www.publish.csiro.au/journals/asb *Australian Systematic Botany* **21**, 1–11

Genetic structuring in the spotted gum complex (genus *Corymbia*, section *Politaria*)

Mervyn Shepherd[A,B], *Shabana Kasem*[A], *Gary Ablett*[A], *Joel Ochieng*[A] *and Allison Crawford*[A]

[A]Centre for Plant Conservation Genetics and Subtropical Forestry Centre, Southern Cross University, PO Box 157, Lismore, NSW 2480, Australia.

[B]Corresponding author. Email: mervyn.shepherd@scu.edu.au

Abstract. Spotted gums (genus *Corymbia*, section *Politaria*) occur as a species replacement series along the eastern seaboard of Australia, their distributions marked by regions of disjunction and sympatry. Their taxonomy remains controversial, with species assignment often challenging and reliant on knowledge of geographic origin as well as subtle morphological or leaf-oil variation. In the present paper, we explore a classification for spotted gums, without assuming predefined geographic or taxonomic groups but instead using genetic structure at microsatellite marker loci ($n = 9$) and a Bayesian model-based clustering approach implemented in STRUCTURE software. The *C. torelliana* outgroup ($n = 21$; section Cadagaria) formed a well resolved cluster (minimum pairwise $F_{st} = 0.19$). Four populations were evident within the spotted gums ($n = 93$) but structure was weak (pairwise F_{st} range 0.13–0.05). Geographic distance, topography and distribution disjunction were major determinants of structure, with migration among populations approximating a linear stepping-stone model. *Corymbia maculata* was resolved as a taxon and had the greatest genetic distance from any other population (minimum pairwise F_{st} 0.08). Three clusters were evident within the northern taxa but alignment with taxonomic groupings was poor. *C. citriodora* material from north of a major disjunction in central Queensland formed a Northern population. *C. citriodora*, *C. variegata* and *C. henryi* material south of this disjunction but north of the Border Range, formed a Central population, whereas a Southern population comprised *C. variegata* and *C. henryi* from predominately south of the Border Range.

Introduction

Describing population structure and its conversion into evolutionary or management units provides a framework for conservation management and sustainable utilisation (Moritz 1994). Populations within species can be identified by multilocus allele frequencies and statistical approaches for clustering individuals or populations (Allendorf and Luikart 2007). A population approach imposes some pre-existing grouping, either geographic or phenetic, on the data to compute population allele frequencies and genetic distances among populations. Populations are identified as clusters in population dendrograms or by multidimensional representation such as principle components analysis. Relying on taxonomic groups may be misleading, however, because phenetically similar taxa may not represent breeding populations, and conversely, highly polytypic species may be interbreeding (Ridley 2004). Grouping individuals by geographic locality can also be problematic because boundaries may be arbitrary and population allele frequencies may be biased by transitory migration and inter-population hybridisation (Allendorf and Luikart 2007). In contrast, individual-based methods for determining population structure make no *a priori* assumptions about how many populations exist or their boundaries. In this regard, the use of individual-based approaches for determining population structure may be particularly appropriate for studying groups of closely related sympatric species that are very similar morphologically.

Spotted gums are 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). Their taxonomy remains controversial as classification often relies on subtle variation in morphology and leaf oils, and knowledge of the geographic origins of a specimen (McDonald and Bean 2000). Four species that occur as a latitudinal replacement series were recognised by Hill and Johnson (1995) in their revision of the genus *Corymbia*. Isozyme analyses have shown distinct geographic structuring within spotted gums, with the following two distinct genetic alliances suggested: *C. citriodora*–*C. variegata* and *C. henryi*–*C. maculata* (McDonald *et al*. 2000). Subsequent reanalysis of these data, however, revealed that selection at one locus, *PGM2*, had inflated genetic-distance estimates between *C. henryi* and the other northern taxa, with the effect of exaggerating the affinity between *C. henryi* and *C. maculata* (King 2004). Analysis of the diversity and distribution of chloroplast DNA haplotypes confirmed low levels of differentiation among northern spotted gum taxa, the distinctness of *C. maculata* from northern taxa and population subdivision within *C. maculata* (King 2004).

 10.1071/SB07028 1030-1887/08/010001

The aim of the present study was to develop a classification for spotted gum on the basis of genetic structure. One potential advantage of a genetic classification system for spotted gum might be the ability to classify individuals without information on its geographical origin, provided a genetic group or population can be associated with a geographic location. Furthermore, we were interested in the relationship between genetic and taxonomic groupings, and whether by comparing the alignment of genetic and taxonomic groupings, some of the ongoing uncertainty surrounding the taxonomy of the group might be resolved. Here, we describe genetic structuring at microsatellite markers by an individual-based approach that uses variation to avoid the limitations of imposing predetermined taxonomic or geographic structure on the analysis. We discuss the congruence in genetic structure on the basis of microsatellite markers, with the population structuring determined in previous genetic-marker studies, and its implications for taxonomy of the section.

Materials and methods

Study taxa, distributions and sampling

The four species of spotted gums, *C. citriodora* (Hook.) K.D.Hill & L.A.S.Johnson, *C. variegata* (F.Muell.) K.D.Hill & L.A.S.Johnson, *C. henryi* (S.T.Blake) K.D.Hill & L.A.S.Johnson and *C. maculata* (Hook.) K.D.Hill & L.A.S.Johnson belong to the section *Politaria* of the genus *Corymbia* (Hill and Johnson 1995). *C. citriodora* is distributed from west of Cooktown to south of Gladstone, west to the Great Dividing Range west of Springsure, Queensland, with a major disjunction of 300 km separating northern and southern populations (Hill and Johnson 1995; McDonald *et al*. 2000). *C. variegata* has a wide range from Carnarvon and Dawes Ranges north of Monto (Queensland), contracting southwards to subcoastal regions south to Nymboida River and north-west of Coffs Harbour in New South Wales (NSW) (Hill and Johnson 1995). Intergrades and hybrids with *C. citriodora* are believed to occur in the north-east of *C. variegata*'s range (Hill and Johnson 1995). *C. henryi* tends to occur on less fertile low-lying country from Brisbane (Queensland) in the north, south to Glenreagh near Grafton in NSW (Hill and Johnson 1995). *C. henryi* is presumed to hybridise with *C. variegata* in regions of sympatry but this is yet to be confirmed (Hill and Johnson 1995). *C. maculata* occurs mainly along the coast of NSW, from the Manning River valley in the north to near Bega in the south, with outlying occurrences near Orbost in eastern Victoria (Hill and Johnson 1995). *C. torelliana* (F.Muell.) K.D.Hill & L.A.S.Johnson (Cadagi), which belongs to a closely related section, *Cadagaria*, was also included in the study as an outgroup. It occurs in a restricted distribution in rainforest margins in northern Queensland and is known to spontaneously hybridise with spotted gums (Hill and Johnson 1995).

