Conserv Genet (2010) 11:45–56
DOI 10.1007/s10592-009-0001-3

RESEARCH ARTICLE

Two sympatric spotted gum species are molecularly homogeneous

Joel W. Ochieng · Mervyn Shepherd · Peter R. Baverstock · Garth Nikles · David J. Lee · Robert J. Henry

Received: 15 December 2008 / Accepted: 29 September 2009 / Published online: 17 October 2009

Abstract Large fruited spotted gum eucalypt *Corymbia henryi* occurs sympatrically with small fruited spotted gum *Corymbia citriodora* subspecies *variegata* over a large portion of its range on the east coast of Australia. The two taxa are interfertile, have overlapping flowering times and share a common set of insect and vertebrate pollinators. Previous genetic analysis of both taxa from two geographically remote sites suggested that the two were morphotypes rather than genetically distinct species. In this study we further explore this hypothesis of genic species by expanding sampling broadly through their sympatric locations and examine local-scale spatial genetic structure in stands that differ in species and age composition. Delineation of populations at five microsatellite loci, using an individual-based approach and Bayesian modelling, as well as clustering of individuals based on allele frequencies showed the two species to be molecularly homogeneous. Genetic structure aligned largely with geographic areas of origin, and followed an isolation-by-distance model, where proximal populations were generally less differentiated than more distant ones. At the stand level, spotted gums also generally showed little structure consistent with the high levels of gene flow inferred across the species range. Disturbances in the uniformity of structuring were detected, however, and attributed to localised events giving rise to even aged stands, probably due to regeneration from a few individuals following fire.

Keywords Genic species · Genetic structure · Spotted gum · Isolation-by-distance

J. W. Ochieng (✉) · M. Shepherd · P. R. Baverstock · R. J. Henry
Centre for Plant Conservation Genetics, Southern Cross University, P.O. Box 157, Lismore, NSW 2480, Australia
e-mail: jochieng@uonbi.ac.ke

G. Nikles
Department of Primary Industries and Fisheries, 80 Meiers Road, Indooroopilly, QLD, Australia

D. J. Lee
Department of Primary Industries and Fisheries, LB 16 Fraser Road, Gympie, QLD 4570, Australia

J. W. Ochieng
Centre for Biotechnology and Bioinformatics, University of Nairobi, Box 29053 (00625), Nairobi, Kenya

Introduction

Conservation and management of natural populations is aided by the identification of evolutionarily significant or management units (Moritz 1994). Taxonomic assignments (essentially hypotheses of the diagnosability of taxa) often guide these decisions, but are mostly based on phenotypes, which may not represent genetic populations. Similarly, many workers often use sampling locations to define populations in studies aimed at managing genetic resources (Pearse and Crandall 2004). A misleading interpretation can result when population boundaries do not strictly follow a species physical distribution patterns, such as the case of mobile species or species with complex dispersal patterns. In such a case, an individual-based method for delineating genetically distinct management units without regard to species boundaries or sampling locations (Pritchard et al. 2000; Evanno et al. 2005) may be particularly appropriate for studying groups of closely related sympatric species that are suspected to interbreed in nature.

Two closely related spotted gum eucalypts: *Corymbia variegata* and *Corymbia henryi* (Hill and Johnson 1995)

occur in sympatry in the east coast of Australia. Spotted gums (Section *Politaria*) form a complex of closely related taxa that occur along the eastern seaboard of Australia from approximately latitude 16°S in northern Queensland to latitude 37°S in eastern Victoria (Hill and Johnson 1995). Four species (*C. maculata*, *C. henryi*, *C. variegata*, and *C. citriodora*) that occur as a latitudinal replacement series were recognised by Hill and Johnson (1995) in their revision of the genus *Corymbia*. However, their taxonomy remains controversial as classification often relies on variation in morphology and leaf oils, and knowledge of the geographic origins of a specimen (McDonald and Bean 2000). An alternative classification based on isozyme variation subsumes *C. variegata* into a subspecies of *C. citriodora*, i.e., *C. citriodora* ssp. *citriodora* and *C. citriodora* ssp. *variegata* (McDonald and Bean 2000). In this manuscript, we adopt Hill and Johnson (1995) classification for simplicity. The two taxa *C. variegata* and *C. henryi* (Hill and Johnson 1995) are regarded as separate species based on divergence at quantitative traits, such as fruit (capsule) and foliage morphology (Hill and Johnson 1995). They occur sympatrically over a large portion of their ranges, have overlapping flowering times and share a common set of insect and vertebrate pollinators (Hill and Johnson 1995; Southerton et al. 2004). Earlier attempts to clarify the relationships amongst the spotted gums by molecular approaches (McDonald et al. 2000; King 2004) provided conflicting results on the affinity of these two species.

Some clarification of the genetic relationships within the spotted gum complex was recently attained with a study of genetic variation across the four taxa in the complex using microsatellites (Shepherd et al. 2008). This study showed that there was a stronger affinity among all three northern taxa, subsp. *citriodora*, subsp. *variegata* and *C. henryi*, than with their southern counterpart, *C. maculata*. A further study was conducted to distinguish among possible evolutionary scenarios that might explain the persistence of what were believed to be interfertile and clearly delineated morphological species, subsp. *variegata* and *C. henryi* (Ochieng et al. 2008). Here, as a first step, the level of molecular divergence between the two species at two separate (approximately 300 km apart) locations where the species occurred in sympatry was examined using 12 microsatellite markers. It was reasoned that if more variation was found between taxa at each site than within a taxa across the two sites, this would support the recognition of two species (Muir et al. 2000). However, no species-specific alleles or significant allele frequency differences were detected within sites, suggesting high recurrent local gene flow between the two species.

The purpose of this study was to test whether this observation held more broadly across the sympatric range of the two taxa; in which case it may be reasonable to conclude that the two taxa are genic species, where segregation at a gene(s) with large effects or selection (Wu 2001; Lexer et al. 2007) might be the source of their phenotypic differences without a corresponding genomic differentiation. Therefore, in this study, sampling is expanded to cover an additional four locations. Further, this study differs from the previous one (Ochieng et al. 2008) in that here we examine spatial genetic structure at three different scales: within neighbourhoods, between neighbourhoods within a locality, and among localities. Sampling was structured so that multiple sites were sampled within a stand (location) as well as in transects through forests varying in species and age composition. This allowed patterns of gene flow at the landscape and neighbourhood level to be described. Unlike the preceding study, this study does not use sampling locations to define populations; rather, an individual-based method is used to delineate genetically distinct management units without regard to species boundaries or sampling locations.

