Developing a genetic classification for gene pool management of spotted gums

Mervyn Shepherd*, Shabana Kasem, Gary Ablett, Joel Ochieng and Allison Crawford

Centre for Plant Conservation Genetics and Subtropical Forestry Centre, Southern Cross University, PO Box 157 Lismore, NSW, 2480

*Corresponding Authors Address

Centre for Plant Conservation Genetics

Southern Cross University

PO Box 157

Lismore, NSW 2480, Australia.

Ph 02 66 203412

e-mail mervyn.shepherd@scu.edu.au

Keywords: population structure; *Corymbia;* eucalypt; microsatellite; geographic variation; *torelliana*; *variegata*; *henryi; maculata*

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 upon knowledge of geographic origin as well as subtle morphological or leaf oil variation. In this paper we explore a classification for spotted gums without assuming predefined geographic or taxonomic groups, 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 (min. pairwise Fst = 0.19). Four populations were evident within the spotted gums (n=93) but structure was weak (pairwise Fst range 0.13 -0.05). Geography, both distance and topography 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 to any other population (min pairwise Fst 0.08). Three clusters were evident within the northern taxa but alignment with taxonomic groupings was poor. *Corymbia citriodora* material from north of a major disjunction in Central Queensland formed a Northern population. *Corymbia citriodora*, *C. variegata* and *C. henryi* material from below this disjunction but north of the Border Ranges, formed a Central population, whereas a Southern population was comprised of *C. variegata* and *C. henryi* from predominately south of the Border Ranges. Fewer ambiguous assignments occurred using genetic rather than taxonomic groups for self classification of the spotted gum reference population.

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 using 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 many 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, by contrast, make no *a prior* 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 sympatry species that are very similar morphologically.

Spotted gums are a complex of closely related taxa that occur along the eastern seaboard of Australia from around latitude 16° in north Queensland to latitude 37° 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, which 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 two distinct genetic alliances suggested: *C. citriodora-C. variegata* and *C. henryi-C. maculata* (McDonald *et al.*, 2000). Subsequent reanalysis of this 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 exaggerated 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).

The expansion of the spotted gum plantations in the north of Australia, within their native range, presents a challenge for gene pool management. Spotted gums are a principle hardwood plantation taxa representing around a 1/3 of the young plantation estate. Although none of the spotted gums are regarded as endangered and each is relatively widespread and abundant, the impact of the mixing of species and provenances that have been historically isolated upon the genetic structure of spotted gums in native forests is unknown.

In this study we describe genetic structuring at microsatellite loci in spotted gums and consider its potential for developing a classification system and its applications for gene pool management. One potential advantage of a genetic classification system for spotted gums will be the ability to classify without information on geographical origins. A key criterion for evaluating a system based on microsatellites therefore will be whether allelic variation has a geographic distribution. We consider the soundness of a genetic classification system to reflect natural groups by examining the congruence of genetic structures revealed by different samples and different genetic markers. Finally, because of ongoing uncertainty surrounding the taxonomy of spotted gums, we examine the alignment between taxonomic and genetic groupings and include *C. torelliana* (Section Cadageria), a close relative of the spotted gums in our analysis in order to study the relationships among spotted gums.

An individual-based approach was used to explore structure to avoid limitations of imposing predetermined taxonomic or geographic structure. A Bayesian model is implemented in STRUCTURE software which introduces structure to promote within-population Hardy-Weinberg and linkage equilibrium (Pritchard *et al.*, 2000). A further advantage of the approach is that candidate products of interspecific or inter-population hybridisation are identified as individuals of population admixture. We explore whether classification power is increased when using newly defined genetic groups and eliminating admixed individuals compared with taxonomic groups.

Methods

Study taxa, distributions and sampling

The four species of spotted gums, *C. citriodora*, *C. variegata*, *C. henryi* and *C. maculata* belong to the Politaria Section 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 (Qld), 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 (Qld), contracting southwards to subcoastal regions south to Nymboida River and northwest of Coffs Harbour in NSW (Hill and Johnson, 1995). Integrades and hybrids with *C. citriodora* are believed to occur in the northeast of *C. variegata*'s range (Hill and Johnson, 1995). *C. henryi* tends to occur on less fertile low lying country from Brisbane (Qld) in the north to near Glenreagh south of Grafton in NSW in the south (Hill and Johnson, 1995). Hybrids or intergrades are expected with *C. variegata* where they co-occur (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 Nowa in eastern Victoria (Hill and Johnson, 1995). *Corymbia torelliana* (Cadagi) which belongs to a closely related Section, Cadageria was also included in the study as an outgroup. It occurs in a restricted distribution in rainforest margins in North Queensland and is known to spontaneously hybridise with spotted gums (Hill and Johnson, 1995).