Samples were selected from collections of bark and foliage tissue held at the Centre for Plant Conservation Genetics, Southern Cross University, to provide broad representation across the four taxa of spotted gums and *C. torelliana*. A total of 118 individuals was chosen, comprising 25 individuals from six populations of *C. citriodora* (ranging from 1 to 5 trees per locality); 22 *C. henryi* individuals from eight locations (1–6 individuals per location); 24 *C. maculata* specimens from six locations (3–6 trees per location); 21 *C. torelliana* from three locations plus three trees of unknown origin, and 26 *C. variegata* from eight locations (1–12 trees per location) (Fig. 1, Table 1). In the cases of *C. henryi* and *C. variegata*, bark and/or foliage were collected from trees from native forests. Field identifications were later verified on the basis of a herbarium specimen by a botanist at the Queensland Herbarium. One *C. variegata* sample, the GYMP sample, was obtained from a Queensland Department of Primary Industries and Fisheries (QDPIF) trial. Specimens for the remaining taxa were obtained from seedlings in the Forests NSW Grafton nursery, except three *C. torelliana* samples of unknown origin which were obtained from ornamental plantings at Gympie and provided by Dr D. Lee (QDPIF), and one *C. citriodora* sample (DUAR) which was obtained from a QDPIF trial.

DNA extraction and microsatellite-marker analysis

DNA extraction and microsatellite genotyping were conducted as described by Shepherd *et al*. (2006). Nine microsatellite loci were selected on the basis of their transferability across species, polymorphism, lack of null alleles and because they were known to be unlinked or loosely linked from genetic mapping studies (Shepherd *et al*. 2006. The loci used were EMCRC 27, 93, 38, 37, 40, 36, 34, 35 and 46 (Jones *et al*. 2001; Shepherd *et al*. 2006).

Determining optimal clustering with STRUCTURE *software*

STRUCTURE uses a Bayesian model-based approach to group individuals by multi-locus genotypes (Pritchard *et al*. 2000). Successive iterations of the model attempt to optimise clustering by introducing structure to account for Hardy–Weinberg and linkage disequilibria. Determining the optimal number of populations (K) is an *ad hoc* process based on testing a range of models with different K values and inspecting the estimated log probability of data Pr(X|K) to identify the most likely model (Pritchard *et al*. 2000; Pritchard and Wen 2004). In simulation studies it has also been observed that as the real K is reached, likelihoods for larger Ks plateau and the variance increase (Latch *et al*. 2006). The approach of Evanno *et al*. (2005) which uses an alternate optimisation criterion—delta K, related to the second-order rate change in the log probability of the data—was also used in the present study. A hierarchical approach was used to successively analyse structure, with the entire dataset examined initially and identified populations then analysed to identify further substructure. Model parameters followed those described by Evanno *et al*. (2005). Run parameters were 10 000 iterations each for the 'burn-in' and Markov Chain Monte Carlo (MCMC) stages and a 'correlated allele frequency model' for ancestry. 'No admixture' models were used for detecting primary populations because limited gene flow was expected at this higher order of structure between sections and species. Both 'admixture' and 'no admixture' models were investigated when testing for substructure within primary populations. Ten runs for each K (for K = 1–10) were used to calculate means and standard deviations for posterior probability of the data for a given K (L(K)) and delta K. These parameters were found to

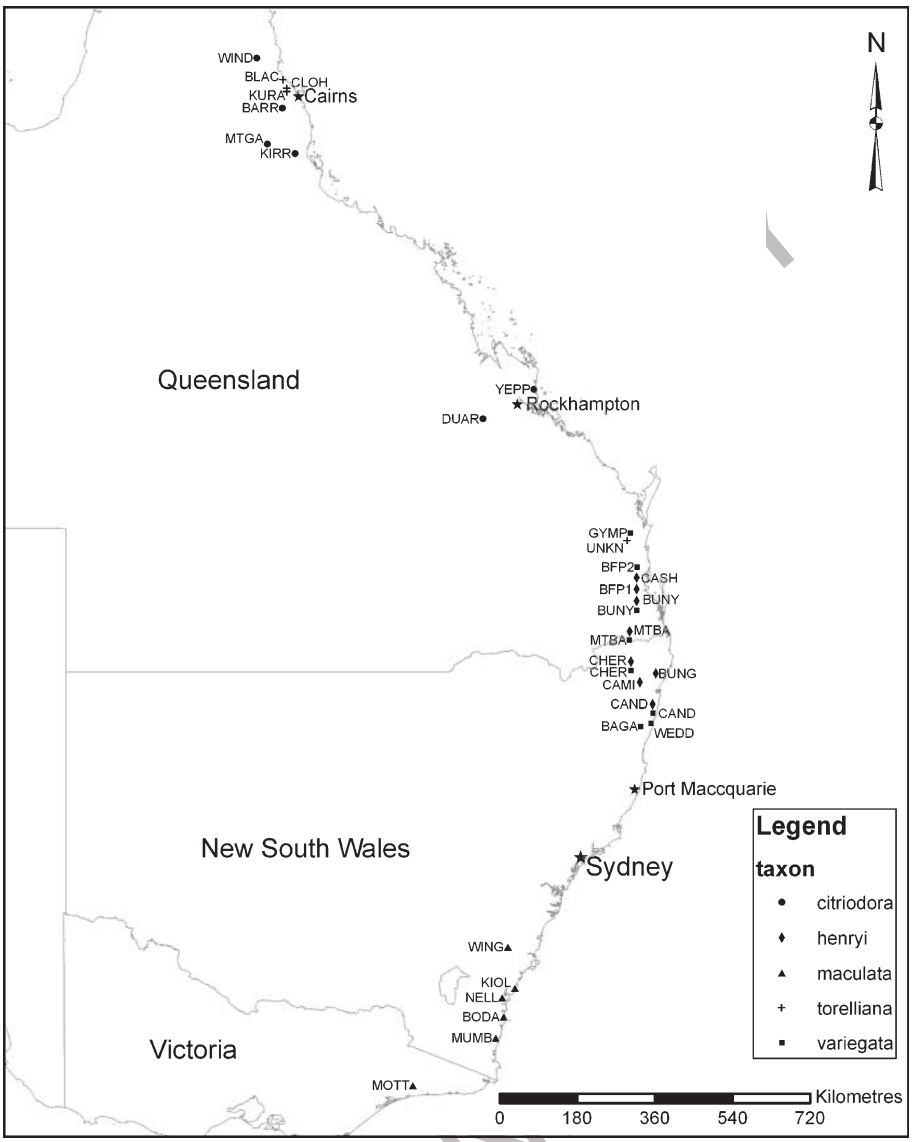

Fig. 1. Locations of sampling sites for the spotted gum reference population (locality abbreviations are defined in Table 1). The Border Range is located between the Mount Barney locality (MTBA) and Cherry Tree State Forest (CHER).

provide stable optimisation criteria for replicate runs of our dataset when detecting primary populations. A lack of structure was assumed when plots of delta K provided little evidence of a clear peak for any K value (Evanno *et al*. 2005). Further corroboration of lack of structure was sought by examining assignments at most likely K values. Typically, the proportion of samples assigned to each population is symmetrical when there is no structure and most individuals are admixed (Pritchard and Wen 2004). Alpha values also tend not to stabilise in this scenario.

