Multiple evolutionary processes drive the patterns of genetic differentiation in a forest tree species complex

Rebecca C. Jones[1,2], Dorothy A. Steane[1,2,3], Martyn Lavery[4], René E. Vaillancourt[1,2] & Brad M. Potts[1,2]

[1]School of Plant Science, University of Tasmania, Private Bag 55, Hobart, Tasmania 7001, Australia
[2]CRC for Forestry, Private Bag 12, Hobart Tasmania 7001, Australia
[3]Faculty of Science, Health, Education and Engineering, University of the Sunshine Coast, Sippy Downs, Queensland, Australia
[4]Arianda Pty Ltd, PO Box 732, Williamstown, Victoria 3016, Australia

Abstract

Forest trees frequently form species complexes, complicating taxonomic classification and gene pool management. This is certainly the case in *Eucalyptus*, and well exemplified by the *Eucalyptus globulus* complex. This ecologically and economically significant complex comprises four taxa (sspp. *bicostata*, *globulus*, *maidenii*, *pseudoglobulus*) that are geographically and morphologically distinct, but linked by extensive "intergrade" populations. To resolve their genetic affinities, nine microsatellites were used to genotype 1200 trees from throughout the natural range of the complex in Australia, representing 33 morphological core and intergrade populations. There was significant spatial genetic structure ($F_{ST} = 0.10$), but variation was continuous. High genetic diversity in southern ssp. *maidenii* indicates that this region is the center of origin. Genetic diversity decreases and population differentiation increases with distance from this area, suggesting that drift is a major evolutionary process. Many of the intergrade populations, along with other populations morphologically classified as ssp. *pseudoglobulus* or ssp. *globulus*, belong to a "cryptic genetic entity" that is genetically and geographically intermediate between core ssp. *bicostata*, ssp. *maidenii*, and ssp. *globulus*. Geography, rather than morphology, therefore, is the best predictor of overall genetic affinities within the complex and should be used to classify germplasm into management units for conservation and breeding purposes.

Introduction

The most widely accepted species concept defines species as reproductively isolated units (Coyne and Orr 2004). However, for many organisms, morphological and ecological differences among recognized species are not always accompanied by reproductive isolation (Stebbins 1989) and morphological boundaries between species can be blurred by hybridization and introgression, resulting in species complexes that are taxonomically challenging. This is seen in diverse plant groups ranging from orchids and sunflowers (Dressler and Dodson 1960; Rieseberg et al. 1991) to forest trees such as the oaks and eucalypts (Johnson 1976; Manos and Fairbrothers 1987; Griffin et al. 1988). The use of morphological traits to define taxa in species complexes can also be problematic, as homology (similarity due to common ancestry) and homoplasy (similarity due to convergence) cannot be distinguished. Hence, in the last few decades, molecular markers have been used to assess genetic differentiation and evolutionary histories of species complexes.

Molecular studies have often identified low levels of genetic divergence among taxa in species complexes, implying recent divergence and/or hybridization and introgression. Particularly well-studied species complexes include the sunflowers (e.g., Beckstrom-Sternberg et al. 1991; Rieseberg et al. 1991; Yatabe et al. 2007) and oaks (e.g., Hokanson et al. 1993; Hess and Stoynoff 1998; Gomory et al. 2001; Kashani and Dodd 2002; Aldrich et al. 2003). Molecular markers have been used to elucidate phylogenetic relationships among closely related taxa in other forest tree species complexes such as *Abies magnifica–Abies procera* (Oline 2008), *Melaleuca quinquenervia* (Cook et al. 2008), *Astronium* (Caetano et al.

2008), *Fagus sylvatica* (Gomory et al. 2007), *Acacia* (Byrne et al. 2002; Millar et al. 2011), and *Eucalyptus* (reviewed by Grattapaglia et al. 2012). Recent advances in Bayesian assignment procedures can prove useful in defining taxonomic units in species complexes (e.g., Gomory et al. 2007; Caetano et al. 2008; Duminil and Di Michele 2009; Millar et al. 2011). Duminil et al. (2006) even argued that the best approach when studying closely related taxa is to use only Bayesian procedures for species delimitation, without reference to morphological data.

