



OPEN

Genetic assessment of the value of restoration planting within an endangered eucalypt woodland

Natalie L. Rosser¹, Anthony Quinton¹, Huw Davey², David J. Ayre¹ & Andrew J. Denham^{1,3}✉

Assessment of woodland restoration often focusses on stand demographics, but genetic factors likely influence long-term stand viability. We examined the genetic composition of Yellow Box (*Eucalyptus melliodora*) trees in endangered Box-Gum Grassy Woodland in SE Australia, some 30 years after planting with seeds of reportedly local provenance. Using DArT sequencing for 1406 SNPs, we compared genetic diversity and population structure of planted *E. melliodora* trees with remnant bushland trees, paddock trees and natural recruits. Genetic patterns imply that natural stands and paddock trees had historically high gene flow (among group pairwise $F_{ST} = 0.04$ – 0.10). Genetic diversity was highest among relictual paddock trees ($H_e = 0.17$), while diversity of revegetated trees was identical to natural bushland trees ($H_e = 0.14$). Bayesian clustering placed the revegetated trees into six genetic groups with four corresponding to genotypes from paddock trees, indicating that revegetated stands are mainly of genetically diverse, local provenance. Natural recruits were largely derived from paddock trees with some contribution from planted trees. A few trees have likely hybridised with other local eucalypt species which are unlikely to compromise stand integrity. We show that paddock trees have high genetic diversity and capture historic genetic variety and provide important foci for natural recruitment of genetically diverse and outcrossed seedlings.

In many parts of the world, woodland ecosystems have been extensively cleared for agriculture and urban development, often leaving only small remnants. As a result, trillions of dollars are spent on ecological restoration and revegetation planting programs each year to restore biodiversity to degraded landscapes^{1–3}. The goal of restoration is to create a self-supporting ecosystem that is resilient to adversity⁴, but the effectiveness of ecological restoration programs is dependent upon a suite of population genetic issues that influence the success of ecological restoration and revegetation planting programs^{5–7}. First, the level of population subdivision among remnant or revegetated populations has implications for the level and distribution of genetic diversity, and the likelihood that restoration will exacerbate inbreeding or outbreeding and may produce inbreeding or outbreeding depression^{8–11}. In the long term, natural genetic subdivision for any species will reflect the dynamics of pollen and seed dispersal and patterns of extinction and colonisation. However, in restoration programs, fragmentation will initially be determined by patterns of planting and consequently by potentially atypical gene flow (e.g.¹¹). If there are high levels of subdivision and low gene flow among natural stands, restoration planning should consider the likelihood of site-specific adaptation in sourcing material for transplantation¹². Second, remnant or revegetated populations may lack sufficient or appropriate genetic diversity as a result of genetic drift, founder effects, inbreeding and inappropriate sourcing of material for transplants. Fitness can be reduced in inbred offspring either because of loss of beneficial heterozygous traits, or by the expression of recessive deleterious traits, and can include reduced seed crops, smaller seeds, poor seedling survival and lower capacity to respond to environmental change (reviewed in¹³). Restoration programs should typically aim to avoid elevated inbreeding due to small stand sizes and/or sourcing seed or seedlings from small numbers of closely related individuals. Such programs should also ensure that revegetation programs have at least natural levels of genetic diversity because it is positively associated with population fitness in the short-term, and adaptive capacity and resilience to environmental change in

¹Centre for Sustainable Ecosystem Solutions, School of Earth, Atmospheric and Life Sciences, University of Wollongong, Wollongong, NSW, Australia. ²Independent Researcher, Wollongong, NSW, Australia. ³Science, Economics and Insights Division, NSW Department of Planning and Environment, Parramatta, NSW, Australia. ✉email: andrew.denham@environment.nsw.gov.au

the long-term^{14,15}. Third, many species that are targets for woodland restoration have been shown to hybridize with native and exotic species^{16–20} which can be a problem for restoration programs since first or later generation hybrids may have lower viability than pure-bred individuals¹⁶, and hybridization will also limit the value of restoration programs if conserving pure gene pools is a priority²¹. It is important to note however, that a range of studies have argued the benefits of utilizing first generation hybrids in restoration programs if they possess beneficial traits such as drought tolerance or elevated resistance to pests or climate change^{22–24}.

Eucalypt woodlands are a key component of the Australian landscape and were once widespread in southern Australia. However, large areas have been extensively cleared since European settlement, and in many landscapes where clearing has reduced the extent of remnant woodlands, remnant paddock trees are a key resource for woodland restoration^{13,25}. Eucalypts have a mixed mating system, but are predominantly outcrossed via insect pollination with outcrossing rates typically between 0.44 and 0.96²⁶. Interestingly, the degree of outcrossing within a single tree can vary from the top down due to pollinator foraging patterns as they move through the canopy²⁷. While eucalypt pollen can travel up to 2 km²⁸, more typical dispersal distances are around 200 m^{13,29,30}. These characters imply that even moderate spatial isolation will increase the likelihood of selfing or inbreeding for paddock trees and remnant stands and that populations will display evidence of subdivision. Planning or evaluating the success of restoration programs is further complicated by the fact that eucalypts can sometimes hybridize^{16,17}, though the likelihood of hybridization can vary dramatically with a range of factors including the genetic distance between potentially hybridizing taxa³¹, their relative and absolute density, and pollen dispersal distances^{16,32}.

In eastern Australia, the White Box-Yellow Box-Blakely's Red Gum Grassy Woodland ecosystem (known as Box-Gum Grassy Woodland) is critically endangered and has been the target of restoration programs that have been largely unassessed (but see^{13,33,34}. Broadhurst's study¹³ using genetic (microsatellite) data to assess the diversity and mating systems of small numbers of scattered and revegetated *Eucalyptus melliodora* (Yellow Box) trees is an important exception. Broadhurst¹³ collected samples from five restoration sites each within 200 km of Canberra in the Australian Capital Territory. Despite considerable among-site variation, both in the design of the restoration plantings and proximity to paddock trees, that study highlighted the importance of paddock trees as sources of genetic diversity and as pollen sources for seed on revegetated trees¹³. Broadhurst¹³ also found considerable differentiation among paddock and planted trees and concluded that many were not of local provenance. We set out to build on Broadhurst's work by evaluating, from a genetic perspective, the success of the restoration program in Warrumbungle National Park (WNP) where there is a set of *E. melliodora* plantings that were carried out over approximately 10 years to supplement a small number of remnant paddock trees that are potentially isolated from surrounding natural stands and that have formed foci for some natural regeneration. The restoration program is located within a 500 ha valley that had been extensively cleared for pastoral activity and grazed from the 1900s until the 1960s. Although we have observed in situ seed set and recruitment in the valley area, it is not known whether there is sufficient genetic diversity to maintain populations in the long term, or whether inbreeding is likely to constrain their persistence. Moreover, plantings are thought to have been locally sourced, but this has not been confirmed genetically, nor is anything known of the genetic make-up of either natural or revegetated stands, their mating systems or extent of genetic connectivity throughout the system.