Ancestry assignments generated by STRUCTURE were used to classify individuals into primary populations. Individuals with population admixture were assigned to a single population if they possessed >90% ancestry from one population; otherwise they were considered to be population hybrids. Pairwise F_{st} were used to examine relationships among populations determined by both taxonomic and genetic classifications. Pairwise F_{st} and average genetic diversity were estimated with ARLEQUIN software (Excoffier *et al*. 2005), using the standard data model. Population admixture individuals were excluded from pairwise F_{st} estimates between genetic groups

Comparison of genetic and phenetic groupings by assignment tests and by testing for isolation by distance

The proportion of individuals with incorrect or ambiguous assignments was used to test the relative performance of taxonomic and genetic groupings by using self-classification of the entire dataset ($n = 118$). Therefore, in the case of genetic

Table 1. Sampling locations for 118 *Corymbia* sp. individuals arranged by taxon and locality

Locality descriptor consists of the state, land tenure and locality name: QLD = Queensland; NSW = New South Wales; SF = State Forest with SF number if known; NP = National Park; O = Land tenure other than SF or NP

Taxon	Locality	Locality code	Location	Taxon count	Locality count
C. citriodora	QLD SF Windsor SF144	WIND	16°13′S, 144°58′E	25	5
	QLD SF Barron SF194	BARR	17°15′S, 145°30′E		5
	QLD O Mt Garnet—35km south	MTGA	18°00′S, 145°11′E		4
	QLD SF Kirrama SF344	KIRR	18°12′S, 145°46′E		5
	QLD O Yeppoon	YEPP	23°07′S, 150°44′E		5
	QLD O Duaringa (Ex QDPIF trial)	DUAR	23°44′S, 149°40′E		1
C. henryi	QLD O Cashmere/Strathpine	CASH	27°18′S, 152°54′E	22	1
	QLD SF Bunyaville	BUNY	27°22′S, 152°57′E		5
	QLD BFP1 Bris. Forest Park—Cabbage Range Rd	BFP1	27°23′S, 152°55′E		1
	QLD NP Mt Barney	MTBA	28°18′S, 152°43′E		1
	NSW SF Cherry Tree SF169	CHER	28°55′S, 152°45′E		6
	NSW SF Bungawalbin SF152	BUNG	29°02′S, 153°16′E		4
	NSW SF Camira SF	CAMI	29°13′S, 152°56′E		1
	NSW SF Candole SF25	CAND	29°46′S, 153°12′E		3
C. maculata	NSW O Wingello	WING	34°44′S, 150°11′E	24	3
	NSW SF Kioloa	KIOL	35°35′S, 150°20′E		5
	NSW O Nelligen	NELL	35°47′S, 150°04′E		6
	NSW SF Bodalla	BODA	36°11′S, 150°06′E		4
	NSW SF Mumbulla	MUMB	36°38′S, 149°56′E		3
	VIC O Mottle Range	MOTT	37°37′S, 148°13′E		3
C. torelliana	UNKNOWN (Ornamental plantings at Gympie)	UNKN	-	21	3
	QLD O Black Mountain Rd—from south	BLAC	16°40′S, 145°31′E		5
	QLD O Kuranda, west of, on Kennedy Hwy	KURA	16°51′S, 145°36′E		1
	QLD O Clohesy River Rd	CLOH	16°55′S, 145°36′E		12
C. variegata	QLD O Gympie (Woolvi/Wondum) (Ex DPIF trial)	GYMP	26°10′S, 152°45′E	26	1
	QLD SF Bunyaville	BUNY	27°22′S, 152°57′E		6
	QLD BFP2 Bris. Forest Park—Lake Manchester	BFP2	27°28′S, 152°44′E		1
	QLD NP Mt Barney	MTBA	28°18′S, 152°43′E		1
	NSW SF Cherry Tree SF168	CHER	28°55′S, 152°45′E		12
	NSW SF Candole SF24	CAND	29°46′S, 153°12′E		3
	NSW SF Wedding Bells	WEDD	30°04′S, 153°10′E		1
	NSW SF Bagawa	BAGA	30°08′S, 152°57′E		1
Totals				118	118

grouping, individuals were grouped by genetic groups and in the case of taxonomic grouping, individuals were grouped by taxonomic groups. Assignment tests were conducted with GENECLASS v 1.02 (Cornuet *et al*. 1999), with Bayesian simulation of allele frequencies and program default parameters (10 000 simulated individuals and a *P*-value threshold of 0.01). Ambiguous individuals were assigned to more than one group at the stated *P*-value. An incorrect assignment was declared when the lowest likelihood value indicated a group other than the classification group. *F*-statistics were also estimated for genetic and taxonomic grouping, for comparison to other studies. Analysis of molecular variance (AMOVA) and pairwise F_{st} values were estimated with ARLEQUIN v2 software (Excoffier *et al*. 2005). A Mantel test was carried out in ARLEQUIN to test for isolation by distance effects, by using the genetic groups determined in the STRUCTURE analysis. A geographic distance matrix was derived from measurements of the distance between the latitudinal mid-points of each population identified (i.e. *C. maculata*, Northern, Central and Southern populations).

Results

Hierarchical analysis cleaves two populations corresponding to C. torelliana *and* C. maculata *from northern spotted gum taxa*

Inspection of the delta K plot for models with a range of K values from 1 to 10 on the basis of the entire dataset ($n = 118$), including *C. torelliana*, revealed a distinct peak at K = 2 (Fig. 2). This indicated a structure model with two populations was optimal when all taxa were considered together. Examination of ancestry assignments for individuals by the K = 2 models showed that all *C. torelliana* were assigned entirely to one population and all spotted gums were assigned entirely to the second population (data not shown).