The speciose Australian tree genus *Eucalyptus* includes many species complexes, where poor morphological resolution of recognized taxa is thought to be due, at least in part, to recent and ongoing speciation in some lineages (Byrne 2008), coupled with contemporary and historic hybridization (Griffin et al. 1988; McKinnon et al. 2001). We here report a molecular study of the *Eucalyptus globulus* species complex (Fig. 1) aimed at (1) understanding the pattern of genetic variation and the evolutionary processes that have shaped genetic differentiation and (2) providing a framework for the management of this important genetic resource. This complex is of ecological significance as it dominates many lowland forests of southeastern Australia. It is a foundation species in these forests and many insects and birds, including the endangered swift parrot (*Lathamus discolor*), feed on the nectar of its flowers (Hingston et al. 2004). The species complex is also of economic importance to the global pulp and paper industry because it includes the world's most widely planted temperate hardwood species (Eldridge et al. 1993).

Conserving the maximum amount of genetic diversity in a species, thereby maximizing its long-term adaptive flexibility (Frankel and Soule 1981), has been a priority of
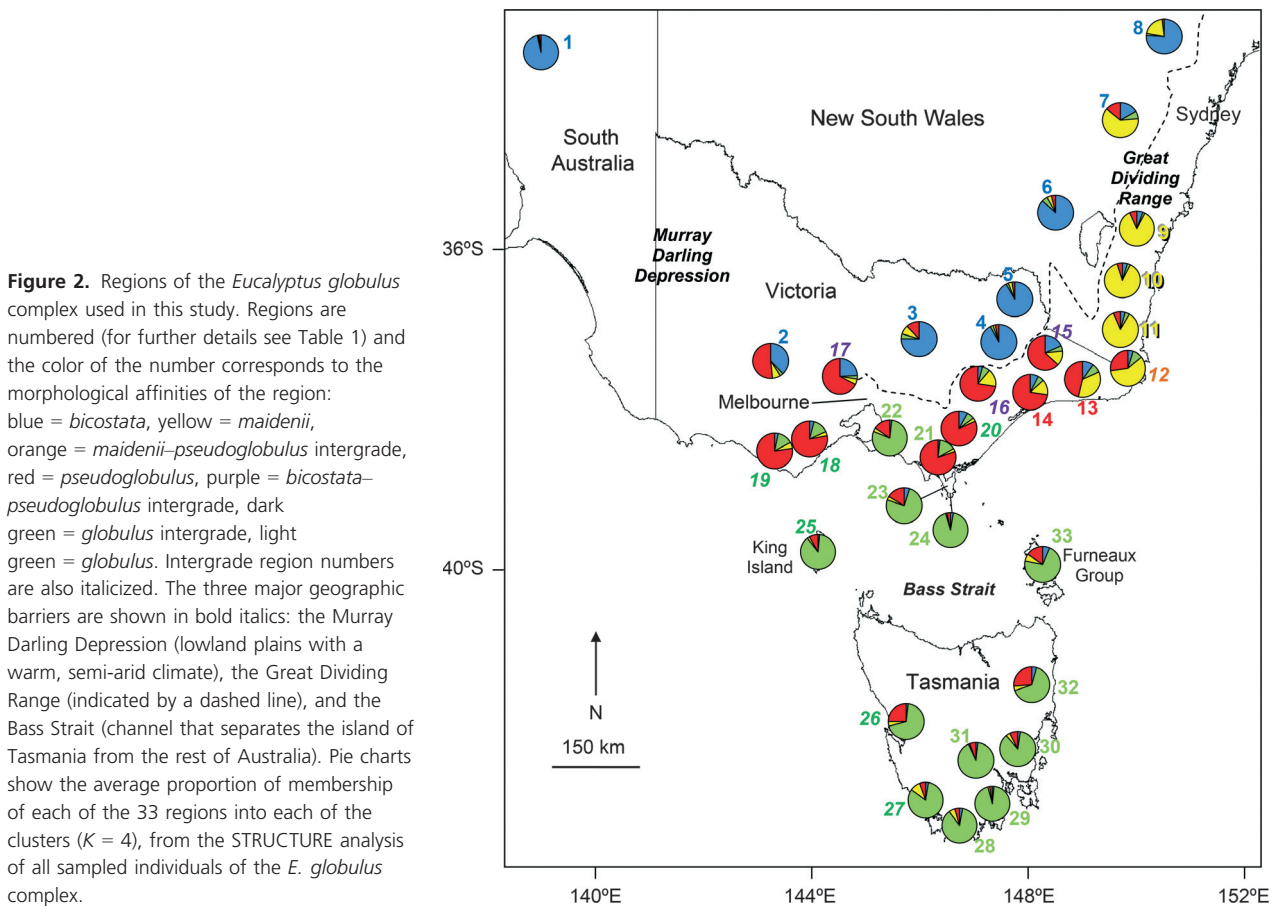