Methods

Sampling design

Sampling for this study was designed to test for genetic structure in a species complex comprising *C. variegata* and *C. henryi* in northern New South Wales and southern Queensland following the approach of Bacles et al. (2005). The design allowed testing for structure at three different scales: 1, within a neighbourhood (site); 2, within a locality between neighbourhoods and 3, among localities (range-wide). Material of both taxa was collected in approximately equal proportions where possible, as well as putative hybrid material. To expand the range from an earlier study based on Cherry Tree and Bunyaville, five new localities were selected to give coverage across their distribution and to include locations where both taxa were known to occur in sympatry as well as allopatric locations. The seven locations are shown in Fig. 1 and names and numbers of sites at each location described in Table 1. The most southerly location sampled was Candole State Forest (SF) near Grafton NSW Australia (Lat. 29°46′; Long. 153°12′). It is a sympatric case, although *C. variegata* can be found only in a small patch toward the northern end of the forest. A nearby location is Ewingar SF, NSW (Lat. 29°04′; Long. 152°30′). The spotted gums found in Ewingar are predominantly *C. henryi*; but *C. variegata* was found to occur in a narrow patch (20 m radius) and was represented by less than 10 trees. Bungawalbin SF, another locality in NSW (Lat. 29°02′; Long. 153°16′), was characterised by a single species (*C. henryi*) stand with no trees assignable as

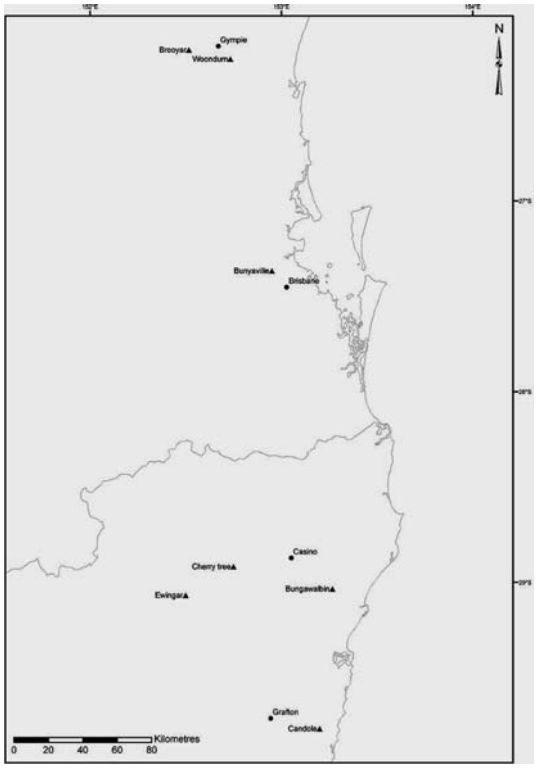


Fig. 1 Map of the East coast of Australia showing the provenances from which samples of *C. variegata* and *C. henryi* were obtained. *Circles* are major towns while *triangles* are State Forests where samples were obtained

C. variegata. Cherry Tree SF near Casino, NSW (Lat. 28°55′; Long. 152°45′) and Bunyaville SF in Brisbane, Queensland (Lat. 27°22′; Long. 152°57′) were sampled again for this study. The samples from both Cherry Tree SF and Bunyaville SF used in the previous analysis were also included in the present study. Samples from two other locations both near Gympie, Queensland, were also used in this study: Woondum (Lat. 26°15′; Long. 152°48′) and Brooyar SFs (Lat. 26°13′; Long. 152°32′).

Between two and four neighbourhoods were sampled per locality except Brooyar and Woondum. A neighbourhood comprised between 20 and 30 sampled adult trees within a radius of 20–30 m. Distance between neighbourhoods within a locality ranged between 1 and 3 km. At some localities, stand composition varied considerably with respect to taxa abundance between sites. The most extreme cases were at Ewingar SF and Candole SF where sites of apparently highly pure and even aged stands of *C. henryi* occurred as well as one site in each case (Ewingar site 2; Candole site 3) where a few (<20) *C. variegata* trees occurred in a narrow strip alongside a pure stand of *C. henryi*. Age structure also varied among stands. Whereas some sites were characterised by trees of a more or less uniform age, other sites tended to be much more variable in age. Cherry Tree site 1 and Candole site 3 were distinguished in this way, both with apparently uniform age class, probably representing young regrowth stands regenerated from a few older trees following a fire.

In addition to this sampling design, transect sampling was conducted at two localities, Cherry Tree SF (New South Wales) and Bunyaville (Queensland). The purpose of these transects was to test for genetic structure due to taxon related differences. Transects were designed to transgress forests where the abundance of *C. henryi* varied in relation to *C. variegata*. At the Cherry Tree SF locality, samples were collected along Cherry Tree Road over a distance of about 16 km. The road dissects through the State Forest in an approximate North–South direction gradually declining in elevation from approximately 400 to 100 m above sea level. It is not clear whether altitude would play a significant role at these micro-scales, however, these slight changes in altitude might correlate with soil differences (not investigated) or some other parameter such as pollinator behaviour. The road sequentially traverses through spotted gum dominated forest types containing pure stands of C. *variegata* at the Northern end, admixtures of *C. variegata* and *C. henryi*, and largely pure stands of *C. henryi* at the southern-most and lower altitudes. At Bunyaville, sample collections transacted altitudinal variation ranging from 84 to 143 m. Transect sampling at both locations obtained material from trees at least 100 m apart with a total of 82 and 56 trees collected at Cherry Tree and Bunyaville localities, respectively. Finally, a set of 24 *C. maculata* and 12 *C. citriodora* (spotted gum) samples collected from diverse locations were included in cluster analysis.

Sample collection and molecular characterization

Trees with reproductive material were targeted where possible to allow more confident taxonomic assignments. Bark was collected from each tree for the purpose of DNA extraction. A bark plug was obtained by using a punch and hammer, then dried using silica gel and stored at room temperature. In addition to bark samples for DNA extraction, herbarium specimens (capsules, flowers, buds, leaves) were collected from each individual for a majority of the samples. Field identifications were confirmed at the Queensland Herbarium and identified specimens were deposited with the Southern Cross University Herbarium. A GPS position and altitude was recorded for each tree and photographs were taken for most trees. DNA extraction, PCR amplifications and genotyping were performed as previously described (Ochieng et al. 2008). Trees were genotyped at five microsatellite loci: EMCRC45, EMCRC46, EMCRC51, EMCRC55, and EMCRC93 (Shepherd et al. 2006).