A distinct peak at K = 2, when the analysis of the delta K plot for population models for K = 1–10 was repeated, on the basis of the subset of spotted gum samples ($n = 94$) only, again indicated that a model with two populations was optimal (Fig. 3). Ancestry assignments for the K = 2 models showed that all *C. maculata* individuals were assigned to one population, whereas northern

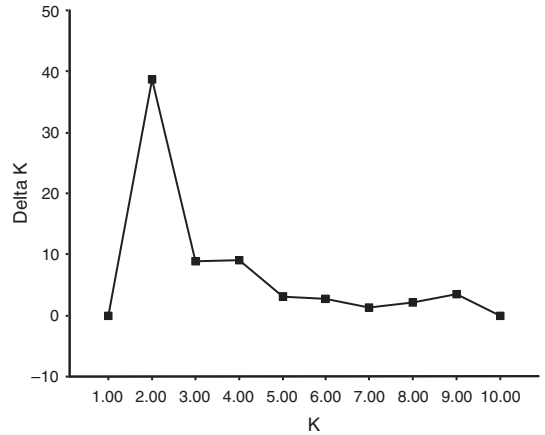

Fig. 2. Magnitude of delta K as a function of K (mean of 10 replicates) for K = 1–10, on the basis of the dataset derived from 115 spotted gum and *Corymbia torelliana* individuals. A distinct peak at K = 2 was indicative that a model with two populations was optimal. All spotted gums grouped to one population and *C. torelliana* to the second.

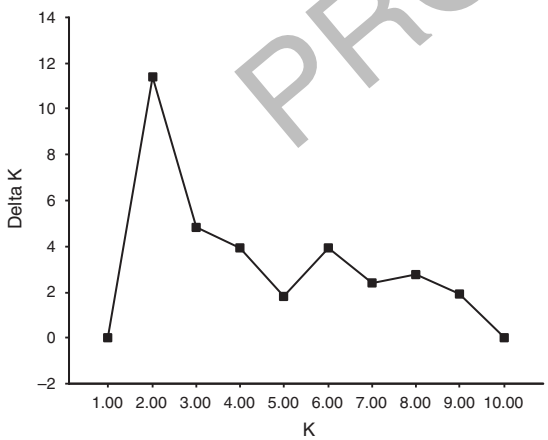

Fig. 3. Magnitude of delta K as a function of K (mean of 10 replicates) for K = 1–10, on the basis of the dataset derived from 94 spotted gum individuals. A distinct peak at K = 2 was indicative that a model with two populations was optimal. All *Corymbia maculata* individuals clustered to one population.

taxa of spotted gums were assigned to the second population (data not shown).

Three populations were optimal for the three northern spotted gum taxa, but groups had poor alignment with taxonomy

A three-population model was indicated as optimal by the peak at K = 3 on the plot of delta K for models where K ranged from 1 to 10 on a dataset containing the three northern spotted gum taxa ($n = 73$), *C. citriodora*, *C. variegata* and *C. henryi* (Fig. 4). Ancestry assignments indicated poor alignment between genetic populations and taxonomic groupings, with boundaries between populations tending to align more with geographic features or distribution disjunctions (Fig. 5, Table 2). One population

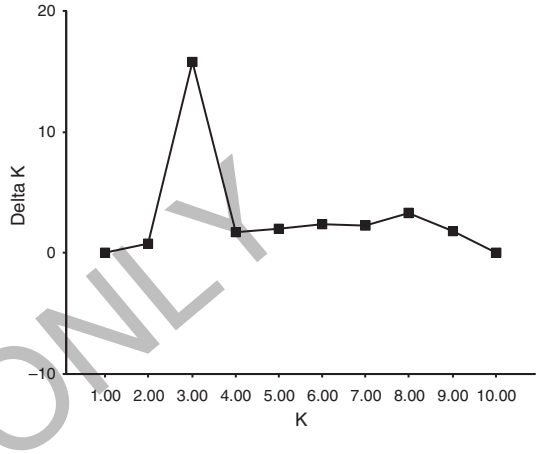

Fig. 4. Magnitude of delta K as a function of K (mean of 10 replicates) for K = 1–10, on the basis of the dataset derived from 73 spotted gum individuals from the three northern taxa, *Corymbia citriodora*, C. *variegata* and *C. henryi*. A peak at K = 3 was indicative that a model with three populations was optimal.

(Northern) corresponded to *C. citriodora* from localities north of the 300-km disjunction in central Queensland. The second population (Central) encompassed individuals from all three taxa from localities south of the central Queensland disjunction but typically north of the Border Range between Queensland and NSW. This included *C. citriodora* taxa from the Yeppoon locality on the coast in central Queensland, *C. variegata* from the Bunyaville locality north of Brisbane, Brisbane Forest Park and Mount Barney (Queensland). This second population also included *C. henryi* individuals from Mount Barney, Bunyaville and Candole State Forest (NSW). The third population (Southern) comprised *C. variegata* and *C. henryi* mostly from localities south of the Border Range. Samples of both taxa from Cherry Tree State Forest belonged predominantly to this population as did most material from other NSW localities, including Candole State Forest, Wedding Bells State Forest and Bungawalbin State Forest. The strongest exception to a strictly latitudinal structuring of genetic variation existed in this population because some *C. henryi* individuals with this population ancestry were obtained at the Bunyaville locality north of Brisbane. Curiously, a single *C. citriodora* representative from an inland central Queensland locality, Duaringa, was also assigned to this population. Further sampling will be required to validate this relationship, however, as only one sample (from a planted source) from this locality was studied.

Admixture individuals between the Central and Southern populations were common

Thirteen (17%) of ninety-four spotted gum individuals were declared to be of a population admixture (i.e. <90% ancestry from a single population) and these were typically an admixture of the Central and Southern populations (Table 2). Most (8) had been taxonomically classified as *C. henryi*, with the remainder as *C. variegata*. Individuals of mixed ancestry suggest recent inter-population hybridisation and were identified from most localities where *C. henryi* and