Figure 1. *Eucalyptus globulus* ssp. *globulus* at Coles Bay, on the east coast of Tasmania.

conservation strategies, and recent papers have called for evolutionary principles to be incorporated into conservation efforts (Sgrò et al. 2010), particularly in the era of climate change (Mergeay and Santamaria 2012). In the case of foundation species such as *E. globulus*, these strategies will also help to conserve associated biodiversity (Barbour et al. 2009; Whitham et al. 2010). However, a broad-based molecular genetics framework is also required to guide the management of this native genetic resource, particularly with respect to the risk of genetic contamination from plantation establishment (Potts et al. 2003; Barbour et al. 2008). The risk of contamination of native forest gene pools from interspecific hybridization with locally exotic plantation species has been well studied (Laikre et al. 2010). F_1 hybrids have been detected in open-pollinated seed collected from native species adjacent to *E. globulus* plantations and, while rare at this stage, F_1 hybrids have also been found established in the wild (Barbour et al. 2008). Interspecific hybridization is often readily detected through simple morphological observations (Barbour et al. 2005). However, assessing the risk of gene flow from *E. globulus* plantations established within the native range of the species complex is more difficult and, as a first step, requires an understanding of the differentiation between the germplasm used in plantations and the adjacent native forest.

The *E. globulus* complex comprises four taxa of forest trees that form a monophyletic group (McKinnon et al. 2008; Steane et al. 2011) and are variously treated as species or subspecies (*bicostata*, *globulus*, *maidenii*, and *pseudoglobulus*; Pryor and Johnson 1971; Kirkpatrick 1975; Jordan et al. 1993; Brooker 2000), but most recently as subspecies (Slee et al. 2006). These taxa will hereafter be referred to by their subspecific names, hence *globulus* refers to the subspecies and *E. globulus* refers to the species complex. The *E. globulus* complex has a largely continuous distribution in southeastern Australia, although there are several small, disjunct populations at the periphery of the species' range (Brooker and Kleinig 2006; Fig. 2). Three major geographic barriers provide significant contemporary barriers to dispersal: the Bass Strait, the Great Dividing Range, and the Murray Darling Depression (Fig. 2). However, while the core distributions of the four *E. globulus* taxa are geographically separated, the taxa – as defined taxonomically – span these major contemporary barriers. Differences in flowering time are also expected to provide barriers to gene flow among populations (Jones et al. 2011).

Morphologically, the four *E. globulus* taxa are differentiated mainly by capsule traits: *maidenii* has the smallest fruit in the complex and up to seven fruit per umbel; *bicostata* and *pseudoglobulus* are both three-fruited, with *pseudoglobulus* having smaller fruit, longer pedicels, and

Figure 2. Regions of the *Eucalyptus globulus* complex used in this study. Regions are numbered (for further details see Table 1) and the color of the number corresponds to the morphological affinities of the region: blue = *bicostata*, yellow = *maidenii*, orange = *maidenii–pseudoglobulus* intergrade, red = *pseudoglobulus*, purple = *bicostata–pseudoglobulus* intergrade, dark green = *globulus* intergrade, light green = *globulus*. Intergrade region numbers are also italicized. The three major geographic barriers are shown in bold italics: the Murray Darling Depression (lowland plains with a warm, semi-arid climate), the Great Dividing Range (indicated by a dashed line), and the Bass Strait (channel that separates the island of Tasmania from the rest of Australia). Pie charts show the average proportion of membership of each of the 33 regions into each of the clusters (*K* = 4), from the STRUCTURE analysis of all sampled individuals of the *E. globulus* complex.