Samples were selected from collections 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 were chosen, comprised 25 individuals from 6 populations of *C. citriodora* (ranging from 1 to 5 trees per locality); 22 *C. henryi* from 8 locations (1-6 individuals per location); 24 *C. maculata* from 6 locations (3 to 6 trees per location); 21 *C. torelliana* from 3 locations plus 3 trees of unknown origin and 26 *C. variegata* from 8 locations (1-12 trees per location) (Table 1). In the case of *C. henryi* and *C. variegata* foliage for DNA extraction and herbarium specimens were collected from native forests and later classified at the Queensland Herbarium, except one GYMP sample obtained from a Qld DPIF trial.

Specimens for the remaining taxa were obtained from seedling in the Forests NSW Grafton nursery, except three *C. torelliana* of unknown origin obtained from ornamental plantings at Gympie and provided by Dr D. Lee (DPIF), and a *C. citriodora* sample (DUAR) obtained from a Qld DPIF trial.

A map of the localities

DNA extraction and Microsatellite marker analysis

DNA extraction and microsatellite genotyping were conducted as described in (Shepherd *et al.*, 2006). Nine microsatellite loci were selected on the basis of their transferability across species, polymorphism and lack of null alleles (Crawford *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 using STRUCTURE software

STRUCTURE uses a Bayesian model-based approach to group individuals using multi-locus genotypes (Pritchard *et al.*, 2000). Successive iterations of the model attempt to optimise clustering by introducing structure to account for Hardy-Weinberg or linkage disequilibrium. 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 increases (Latch *et al.*, 2006). The approach of Evanno *et al.*, (2005) using an alternate optimisation criterion - delta K, related to the second order rate change in the log probability of the data - was also used in this study. A hierarchical approach was used to successively analyse structure, with the entire data set examined initially and identified populations then analysed to identify further substructure. Model parameters followed that described by Evanno *et al.*, (2005). Run parameters were 10,000 iterations each for the burn-in with MCMC 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 SD for L(K) and delta K. These parameters were found to provide stable optimisation criteria for replicate runs of our data set 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. Alpha values also tend not to stabilise in this scenario (Pritchard and Wen, 2004).

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 Fst were used to examine relationships among populations determined by both taxonomic and genetic classifications. Pairwise Fst, and average genetic diversity were estimated using ARLEQUIN software (Excoffier *et al.*, 2005) using the standard data model. Population admixture individuals were excluded from pairwise Fst estimates between genetic groupings

Comparing genetic and phenetic groupings by assignment tests.

The proportion of individuals with incorrect or ambiguous assignments was used to test the relative performance of taxonomic and genetic groupings using self classification of the entire data set (n=118). Assignment tests were conducted using Geneclass v 1.02 with Bayesian simulation of allele frequencies and program default parameters (10000 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.

Results

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

Models with a range of K values from 1 to 10 were tested on the entire dataset including *C. torelliana* (n=118). Inspection of the delta K plot revealed a distinct peak at K=2 which indicated a structure model with two populations was optimal (Figure 1). Examination of ancestry assignments for individuals in 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).

Testing of models for K = 1 to 10 was repeated on the subset of spotted gum samples (n=94). Examination of the delta K plot indicated a distinct peak at K=2, again indicating a model with two populations was optimal (Figure 2). Ancestry assignments for the K=2 models showed that all *C. maculata* individuals were assigned to one population whereas northern 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.