Our specific objectives were to: (a) assess and compare the genetic differentiation and mating systems of natural stands surrounding the restoration area with that of surviving paddock trees, recruits and planted stands to estimate levels of connectedness; (b) determine whether paddock trees, recruits and planted trees capture similar levels of genetic diversity to the natural stands; and (c) use these data to infer the provenance of planted trees and to test for evidence of hybridization.

Methods

Study system. This study focused on Yellow Box (*Eucalyptus melliodora*) in Box-Gum Grassy Woodlands in WNP. Box-Gum Grassy Woodlands are critically endangered and are protected by Australian Commonwealth³⁵ and NSW state legislation³⁶. Due to the occurrence of these woodlands on fertile soils, they were extensively cleared for agriculture, and intact remnants are now extremely rare. Current estimates indicate that only 405,000 ha of the woodland ecosystem remain (~10%), with few high-quality remnants anywhere across its former range³⁷. These woodlands are characterised by a dominance or prior dominance of White Box (*Eucalyptus albens*) and/or Yellow Box (*E. melliodora*) and/or Blakely's Red Gum (*E. blakelyi*) trees. Box-Gum Grassy Woodland was previously abundant within the central valley of WNP (31°17'S, 149°00'E), but extensive land clearing led to the widespread removal of trees prior to the establishment of the reserve. Scattered relictual *E. melliodora* paddock trees remain in the central valley. These trees were often solitary and were distributed broadly across the valley.

Warrumbungle National Park covers an area of 23,312 ha on the north-west slopes of NSW, covering the western end of the Warrumbungle Range. A series of revegetation projects were conducted within the central valley of the reserve from 1983 to restore the previously abundant Box-Gum Grassy Woodland. The seeds for these plantings are believed to have been sourced from existing trees within the reserve. However, the number of paddock trees used or how broadly seeds were collected across the valley, is not known.

Eucalyptus melliodora sampling. We collected a total of 221 leaf samples of *E. melliodora* trees from 14 sites across the central valley and the uncleared valley sides (Fig. 1). We selected samples from six remnant populations on the edge of valley (n = 60; termed "natural stands"); from approximately half of the paddock trees in the valley (n = 36; termed "paddock"); from two populations of recruits, each one growing near one of the paddock trees (n = 48; termed "recruits"); and from five populations of planted trees that were planted at different years in the restoration project (n = 77; termed "planted").

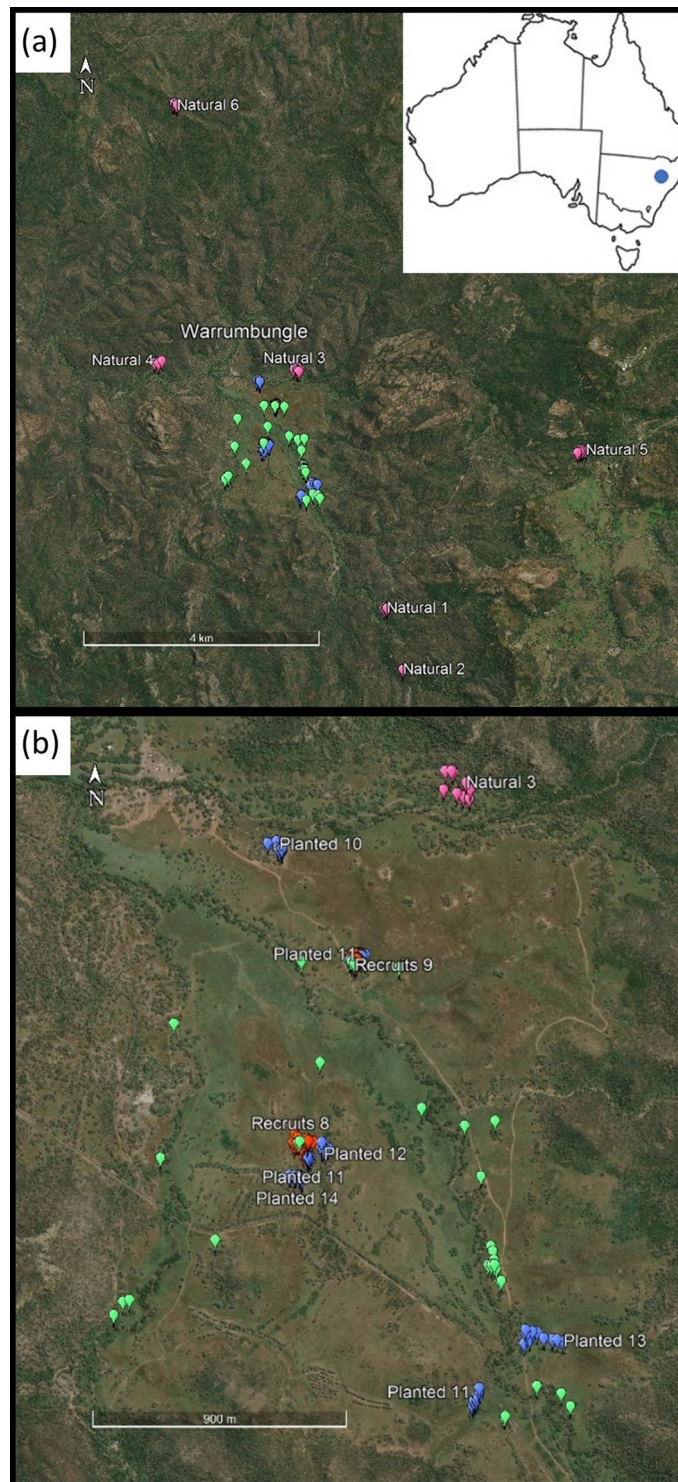


Figure 1. Location of sampled *Eucalyptus melliodora* plants in Warrumbungle National Park. Different colours represent different population types: Natural populations—pink; Paddock trees—green; Planted trees—blue; Recruits—orange. (a) Overview showing the sample locations of natural populations and the previously cleared central valley where planting occurred. The insert shows the location of the study site within NSW, Australia. (b) Close up of the central valley showing the sampled planted populations (grouped by planting year), the two groups of recruits and relictual paddock trees. Map created using Google Earth Pro 7.3.4.8642 (www.google.com/earth/versions/#earth-pro).

We selected samples from remnant populations on the valley sides at a distance > 1 km from plantings on the uncleared slopes of the valley or separate to the central valley, with a preference for larger (likely older) trees, to increase the probability that these trees pre-date the germination of planted trees and hence provide a better reference for remnant *E. melliodora* within the region. Relictual paddock trees were distinguished from planted trees by comparing aerial photos taken prior to the establishment of the reserve (Supplementary Fig. S1) with contemporary imagery. Paddock trees were easily relocated in the cleared central valley as they are large (20–30 m tall) and often have associated woody debris, while planted trees were much smaller than paddock trees (6–15 m tall) and were generally planted in rows, often with adjacent timber posts. The five planted populations were planted in different years, with population 11 planted in 1995, population 12 in 1998 and population 14 planted in 1992, while the dates that the other two populations (10 and 13) were planted are unknown. Recruits were smaller still (0.5–5 m tall) and occurred in scattered locations, often in a ‘shadow’ surrounding adult paddock trees. Field identification was done by one of us with relevant training and experience (AJD). Voucher specimens were lodged at the Janet Cosh Herbarium at the University of Wollongong with identification formally confirmed by Patricia Nagle (WOLL12125 *E. albens* and WOLL12124 *E. melliodora*).