fewer ribs on the capsules than *bicostata*; *globulus* has large, solitary capsules. The geographic cores of these four taxa are morphologically distinct, but are linked by "intergrade" populations that are intermediate in morphology (Kirkpatrick 1975). These intergrade populations complete a continuum in capsule morphology across the geographic range of *E. globulus*. The origins of the intergrade populations, as well as the origin of the species complex and migration pattern, have been the subjects of considerable debate (Kirkpatrick 1975; Jordan et al. 1993; Wallis et al. 2011). The intergrade populations occur across a wide area (Fig. 2), representing a significant proportion of the distribution of the species complex, and are therefore of ecological as well as evolutionary importance. Furthermore, their diverse and intermediate morphologies confound taxonomic classification and cause difficulties in managing these forests and genetic resources. The intergrade populations may originate from: (1) primary differentiation (divergence within a continuous series of populations along a selective gradient), or (2) secondary intergradation (hybridization and introgression between previously isolated gene pools), although the possibility of a more complex scenario involving both

processes cannot be dismissed (Jordan et al. 1993; Wallis et al. 2011). These two evolutionary scenarios are difficult to distinguish based on current patterns of variation (Endler 1977). However, a zone of recent secondary intergradation would be expected to display increased diversity due to gene flow from differentiated populations (Rieseberg and Wendel 1993). For example, molecular data show that the genetic diversity within a *Populus fremontii–Populus angustifolia* hybrid zone is much higher than in the pure species zones (Whitham et al. 1999). In addition, a recent hybrid zone would be expected to have correlated morphological and molecular affinities, as was the case in the *Quercus crassifolia–Quercus crassipes* hybrid zone in Mexico (Tovar-Sanchez and Oyama 2004), because linkage disequilibrium is expected to decay slowly in long-lived organisms.

Genetic variation across parts of the *E. globulus* species complex has been investigated in several taxonomic, quantitative, and molecular genetic studies (reviewed by Potts et al. 2004; see also McKinnon et al. 2005; Steane et al. 2006; Hamilton et al. 2011; Kulheim et al. 2011; Stackpole et al. 2011; Wallis et al. 2011; Yeoh et al. 2012; Dutkowski and Potts 2012). However, most of these

studies have focused on *globulus* and its intergrade populations, as these have been of most interest to breeding programs and the plantation industry. Very few, or no, samples of *pseudoglobulus*, *bicostata*, or *maidenii* or their intergrade populations were included in these genetic studies. In addition, selection acting on morphological and other quantitative traits (e.g., Steane et al. 2006) could cause a misrepresentation of true genetic relationships in the species complex. This study therefore uses putatively neutral nuclear DNA markers (microsatellites) to understand the pattern of genetic variation and the evolutionary processes that have shaped genetic differentiation in the *E. globulus* gene pool. To achieve this, we genotyped one of the largest samples yet of a forest tree species and coupled these results with morphological measurements to determine the evolutionary origin of intergrade populations.

Materials and Methods

Plant material

Leaf tissue for DNA extraction, a herbarium specimen (including buds and capsules), and open-pollinated seed (where available) were sampled from 1200 adult trees representing 103 natural populations, 33 regions, and the four taxa of the *E. globulus* species complex, including their intergrades (Fig. 2, Table S1). Trees sampled were at least 100 m apart, to avoid sampling closely related trees (see Skabo et al. 1998; Jones et al. 2007). Localities of *globulus* and its intergrades were grouped into races (as defined by Dutkowski and Potts 1999), hereafter referred to as "regions"; the other taxa and their intergrades were grouped into geographic regions based on overall morphology (Table S1). Populations were classified into a pure taxon or intergrade type following the Brooker and Kleinig (2006) field guide and/or the collections of Kirkpatrick (1975) and Jordan et al. (1993) (Table S1).

Molecular methods

DNA extraction and PCR

Total genomic DNA was extracted from fresh or frozen leaf tissue using a CTAB method (Doyle and Doyle 1990) with several modifications (McKinnon et al. 2004b). DNA quality and quantity were assessed by gel electrophoresis and comparison with a Lambda *Hin*dIII molecular weight standard.