Table 1 Predefined populations, sampling designs, number of samples (N), expected heterozygosities (H_E; H_O), mean number of alleles (MNA) and inbreeding coefficient (F_{IS}; Wright 1951) across loci per population analysed

Provenance	Lat.	Long.	Site and number of samples	Species	H_E [a]	H_O [a]	MNA[a]	F_{IS} [a]
Brooyar ($N = 4$)	26°13′	152°32′	Site 1 = 4	*C. variegata*	0.750	0.900	4.6	−0.241
Woondum ($N = 16$)	26°15′	152°48′	Site 1 = 16	*C. variegata*	0.832	0.875	8.6	−0.054
Bunyaville ($N = 56$)	27°22′	152°57′	Transect = 28	*C. henryi*	0.849	0.574	12.6	0.334
			Transect = 28	*C. variegata*	0.913	0.641	14.8	0.292
Cherry tree ($N = 112$)	28°55′	152°45′	Site 1 = 17	*C. henryi*	0.903	0.796	15.2	0.120
			Site 1 = 3	*C. variegata*				
			Site 3 = 15	*C. henryi*				
			Site 3 = 10	*C. variegata*				
			Site 3 = 5	Put. hybrids				
			Transect = 31	*C. henryi*				
			Transect = 31	*C. variegata*				
Bungawalbin ($N = 60$)	29°02′	153°16′	Site 1 = 20	*C. henryi*	0.867	0.827	15.8	0.047
			Site 2 = 20	*C. henryi*				
			Site 4 = 20	*C. henryi*				
Ewingar ($N = 40$)	29°04′	152°30′	Site 2 = 15	*C. henryi*	0.824	0.834	12.6	0.019
			Site 2 = 5	*C. variegata*				
			Site 3 = 20	*C. henryi*				
Candole ($N = 60$)	29°46′	153°12′	Site 1 = 20	*C. henryi*	0.828	0.768	13.4	0.073
			Site 2 = 20	*C. henryi*				
			Site 3 = 8	*C. henryi*				
			Site 3 = 12	*C. variegata*	0.614	0.750	3.8	−0.233
Global ($N = 348$)					0.923	0.776	27.8	0.159

[a] Genetic parameter per provenance

Statistical analyses

Population diversity estimates

Possible non-random allelic associations between pairs of loci were assessed within each predefined population and globally using Fisher's exact test on contingency tables, implemented in *GENEPOP on the web*, version 3.3 (Raymond and Rousset 1995). The significance level for each test was adjusted using sequential Bonferroni procedure (Rice 1989). Due to the neighbourhood sampling design, some inbreeding within sites might be expected to cause non-random allelic associations. As such, a more stringent confidence level (99%) for the null hypothesis of allelic associations was set. Gene diversity (Nei 1987), inbreeding coefficient F_{IS}, observed heterozygosity (H_O) and the mean number of alleles (MNA) per population were calculated using GENEPOP.

Population differentiation and clustering with predefined structure

Levels of genetic differentiation (F_{ST}; Weir and Cockerham 1984) and relationships between species occurring in different geographic regions was estimated in GENEPOP and their significance assessed by the log-likelihood exact G-test on genotypes (Goudet et al. 1996). A MANTEL test for populations (test for isolation-by-distance) was implemented in FSTAT. Cavalli-Sforza and Edward's (1967) chord distance was used to estimate genetic distances because it models short-term divergence and is thus most suitable for comparing closely related groups. Neighbour joining (NJ) clustering of distances was done in the NEIGHBOR program of the PHYLIP package (Felsenstein 1995), using *C. maculata* as outgroup.

Determination of genetically homogeneous clusters

Genetically homogeneous clusters of individuals were determined using a Bayesian model-based method implemented in the programme STRUCTURE Version 2.1 (Pritchard et al. 2000), which apportions population ancestry to individuals based on the model chosen, i.e., K value. This was done without assuming predefined structure. To simulate a more likely number of clusters (K), we used an alternative optimisation criterion—delta-K, which relates to the second order rate change in the log probability of the data (Evanno et al. 2005). A hierarchical approach was used to

successively analyse structure: structure was initially examined for the entire dataset and genetically homogeneous subsets identified, then further substructure was probed using STRUCTURE. We adopted the following model parameters: 10,000 iterations each for the "burn-in" and Markov Chain Monte Carlo (MCMC) stages and a "correlated allele frequency model" for ancestry. Both "admixture" and "no admixture" models were investigated when testing for structure within populations. Ten runs for each K (for $K = 1$–10) were used to calculate means and standard deviation (SD) for posterior probability of the data for a given K ($L(K)$) and delta-K. These parameters provided generally conservative optimisation criteria for replicate runs when detecting populations. Presence of structure was assumed when plots of delta-K provided evidence of a clear peak for K values (Evanno et al. 2005). To validate a lack of structure, assignments of individuals at the most-likely K values were further investigated. Symmetry in the proportion of samples assigned to each population and an apparent admixture of most individuals was interpreted to mean an absence of genetic structure, following Pritchard and Wen (2004). The proportion of each population ancestry for the number of populations set by the K value was used to classify individuals. Individuals were considered well assigned if they possessed >90% of ancestry from a single population, otherwise they were considered admixed. An incorrect assignment was concluded when the predefined source population showed the lowest likelihood value of being the source. The proportion of individuals with incorrect or ambiguous assignments was used to infer immigration between species and provenances. Finally, levels of genetic differentiation (F_{ST}; Weir and Cockerham 1984) were estimated for posterior genetic clusters in GENEPOP.

Individual specific clustering

Due to some limitations of STRUCTURE relative to analysis of F_{ST} when dealing with predefined genetic groups and at a low number of loci (5), a cluster analysis on individuals was conducted using MICROSAT (version 1.5d) and PHYLIP (version 3.57c). The matrix created by MICROSAT based on Cavalli-Sforza and Edward's (1967) chord distance was used by PHYLIP to plot a phylogenetic tree.

Results

Marker variation was high, providing power to test for structure

Genetic diversity across the five microsatellites was high (average number of alleles per locus = 28.0; Table 1). Observed heterozygosity (H_O), expected heterozygosity (H_E) and mean number of alleles (MNA) per predefined population are given in Table 1.