DART genotyping. We used DARTseq services (Diversity Arrays Technology, Canberra) for both DNA extraction from leaf material, and sequencing of SNPs (Single Nucleotide Polymorphisms) using a high-density microarray developed for eucalypts^{38,39}. Extraction of genomic DNA was conducted in accordance with a modified CTAB protocol produced by DART.

Complexity reduction was conducted through a *PstI/TaqI* based method developed by³⁸, using enzymatic breakdown to select for more active genomic regions and remove repeat sequences. Next Generation Sequencing (NGS) was used to detect SNP polymorphisms across the genome of each sample against a library developed for eucalypts³⁸ detecting dominant, biallelic SNP loci. Polymorphic SNP loci were sequenced across the genome all leaf samples, using the DARTseq microarray at a ‘high intensity’ run. Sequencing was carried out on an Illumina HiSeq 2,500 using 75-cycle single end reads. Raw reads were processed using DART’s proprietary variant calling pipeline, DARTsoft-14.

Data analysis. We filtered the primary dataset at a stringent level to ensure only high-quality markers were retained, and therefore genotypes were accurate. Loci were filtered in DART⁴⁰ using the following filters: secondaries removed, read depth ≥ 10 , call rate ≥ 0.9 , minor allele frequency ≥ 0.01 , reproducibility rate ≥ 0.99 , monomorphic loci removed and individuals with 20% missing data were removed. In addition to the excluding loci with a reproducibility rate of < 0.99 , we also included 10 duplicate samples in the sequencing run (i.e. 10 individuals were sequenced twice) to check the reliability of the sequencing results. For each duplicate we compared the sequencing result at each locus (0, 1, 2 or ‘-’ for no result) and counted the number of times this result differed between the replicates, then divided this by the total number of loci.

To analyse population structure, we calculated pairwise F_{ST} in DART⁴⁰, and conducted a Principal Components Analysis (PCA) in Adegenet⁴¹ together with a Bayesian cluster analysis in the software *ParallelStructure* 2.3.4 on the CIPRES portal⁴² to identify the number of genetic clusters in the dataset. We used the admixture model of ancestry with correlated allele frequencies, a burn-in length of 80,000 followed by 120,000 MCMC replicates after burn-in, and we conducted 5 iterations each of $K = 1$ to $K = 14$. To estimate the number of genetic clusters (K), we used *StructureSelector*⁴³, and considered DeltaK⁴⁴ and MedMeaK, MaxMeaK, MedMedK and MaxMedK⁴⁵. A subpopulation is considered to belong to a cluster if its arithmetic mean (for MedMeaK and MaxMeaK) or its median (for MedMedK and MaxMedK) membership coefficient to that cluster is greater than a threshold value (set to 0.5⁴⁵).

To estimate genetic diversity and mating system, we calculated observed heterozygosity and expected heterozygosity under Hardy–Weinberg equilibrium, and F (inbreeding coefficient) in GenAEx 6.51b⁴⁶. To detect hybrids, we compared the genotypes of a subset of *E. melliodora* trees with other trees in the study site. Included in this analysis were 61 samples of *E. melliodora* selected randomly from all populations in the main dataset (except recruits) including natural stands, paddock trees, and planted trees, together with 42 samples of *E. albens*, and 11 samples from other species growing in close proximity including *E. crebra* and *E. blakelyi* (labelled “other spp.”). *Eucalyptus melliodora*, *E. albens* and *E. crebra* all belong in *Eucalyptus* Section *Adnataria* and are relatively likely to hybridise with each other^{31,47}. *Eucalyptus blakelyi* is from Section *Exsertaria* and is less likely to hybridise with these species³¹. Loci were filtered as above (read depth of ≥ 10 , a call rate of ≥ 0.9 , and minor allele frequency of ≥ 0.01 and a reproducibility rate of ≥ 0.99), and a genetic cluster analysis was conducted in *ParallelStructure*. We used the admixture model of ancestry with correlated allele frequencies, a burn-in length of 80,000 followed by 120,000 MCMC replicates after burn-in, and we conducted 5 iterations each of $K = 1$ to $K = 6$. To estimate the number of genetic clusters (K), we used the Puechmaille method⁴⁵ on *StructureSelector*⁴³, which calculates MedMeaK, MaxMeaK, MedMedK and MaxMedK.

Ethical approval. The experimental research and field studies on plants (either cultivated or wild), including the collection of plant material, complied with relevant institutional, national, and international guidelines and legislation. Samples were collected under NSW Scientific Licence SL102262.

Results

Population structure of *E. melliodora* trees. The DARTseq high-throughput microarray yielded 64,416 polymorphic SNP loci across the 221 samples. This data was of a high quality with 35.2% of loci at a call rate $\geq 85\%$, and an average call rate of 63.5%. A total of 219 individuals and 1406 loci were retained under our high stringency filters. Out of the 10 duplicate samples, the average similarity was 99.4% (9 duplicates had an

average of 99.97% and one had an average of 94.3%). In all cases, the difference between the duplicates was due to a difference in the level of missing data (i.e. the number of loci that were scored), with no instances of loci being genotyped differently between the duplicates.

In combination, the results of our pairwise F_{ST} analysis, PCA and Bayesian clustering analysis in *Structure* indicated that there was moderate genetic subdivision across the entire set of sampled *E. melliodora* trees. Global F_{ST} across all sites was 0.11, while pairwise F_{ST} values ranged from 0.01 to 0.32 and all were significantly different from zero ($p < 0.05$; Table 1). The number of genetic clusters (K) estimated from DeltaK, LnPK and MedMeaK, MaxMeaK, MedMedK and MaxMedK values was approximately 10 (Fig. 2). At the broadest level, the greatest amount of genetic differentiation was evident in the planted populations 12 and 14, some of the paddock trees and their respective recruits (Supplementary Figs. S2 and S3). Fine-scale population structure illustrated by the 10 genetic clusters in the *Structure* plot (Fig. 3) is described below.