Nine microsatellite loci were selected; six (EMCRC) designed by Steane et al. (2001), two (EMBRA11 and EMBRA19) by Brondani et al. (1998), and one (EMBRA30) by Brondani et al. (2002), and the primer

sequences are available in each of these references. These loci have been mapped and there is no evidence for linkage between the loci (J. S. Freeman, pers. comm.). Forward primers were synthesized to include a WellRED fluorescent label (Sigma-Aldrich Pty Ltd, Castle Hill, NSW, Australia) that enables detection on a CEQ^{TM} fragment analysis system (Beckman Coulter Inc., Fullerton, California). PCR conditions followed Brondani et al. (1998) for the three EMBRA loci, and Jones et al. (2002) for the EMCRC loci except that 0.28 μmol/L of each primer was used in a total reaction volume of 12.5 μL. PCRs were performed separately and PCR products were co-loaded in two sets (Set 1: EMCRC5, EMCRC6, EMCRC7, EMBRA19; Set 2: EMCRC2, EMCRC10, EMCRC11, EMBRA11, EMBRA30) and sized by comparison with the CEQ^{TM} DNA Size Standard-400 using the Beckman Coulter Inc. Fragment Analysis software for allele binning. To allow data sets from Jones et al. (2002), Steane et al. (2006), Jones et al. (2007), and Foster et al. (2007) to be combined with results from this study, control samples were included (see Data S1; repeatability and missing data percentages are also given in Data S1). Only samples that yielded results for six or more loci were retained, resulting in a final data set of 1200 individuals representing 103 localities, 33 regions, and all four taxa of the *E. globulus* species complex, including their intergrades, across the entire natural range of the species complex (Fig. 2, Table S1).

Molecular data analysis

All statistics calculated at the locality or regional level were based on the 103 localities or the 33 regions shown in Fig. 2 and listed in Table S1.

A variety of genetic diversity statistics were used to assess and compare genetic variation within populations. GDA 1.1 (Lewis and Zaykin 2001) was used to calculate the following genetic diversity parameters, averaged for each locus and region: number of alleles (*A*), observed and expected heterozygosity (H_O and H_e, respectively), and Wright's Fixation Index (*F*). Null allele frequencies in each region were estimated using the Population Inbreeding method in INEst (Chybicki and Burczyk 2009). Allelic richness (El Mousadik and Petit 1996) was calculated at the region and locality levels, using FSTAT (Goudet 2002). Contours of allelic richness per locality were plotted on a map using 3DField 2.9.6 (Galouchko 2007).

To determine the relationships among regions, F-statistics (FIS, FIT, FST, Weir and Cockerham 1984) and Nei's (1972) genetic distance were calculated. A pairwise matrix of Nei's (1972) genetic distance among regions, calculated using GDA 1.1, was used to construct a Neighbor-Joining

(Saitou and Nei 1987) radial tree with proportional branch lengths, using T-REX 4.0 (Makarenkov 2001). *F*-statistics were calculated for each locus at the regional level, using GDA 1.1. F_{ST} was also calculated for each taxon at the regional level.

To determine the relationships among the four taxa, pairwise F_{ST} values among taxa were calculated using FSTAT (Goudet 2002), based on the "core" localities indicated in Table S1. The cores were based on morphology, geography, and prior analyses with STRUCTURE (i.e., individuals from geographically disjunct localities, morphological intergrade localities, and intermediate localities from STRUCTURE analyses were excluded from the analysis).

The number of groups of genetically similar individuals (*K*) in the *E. globulus* complex, and the affinities of individuals to these groups (*Q*) independent of *a priori* morphological assumptions, were determined using STRUCTURE 2.2.3 (Pritchard et al. 2000) and the Δ*K* method described by Evanno et al. (2005). Assuming no prior population groupings and using the admixture model, the estimated *K* was determined by comparing the estimated log probability of data at different values of *K* (from *K* = 1 to *K* = 20), using 100 000 MCMC repetitions following a burnin of 120 000 repetitions (at which point stationarity had been reached). Twenty independent runs for each value of *K* were performed, and results were interpreted using the five runs with the highest likelihoods. The *Greedy* algorithm in CLUMPP (Jakobsson and Rosenberg 2007) was used to derive a single output from the five independent runs at each *K* from *K* = 2 to *K* = 6. DISTRUCT (Rosenberg 2004) was used to display the probability of membership (*Q*) of each individual into each of the inferred clusters. The average proportion of membership of each of the 33 regions into the *K* = 6 clusters, from the analysis of all sampled individuals of the *E. globulus* complex, was plotted on a map.