Species were not differentiated but geographic clusters were identified

Inspection of the delta-K plot for models with a range of K values from 1 to 10 for the entire dataset revealed a distinct peak at $K = 2$ and another at $K = 6$ (Fig. 2). The peak at $K = 2$ was stronger, indicating that the highest hierarchy of structure had two populations when all samples were considered. In the two-cluster model, the first group comprised both *C. variegata* and *C. henryi* individuals from Bunyaville, Cherry Tree transect, and Cherry Tree site 3. The second group comprised *C. variegata* and *C. henryi* individuals from Candole, Ewingar, Bungawalbin, Cherry Tree site 1, and *C. variegata* from Woondum and Brooyar (figure not shown). Despite the stronger statistical support for a two clusters model, this outcome may have been artefactual (see "Discussion"). The alternative structure model with six genetically homogeneous clusters ($K = 6$) delineated groups that corresponded with sampling locations: Brooyar–Woondum, Bunyaville, Cherry Tree, Bungawalbin, and Ewingar–Candole and a sixth group comprising two neighbourhoods, one at Cherry Tree (site 1) and the other at Candole (site 3), which were separated from other neighbourhoods within their localities (Fig. 3). In all tested models ($K = 1$–10) using the default settings or model parameters, *C. variegata* and *C. henryi* individuals showed no taxonomic structure; instead, they were consistently homogenised in all identifiable genetic clusters (Fig. 3).

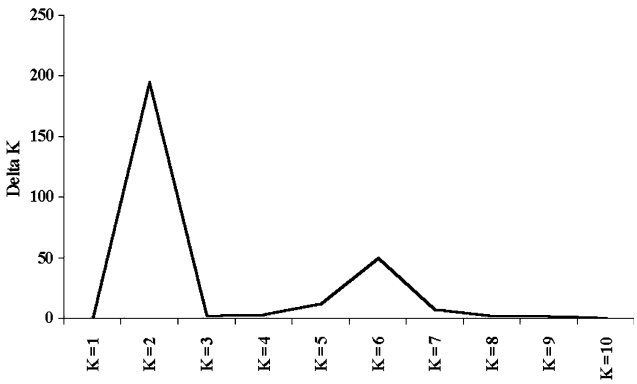
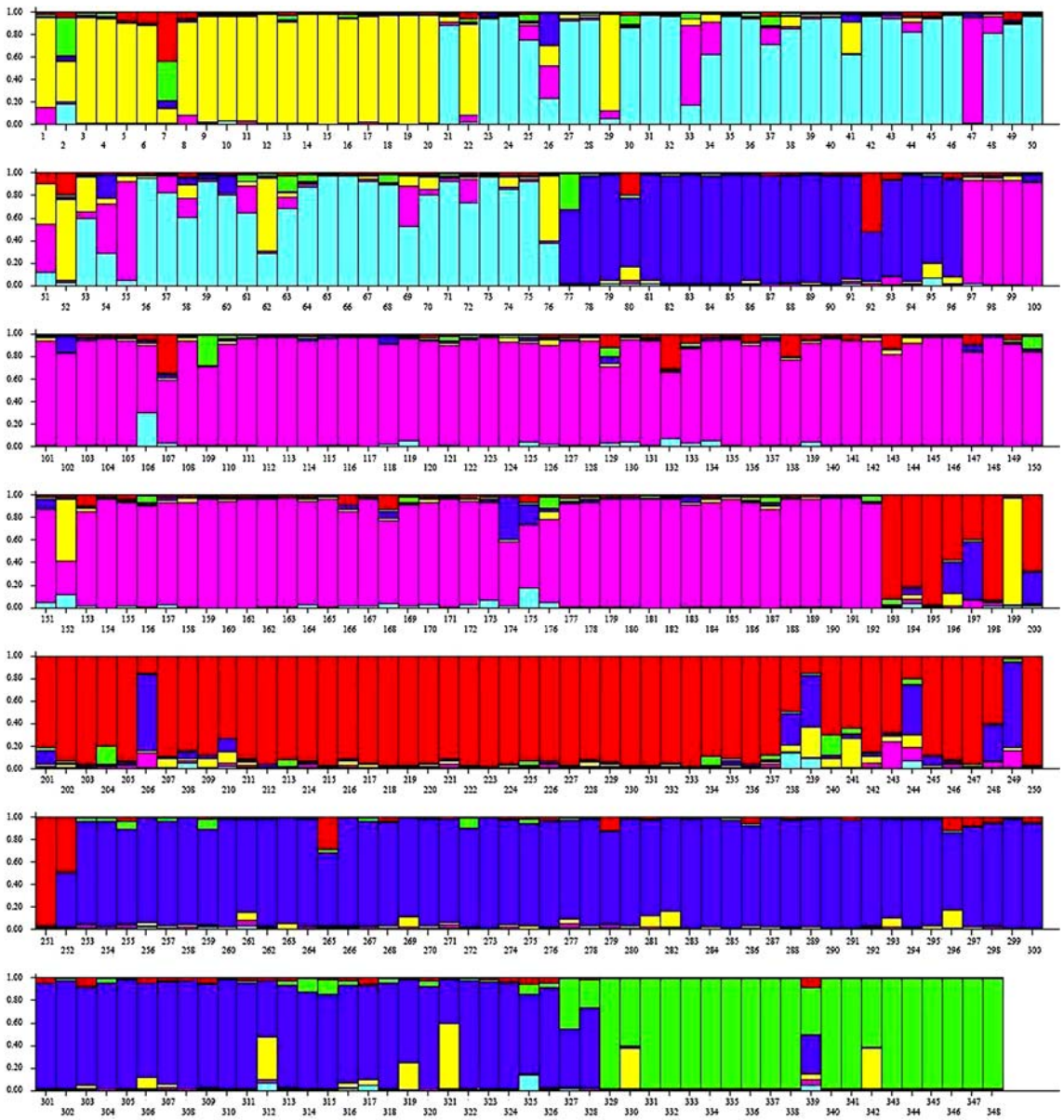


Fig. 2 Magnitude of ΔK as a function of K (mean ± SD over 10 replications) calculated using the procedure outlined in the methods. K is a model representing the most likely number of clusters, while delta-K is the optimal probability of the data for a given K



Samples 001-004	Brooyar	Samples 117-192	Cherry tree transect
Samples 005-020	Woondum	Samples 193-253	Bungawalbin neighbourhoods 1, 2 and 3
Samples 021-076	Bunyaville transect	Samples 254-291	Ewingar neighbourhoods 1 and 2
Samples 077-096	Cherry Tree neighbourhood 1	Samples 292-327	Candole neighbourhhods 1 and 2
Samples 097-116	Cherry Tree neighbourhood 3	Samples 328-348	Candole neighbourhhod 3

Fig. 3 Ancestry assignments for 348 *C. henryi* and *C. variegata* individuals from seven sampling locations determined by STRUCTURE using the $K = 6$ model. The *X*-axis is the separate individuals, while the *Y*-axis shows the proportion of ancestry for each group as represented by the different colours

Cluster analysis of individuals showed no evidence for species differentiation

Clustering of individuals based on allele frequencies (Fig. 4) showed that the demarcation and assignment of individuals to species lacks support at the neutral loci studied. However, unlike in STRUCTURE analysis, the demarcation of geographic clusters is not perfectly clear, suggesting homogeneity among populations and a very weak geographic structure.