Models with K from 1 to 10 were tested on a dataset containing the three northern spotted gum taxa (n=73), *C. citriodora*, *C. variegata* and *C. henryi*. A plot of Delta K showed a distinct peak at K=3 indicating three populations were optimal (Figure 3). 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 (Figure 4 and Table 2). One population (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 Ranges between Queensland and New South Wales. 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 Mt

Barney (Qld). This second population also included *C. henryi* individuals from Mt Barney, Bunyaville and Candole SF (NSW). The third population (Southern) was comprised largely of *C. variegata* and *C. henryi* from localities south of the Border Ranges. Samples of both taxa from Cherry Tree SF belonged predominantly to this population as did most material from other NSW localities, Candole SF, Wedding Bells SF and Bungawalbin SF. 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 out of 94 spotted gum individuals (17%) 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 *C. variegata* occur in sympatry; 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 5 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 5 primary populations as indicated by a lack of pronounced peaks in plots of the delta K criterion, instability in alpha 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 distinct north coast New South Wales and a south coast NSW region subgroups in a previous studies of *C. maculata* (King, 2004; McDonald *et al.*, 2000), was most likely due to the lack of samples from the north coast NSW in the current study. The most northerly *C. maculata* sampled in this study was Wingello, which is the most northern population of the south coast NSW subgroup sampled in previous studies (McDonald *et al.*, 2000).

Relationships among genetic groups and 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 around half (8.6%) when the four spotted gum populations were considered, indicating a low degree of differentiation among the spotted gums. Comparatively more structure was evident when individuals were grouped by taxonomy (13%; CRAWFORD *et. al* 2006). Pairwise Fst values for genetic groups suggested isolation by distance effects among the spotted gums, with adjacent populations showing higher affinity (Table 3). *Corymbia maculata* was the most strongly resolved population and was genetically the closest to the Southern population (pairwise Fst = 0.08). Among the northern taxa, there was a similar degree of differentiation across the group (Northern versus Southern pairwise Fst = 0.09), but between neighbouring pairs, Northern and Central, and Central and Southern, differentiation was lower (both pairwise Fst = 0.05).

Compared to taxonomic groups (and equating genetic grouping with their approximate taxonomic equivalent ie. North = *C. citriodora*; Central = *C. variegata* and South=*C. henryi*), the corresponding pairwise Fst were the same or increased slightly, indicating clusters determined using STRUCTURE were more highly resolved as a result of redefining groups and removing admixture 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 & 4). Gene diversity in *C. henryi* also remained unchanged. Migrating Southern *C. citriodora* localities to the Cental population lead to a reduction and increase in diversity to *C. citriodora* and *C. variegata*, respectively (Tables 3 and 4).

Relative performance of genetic and taxonomic groupings

Fewer ambiguous self assignments were obtained using genetic rather than taxonomic groupings. Thirty-three percent of self assignments were ambiguous using taxonomic groupings whereas only twenty-two percent were ambiguous using genetic classification (Table 5). *Corymbia variegata* and *C. henryi* taxa, and admixture and Southern groups, contained the highest numbers of ambiguously assigned individuals in the taxonomic and genetic classifications, respectively. Two incorrect assignments occurred using 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. In the second case, the individual in question had the greatest amount of missing data with genotypes for only 3 out of the 9 loci available, therefore there was lower power for assignment testing. Using genetic groupings, only the CV8642CAND was incorrectly assigned, again probably reflecting the low power for assignment.

Discussion

A model of four populations, distributed upon a latitudinal cline and determined by geography applied to spotted gums

Four populations were identified within spotted gums, a population corresponding to *C. maculata*, and three populations among the three northern taxa; Northern, Central and Southern populations, corresponding with their latitudinal position down the east coast of Australia. A clear separation of *C. citriodora* into northern and southern regions corresponded with a major disjunction of around 300 km in its distribution in Central Queensland (Figure 4 and Table 2). A second geographical feature, the Border Ranges between Queensland and New South Wales, 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 gums and the pattern of genetic distances among them, suggested geography was a dominant factor shaping genetic structure and a stepping-stone model of migration was appropriate. Neighbouring populations showed lower pairwise Fst (Table 3) than more distant populations, suggesting gene flow is restricted by geographic distance. A stepping-stone model of migration was 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 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 Ranges between Queensland and New South Wales, 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. Evidence that Border Ranges acts as a barrier was evident in *C. variegata* because samples from Mt Barney (a western Qld 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 the more proximal site of Cherry Tree south of the border. Furthermore, *C. henryi* from Mt Barney also clustered in the Central population, rather than the *C. henryi* dominated Southern population. Some gene flow across the Range probably occurs however, as a small proportion of *C. variegata* from the Cherry Tree site had Central population ancestry.