A second STRUCTURE analysis was undertaken to directly compare molecular and morphological affinities of individuals. This analysis used a model with *a priori* population information (i.e., POPFLAG = 1) for populations of "core" *globulus*, *bicostata*, *pseudoglobulus*, and *maidenii* (based on morphology and geography, see Table S1). An admixture model was used to assign individuals from intergrade and geographically outlying populations into the clusters defined as core *globulus*, *bicostata*, *pseudoglobulus*, and *maidenii*, and the original "core" individuals were also reassigned to the four clusters using an admixture model (i.e., POPFLAG = 0). These core individuals were therefore duplicated in the analysis. Ten independent runs at *K* = 4 were performed using 50 000 MCMC repetitions and a burnin of 50 000 repetitions,

and the individual *Q*-values from the run with the highest likelihood were used to test for correlations with morphological affinities (see below) and were not presented graphically. Only individuals for whom both morphological and molecular data were available were included in this analysis.

Morphological methods

For 446 of the trees sampled in this study, 12 capsule, inflorescence, and leaf variables (defined in Fig. 3) were measured following the descriptions of Kirkpatrick (1975), Jordan et al. (1993), and Jones et al. (2002). Five average-sized umbels with all parts intact were selected from the youngest age class (1 year old) of mature capsules, and the middle capsule of each umbel (where applicable) was measured. Inflorescence data were collected from up to 70 umbels per tree. Where abortion scars were present on the umbel, or capsules had obviously dislodged during sampling, capsules were scored as present. Leaf measurements were taken on a
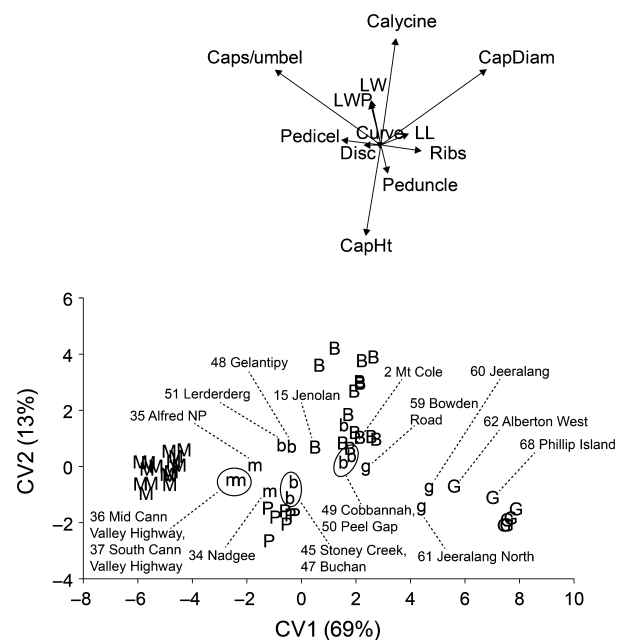

Figure 3. Plots of *Eucalyptus globulus* locality mean canonical variate scores in the space defined by the region level analysis, using 12 capsule, inflorescence and leaf variables. CapDiam, capsule diameter; CapHt, total capsule height; Disc, disc height; Calycine, calycine ring height; Ribs, number of ribs per capsule; Pedicel, pedicel length; Peduncle, peduncle length; Caps/umbel, mean number of capsules per umbel; LW, lamina width; LWP, length to widest point of leaf; LL, lamina length; Curve, leaf curvature (length of a line drawn from the primary vein perpendicular to a line connecting the lamina base to the lamina tip).

single representative leaf. Measurements were combined with those from Jones et al. (2002).

Analyses were undertaken using the means of the five capsule replicates from each tree. The variables were transformed to optimize population normality and remove the correlation between locality or region means and variances. The transformations were the same as in Jordan et al. (1993), with four exceptions: length to widest point of leaf (LWP) (log), Pedicel (square root), Ribs (square root), and Curve ($x^{0.75}$). Multivariate analyses were undertaken at the region level using Canonical Discriminant Analysis, an ordination technique that maximizes differences between groups relative to within-group variation, using the PROC DISCRIM procedure in the statistical package SAS (Version 9.1; SAS Institute, Cary, North Carolina). The locality level mean canonical variate scores were placed within the multivariate discriminant space defined by the region level analysis. The vectors indicating the direction of variation in each variable were constructed from the standardized canonical coefficients. The lengths of the vectors were proportional to the analysis of variance (ANOVA) *F*-values (see Jordan et al. 1993), calculated between groups for each variable. Pairwise Mahalanobis (morphological) distances among regions and significance ($P < 0.05$) were calculated using PROC DISCRIM.

