunningham RB, Tongway D, Lindenmayer DB. 2020. Woodlands and woody debris: understanding structure and composition to inform restoration. PLoS ONE. 15 (3):e0224258. doi: 10.1371/journal.pone.0224258.

Mantel N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Res. 27(2 Part 1):209–220.

Meirmans PG. 2015. Seven common mistakes in population genetics and how to avoid them. Mol Ecol. 24 (13):3223–3231. doi: 10.1111/mec.13243.

Mora F, Arriagada O, Ballesta P, Ruiz E. 2017. Genetic diversity and population structure of a drought-tolerant species of *Eucalyptus*, using microsatellite markers. J Plant Biochem Biotechnol. 26(3):274–281. doi: 10.1007/s13562-016-0389-z.

Narum SR. 2006. Beyond Bonferroni: less conservative analyses for conservation genetics. Conserv Genet. 7 (5):783–787. doi: 10.1007/s10592-005-9056-y.

Nicholas I, Millen P. 2012. Durable eucalypt leaflet series: *Eucalyptus bosistoana*. Marlborough (New Zealand): NZDFI, Ed.

Peakall R, Smouse PE. 2006. Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes. 6(1):288–295. doi: 10.1111/j.1471-8286.2005.01155.x.

Peakall R, Smouse PE. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics. 28(19):2537–2539. doi: 10.1093/bioinformatics/bts460.

Phillips M, Brown A. 1977. Mating system and hybridity in *Eucalyptus pauciflora*. Aust J Biol Sci. 30(4):337–344. doi: 10.1071/BI9770337.

Pina-Martins F, Silva DN, Fino J, Paulo OS. 2017. Structure_threader: an improved method for automation and parallelization of programs structure, fastStructure and MavericK on multicore CPU systems. Mol Ecol Resour. 17(6):e268–e274. doi: 10.1111/1755-0998.12702.

Potts BM, Dungey HS. 2004. Interspecific hybridization of *eucalyptus*: key issues for breeders and geneticists. New For. 27(2):115–138. doi: 10.1023/A:1025021324564.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics. 155(2):945–959. doi: 10.1093/genetics/155.2.945.

Puechmaille SJ. 2016. The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. Mol Ecol Resour. 16(3):608–627. doi: 10.1111/1755-0998.12512.

Rajapaksha C, Apiolaza L, Squire M, Altaner C. 2023a. Genetic parameters of essential-oil traits for *Eucalyptus bosistoana*. Aust For. 86(2):90–99. doi: 10.1080/00049158.2023.2270681.

Rajapaksha C, Greaves P, Altaner CM. 2023b. Economic potential of essential oil production from New Zealand-grown *Eucalyptus bosistoana*. Sci Rep. 13 (1):14083. doi: 10.1038/s41598-023-40632-5.

R Core Team. 2021. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW. 2002. Genetic structure of human populations. Science. 298(5602):2381–2385. doi: 10.1126/science.1078311.

RStudio Team. 2021. Rstudio: Integrated development environment for R. Boston (MA): RStudio, PBC.

Sale M, Potts B, West A, Reid J. 1996. Molecular differentiation within and between *Eucalyptus risdonii*, *E. amygdalina* and their hybrids using RAPD markers. Aust J Bot. 44 (5):559–569. doi: 10.1071/BT9960559.

Schmidt TL, Jasper ME, Weeks AR, Hoffmann AA. 2021. Unbiased population heterozygosity estimates from genome-wide sequence data. Methods Ecol Evol. 12 (10):1888–1898. doi: 10.1111/2041-210X.13659.

Sharma M, Dijkstra SM, Garrill A, Collings DA, Altaner CM. 2024. Assessing wood grain and twist in a 2-year-old *Eucalyptus bosistoana* breeding population. Trees. 38 (6):1481–1489. doi: 10.1007/s00468-024-02567-w.

Slee A, Brooker M, Duffy S, West J. 2006. Euclid - eucalypts of Australia. Melbourne: CSIRO. accessed 2025 May 11. https://apps.lucidcentral.org/euclid/text/intro/index.html.

Valbuena-Urena E, Soler-Membrives A, Steinfartz S, Alonso M, Carbonell F, Larios-Martin R, Obon E, Carranza S. 2017. Getting off to a good start? Genetic evaluation of the ex situ conservation project of the critically endangered Montseny brook newt (*Calotriton arnoldi*). PeerJ. 5:e3447. doi: 10.7717/peerj.3447.

Werner CR, Gaynor RC, Gorjanc G, Hickey JM, Kox T, Abbadi A, Leckband G, Snowdon RJ, Stahl A. 2020. How population structure impacts genomic selection accuracy in cross-validation: implications for practical breeding. Front Plant Sci. 11:592977. doi: 10.3389/fpls.2020.592977.