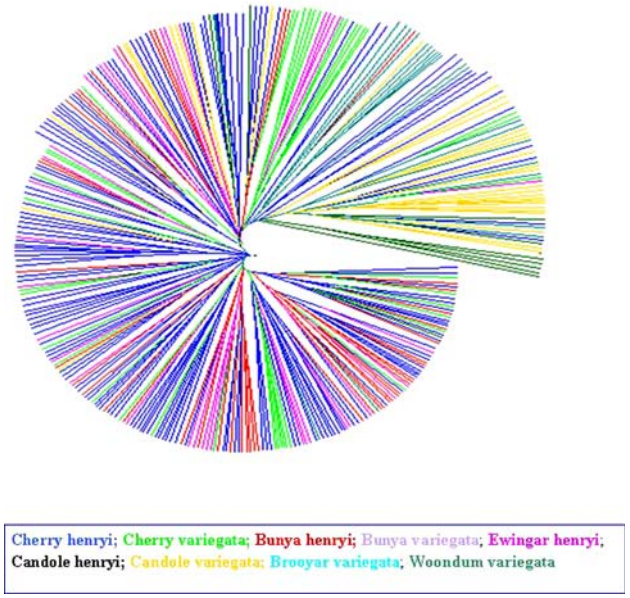


Fig. 4 Unrooted UPGMA dendrogram constructed from allele sharing statistics among 332 individuals from nine populations. Each population is represented in a different font colour

No evidence for substructure

Examination of substructure, implemented for each of the six subsets revealed no evidence for genetic substructure. In hierarchical analysis, Cherry Tree site 1 was delineated from Candole, with which it appeared homogeneous when the entire dataset was considered. Similar to the results of the entire dataset, there was no evidence for taxonomic structure at this finer scale i.e., for all runs and models, there was no single case in which *C. variegata* and *C. henryi* could be distinguished on the basis of their multilocus structure. This includes the transect samples from Cherry Tree SF, comprising both species in equivalent proportions sampled along a 16 km length of the forest. Individuals from this locality formed a completely homogeneous subset (Fig. 5). All individuals were symmetrically admixed, a phenomenon interpreted to code for an absence of genetic structure (Pritchard and Wen 2004). Given limited seed dispersal in forest trees, neighbourhoods might be expected to show some level of structure. However, these data showed that neighbourhoods were generally homogenised within each locality, except for two cases (Cherry Tree site 1 and Candole site 3).

Isolation-by-distance model lacked statistical support

Genetic differentiation between the six homogeneous clusters obtained from STRUCTURE mirrored geographically separated sampling localities. Neighbouring populations showed a generally lower pairwise F_{ST} values (Table 3) than more distant populations, and suggested gene flow was restricted by geographic distance. MANTEL tests for an isolation-by-distance model based on linear genetic distances and log transformed geographic distances were, however, not significant ($r = 0.068$; $P = 0.294$). The use of linear geographic distances gave an even less significant relationship ($P = 0.971$).

Population based approach was consistent with genetic structure

Neighbour-Joining trees of predefined populations clustered using Cavalli-Sforza and Edward's (1967) chord distance, is shown in Fig. 6. The *C. maculata* outgroup was well separated from the in-group topology with a substantially long branch. The next to split was *C. citriodora*, while *C. variegata* and *C. henryi* formed a clade. Terminal

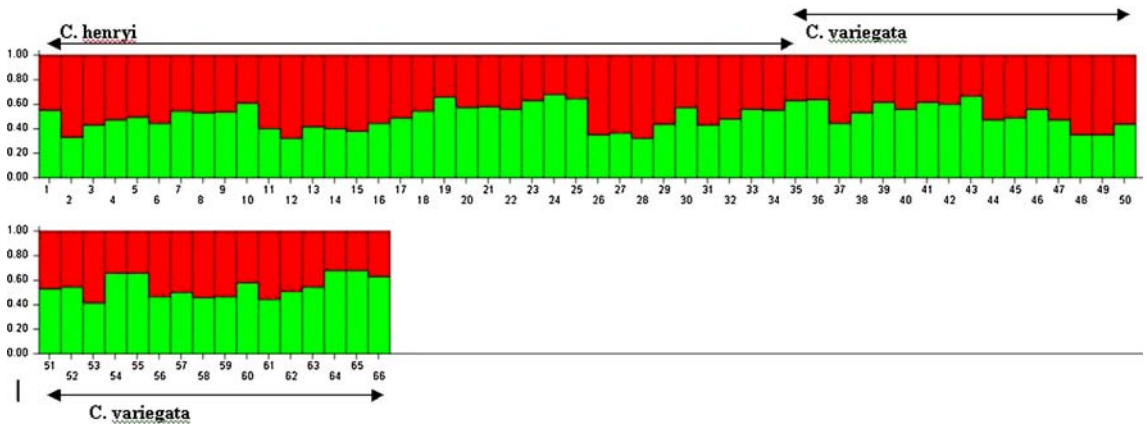


Fig. 5 Ancestry assignments for 66 *C. henryi* and *C. variegata* individuals from Cherry Tree State Forest determined by STRUCTURE using the $K = 2$ model. The *X*-axis is the separate individuals, while the *Y*-axis shows the proportion of ancestry for each group as represented by the different colours