As a gene flow barrier, the Border Range did not impact on overall structure in *C. henryi* to the same degree as *C. variegata*, probably because its distribution predominantly lies to the east of the Range on the coastal plain, where it is more continuous. Nevertheless, like other species of spotted gums, there was evidence for geographic sub-structuring 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 which showed that although *C. henryi* was genetically diverse, it had a high degree of cohesion across sites (McDonald *et al.*, 2000). The isozyme study however, did not include populations of *C. henryi* north of Ewingar in NSW (Lat 29º 01' Long 152º 29') whereas the present study included *C. henryi* from several localities in Queensland as well as New South Wales. This suggests that when the full geographic extent of *C. henryi* is assessed, the same geographic isolation effects 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%; Crawford *et al.*, (2006) it was similar to isozymes (15% among species variation in McDonald *et al.*, (2000)). Based on microsatellite variation, *C. maculata* was the most strongly resolved population with a minimum pairwise Fst of 0.08 to any other population. *C. maculata* was also the most strongly differentiated taxa by isozymes (D_{st}=0.035 Table 5 (McDonald *et al.*, 2000)) and was also the most distant taxa by cpDNA analysis, as it shared no cpDNA haplotypes with any other taxa (Table 2.3. p31 (King, 2004)).

The three genetic studies were also generally congruent in finding less differentiation among the northern taxa, and poor alignment between genetic and taxonomic groupings. Neighbouring populations among the northern taxa were less differentiated from each

other (max. pairwise Fst 0.05) than *C. maculata*, but the overall degree of differentiation across the northern taxa was similar to their differentiation to *C. maculata*. This was consistent with evidence from cpDNA, as haplotypes were shared among northern taxa and showed little alignment with taxa groupings (Figure 2.5 p36 (King, 2004)) and population dendrograms based on isozymes (Figure 2.7 p40; King, (2004)).

Evidence for division of *C. citriodora* into two populations was also found in a study of seedling morphology (Larmour *et al.*, 2000). They found that northern *C. citriodora* have a distinct leaf size and shape compared 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 systems are probably attributable to sampling

Differences in the degree of sub-structure within *C. henryi* were discussed earlier and were attributed to differences in sampling between studies. Differences in sub-structuring of *C. maculata* were also evident between studies using genetic systems and also appear due to different samples. Substructure in *C. maculata* was identified in both cpDNA and isozymes but no substructure was evident in our study of microsatellites. Populations from the north coast region possessed a distinct cpDNA haplotype from those from the south coast region and also different alleles at the *PGM-3* locus (p31; King, (2004); p495 McDonald *et al.*, (2000)). 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 (p497 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.

Corroborating genetic models differ from taxonomic grouping

The consensus emerging from analysis of genetic structure is that there is a single major division within the spotted gums 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; 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 be best considered as morphotypes. 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 individual species is less defensible and has little support from genetics. The detection of admixture 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, has shown that the genetic distance within each taxon (across sites) exceeds the genetic distance between taxa (at each site) (Ochieng *et al.*, (2006) and J. Ochieng unpublished). 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 *et al.*, 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 genetic) barriers.

Gene pool management of plantation Corymbia

There are two immediate challenges for *Corymbia* gene pool management. First, a method to identify the origins of planted material of unknown origin is required. The first challenge arises because there is widespread planting of spotted gums but often their origins have been poorly recorded or are unknown (p447; Larmour *et al.*, (2000)). Second, a method to monitor gene flow from intersectional hybrids with *C. torelliana* is also necessary. Identifying the origins of planted material would provide valuable information on provenance performance for breeding and for identifying plantings of

exotic species and provenances to aid in conserving and managing base populations (Potts *et al.*, 2001).

Here, we have demonstrated that by using a genetic classification it is possible to assign material broadly to regions without reference to their origins and with fewer ambiguous assignments compared with taxonomic groups. A genetic classification approach was possible because of the broad-scale geographic structuring of microsatellite variation. This, together with the ability to identify and exclude individuals of population admixture, may explain the increased power for assignment testing. More extensive sampling throughout the ranges of *C. variegata* and *C. henryi* has also been conducted and will be used to investigate the feasibility of more precise localisation of material from these taxa.