PROOF ONLY

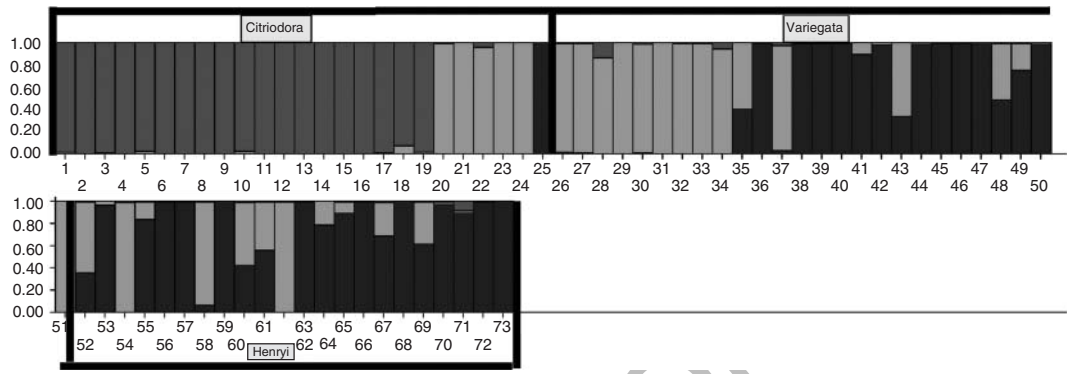

Fig. 5. Ancestry assignments for 73 individuals from the northern spotted gum taxa determined in a STRUCTURE analysis by using a K = 3 model. Ancestry was assigned proportionally to one or more of the three populations indicated by the three shades of grey. Individuals are arranged by taxa (*Corymbia citriodora*; *C. variegata* then *C. henryi*—a vertical black line denotes divisions between taxa). Within taxa, individuals are arranged according to their locality, sequentially from the most northern locality; e.g. individual no. 1 belongs to the *C. citriodora* taxon and originated in the most northerly locality, whereas individual no. 25 also belongs to the *C. citriodora* taxon but was obtained from the most southerly locality sampled. Individuals are numbered sequentially and their number corresponds to their index number in Table 2.

C. variegata occur in sympatry, namely Bungawalbin, Cherry Tree, Candole, Brisbane Forest Park and Bunyaville localities. The high proportion of putative hybrids suggests high levels of contemporary gene flow between the Central and Southern populations.

Substructure within primary populations

Each of the five primary populations (1, *C. torelliana*; 2, *C. maculata*; 3, Northern; 4, Central; and 5, Southern) were tested separately for substructure as described above, except that both 'no admixture' and 'admixture' models were tested. There was no evidence of substructure within any of the five primary populations as indicated by a lack of pronounced peaks in plots of the delta K criterion, instability in α parameter plots and symmetrical assignments of ancestry to all populations for all or most individuals (data not shown).

A lack of substructure in *C. maculata*, despite the detection of a distinct north coast NSW and a south coast NSW region subgroups in previous studies of *C. maculata* (McDonald *et al*. 2000; King 2004), was most likely due to the lack of samples from the north coast NSW in the present study. The most northerly *C. maculata* sampled in the present study was Wingello (WING in Fig. 1), which is the most northern population of the south coast NSW subgroup sampled in previous studies (McDonald *et al*. 2000).

Relationships among genetic groups, compared with taxonomic groupings

The amount of variation among the five populations identified by STRUCTURE analysis, (including the *C. torelliana* population) was 15.0% (AMOVA not shown). This variation declined to about half (8.6%) when *C. torelliana* was excluded and the four spotted gum populations were considered alone, indicating a low degree of differentiation among the spotted gums. Comparatively more variation was evident between populations when individuals were grouped by taxonomy (13%; AMOVA not shown). Pairwise F_{st} values for genetic groups suggested isolation by distance effects among the spotted gums, with adjacent populations showing higher affinity (Table 3). *C. maculata* was the most strongly resolved population and was genetically the closest to the Southern population (pairwise $F_{st} = 0.08$). Among the northern taxa, there was a similar degree of differentiation across the group (Northern *v.* Southern, pairwise $F_{st} = 0.09$); however, between neighbouring pairs, Northern and Central, and Central and Southern, differentiation was lower (for both, pairwise $F_{st} = 0.05$). A Mantel test for matrix correlation also tended to support geographic distance as a major determinant of genetic structure, as the matrix correlation coefficient for genetic distances (pairwise F_{st}) based on genetic groups identified by the STRUCTURE analysis and geographic distance was 0.77 (P-value = 0.089).

When compared with its approximate taxonomic equivalent (i.e. North cf. *C. citriodora*; Central cf. *C. variegata* and South cf. *C. henryi*), the pairwise F_{st} values were the same or slightly higher, indicating that clusters identified by STRUCTURE were resolved better as a result of redefining groups and removing admixed individuals. As *C. torelliana* and *C. maculata* were not redefined as a result of STRUCTURE analysis, their average gene diversity did not change (Tables 3 and 4). Genetic diversity in *C. henryi* also remained unchanged. Regrouping Southern *C. citriodora* localities with the Central population lead to a reduction and increase in diversity of *C. citriodora* and *C. variegata*, respectively (Tables 3 and 4).

Relative performance of genetic and taxonomic groupings

Fewer ambiguous self-assignments were obtained with genetic rather than taxonomic groupings. Thirty-three per cent of self-assignments with taxonomic groupings were ambiguous, whereas only 22% were ambiguous with genetic classification (Table 5). With the morphological taxonomy, *C. variegata* and *C. henryi* taxa contained the highest number of ambiguously assigned individuals. In the 'genetic' groupings, the Admixed and Southern groups contained the highest numbers of ambiguously assigned individuals. Two incorrect

Table 2. Proportional ancestry assignments for 73 spotted gums from the three northern taxa

Assignments generated in STRUCTURE by using a K = 3 model. Individuals assigned to a single population if proportion of ancestry was = >90% for one population. Population 1 = Northern; 2 = Central; and 3 = Southern. Individual identity code comprises a taxon, database record number and locality element: e.g. CH8855CAMI is an individual classified as *C. henryi* with entry number 8855 in the CPCG Forestry database sampled at the Camira locality. Locality codes as per Table 1. CV = *C. variegata*; CH = *C. henryi* and CC = *C. citriodora*