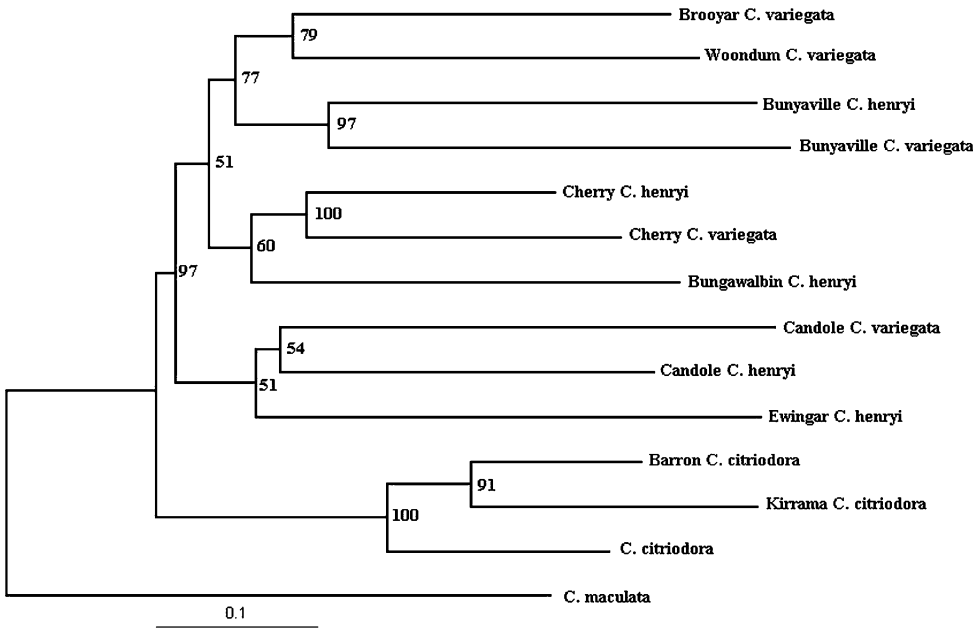


Fig. 6 Neighbour-joining cluster of predefined populations based on allele size variation at five microsatellite markers

Table 2 Pairwise genetic differentiation (F_{ST}) among predefined populations of *C. variegata* and *C. henryi*

	Woo V	Buny H	Buny V	Cherr H	Cherr V	Bunga H	Ewin H	Ewin V	Cand H	Cand V
Woondum V										
Bunyaville H	0.092									
Bunyaville V	0.071	0.041								
Cherry tree H	0.078	0.063	0.038							
Cherry tree V	0.096	0.076	0.052	0.011						
Bungawalbin H	0.084	0.092	0.076	0.041	0.069					
Ewingar H	0.124	0.120	0.085	0.089	0.096	0.095				
Ewingar V	0.123	0.136	0.088	0.103	0.108	0.105	0.039			
Candole H	0.131	0.119	0.091	0.097	0.105	0.095	0.007	0.051		
Candole V	0.129	0.110	0.089	0.094	0.106	0.098	0.009	0.027	0.021	

All F_{ST} values are significant at $P = 0.01$

V—*C. variegata*; H—*C. henryi*; abbreviated location names on column headings are the same as in raw headings

topology, however, showed populations of the two species (*C. variegata* and *C. henryi*) to be paraphyletic, with nodes representing geographic origin rather than taxonomy (Fig. 6), consistent with differentiation (F_{ST}) among the predefined populations (Table 2). The *C. variegata* and *C. henryi* from Queensland (Bunyaville, Woondum, Brooyar) were clustered together, as was Cherry Tree and Bungawalbin. These locations are geographically proximal (Fig. 1). Exceptionally, populations at Ewingar were allied to those of Candole, albeit with considerably low statistical support (51% bootstrap; Fig. 6). These are not the most proximal locations (Fig. 1). Similarly, Cherry Tree–Bungawalbin group was clustered with the Queensland group with a low statistical support (bootstrap 51%).

Discussion

This study showed that broadly across the range of *C. henryi*, it was genetically homogenous with *C. variegata*, confirming earlier findings based on two sites (Ochieng et al. 2008). This was strong evidence for the genic species hypothesis (Wu 2001), one of three recently mooted (Ochieng et al. 2008; Shepherd et al. 2008) to explain the presence of two sympatric species with distinctive morphology but that were apparently fully interfertile. The current study also found spatial uniformity at the stand level generally in spotted gums, consistent with the high levels of gene flow hypothesised across sites and hence relatively weak geographic structuring (Ochieng

et al. 2008; Shepherd et al. 2008). There were exceptions to this uniformity at some sites, however, with evidence that some even aged stands (Cherry Tree site 1; Candole site 3) tended to separate out in a STRUCTURE analysis based on their allele frequencies, and tended to have higher levels of relatedness than on average (see below). We propose that within-site substructure emanate from localised events such as regeneration following fire from a few surviving trees, rather than structure resulting from micro niche adaptation.

Species were genetically homogeneous

Genetic structure in two sympatric and widespread forest trees, *C. variegata* and *C. henryi* was assessed. Microsatellite markers revealed high variation, hence ample power to detect genetic structure within and between species. In both STRUCTURE and clustering based on allele frequency of individuals, genetic structuring was found to transcend taxonomic boundaries, such that the two species appeared completely genetically homogenised at each sampling location, i.e., all identifiable genetic clusters comprised both species regardless of sampling design, models and parameters used (Figs. 3, 4). However, the number of loci used is probably too low to detect substructuring at the location level, and certainly insufficient to discount the possibility that taxonomic structure might have occurred in sympatric locations. Moreover, it is notable that population-based analysis showed very low, but significant F_{ST} between species within populations. But also worth notice are the F_{ST} values between species within locations which are much lower than that within species between locations, consistent with our earlier findings (Ochieng et al. 2008). UPGMA clustering of individuals did not show any differences among populations. Both individual-based clustering and F_{ST} estimates rely on allele frequencies; thus differentiation between species within localities, though significant in a statistical framework, is insufficient to invalidate the homogeneity paradigm. If neutral molecular variation were used to classify these trees, then the two spotted gums merit no species recognition. If the traits used to identify the species are under the control of a few genes with categorical effects (Doebley and Lukens 1998), however, as is the case for some genes with major effect on domestication traits in plants (Doebley and Stec 1991), then unlinked microsatellite loci in the two genomes may not be differentiated. Similarly, natural selection may bring about an analogous situation, where there may be no neutral marker variation if 'species' consisted of regions in the genome that, by virtue of their phenotypic effects, are under strong disruptive selection and thus have high differentiation, whereas the remainder of the genome behave as a single population (Beaumont 2005).