Resistance to Ramularia Shoot blight as well as other beneficial traits such as vigorous growth and ease of propagation via cloning has driven the development of *C. torelliana* hybrids for commercial forestry in Queensland (Lee, 2007). However, the invasive potential of *C. torelliana* due to its vigour, fecundity and unique seed dispersal (Wallace *et al.*, 2005; Wallace and Trueman, 1995) has led to its declaration as an environmental weed in some shires in northern NSW, and raised concerns about the potential for gene flow to native forests from plantings of the hybrid. As plantings of the *C. torelliana* hybrid expand, there will be an increasing need to monitor for genetic invasion. We are currently testing material from control crosses and putative hybrids to evaluate the power for identifying these hybrids and determining genealogical classes.

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, Qld 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, G Nikles, D. Lee and C. Nock for helpful discussions regarding spotted gum taxonomy, distribution and STRUCTURE analysis.

References

Allendorf FW, Luikart G (2007) Units of conservation. "Conservation and the Genetics of Populations". Blackwell, Malden MA, p. 661

Asante K, Brophy J, Doran J, *et al.* (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

Crawford A, Kasam S, Ablett G, Shepherd M (2006) ARC TR07 Microsatellite markers for gene pool management of spotted gums and their hybrids with *Corymbia torelliana* (Stage 3 report). p. 66. Centre for Plant Conservation Genetics, Lismore

Dutkowski GW, Potts BM (1999) Geographical patterns of genetic variation in *Eucalyptus globulus* ssp. *globulus* and a revised racial classification. *Australian Journal of Botany 47*:237-263

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

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, *et al.* (2001) Isolation of microsatellite loci from spotted gum (*Corymbia variegata*), and cross-species amplification in *Corymbia* and *Eucalyptus*. *Molecular Ecology Notes 1*:276-278

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

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

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

Lee D (2007) Development of *Corymbia* species and hybrids for plantations in eastern Australia. *Australian Forestry 70*:11-17

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

Moritz C (1994) Defining "Evolutionary Significant Units" for conservation. *Tree 9*:373-375

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

Potts BM, Barbour RC, Hingston AB (2001) Genetic Pollution from Farm Forestry. p. 119. RIDC, Hobart

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

Pritchard JK, Wen W (2004) Documentation for *structure* software: Version 2.

Ridley M (2004) Species concepts and intraspecific variation. "Evolution". Blackwell, Oxford, p. 355

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

Wallace H, Stokoe R, Trueman SJ, Lee D (2005) *Corymbia torelliana* hybrids for hardwood forestry - without the weediness. In: Lee D (ed) Corymbia Research Meeting: Underpinning development of a profitable hardwood plantation industry in northern Australia by research into Corymbia species and hybrids. Gympie, Queensland, pp 7-8

Wallace HM, Trueman SJ (1995) Dispersal of *Eucalyptus torelliana* seeds by the resin-collecting stingless bee, *Trigona carbonaria*. *Oecologia 104*:12-16

Table 1. Sample distribution for 118 *Corymbia* sp. individuals by taxa and locality.

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

Table 2. Proportional ancestry assignments for 73 spotted gums from the three northern taxa. Assignments generated in STRUCTURE using a K=3 model. Individuals assigned to a single population if proportion of ancestry was =>90% for one population.