Fine-scale population differentiation illustrated by the 10 genetic clusters in the *Structure* plot showed that the six natural stands sampled from uncleared vegetation around the valley were moderately differentiated from

Population	Natural stands						Paddock trees	Recruits		Planted trees					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
1	0.000														
2	0.108	0.000													
3	0.051	0.136	0.000												
4	0.083	0.175	0.102	0.000											
5	0.048	0.126	0.078	0.098	0.000										
6	0.046	0.121	0.075	0.107	0.067	0.000									
7	0.020	0.099	0.040	0.067	0.041	0.036	0.000								
8	0.065	0.143	0.080	0.112	0.083	0.089	0.038	0.000							
9	0.078	0.149	0.094	0.126	0.104	0.100	0.059	0.095	0.000						
10	0.052	0.140	0.088	0.113	0.074	0.078	0.041	0.091	0.104	0.000					
11	0.029	0.114	0.047	0.081	0.055	0.052	0.011	0.047	0.031	0.056	0.000				
12	0.179	0.241	0.200	0.222	0.181	0.185	0.150	0.156	0.204	0.199	0.153	0.000			
13	0.071	0.170	0.107	0.133	0.101	0.096	0.061	0.104	0.124	0.099	0.076	0.217	0.000		
14	0.203	0.289	0.211	0.251	0.227	0.221	0.134	0.214	0.231	0.231	0.128	0.328	0.245	0.000	

Table 1. Pairwise F_{ST} comparisons between different populations of *E. melliodora* across the four groups of samples.

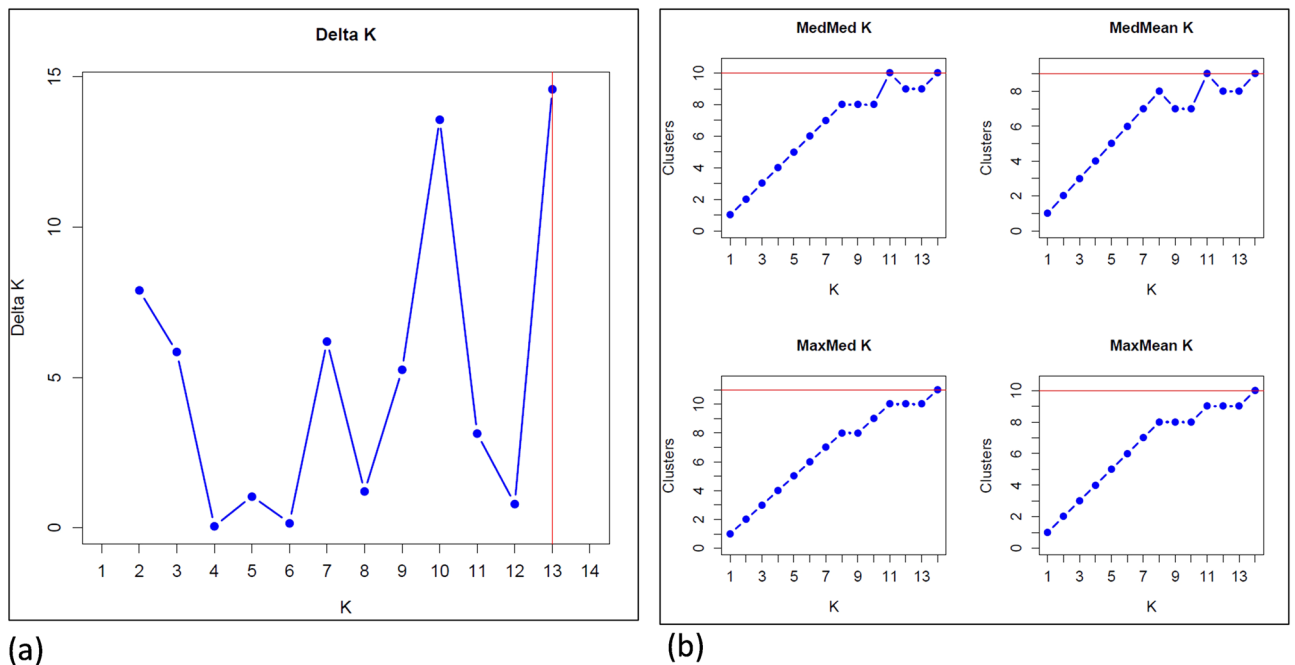


Figure 2. Number of genetic clusters (K) among the *E. melliodora* dataset based on 1406 loci estimated with (a) the Evanno method (DeltaK)⁴⁴ and (b) the Puechmaille method⁴⁵.

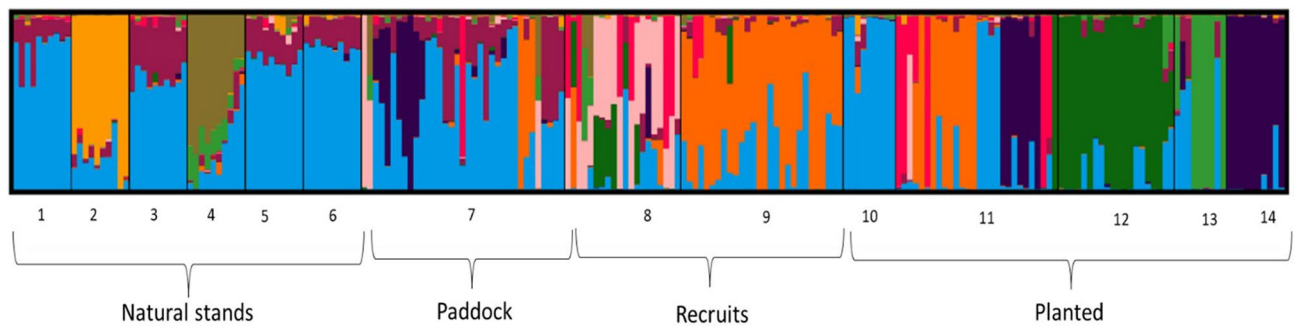


Figure 3. *Structure* plot from all populations of *E. melliodora* trees, including natural stands, paddock trees, recruits and planted trees, showing genetic clustering when $K = 10$.

one another (F_{ST} values in this group ranged from 0.02 to 0.17; Table 1). Four of these widely scattered stands formed a single undifferentiated cluster (dominated by light blue in Fig. 3; mean $F_{ST} < 0.060$) that was well represented within the paddock trees, while the remaining two natural stands (locations 2 and 4), were distinct and not represented among paddock trees, planted trees or recruits. The widely scattered paddock trees showed some differentiation from the natural stands (F_{ST} 0.029–0.097; Table 1) and also displayed remnant genotypes that were not present in the natural stands of the surrounding valley sides (Fig. 3).

The planted populations showed the highest level of differentiation from other groups (F_{ST} values ranged from 0.03 to 0.33; Table 1), and the fine-scale *Structure* plot identified six genetic clusters within this group, two of which were not represented elsewhere among the natural stands or paddock trees (populations 12 and 13, Fig. 3). Planted population 10 was genetically similar to the natural stands, while trees in population 11 appear to constitute a seed mix from the native stands and paddock trees (supported by low F_{ST} values between population 7 and 11; $F_{ST} = 0.01$, Table 1). Population 14 was the most differentiated population from all other populations ($F_{ST} = 0.13$ –0.33; Table 1). However, the *Structure* plot suggests it is most closely related to some of the paddock trees (Fig. 3) which is supported by pairwise F_{ST} values, as populations 14 and 7 (paddock trees) have the lowest pairwise F_{ST} of all other comparisons with population 14 ($F_{ST} = 0.13$; Table 1).