Index	Individual	Proportion of ancestry			Population	Index	Individual	Proportion of ancestry			Population
		Northern	Central	Southern	assignment			Northern	Central	Southern	assignment
28	CV8690BUNY	0.13	0.87	0.00	0	26	CV7799GYMP	0.00	0.98	0.02	2
35	CV8433CHER	0.00	0.59	0.41	0	27	CV8684BUNY	0.00	0.99	0.01	2
43	CV8505CHER	0.00	0.66	0.34	0	29	CV8694BUNY	0.00	1.00	0.00	2
48	CV8636CAND	0.00	0.51	0.49	0	30	CV8705BUNY	0.01	0.98	0.01	2
49	CV8642CAND	0.01	0.24	0.75	0	31	CV8707BUNY	0.00	1.00	0.00	2
52	CH8637CAND	0.01	0.64	0.36	0	32	CV8719BUNY	0.00	1.00	0.00	2
55	CH8847CASH	0.00	0.16	0.84	0	33	CV7797BFP2	0.01	0.99	0.00	2
60	CH8725BUNY	0.01	0.56	0.43	0	34	CV8885MTBA	0.06	0.94	0.00	2
61	CH8881BFP1	0.00	0.44	0.56	0	37	CV8435CHER	0.03	0.94	0.03	2
64	CH8440CHER	0.00	0.21	0.79	0	51	CV8845BAGA	0.00	1.00	0.01	2
65	CH8453CHER	0.00	0.11	0.89	0	54	CH8661CAND	0.01	0.99	0.00	2
67	CH8483CHER	0.01	0.30	0.69	0	58	CH8701BUNY	0.00	0.93	0.07	2
69	CH8353BUNG	0.00	0.38	0.62	0	62	CH8886MTBA	0.00	1.00	0.00	2
1	CC8922WIND	0.98	0.02	0.00	1	25	CC7792DUAR	0.00	0.00	1.00	3
2	CC8924WIND	1.00	0.00	0.00	1	36	CV8434CHER	0.00	0.00	1.00	3
3	CC8926WIND	0.99	0.01	0.00	1	38	CV8475CHER	0.00	0.00	1.00	3
4	CC8928WIND	1.00	0.00	0.00	1	39	CV8476CHER	0.00	0.00	1.00	3
5	CC8930WIND	0.98	0.03	0.00	1	40	CV8477CHER	0.00	0.00	1.00	3
6	CC8896BARR	1.00	0.00	0.00	1	41	CV8478CHER	0.00	0.10	0.90	3
7	CC8898BARR	1.00	0.00	0.00	1	42	CV8504CHER	0.00	0.02	0.98	3
8	CC8900BARR	1.00	0.00	0.00	1	44	CV8506CHER	0.00	0.01	0.99	3
9	CC8904BARR	1.00	0.00	0.00	1	45	CV8507CHER	0.00	0.00	1.00	3
10	CC8906BARR	0.97	0.03	0.00	1	46	CV8852CHER	0.00	0.00	1.00	3
11	CC8934MTGA	1.00	0.00	0.00	1	47	CV8633CAND	0.00	0.00	1.00	3
12	CC8936MTGA	1.00	0.00	0.00	1	50	CV4889WEDD	0.00	0.01	0.99	3
13	CC8938MTGA	1.00	0.00	0.00	1	53	CH8648CAND	0.00	0.04	0.97	3
14	CC8940MTGA	1.00	0.00	0.00	1	56	CH8687BUNY	0.00	0.00	1.00	3
15	CC8908KIRR	1.00	0.00	0.00	1	57	CH8700BUNY	0.00	0.00	1.00	3
16	CC8910KIRR	1.00	0.00	0.00	1	59	CH8715BUNY	0.00	0.00	1.00	3
17	CC8912KIRR	0.99	0.01	0.00	1	63	CH8437CHER	0.00	0.00	1.00	3
18	CC8914KIRR	0.93	0.07	0.00	1	66	CH8461CHER	0.00	0.00	1.00	3
19	CC8916KIRR	0.98	0.02	0.00	1	68	CH8851CHER	0.00	0.00	1.00	3
20	CC8944YEPP	0.00	1.00	0.00	2	70	CH8371BUNG	0.00	0.03	0.97	3
21	CC8946YEPP	0.00	1.00	0.01	2	71	CH8373BUNG	0.08	0.02	0.90	3
22	CC8948YEPP	0.04	0.96	0.00	2	72	CH8393BUNG	0.00	0.00	1.00	3
23	CC8952YEPP	0.00	1.00	0.00	2	73	CH8855CAMI	0.00	0.00	1.00	3
24	CC8954YEPP	0.00	1.00	0.00	2						

Table 3. Distances among genetic groupings in spotted gums and *Corympia torelliana*

Five populations were identified in a STRUCTURE analysis, the populations and sample sizes were CT, *C. torelliana* (21); CM, *C. maculata* (24); N, Northern (19), C, Central (18) and S, Southern (23). Below diagonal: pairwise F_{st} all significant at 95%. Average gene diversity (s.d.) along diagonals. Parameters estimated using ARLEQUIN

	CT	CM	N	C	S
CT	0.68 (0.36)				
CM	0.26	0.77 (0.41)			
N	0.21	0.13	0.79 (0.42)		
C	0.19	0.11	0.05	0.80 (0.43)	
S	0.22	0.08	0.09	0.05	0.67(0.36)

Table 4. Distances among taxonomic groupings in spotted gums and *Corympia torelliana*

Below diagonal: Pairwise F_{st} all significant at 95%; Average gene diversity ± s.d. over loci along diagonals. Parameters estimated using ARLEQUIN. Taxa codes; CT = *C. torelliana*, CM = *C. maculata*, CC = *C. citriodora*, CV = *C. variegata*, CH = *C. henryi*

	CT	CM	CC	CV	CH
CT	0.68 (0.36)				
CM	0.26	0.77 (0.41)			
CC	0.19	0.12	0.82 (0.43)		
CV	0.19	0.07	0.02	0.77 (0.41)	
CH	0.22	0.07	0.04	0.01	0.67 (0.36)

assignments occurred with taxonomic groupings, for the samples CC7792DUARA and CV8642CAND. The identity of the first individual remains under question and may represent a case of mistaken identity rather than mis-classification. With genetic groupings, only the CV8642CAND was incorrectly assigned. This individual was exceptional in that it had the

Table 5. Higher rates of ambiguous assignments occurred using taxonomic compared with genetic groups
Ambiguous assignments were those that were not assigned back to their group with a $P < 0.01$ for self-assignment with GENECLASS

Taxonomic classification			Genetic classification		
Group	No. of ambiguous assignments	No. in group	Group	No. of ambiguous assignments	No. in group
			Admix	8	13
CC	4	25	Nth	2	19
CV	14	26	Ctl	2	18
CH	17	22	Sth	14	23
CM	4	24	CM	1	24
CT	0	21	CT	0	21
Total	39	118		27	118
Percentage ambiguous		33.05			22.88

greatest amount of missing data of any individual tested, with data for only three of the nine loci available; therefore, there was lower power for assignment testing. For the dataset overall, however, missing data was low at 0.06%, and three other individuals with data for only four loci were correctly assigned.

Discussion

Geography was a major determinant of population structure in spotted gums

Four populations were identified within spotted gums, including a population corresponding to *C. maculata*, and three populations among the three northern taxa, namely Northern, Central and Southern populations, that corresponded to their latitudinal positions down the east coast of Australia. A clear separation of *C. citriodora* into northern and southern regions corresponded with a major disjunction of ~300 km in its distribution in central Queensland (Fig. 5, Table 2). A second geographical feature, the Border Range between Queensland and NSW, corresponded with a discontinuity in the west of the distribution of spotted gums at this latitude, and a division of *C. variegata* into the Central and Southern populations.