Index	*Individual*	*Proportion of ancestry*			*Population Assignment*
		Pop1	Pop2	Pop3	
28	CV8690BUNY	0.13	0.87	0.00	0
35	CV8433CHER	0.00	0.59	0.41	0
43	CV8505CHER	0.00	0.66	0.34	0
48	CV8636CAND	0.00	0.51	0.49	0
49	CV8642CAND	0.01	0.24	0.75	0
52	CH8637CAND	0.01	0.64	0.36	0
55	CH8847CASH	0.00	0.16	0.84	0
60	CH8725BUNY	0.01	0.56	0.43	0
61	CH8881BFP1	0.00	0.44	0.56	0
64	CH8440CHER	0.00	0.21	0.79	0
65	CH8453CHER	0.00	0.11	0.89	0
67	CH8483CHER	0.01	0.30	0.69	0
69	CH8353BUNG	0.00	0.38	0.62	0
1	CC8922WIND	0.98	0.02	0.00	1
2	CC8924WIND	1.00	0.00	0.00	1
3	CC8926WIND	0.99	0.01	0.00	1
4	CC8928WIND	1.00	0.00	0.00	1
5	CC8930WIND	0.98	0.03	0.00	1
6	CC8896BARR	1.00	0.00	0.00	1
7	CC8898BARR	1.00	0.00	0.00	1
8	CC8900BARR	1.00	0.00	0.00	1
9	CC8904BARR	1.00	0.00	0.00	1
10	CC8906BARR	0.97	0.03	0.00	1
11	CC8934MTGA	1.00	0.00	0.00	1
12	CC8936MTGA	1.00	0.00	0.00	1
13	CC8938MTGA	1.00	0.00	0.00	1
14	CC8940MTGA	1.00	0.00	0.00	1
15	CC8908KIRR	1.00	0.00	0.00	1
16	CC8910KIRR	1.00	0.00	0.00	1
17	CC8912KIRR	0.99	0.01	0.00	1
18	CC8914KIRR	0.93	0.07	0.00	1
19	CC8916KIRR	0.98	0.02	0.00	1
20	CC8944YEPP	0.00	1.00	0.00	2
21	CC8946YEPP	0.00	1.00	0.01	2
22	CC8948YEPP	0.04	0.96	0.00	2
23	CC8952YEPP	0.00	1.00	0.00	2
24	CC8954YEPP	0.00	1.00	0.00	2
26	CV7799GYMP	0.00	0.98	0.02	2
27	CV8684BUNY	0.00	0.99	0.01	2
29	CV8694BUNY	0.00	1.00	0.00	2
30	CV8705BUNY	0.01	0.98	0.01	2
31	CV8707BUNY	0.00	1.00	0.00	2
32	CV8719BUNY	0.00	1.00	0.00	2
33	CV7797BFP2	0.01	0.99	0.00	2
34	CV8885MTBA	0.06	0.94	0.00	2
37	CV8435CHER	0.03	0.94	0.03	2
51	CV8845BAGA	0.00	1.00	0.01	2
54	CH8661CAND	0.01	0.99	0.00	2
58	CH8701BUNY	0.00	0.93	0.07	2
62	CH8886MTBA	0.00	1.00	0.00	2

25	CC7792DUAR	0.00	0.00	1.00	3
36	CV8434CHER	0.00	0.00	1.00	3
38	CV8475CHER	0.00	0.00	1.00	3
39	CV8476CHER	0.00	0.00	1.00	3
40	CV8477CHER	0.00	0.00	1.00	3
41	CV8478CHER	0.00	0.10	0.90	3
42	CV8504CHER	0.00	0.02	0.98	3
44	CV8506CHER	0.00	0.01	0.99	3
45	CV8507CHER	0.00	0.00	1.00	3
46	CV8852CHER	0.00	0.00	1.00	3
47	CV8633CAND	0.00	0.00	1.00	3
50	CV4889WEDD	0.00	0.01	0.99	3
53	CH8648CAND	0.00	0.04	0.97	3
56	CH8687BUNY	0.00	0.00	1.00	3
57	CH8700BUNY	0.00	0.00	1.00	3
59	CH8715BUNY	0.00	0.00	1.00	3
63	CH8437CHER	0.00	0.00	1.00	3
66	CH8461CHER	0.00	0.00	1.00	3
68	CH8851CHER	0.00	0.00	1.00	3
70	CH8371BUNG	0.00	0.03	0.97	3
71	CH8373BUNG	0.08	0.02	0.90	3
72	CH8393BUNG	0.00	0.00	1.00	3
73	CH8855CAMI	0.00	0.00	1.00	3

Table 3 Distances among genetic grouping in spotted gums and *C. torelliana*. Five populations were identified in a STRUCTURE analysis, the populations and sample sizes were; 1. *C. torelliana* (21); 2. *C. maculata* (24); 3 Northern (19), 4. Central (18) and 5. Southern (23). Below diagonal: Pairwise Fst all significant at 95%. Average gene diversity ±SD along diagonals. Parameters estimated using Arlequin.