All recruits were sampled near a pair of adult paddock trees, and they showed genotypic similarity to each of these pairs of paddock trees. Paddock trees R001 and R002 and all nearby recruits show some level of association with the maroon cluster, while paddock trees R077 and R078 and all nearby recruits showing some level of association with the dark green cluster (Fig. 3). In each case, only one other paddock tree showed a partial association with these clusters and there was no association with any of the sampled valley-side stands implying that recruits were the offspring of one or both of the neighbouring paddock trees (Fig. 3). Similarly, almost half of the recruits surrounding paddock trees R001 and R002 also showed a partial association with the navy-blue cluster which was almost unique to planted stand 12 located only 30 to 120 m from these recruits. Taken together, these data imply that planted individuals had successfully cross-pollinated with paddock trees and that most matings involved adult trees separated by only tens of metres.

Genetic diversity and mating system among *E. melliodora* trees. The relictual paddock trees had the highest genetic diversity of any of the groups (mean $H_e = 0.170$; Table 2). The six natural stands were all similarly genetically diverse (mean $H_e = 0.141$; Table 2), but were less diverse than the paddock trees (Table 2). The planted trees had a similar level of genetic diversity to the natural stands (mean $H_e = 0.141$; Table 2; no significant difference $p > 0.05$, Student's *t*-test), while the two populations of recruits had higher diversity than the planted and natural stands (mean $H_e = 0.159$).

Most natural stands displayed heterozygous deficiencies consistent with a mixed mating system and low to moderate levels of inbreeding. Inbreeding coefficients (F_{IS}) for the valley side stands averaged 0.203, while the relatively isolated paddock trees and surrounding recruits had mean inbreeding coefficients of 0.194 ± 0.004 and 0.172 ± 0.004 respectively (Table 2). The planted trees displayed inbreeding coefficients that ranged from 0.06 to 0.229 likely reflecting variable practices of seed collection but with an average inbreeding coefficient almost identical to the mean for the natural stands (Table 2).

Detection of putative hybrids. After loci were filtered, genotypes were available for 1137 loci which comprised *E. melliodora* ($n = 61$), *E. albens* ($n = 42$) and individuals of other species ($n = 11$) representative of the seven common eucalypt species within the valley. The MedMeaK, MaxMeaK, MedMedK and MaxMedK values indicated that the most likely number of clusters was $K = 3$. Our analysis of these data using *ParallelStructure* 2.3.4 showed that both the *E. melliodora* and *E. albens* trees formed clear clusters differentiated from the other species (Fig. 4). Hybridization was rare with four apparent hybrid individuals detected in the sample of *E. melliodora* and *E. albens* trees that we analysed. These comprised two *E. albens* X *melliodora* hybrids (one paddock tree and one from within a natural stand), one *E. melliodora* X other species hybrid and one *E. albens* X other species hybrid. Both of these were sampled from natural stands.

Type	Population	N individuals	N loci	Ho ± SE	He ± SE	F _{IS}
Natural	1	10	1406	0.113 ± 0.004	0.144 ± 0.004	0.258
	2	10	1406	0.127 ± 0.004	0.131 ± 0.005	0.079
	3	10	1406	0.122 ± 0.004	0.147 ± 0.004	0.211
	4	10	1406	0.112 ± 0.004	0.134 ± 0.004	0.208
	5	10	1406	0.118 ± 0.004	0.146 ± 0.004	0.234
	6	10	1406	0.116 ± 0.006	0.143 ± 0.005	0.231
	Mean			0.118 ± 0.004	0.141 ± 0.004	0.203
Paddock	7	34	1406	0.121 ± 0.003	0.170 ± 0.004	0.194
Recruits	8	30	1406	0.136 ± 0.004	0.161 ± 0.004	0.179
	9	28	1406	0.134 ± 0.004	0.158 ± 0.004	0.165
	Mean			0.135 ± 0.004	0.159 ± 0.004	0.172
Planted	10	9	1406	0.112 ± 0.004	0.136 ± 0.005	0.229
	11	28	1406	0.132 ± 0.004	0.173 ± 0.004	0.255
	12	19	1406	0.124 ± 0.003	0.129 ± 0.004	0.063
	13	10	1406	0.115 ± 0.004	0.134 ± 0.004	0.185
	14	10	1406	0.129 ± 0.004	0.132 ± 0.004	0.069
	Mean			0.122 ± 0.004	0.141 ± 0.004	0.160

Table 2. Genetic diversity measures of *E. melliodora* populations with standard errors. *Ho* observed heterozygosity, *He* expected heterozygosity, *F_{IS}* inbreeding coefficient.

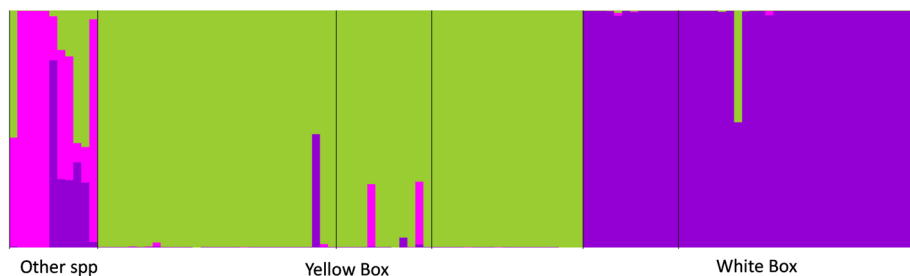


Figure 4. Structure plot showing K = 3 genetic clusters based on samples from 114 trees comprised of Yellow Box (*E. melliodora*), White Box (*E. albens*) and other eucalypt species.

Discussion

Our study of the population genetics of natural and revegetated *Eucalyptus melliodora* populations in WNP is one of the few studies that assess the genetics of woodland tree restoration by comparing the genetic structure of revegetated stands to both paddock trees and intact natural stands (e.g.⁷). We found that at a broad level, paddock trees and surrounding natural stands form comparably diverse and well-connected populations; moreover, the 30-year old Warrumbungle NP restoration program has been largely successful in adding genetically diverse stands derived from locally sourced seeds. In addition, our study further supports earlier work with both *E. melliodora* and other eucalypts in highlighting the value of paddock trees as both sources of existing genetic diversity and foci for recruitment^{13,25,29}. This study builds on the work of Broadhurst¹³ who emphasised both the long and short-term value of paddock trees as reservoirs of genetic diversity and as potentially important pollen sources that minimise the effects of inbreeding within seed crops on planted trees. Our data show that paddock trees include genetic diversity not conserved elsewhere in the nearby stands, and that they were clearly parents of the majority of our sampled recruits. However, our data also show that similar to Broadhurst's study¹³, a proportion of planted trees were not locally sourced and at least one paddock tree was a recent generation hybrid. Overall, our data suggest that the WNP Box-Gum Grassy Woodland restoration program has been particularly successful in capturing the genetic diversity of the broader population of *E. melliodora* within WNP, and the combination of this restoration and recruitment centred on paddock trees should ensure that the population retains diversity for future generations.