Results of this study are similar to those observed in other eucalypt species: Turner et al. (2000) analysed genetic variation within and between *Eucalyptus risdonii* and its more widespread sister species, *E. tenuiramis* using allozymes, and found that the Southern *E. tenuiramis* had greater genetic affinity with *E. risdonii* than with northern *E. tenuiramis*. In studies of a natural hybrid swarm between *Eucalyptus amygdalina* and *E. risdonii* and nearby allopatric stands using RAPD markers revealed that, despite clear morphological differences, variation was greatest within populations and lowest between species (Sale et al. 1996). Using field trials, Dutkowski and Potts (1999) partitioned *Eucalyptus globulus* species into 13 geographic races based on the quantitative genetic variation (morphological and developmental traits) among progenies. In a more recent analysis of racial differentiation based on quantitative (Q_{ST}) and molecular (F_{ST}) data, Steane et al. (2006) detected spatial differentiation across the distribution range of *E. globulus*. They found that when molecular markers (SSRs) were used, geographically proximal races tended to be more closely related than geographically distant races, contrasting markedly with analyses based on quantitative genetic data, where some races appeared to be highly divergent from their geographically closest neighbours. They found that some of the quantitative traits used for defining races of *E. globulus* have been influenced by natural selection, resulting in cases of both phenotypic divergence of parapatric races and phenotypic convergence of allopatric races. These studies suggested that morphological differences does not necessarily reflect neutral genetic (DNA) variation and in some cases detectable DNA differences between spatially distant populations of the same species may be as great as or greater than the differences between species. This scenario has also been observed in non-plant systems—the genus *Geospiza*, one of the genera comprising Darwin's finches (Freeland and Boag 1999).

A structure model with six genetic clusters was optimal

In this study, six genetic clusters corresponding to sampling locations were identified. Geographically proximal clusters showed less differentiation compared to the more distant ones, consistent with a stepping-stone model of migration. A search for the optimal number of clusters (K) revealed a distinct peak at $K = 2$ and another at $K = 6$ (Fig. 2). The first peak was stronger, suggesting that a two-cluster model characterised by within-population substructure was optimal. The model with six populations ($K = 6$) delineated groups that corresponded with sampling localities, except for two cases of within-locality substructure (Fig. 3). This model was congruent with the population-based clusters (Neighbour-Joining tree; Fig. 6).

Whereas STRUCTURE methods were based on the individuals, the population approach was based on geographic groups. The most likely cause for the detection of ($K = 2$) in this study was partial sampling, which has been demonstrated to confound the detection of true K (Evanno et al. 2005).

Structuring followed a stepping-stone model of migration

Five of the six homogeneous clusters identified corresponded to their latitudinal positions down the east coast of Australia: Brooyar and Woondum near Gympie, Bunyaville in Brisbane, Cherry Tree, which stretches further inland from Casino, Bungawalbin, located south of Casino, and Candole, on the coastal plain at Grafton. Genetic differentiation between clusters showed a tendency to follow geographic distances, with neighbouring populations being less differentiated than more distant ones (Table 3), suggesting that gene flow was restricted by geographic distance. A stepping-stone model of migration (Kimura and Weiss 1964) is proposed when populations are arranged linearly so that genes migrate more frequently between adjacent populations than between more distant populations. The generally linear, latitudinal arrangement of populations within our sample of these taxa and the pattern of genetic distances among them, suggested that geography was a dominant factor shaping genetic structure and reflected a stepping-stone model of migration (Kimura and Weiss 1964), as proposed earlier for the spotted gum complex of four taxa (Shepherd et al. 2008).

Some within-locality structure was identified

The optimal number of genetically distinct populations largely corresponded with sampling localities. However, two cases of within-locality structure were identified, one within Cherry Tree SF (site 1) and the other within Candole SF (site 3). These neighbourhoods were delineated from their locales whether the entire dataset or locality-specific datasets were used. A common feature for both neighbourhoods in our study was the low elevation—both neighbourhoods were located on the escarpments of a ridge; lower in elevation, and at Cherry Tree, the group was characterised by a young stand of spotted gums, possibly a regeneration following a fire. These two sites therefore differed from other sites in our study as well as those reported by King (2004). Unlike our study, where neighbourhoods comprised trees within a 30-m radius, the trees sampled in the study of King (2004) were spaced farther apart (50 m)—hence less power to detect neighbourhood structure. Our results suggest that the genetic structuring is not uniform in spotted gums throughout the range; variation exists among provenances and the type of structuring detected is susceptible to the sampling design.

Implications for tree management and forestry

Ideally, conservation of species would rely on knowledge of reproductive isolation, species boundaries, population structure and substructure. Spotted gums are widespread (Hill and Johnson 1995); none are threatened nor endangered (Barbour et al. 2008), however, the maintenance of genetic diversity both within native forests and plantations may be essential for their long-term utilization (as sources of fibre from both plantations and native forests in subtropical Australia; Lee 2007). Three main conclusions can be drawn from this study: (1) the two ‘species’ form a genetically homogeneous unit at each location and can thus be treated as a single species. (2) Population structure within the species is weakly determined by geographic isolation but other unidentified factors also contribute to structure; (3) genetic substructure detectable may not be uniform among localities (provenances). Consequently, the two taxa can be treated as one species for gene pool management purposes. To capture greater genetic variation in this complex, one needs to sample from different locations, but not different ‘species’. Similarly, gene flow

Table 3 Genetic differentiation (F_{ST}) among homogeneous subsets (geographic distances in km are given in parentheses)

	Woondum	Bunyaville	Cherry Tree	Cherry S1	Bungawalbin	Ewingar	Candole	Candole S3
Woondum								
Bunyaville	0.068 (125)							
Cherry Tree	0.080 (300)	0.051 (175)						
Cherry site 1	0.131 (300)	0.096 (175)	0.100 (000)					
Bungawalbin	0.084 (320)	0.072 (190)	0.056 (058)	0.115 (058)				
Ewingar	0.120 (320)	0.089 (200)	0.087 (033)	0.043 (033)	0.093 (085)			
Candole	0.131 (400)	0.092 (270)	0.098 (110)	0.047 (110)	0.095 (082)	0.021 (110)		
Candole site 3	0.253 (400)	0.183 (270)	0.154 (110)	0.239 (110)	0.175 (082)	0.217 (110)	0.190 (000)	

All F_{ST} values are significant at $P = 0.01$

between planted and native trees can thus be discussed with regards to locations of origin rather than taxonomy. Gene flow is expected through on-farm deployment between plantings from different geographic regions, which appeared distinct according to these results. However, issues of genetic pollution (gene flow from a domestic, feral, non-native or invasive species to a wild population; Potts et al. 2003) regarding the two 'species' may remain valid if the context of pollution relate to specific genes controlling certain desired quantitative traits, and if the 'species' differ at these traits. In such a case, it may be prudent to still manage them as two 'species' separately. Importantly, the number of loci used in this study (five) may have been too low to detect structuring between species at the location level, as such; discounting the possibility that taxonomic structure occurred in sympatric locations might have serious management implications.