Relationship between morphological, geographic, and molecular affinities

To test for the effects of drift and selection, a Mantel test was conducted using XLSTAT (Version 2006.5; Addinsoft SARL, Paris, France) to test the relationships between Mahalanobis distance (calculated from morphological data), Nei's (1972) genetic distance (calculated from microsatellite data), and geographic distance (calculated from latitude/longitudes in GenAlEx 6.2, Peakall and Smouse 2006) among regions.

Using morphological data, the core *bicostata, globulus, pseudoglobulus,* and *maidenii* individuals (see Table S1) were used to define the multivariate space, and the intergrade/geographically outlying individuals were fitted within this space. The morphological affinities (*M*) of these individuals to each of the core taxa (B, G, P, M) were then calculated using PROC DISCRIM (i.e., M_B, M_G, M_P, M_M) based on the classification functions derived from PROC DISCRIM. PROC CORR was then used to test whether the morphological affinities of individuals were significantly correlated with their molecular affinities, in each region in Victoria and NSW with more than 10 individuals, using Pearson's correlation coefficient. The molecular affinities had been calculated in the STRUCTURE analysis (above) that used *a priori* population information for the same core individuals, and

calculated the affinities of individuals (*Q*) to each of the four core groups (i.e., Q_B, Q_G, Q_P, Q_M).

Results

Genetic diversity

General diversity statistics for the microsatellite loci are given in Table S2 and Data S2. At the population level, genetic diversity, as measured by either expected heterozygosity (H_e) or allelic richness (R_t), was highest in the 11-South East Forests region, in the southern part of the *maidenii* distribution (Table 1). There was a weak but significant negative linear correlation between geographic distance from this region (29-Yurammie, a locality within the 11-South East Forests region, 36.9°S, 149.72°E) and genetic diversity, at both the region and locality level (H_e 33 regions $r^2 = 0.31$, $P = 0.0008$; R_t 33 regions $r^2 = 0.34$, $P = 0.0004$; H_e 93 localities $r^2 = 0.34$, $P < 0.0001$; R_t 93 localities $r^2 = 0.37$, $P < 0.0001$). This was driven partly by the large geographic disjunction and low genetic diversity of the Mt. Bryan locality as well as the lower diversity in Tasmania, but when the Mt. Bryan population and Tasmanian samples were excluded, the correlations were still significant at both the regional and locality levels (data not shown). As the distribution of *E. globulus* is interrupted by geographic barriers (see Fig. 2) and migration is likely be non-linear, contour plotting of allelic richness was used to reveal the areas of high genetic diversity (Fig. 4). The primary center of diversity was in the 11-South East Forests region, with secondary centers on the east coast of Tasmania and on the western side of the Great Dividing Range. The pattern was the same for expected heterozygosity at the locality level, and for both allelic richness and expected heterozygosity at the regional level (data not shown). Genetic diversity in geographically disjunct regions of the species complex was lower than the mean for the species complex (e.g., 1-Mt. Bryan $H_e = 0.57$, 8-Nullo Mountain $H_e = 0.68$, 25-King Island $H_e = 0.65$; mean for *E. globulus* $H_e = 0.78$, Table 1), although some populations in the continuous range also exhibited low levels of genetic diversity (e.g., 24-Wilson's Promontory $H_e = 0.65$, 23-Tidal River $H_e = 0.69$, Table 1).

Recent hybrid origin would be expected to impart intergrade populations with increased genetic diversity. Genetic diversity in these intergrade populations was not higher than in their putative parental populations. For example, 15-Buchan, 16-Mitchell River, and 17-Lerderderg, which were intermediate in morphology between *pseudoglobulus* and *bicostata*, did not have higher diversity ($H_e = 0.86$, 0.81 and 0.82, respectively, Table 1) than the pure populations of *pseudoglobulus* (mean $H_e = 0.86$,

Table 1. Genetic diversity parameters for *Eucalyptus globulus*, calculated at the regional level.