Contrary to our expectation based on Broadhurst¹³ that the revegetated (i.e., planted) individuals of *E. melliodora* would have lower genetic diversity than natural populations due to a low number of source individuals, we found that genetic diversity in the planted population was similar to the surrounding natural populations and the paddock trees, with the planted trees showing no reduction in genetic diversity. Other studies have shown varying results in restoration programs with some restoration populations having higher genetic diversity than reference sites, while others have lower diversity than reference sites. For example, a recent meta-analysis of 26 studies showed that levels of genetic diversity were strongly associated with the number of source sites used,

and most revegetation projects that used multiple source sites showed higher genetic diversity than those that used only one site⁴⁸. The comparable level of genetic diversity between the planted trees and the natural reference population in the surrounding bushland suggests that sourcing seed from six populations for a replanting program of this size would be sufficient to maintain genetic diversity in the revegetated population. From a conservation perspective, the genetic diversity of naturally occurring seedlings and the contribution of genetic material from remnant paddock trees provides stronger evidence of successful restoration. By extending the work of Broadhurst¹³ from looking at potential recruits (seeds), to established naturally occurring seedlings, we can have more confidence in our findings, since these plants have already survived local selective pressures and any genetic load inherited from interactions within and among planted and remnant trees.

As a group, the paddock trees were the most genetically complex and diverse trees of those sampled. The importance of paddock trees in maintaining ecosystem health and biodiversity within agricultural landscapes has been widely reported (reviewed in²⁵ and¹³). Here, we demonstrate that they are a significant source of genetic variation which is important to capture in revegetation programs to increase adaptive potential, retain locally adapted genotypes, enable efficient selection, mitigate inbreeding, and minimise genetic drift^{15,48–50}. The planted trees that were likely sourced from the paddock trees showed similar levels of genetic diversity to natural stands, suggesting that seed collected from paddock trees is of sufficient quality to contribute to population regeneration, as also identified in other studies^{25,30}.

Within WNP, we found that the natural populations of *E. melliodora*, including the paddock trees, showed some genetic differentiation suggesting that while historic gene flow has been sufficient to maintain strong connections among some stands, it has not prevented the development of some local population differentiation. The greatest level of population subdivision across the study site was among the planted trees; however, this reflects patterns of seed collection for the plantings rather than natural levels of connectivity. Our results showed quite high F_{ST} values in comparison to earlier studies^{51,52} that found little differentiation of *E. melliodora* populations spread across much of its geographic range (F_{ST} = 0.029, 0.04 resp. cf. 0.11 in this study). Our study is one of the few studies to use DArT-seq markers in a population genetic study of eucalypts (but see^{53,54}). These studies also found relatively high levels of structure within species (maxima ~ 0.41 for *E. microcarpa* and 0.11 for *E. cunninghamii*, [^{53,54} resp.]). The use of DArT-seq markers, primarily due to the generation of large numbers of SNPs across the genome, appears to improve population differentiation⁵⁵, but caution is advised in comparisons with F statistics derived from other markers⁵⁵.

Bayesian clustering analysis detected a number of unique genetic clusters within the paddock trees which suggests that some genotypes may be favoured within the valley floor rather than valley side populations. Planted population 14 and the relictual paddock tree from which it appears to have been seeded, showed the greatest level of genetic differentiation from other populations, which could indicate that selection is driving micro-site adaptation on the valley floor. A previous study of population genetic differentiation in *E. obliqua* found that local adaptation in response to different elevations can occur over as little as a few hundred meters⁵⁶. This in turn might imply that it is important to maintain strict local provenance in attempting restoration, to maintain site-specific adaptations. However, given that at least some trees have recruited since the clearing associated with pastoral activity, it is also possible that recruits derived from the isolated relictual paddock trees are simply the result of more erratic patterns of long-distance pollination than would occur in valley side stands. Most pollen dispersal in *E. melliodora* and other eucalypts is thought to occur over tens of metres, although intraspecific and interspecific pollination over several kms has been inferred for a range of eucalypt species^{17,28,32,57} including *E. melliodora*¹³.

Our genetic cluster analyses indicated that the planted trees were sourced from at least six different populations, and confirmed that a mixed provenancing strategy was used in this revegetation program. Planted stands 10, 11 and 14 were sourced from local seed from paddock trees and possibly surrounding natural stands, while populations 12 and 13 were probably not sourced locally. There is strong evidence that some of the revegetated trees planted in close proximity to paddock trees have interbred with these paddock trees to contribute to natural recruitment. There is increasing support for adopting strategies of mixed provenancing (i.e., using seed from multiple source sites in revegetation programs) to broaden the genetic base of revegetated areas^{48,53} in order to improve ecological outcomes and increase adaptive potential to environmental change^{49,58}. However, the issue of micro-site genetic variation and adaptation (identified above), also presents a challenge for mixed provenance strategies.

Elevated inbreeding with the consequent risk of inbreeding depression are threats to the long-term survival of both isolated plants and stands restored from small numbers of individuals⁵⁹. However, in this study neither the (often isolated) paddock trees, planted trees or the two sampled stands of recruits showed evidence that inbreeding levels were higher than those seen in the six natural stands. It is noteworthy, however, that Broadhurst¹³, while acknowledging that *E. melliodora* has a mixed mating system, reported an outcrossing rate (t) estimated through maternal adult/seed comparisons that was consistently close to 1.0. In contrast, the genotypic make-up of both adult and recruit stands in this study suggests that all sampled WNP populations were more inbred. Estimation of t assuming a stable equilibrium within our sampled populations ranged from ~ 0.7 to 0.8 (where t at equilibrium = $(1 - F)/(1 + F)$ ⁶⁰).

Perhaps surprisingly, given the relatively low outcrossing rates of the WNP stands and the apparent isolation of the paddock trees, we detected little evidence of hybridisation across the trees sampled in our study. The four recent inter-specific hybrids that we detected were within the scattered paddock trees or natural stands and likely represent a cohort of trees that pre-date agricultural clearing, suggesting historic or even pre-colonial hybridisation. Hybrids among *E. melliodora* and other members of *Eucalyptus* section *Adnataria* are common [47 and references therein]. Section *Adnataria* also includes *E. albens*, *E. crebra* and *E. sideroxylon*, all of which are common within study area and have long and overlapping flowering seasons. The restoration plantings have increased connectivity and relative abundance of multi-species pollen in the restoration area and may increase

the likelihood of hybridisation^{16,17,32}. An assessment of naturally occurring offspring would provide insight into the significance of this phenomenon for restoration practice.

In conclusion, the results of this study indicate that revegetated populations of *E. melliodora* in WNP are largely comprised of genetically diverse, mostly local provenance trees. The high genetic diversity and multiple seed-sources evident in the planted populations are testament to the success of this component of the ecological restoration project. Relictual paddock trees have been an important source of historic genetic diversity for this restoration program, and these trees can continue to be an important source of new recruits in the revegetated area and a valuable source of genetic diversity for restoration programs.

Data availability

The dataset generated during and/or analysed during the current study has been placed in the European Variation Archive (EVA) as project PRJEB59174. It is also in the Dryad repository <https://datadryad.org> with <https://doi.org/10.5061/dryad.hx3ffbgh1>.