	CT	*CM*	*N*	*C*	*S*
CT	0.68 (0.36)				
CM	0.26	0.77 (041)			
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 *C. torelliana*. Four taxa are identified within spotted gums, *C. citriodora*, *C. variegata*, *C. henryi* and *C. maculata*. Below diagonal: Pairwise Fst all significant at 95%; Average gene diversity ±SD over loci along diagonals. Parameters estimated using Arlequin.

	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)

Table 5 Comparison of data self assignment of data set using taxonomic and genetic groupings

Taxonomic classification			**Genetic classification**		
Group	**No. Ambiguous Assignments**	**No. in Group**	**Group**	**No. 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
Totals	39	118		27	118
Percentage Ambiguous		33.05			22.88

Figure 1. Magnitude of delta K as a function of K (mean ±SD over 10 replicates) for K = 1 - 10 using the dataset of 115 spotted gum and *C. torelliana* individuals. A distinct peak at K=2 was indicative that a model with 2 populations was optimal. All spotted gums grouped to one population and *C. torelliana* to the second.

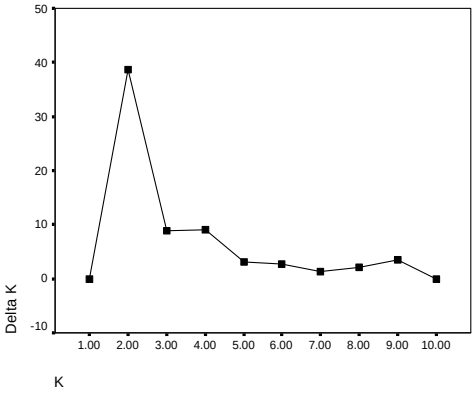

Figure 2. Magnitude of delta K as a function of K (mean ±SD over 10 replicates) for K = 1 to 10 using the dataset of 94 spotted gum individuals. A distinct peak at K=2 was indicative that a model with 2 populations was optimal. All *C. maculata* individuals clustered to one population.

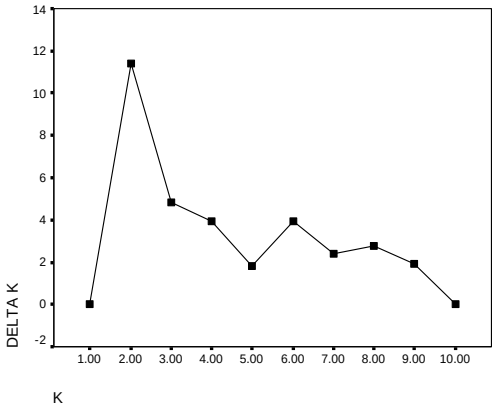

Figure 3. Magnitude of delta K as a function of K (mean ±SD over 10 replicates) for K = 1 to 10 using the dataset of 73 spotted gum individuals from the three northern taxa, *C. citriodora*, C. *variegata* and *C. henryi*. A peak at K=3 was indicative that a model with 3 populations was optimal.

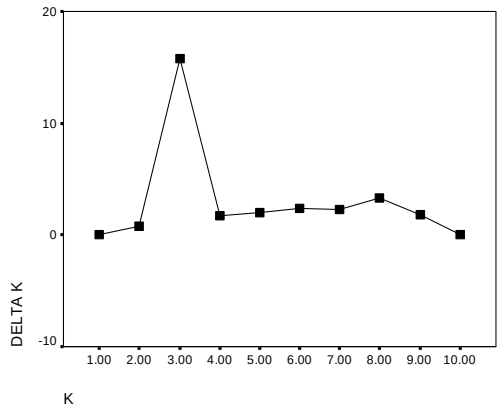

Figure 4. Ancestry assignments for 73 individuals from the northern spotted gum taxa determined by STRUCTURE using the K=3 model. Ancestry is assigned proportionally to one or more of the three populations indicated by the three colours. Individuals are arranged in taxa (*C. citriodora*; *C. variegata* then *C. henryi*- a vertical black line denotes divisions between taxa) and within taxa by localities, sequentially from the most northern locality. Localities of each sample are provided in Appendix 1.