This linear, latitudinal arrangement of populations within spotted gum and the pattern of genetic distances among them, suggested that geography was a dominant factor shaping genetic structure and that a stepping-stone model of migration was appropriate. Neighbouring populations showed lower pairwise F_{st} values (Table 3) than did more distant populations, suggesting gene flow was restricted by geographic distance. A stepping-stone model of migration as proposed by Kimura and Weiss (1964) to model gene exchange when populations are arranged linearly so that genes migrate more frequently between adjacent populations than between more distant populations.

An east–west dimension to the stepping-stone model may be needed to fully account for structuring in C. variegata

A second, east–west dimension may be necessary to fully account for structuring (see below) in *C. variegata*. The Border Range corresponded with a discontinuity in the west of the distribution of spotted gums at this latitude and with a division of *C. variegata* into the Central and Southern populations. Evidence that the Border Range acts as a barrier was evident in *C. variegata* because samples from Mount Barney (a western Queensland locality) on the northern side of the Range, clustered in the Central population with geographically more distal northern sites such as Brisbane Forest Park, Bunyaville and Gympie, rather than with the more proximal site of Cherry Tree, south of the Border Range. Furthermore, *C. henryi* from Mount Barney also clustered in the Central population, rather than in the *C. henryi*-dominated Southern population. Some gene flow across the Range probably occurs, however, because a small proportion of *C. variegata* individuals from the Cherry Tree site were an admixture of Central and Southern population ancestry.

As a gene-flow barrier, the Border Range did not affect *C. henryi* to the same degree as it did *C. variegata*, probably because its distribution lies largely to the east of the Border Range on the coastal plain, where it is more or less continuous. Nevertheless, as for other species of spotted gum, there was evidence of geographic substructuring within *C. henryi*, with individuals clustering into either the Central or Southern population, or as putative inter-population hybrids. In this aspect, our study tended to differ from previous genetic studies that, by using isozymes, showed that although *C. henryi* was genetically diverse, it had a high degree of cohesion across sites (McDonald *et al*. 2000) (Table 6). The isozyme study did not include populations of *C. henryi* from north of Ewingar in NSW (29°01′S, 152°29′E), whereas the present study included *C. henryi* from several localities in Queensland as well as NSW. This suggests that when the full geographic extent of *C. henryi* is assessed, the same geographic isolation effects that are evident in other spotted gum taxa are also evident in *C. henryi*.

Congruence in structure across genetic studies

Genetic structure at microsatellite loci was largely congruent with previous studies of cpDNA and the reanalysis of isozyme data (King 2004). In the present study, although overall differentiation among taxa was low (13%), it was similar to that found with isozymes (15% among-species variation; McDonald *et al*. 2000; Table 6). On the basis of microsatellite variation, *C. maculata* was the most strongly resolved population, with

Table 6. Summary of conclusions and salient features of genetic marker or seedling morphology/leaf-oil studies in spotted gum, relating to population structuring

Reference	Taxa (no. of localities; most northerly and southerly locality)[A]	Notable points about sampling range	Character	G_{st} (%)	Major conclusions in relation to population structuring
(McDonald *et al*. 2000, 2003)	CC (8; Mt Janet—Monto) CV (8; Monto—Bagawa) CH (4; Ewingar—Myrtle Creek) CM (8; Yarrat—Mottle Range) CT (3; Helenvale—Cardwell)	No. CH north of Ewingar	Isozymes	15	Geographic patterns in isozyme variation; Including CT emphasised how closely related spotted gum taxa are; Two alliances were evident, CM-CH and CC-CV. CC and CV were indistinguishable. Northern and southern populations were evident within CM
(King 2004)	As per (McDonald *et al*. 2000) except no CT	No. CH north of Ewingar	cpDNA; reanalysis of isozyme data	24	CH, CC, CV not significantly different; alliance of CH-CM by isozymes biased by one locus, *PGM2*. Two populations within CM
This study	CC (6; Windsor SF—Yeppoon) CV (8; Gympie—Bagawa) CH (8; Strathpine—Candole SF) CM (6; Wingello—Mottle Range) CT (3; Black Mountain—Clohesy River)	No. CM north of Wingello	Microsatellites	13	Four populations within the spotted gums; CM and three within northern taxa;– geographic distance was a barrier to gene flow; poor alignment between taxa and genetic groups in northern taxa; CV and CH indistinguishable; southern CC indistinguishable from CV. One population in CM but northern localities not sampled
(Asante *et al*. 2001)	CC (3; Cheviot Hills—Monto) CV (4; Monto—Richmond Rge) CH (1; Braemar) CM (3; Taree—Kioloa)	Only 1 locality of CH	Seedling leaf-oil composition	—	CM clearly separated from northern taxa; two chemotypes in CM; northern CC well separated from CV but southern CC had strong similarity to CV
(Larmour *et al*. 2000)	CC (2; Herberton—Monto SF) CV (6; Saddler Springs—Paddy Land SF) CH (1; Braemar SF) CM (7; Yarrat—Mottle Range)	Only 1 locality of CH	Frost tolerance; seedling morphology	—	Southern CC and CV had similar morphology

[A]See Table 5 for taxa abbreviations.

a minimum pairwise F_{st} to any other population of 0.08. *C. maculata* was the most strongly differentiated taxon also on the basis of isozymes ($D_{st} = 0.035$; Table 5 in McDonald *et al*. 2000), and the most distant taxon in a cpDNA analysis, as it shared no cpDNA haplotypes with any other taxa (see Table 2.3, p. 31 in King 2004).

The results of the present study were also generally congruent with those of the previous study that used isozymes and cpDNA, in finding less differentiation among the northern taxa, and poor correspondence between genetic and taxonomic groupings. Although the degree of differentiation between the three northern taxa (max. pairwise F_{st} 0.05) was less than their pairwise comparisons to *C. maculata* (min. pairwise F_{st} 0.08), the extent of differentiation across the northern taxa was as just as high between the most distal populations in the northern taxa (pairwise F_{st} 0.09 for Northern *v.* Southern population). This was consistent with the evidence from cpDNA, as haplotypes were shared among northern taxa and showed little alignment with taxa groupings (see fig. 2.5, p. 36 in King 2004) and population dendrograms that were based on isozymes (see fig. 2.7, p. 40 in King 2004).