Received: 31 May 2022; Accepted: 18 April 2023

Published online: 21 April 2023

References

- Cunningham, S. *Rewealth!: Stake Your Claim in the \$2 trillion Development Trend that's Renewing the World* (McGraw-Hill Professional, 2008).
- BenDorr, T. K., Livengood, A., Lester, T. W., Davis, A. & Yonavjak, L. Defining and evaluating the ecological restoration economy. *Restor. Ecol.* **23**, 203–219. <https://doi.org/10.1111/rec.12206> (2015).
- Mappin, B. *et al.* The costs and benefits of restoring a continent's terrestrial ecosystems. *J. App. Ecol.* **59**, 408–419. <https://doi.org/10.1111/1365-2664.14008> (2022).
- Ruiz-Jaen, M. C. & Aide, T. Restoration success: how is it being measured?. *Restor. Ecol.* **13**, 569–577 (2005).
- Mijangos, J. L., Pacioni, C., Spencer, P. B. & Craig, M. D. Contribution of genetics to ecological restoration. *Mol. Ecol.* **24**, 22–37 (2015).
- Breed, M. F. *et al.* The potential of genomics for restoring ecosystems and biodiversity. *Nat. Rev. Genet.* **20**, 615–628 (2019).
- Millar, M. A. *et al.* Evaluating restoration outcomes through assessment of pollen dispersal, mating system, and genetic diversity. *Restor. Ecol.* **29**, e13335 (2021).
- Waser, N. M. & Price, M. V. Optimal outcrossing in *Ipomopsis aggregata*: Seed set and offspring fitness. *Evolution* **43**, 1097–1109 (1989).
- Frankham, R. *et al.* Predicting the probability of outbreeding depression. *Conserv. Biol.* **25**, 465–475 (2011).
- Hufford, K. M., Krauss, S. L. & Veneklaas, E. J. Inbreeding and outbreeding depression in *Styliidium hispidum*: Implications for mixing seed sources for ecological restoration. *Ecol. Evol.* **2**, 2262–2273 (2012).
- Zucchi, M. I. *et al.* Genetic diversity of reintroduced tree populations in restoration plantations of the Brazilian Atlantic Forest. *Restor. Ecol.* **26**, 694–701 (2018).
- Steane, D. A. *et al.* Genomic scans across three eucalypts suggest that adaptation to aridity is a genome-wide phenomenon. *Genome Biol. Evol.* **9**, 253–265 (2017).
- Broadhurst, L. M. A genetic analysis of scattered Yellow Box trees (*Eucalyptus melliodora* A.Cunn. ex Schauer, Myrtaceae) and their restored cohorts. *Biol. Conserv.* **161**, 48–57 (2013).
- Reusch, T. B., Ehlers, A., Hämmerli, A. & Worm, B. Ecosystem recovery after climatic extremes enhanced by genotypic diversity. *Proc. Natl. Acad. Sci. USA* **102**, 2826–2831 (2005).
- Markert, J. A. *et al.* Population genetic diversity and fitness in multiple environments. *BMC Evol. Biol.* **10**, 205 (2010).
- Field, D. L., Ayre, D. J., Whelan, R. J. & Young, A. G. Relative frequency of sympatric species influences rates of interspecific hybridization, seed production and seedling performance in the uncommon *Eucalyptus aggregata*. *J. Ecol.* **96**, 1198–1210 (2008).
- Field, D. L., Ayre, D. J., Whelan, R. J. & Young, A. G. Patterns of hybridization and asymmetrical gene flow in hybrid zones of the rare *Eucalyptus aggregata* and common *E. rubida*. *Heredity (Edinb.)* **106**, 841–853 (2011).
- Goto, S., Iijima, H., Ogawa, H. & Ohya, K. Outbreeding depression caused by intraspecific hybridization between local and nonlocal genotypes in *Abies sachalinensis*. *Restor. Ecol.* **19**, 243–250 (2011).
- van Dijk, K. *et al.* Genomic screening reveals that the endangered *Eucalyptus paludicola* (Myrtaceae) is a hybrid. *Diversity* **12**, 468. <https://doi.org/10.3390/d12120468> (2020).
- Bradbury, D., Binks, R. M. & Byrne, M. Genomic data inform conservation of rare tree species: Clonality, diversity and hybridity in *Eucalyptus* series in a global biodiversity hotspot. *Biodivers. Conserv.* **30**, 619–641. <https://doi.org/10.1007/s10531-020-02106-2> (2021).
- von Takach Dukai, B., Jack, C., Borevitz, J., Lindenmayer, D. B. & Banks, S. C. Pervasive admixture between eucalypt species has consequences for conservation and assisted migration. *Evol. Appl.* **12**, 845–860 (2019).
- Frascaria-Lacoste, N. *et al.* Should forest restoration with natural hybrids be allowed?. *Restor. Ecol.* **19**, 701–704 (2011).
- Janes, J. K. & Hamilton, J. Mixing it up: The role of hybridization in forest management and conservation under climate change. *Forests* **8**, 237. <https://doi.org/10.3390/f8070237> (2017).
- Brennan, A. N., McKenna, J. R., Hoban, S. M. & Jacobs, D. F. Hybrid breeding for restoration of threatened forest trees: Evidence for incorporating disease tolerance in *Juglans cinerea*. *Front. Plant Sci.* <https://doi.org/10.3389/fpls.2020.580693> (2020).
- Ottewell, K. M., Donnellan, S. C. & Paton, S. C. Evaluating the demographic, reproductive, and genetic value of Eucalypt paddock trees for woodland restoration in agricultural landscapes. *Restor. Ecol.* **18**, 263–272 (2010).
- Potts, B. M. & Wiltshire, R. J. E. Eucalypt genetics and genecology. In: *Eucalypt Ecology: Individuals to Ecosystems* (eds Williams, J. E. & Woinarski, J. C. Z.) 56–91 (Cambridge University Press, Cambridge, Oakleigh, Melbourne 1995).
- Patterson, B., Vaillancourt, R. E., Pilbeamb, D. J. & Potts, B. M. Factors affecting variation in outcrossing rate in *Eucalyptus globulus*. *Aust. J. Botany* **52**, 773–780 (2004).
- Sampson, J. F. & Byrne, M. Outcrossing between an agroforestry plantation and remnant native populations of *Eucalyptus loxophleba*. *Mol. Ecol.* **17**, 2769–2781 (2008).
- Byrne, M., Elliott, C. P., Yates, C. & Coates, D. Maintenance of high pollen dispersal in *Eucalyptus wandoo*, a dominant tree of the fragmented agricultural region in Western Australia. *Conserv. Genet.* **9**, 97–105 (2008).
- Ottewell, K. M., Donnellan, S. C., Lowe, A. & Paton, S. C. Predicting reproductive success of insect- versus bird-pollinated scattered trees in agricultural landscapes. *Biol. Conserv.* **142**, 888–898 (2009).
- Larcombe, M. J. *et al.* Patterns of reproductive isolation in *Eucalyptus* - a phylogenetic perspective. *Mol. Biol. Evol.* **32**, 1833–1846 (2015).