Conclusion

We conclude that two spotted gums currently classified as different species: *C. variegata* and *C. henryi* are not differentiated at nuclear microsatellite loci. We view them as morphotypes of a single structured species, where structuring aligns with spatial isolation. We speculate that they are genic species, where phenotypic divergence is maintained by segregation at genes with large effects or strong disruptive selection. The two can therefore be considered as a single gene pool for management purposes. However, some traits such as flowering times, capsule size and growth rates may be important considerations in tree improvement. The two taxa can also be managed separately where purity concerns quantitative traits or divergent genes not assessed in this study.

Acknowledgments This work forms part of J.W. Ochieng's Doctoral Research, supported by a Commonwealth Scholarship, Australian Research Council (ARC) Linkage Grant, Queensland Department of Primary Industries and Fisheries, and Forestry Plantations Queensland. We thank Mr. Mike Henson and the staff of Forestry NSW, as well as Queensland Environment Protection Agency (EPA) for logistical support, Mr. David Kleinig and Ian Bevage for help with sample collections, and Mr. Tony Bean of the Queensland Herbarium for help with sample identification.

References

Bacles CFE, Burczyk J, Lowe AJ, Ennos RA (2005) Historical and contemporary mating patterns in remnant populations of the forest tree *Fraxinus excelsior* L. Evolution 59:979–990

Barbour RC, Crawford AC, Henson M, Lee DJ, Shepherd M (2008) The risk of pollen-mediated gene flow from exotic *Corymbia* plantations into native *Corymbia* populations in Australia. For Ecol Manag 256:1–19

Beaumont MA (2005) Adaptation and speciation: what can F_{ST} tell us? Trends Ecol Evol 20:435–440

Cavalli-Sforza LL, Edwards WF (1967) Phylogenetic analysis: models and estimation procedures. Evolution 32:550–570

Doebley J, Lukens L (1998) Transcriptional regulators and the evolution of plant form. Plant Cell 10:1075–1082

Doebley J, Stec A (1991) Genetic analysis of the morphological differences between maize and teosinte. Genetics 129: 285–295

Dutkowski GW, Potts BM (1999) Geographic patterns of genetic variation in *Eucalyptus globulus* ssp. *globulus* and a revised racial classification. Aust J Bot 47:237–263

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol 14:2611–2620

Felsenstein J (1995) PHYLIP: phylogeny inference package. University of Washington, Seattle

Freeland JR, Boag PT (1999) The mitochondrial and nuclear genetic homogeneity of the phenotypically diverse Darwin's ground finches. Evolution 53:1553–1563

Goudet J, Raymond M, Demeeus T, Rousset F (1996) Testing differentiation in diploid populations. Genetics 144:1933–1940

Hill KD, Johnson LAS (1995) Systematic studies in the eucalypts 7. A revision of the bloodwoods, genus Corymbia (Myrtaceae). Telopea 6:185–504

Kimura M, Weiss GH (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics 49:561–576

King R (2004) Spatial structure and population genetic variation in a Eucalypt species complex. PhD thesis, Griffith University

Lee DJ (2007) Development of *Corymbia* species and hybrids for plantations in eastern Australia. Aust For 70:11–17

Lexer C, Buerkle CA, Joseph JA, Heinze B, Fay MF (2007) Admixture in European Populus hybrid zones makes feasible the mapping of loci that contribute to reproductive isolation and trait differences. Heredity 98:74–84

McDonald MW, Bean AR (2000) A new combination in *Corymbia* section *Politaria*: *C. citriodora* var. *variegata* (Myrtaceae). Austrobaileya 5:735–736

McDonald MW, Butcher PA, Bell JC, Larmour JS (2000) Intra- and interspecific allozyme variation in eucalypts from the spotted gum group, *Corymbia*, section *Politaria* (Myrtaceae). Aust Syst Bot 13:491–507

Moritz C (1994) Defining "evolutionary significant units" for conservation. Trends Ecol Evol 9:373–375

Muir G, Fleming CC, Schlotterer C (2000) Species status of hybridizing oaks. Nature 405:1016

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Ochieng JW, Shepherd M, Baverstock PR, Nikles G, Lee DJ, Henry RJ (2008) Genetic variation within two sympatric spotted gum eucalypts exceeds between taxa variation. Silvae Genet 57:249–256

Pearse DE, Crandall KA (2004) Beyond FST: analysis of population genetic data for conservation. Conserv Genet 5:585–602

Potts BM, Barbour RC, Hingston AB, Vaillancourt RE (2003) Corrigendum to: TURNER REVIEW No. 6. Genetic pollution of native eucalypt gene pools—identifying the risks. Aust J Bot 51:333

Pritchard JK, Wen W (2004) Documentation for structure software: version 2. http://pritch.bsd.uchicago.edu/structure.html

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

Raymond M, Rousset F (1995) An exact test for population differentiation. Evolution 49:1280–1283. http://genepop.curtin.edu.au/

Rice WR (1989) Analyzing tables of statistical tests. Evolution 43:223–225

Sale MM, Potts BM, West AK, Reid JB (1996) Molecular differentiation within and between *Eucalyptus risdonii*, *E. amygdalina* and their hybrids using RAPD markers. Aust J Bot 44:559–569

Shepherd M, Kasem S, Lee D, Henry R (2006) Construction of microsatellite linkage maps for *Corymbia*. Silvae Genet 55:228–238

Shepherd M, Kasam S, Ablett G, Ochieng JW, Crawford A (2008) Developing a genetic classification for gene pool management of spotted gums. Aust Syst Bot 21:1–11

Southerton SG, Birt P, Ford HA (2004) Review of gene movement by bats and birds and its potential significance for eucalypt plantation forestry. Aust For 67:44–53

Steane DA, Conod N, Jones RC, Vaillancourt RE, Potts BM (2006) A comparative analysis of population structure of a forest tree, *Eucalyptus globulus* (Myrtaceae), using microsatellite markers and quantitative traits. Tree Genet Genomes 2:30–40

Turner C, Wiltshire RJE, Potts BM, Vaillancourt RE (2000) Allozyme variation in the *Eucalyptus risdonii* Hook.f–*E. tenuiramis* Miq. complex. Conserv Genet 1:209–215

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370

Wright S (1951) The genetic structure of populations. Annu Eugen 15:323–354

Wu CI (2001) The genic view of the process of speciation. J Evol Biol 14:851–865