The division of *C. citriodora* into two populations observed here was also found in a study of seedling morphology by Larmour *et al*. (2000). These authors found that northern *C. citriodora* have a leaf size and shape distinct to southern populations of *C. citriodora*, and this difference was much greater than the difference between southern *C. citriodora* and adjacent *C. variegata*.

Differences in structure between genetic studies are probably attributable to sampling

Differences between studies in the degree of substructure within *C. henryi* were discussed earlier and were attributed to differences in sampling between studies. Differences between studies in substructuring of *C. maculata* were also evident between studies and appeared also to be due to different sampling strategies. Substructure in *C. maculata* was identified in both cpDNA and isozymes (McDonald *et al*. 2000;

PROOF ONLY

King 2004), but no substructure was evident in our study of microsatellites. Populations from the north coast region possessed a cpDNA haplotype distinct from those from the south coast region and also different alleles at the *PGM2* locus (p. 31 in King 2004; p. 495 in McDonald *et al*. 2000) (Table 6). The absence of substructure in *C. maculata* in our study was most likely due to a lack of samples representing the north coast regions. The most northerly sample we studied was Wingello, which is the most northerly locality of the southern population group (p. 497 in McDonald *et al*. 2000). Clarification of whether there is substructure in *C. maculata* microsatellite variation will require additional testing of populations from this northern region. Future study should also focus on more extensive sampling of southern *C. citriodora* and *C. variegata* from their potential intergrade region in Central Queensland, particularly from poorly studied western provenances.

A single major genetic division within the spotted gums

The consensus emerging from analysis of genetic structure among the spotted gums is that there is a single major division between *C. maculata* and the northern taxa. Further genetic analysis is needed of potential intergrade populations at the northern extremity of the *C. maculata* distribution (p. 391 in Hill and Johnson 1995) but this should not change the distinctness of the core populations of *C. maculata* from northern taxa.

Northern taxa might be better viewed as a single species of either two or three subspecies or races. The study of microsatellite variation supported the recognition of three genetic entities; however, there was poor alignment with current taxonomic groupings—particularly for *C. variegata* and *C. henryi*, which might best be considered as morphotypes (that may be distinguished by capsule size (larger in *C. henryi*)). The genetic distinctness of northern *C. citriodora* reported here, along with uniqueness of seedling morphology and oil characters described by other authors (Asante *et al*. 2001; Larmour *et al*. 2000) argues for at least one division within the northern taxa. Retention of *C. henryi* and *C. variegata* as species is less defensible and has little support from genetics. The detection of admixed individuals of *C. henryi* from six of the eight localities examined (almost exclusively with *C. variegata* background) suggests a high level of contemporary gene flow between these two 'taxa' and few core populations. Other recent, more comprehensive studies of the genetic structure of these two taxa, with broader coverage across the geographic range and greater sampling depth within localities, have shown that the genetic distance within each taxon across sites exceeds the genetic distance between taxa at each site (Ochieng *et al*. 2006; J. Ochieng, unpubl. data). Together with the evidence for high levels of contemporary gene flow between the taxa, this suggests geographic isolation is a stronger determinant of genetic structure than species barriers in this pair. Recognition of the two morphological extremes as races may be valuable if they can be associated with ecological factors or useful traits for breeding, such as flowering time as in the case of *E. globulus* (Dutkowski and Potts 1999). The lack of identifiable core populations tends not to support an alternative hypothesis that *C. henryi* may be a nascent species yet to evolve reproductive or other genetic barriers to gene flow.

Acknowledgements

This study was supported by ARC grants 'Eucalypt gene pool management' LP0455522 and 'Wood properties for spotted gum' LP0348613. The authors are grateful for the assistance and permission of Forests NSW, Queensland DPI&F and Forestry Plantations Queensland for the germplasm used in this study and D. Kleinig for assisting J. Ochieng with collections. The authors also thank T. Bean for identifying herbarium specimens, and G. Nikles, D. Lee, C. Nock and C. Harwood for helpful discussions and comments.

References

Allendorf FW, Luikart G (2007) Units of conservation. In 'Conservation and the genetics of populations'. 661 pp. (Blackwell: Malden, MA)

Asante K, Brophy J, Doran J, Goldsack R, Hibbert D, Larmour J (2001) A comparative study of the seedling leaf oils of the spotted gums: species of the *Corymbia* (Myrtaceae), section *Politaria*. *Australian Journal of Botany* **49**, 55–56. doi: 10.1071/BT00006

Cornuet JM, Piry S, Lurikart G, Estoup A, Solignac M (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* **153**, 1989–2000.

Dutkowski GW, Potts BM (1999) Geographic patterns of genetic variation in *Eucalyptus globulus* spp. *globulus* and a revised racial classification. *Australian Journal of Botany* **47**, 237–263. doi: 10.1071/BT97114

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

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47–50.

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

Jones M, Stokoe R, Cross M, Scott L, Maguire T, Shepherd M (2001) Isolation of microsatellite loci from spotted gum (*Corymbia variegata*), and cross-species amplification in *Corymbia* and *Eucalyptus*. *Molecular Ecology Notes* **1**, 276–278. doi: 10.1046/j.1471-8278.2001.00105.x

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, Queensland.

Larmour JS, Whitfeld SJ, Harwood CE, Owen JV (2000) Variation in frost tolerance and seedling morphology of the spotted gums *Corymbia maculata*, *C. variegata*, *C. henryi* and *C. citriodora*. *Australian Journal of Botany* **48**, 445–453. doi: 10.1071/BT98065

Latch E, Dharmarajan G, Glaubitz J, Rhodes O (2006) Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conservation Genetics* **7**, 295–302. doi: 10.1007/s10592-005-9098-1

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). *Australian Systematic Botany* **13**, 491–507. doi: 10.1071/SB00005

McDonald MW, Butcher PA, Larmour JS, Bell JC (2003) Intra- and interspecific allozyme variation in eucalypts from the spotted gum group, *Corymbia*, section '*Politaria*' (Myrtaceae) (vol. 13, p. 491, 2000)—correction. *Australian Systematic Botany* **16**, 643. doi: 10.1071/SB00005_CO

Moritz C (1994) Defining 'evolutionary significant units' for conservation. *Trends in Ecology & Evolution* **9**, 373–375. doi: 10.1016/0169-5347(94)90057-4

Ochieng JW, Baverstock PR, Henry RJ, Shepherd M (2006) Population panmixia in two sympatric spotted gums (Myrtaceae): implications for association mapping. In '14th IPAG Conference', San Diego, CA. p. P657.

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.

Ridley M (2004) Species concepts and intraspecific variation. In 'Evolution'. 355 pp. (Blackwell: Oxford)

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

Manuscript received 27 June 2007, accepted 1 November 2007