32. Field, D. L., Ayre, D. J., Whelan, R. J. & Young, A. G. The importance of pre-mating barriers and the local demographic context for contemporary mating patterns in hybrid zones of *Eucalyptus aggregata* and *Eucalyptus rubida*. *Mol. Ecol.* **20**, 2367–2379 (2011).
33. Hamonts, K. *et al.* Effects of ecological restoration on soil microbial diversity in a temperate grassy woodland. *Appl. Soil. Ecol.* **117**, 117–128 (2017).
34. Belder, D. *et al.* Is bigger always better? Influence of patch attributes on breeding activity of birds in box-gum grassy woodland restoration plantings. *Biol. Conserv.* **236**, 134–152 (2019).
35. EPBC Act *Environment Protection and Biodiversity Conservation Act 1999*. <https://www.legislation.gov.au/Details/C2014C00506> (1999) (accessed 8th February 2022).
36. BC Act *Biodiversity Conservation Act 2016 No 63*. <https://legislation.nsw.gov.au/view/pdf/asmade/act-2016-63> (2016) (accessed 8th February 2022).
37. Prober, S. M., Thiele, K. R. & Lunt, L. D. Identifying ecological barriers to restoration in temperate grassy woodlands: Soil changes associated with different degradation states. *Aust. J. Botany* **50**, 699–712 (2002).
38. Sansaloni, C. P. *et al.* A high-density diversity arrays technology (DArT) microarray for genome-wide genotyping in *Eucalyptus*. *Plant Methods* **6**, 16 (2010).
39. Petroli, C. D. *et al.* Genomic characterization of DArT markers based on high-density linkage analysis and physical mapping to the *Eucalyptus* genome. *PLoS ONE* **7**, e44684 (2012).
40. Gruber, B., Unmack, P. J., Berry, O. F. & Georges, A. DARTR: An R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Mol. Ecol. Resour.* **18**, 691–699 (2018).
41. Jombart, T. & Ahmed, I. ADEGENET 1.3–1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**, 3070–3071. <https://doi.org/10.1093/bioinformatics/btr521> (2011).
42. Miller, M. A. *et al.* A RESTful API for access to phylogenetic tools via the CIPRES science gateway. *Evol. Bioinform. Online* **11**, 43–48 (2015).
43. Li, Y. & Liu, J. StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. *Mol. Ecol. Resour.* **18**, 176–177 (2018).
44. Evanno, G., Regnault, S. & Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005).
45. Puechmaille, S. J. 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**, 608–627 (2016).
46. Peakall, R. & Smouse, P. E. GenALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* **28**, 2537–2539 (2012).
47. Ferguson, S., Jones, A., Murray, K., Schwessinger, B. & Borevitz, J. O. Interspecies genome divergence is predominantly due to frequent small scale rearrangements in *Eucalyptus*. *Mol. Ecol.* **32**, 1271–1287 (2023).
48. Jordan, R., Breed, M. F., Prober, S. M., Miller, A. D. & Hoffmann, A. A. How well do revegetation plantings capture genetic diversity?. *Biol. Lett.* **15**, 20190460 (2019).
49. Sgrò, C. M., Lowe, A. J. & Hoffmann, A. A. Building evolutionary resilience for conserving biodiversity under climate change. *Evol. Appl.* **4**, 326–337 (2011).
50. Breed, M. F. *et al.* Mating system and early viability resistance to habitat fragmentation in a bird-pollinated eucalypt. *Heredity (Edinb)* **115**, 100–107 (2015).
51. Broadhurst, L. M., Mellick, R., Knerr, N., Li, L. & Supple, M. A. Land availability may be more important than genetic diversity in the range shift response of a widely distributed eucalypt, *Eucalyptus melliodora*. *For. Ecol. Manag.* **409**, 38–46 (2018).
52. Supple, M. E. *et al.* Landscape genomic prediction for restoration of a *Eucalyptus* foundation species under climate change. *Elife* **7**, e31835. <https://doi.org/10.7554/eLife.31835> (2018).
53. Jordan, R., Dillon, S. K., Prober, S. M. & Hoffmann, A. A. Landscape genomics reveals altered genome wide diversity within revegetated stands of *Eucalyptus microcarpa* (Grey Box). *New Phytol.* **212**, 992–1006 (2016).
54. Rutherford, S. *et al.* Speciation in the presence of gene flow: Population genomics of closely related and diverging *Eucalyptus* species. *Heredity* **121**, 126–141. <https://doi.org/10.1038/s41437-018-0073-2> (2018).
55. Steane, D. A. *et al.* Population genetic analysis and phylogeny reconstruction in *Eucalyptus* (Myrtaceae) using high-throughput, genome-wide genotyping. *Mol. Phylogenet. Evol.* **59**, 206–224 (2011).
56. Wilkinson, G. R. Population differentiation within *Eucalyptus obliqua*: implications for regeneration success and genetic conservation in production forests. *Aust. For.* **71**, 4–15. <https://doi.org/10.1080/00049158.2008.10676266> (2008).
57. Yeoh, S. H., Bell, J. C., Foley, W. J., Wallis, I. R. & Moran, G. F. Estimating population boundaries using regional and local-scale spatial genetic structure: an example in *Eucalyptus globulus*. *Tree Genet. Genom.* **8**, 695–708 (2012).
58. Broadhurst, L. M. *et al.* Seed supply for broadscale restoration: maximizing evolutionary potential. *Evol. Appl.* **1**, 587–597 (2008).
59. Charlesworth, D. & Willis, J. H. The genetics of inbreeding depression. *Nat. Rev. Genet.* **10**, 783–796 (2009).
60. Enjalbert, J. & David, J. L. Inferring recent outcrossing rates using multilocus individual heterozygosity: Application to evolving wheat populations. *Genetics* **156**, 1973–1982 (2000).

Acknowledgements

We would like to acknowledge the profound contribution of the late Prof David Ayre to the field of population genetics; he had an incredible mind and fostered interest in population and conservation genetics in many undergraduate and graduate students. He will be dearly missed. We are grateful to Castlereagh Area, NSW National Parks and Wildlife Service for access to Warrumbungle National Park, logistical support and for permission to reproduce Supplementary Fig. S1. This work was funded by a NSW Department of Planning and Environment Saving our Species Science and Research grant to AJD and DJA. We thank Murray Ellis, Jennifer Taylor and Lisa Metcalfe for locating relictual paddock trees, Liz Tasker for helpful comments on an earlier draft of the paper, and the anonymous reviewers whose careful comments greatly improved the manuscript.

Author contributions

N.L.R. (formal analysis, visualization, writing - original draft, writing - review & editing), A.Q. (investigation, data curation, writing - original draft), H.D. (data curation, formal analysis, visualization), D.J.A. (conceptualization, investigation, writing - review & editing, supervision, funding acquisition), A.J.D. (conceptualization, investigation, writing - review & editing, supervision, project administration, funding acquisition).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-33720-z>.

Correspondence and requests for materials should be addressed to A.J.D.

Reprints and permissions information is available at www.nature.com/reprints.

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



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

© The Author(s) 2023