Region (taxon code)	Distribution	n	A	H_e	H_o	R_t	F
1-Mt. Bryan (B)	Disjunct	24.0	4.6	0.57	0.48	4.11	0.17
2-Mt. Cole (B)	Disjunct	18.7	8.6	0.80	0.67	7.15	0.17
3-Hume (B)		29.0	10.8	0.84	0.72	8.13	0.14
4-Omeo (B)		21.9	8.9	0.77	0.66	6.92	0.15
5-Northeastern Victoria (B)		31.2	9.8	0.77	0.60	6.97	0.23
6-Canberra (B)		19.1	7.3	0.76	0.65	6.27	0.16
7-Jenolan (B)	Disjunct	20.1	7.1	0.77	0.65	6.21	0.15
8-Nullo Mountain (B)	Disjunct	19.8	6.8	0.68	0.59	5.63	0.14
bicostata mean		23.0	8.0	0.75	0.63	6.42	0.16
9-Araluen (M)		38.6	14.6	0.85	0.73	9.14	0.14
10-Wadbilliga (M)		112.9	20.2	0.85	0.73	9.44	0.15
11-South East Forests (M)		69.6	17.4	0.87	0.74	9.54	0.15
maidenii mean		73.7	17.4	0.86	0.73	9.37	0.15
12-Alfred-Nadgee (m)		43.8	15.0	0.86	0.68	9.12	0.21
m intergrade mean		43.8	15.0	0.86	0.68	9.12	0.21
13-East Gippsland (P)		25.0	12.7	0.86	0.71	9.40	0.18
14-Lakes Entrance (P)		46.9	13.7	0.87	0.73	8.82	0.16
pseudoglobulus mean		35.9	13.2	0.86	0.72	9.11	0.17
15-Buchan (b)		35.7	13.7	0.86	0.69	9.14	0.21
16-Mitchell River (b)		18.8	9.7	0.81	0.71	7.84	0.12
17-Lerderderg (b)		19.8	9.3	0.82	0.68	7.59	0.17
b intergrade mean		24.7	10.9	0.83	0.69	8.19	0.17
18-Eastern Otways (g Vic)		27.3	10.6	0.79	0.70	7.99	0.12
19-Western Otways (g Vic)		30.4	11.2	0.79	0.70	7.87	0.11
20-Strzelecki Ranges (g Vic)		22.0	9.0	0.79	0.69	7.29	0.13
g intergrade (Vic) mean		26.6	10.3	0.79	0.70	7.72	0.12
21-South Gippsland (G Vic)		27.8	9.9	0.77	0.68	7.15	0.12
22-Phillip Island (G Vic)		23.3	10.2	0.77	0.71	7.51	0.08
23-Tidal River (G Vic)		29.7	7.3	0.69	0.76	5.62	−0.10
24-Wilson's Promontory Lighthouse (G Vic)		28.6	5.8	0.65	0.64	4.67	0.00
globulus (Vic) mean		27.3	8.3	0.72	0.70	6.24	0.03
25-King Island (g Tas)	Disjunct	36.9	7.9	0.65	0.56	5.76	0.14
26-Western Tasmania (g Tas)	Disjunct	31.0	9.1	0.75	0.63	6.69	0.16
27-Port Davey (g Tas)	Disjunct	33.8	9.8	0.74	0.64	6.64	0.13
g intergrade (Tas) mean		33.9	8.9	0.71	0.61	6.36	0.14
28-Recherche Bay (G Tas)		25.4	9.2	0.79	0.68	7.19	0.15
29-Southern Tasmania (G Tas)		23.9	8.9	0.81	0.74	7.08	0.10
30-Southeastern Tasmania (G Tas)		67.7	14.6	0.83	0.69	8.23	0.16
31-Dromedary (G Tas)		17.0	8.1	0.78	0.72	6.96	0.09
32-Northeastern Tasmania (G Tas)		88.6	14.4	0.83	0.69	8.24	0.17
33-Furneaux (G Tas)	Disjunct	43.9	11.1	0.80	0.69	7.49	0.15
globulus (Tas) mean		44.4	11.1	0.81	0.70	7.53	0.14
E. globulus mean		34.9	10.5	0.78	0.68	7.3	0.14

Disjunct regions (isolated from other populations by more than 80 km, and those on the Bass Strait Islands) are indicated; all other regions are within the continuous range of the *E. globulus* distribution.

n = average number of samples per region, A = observed number of alleles per locus, H_e = expected heterozygosity, H_o = observed heterozygosity, R_t = allelic richness, F = Wright's Fixation Index.

Table 1) and *bicostata* (H_e = 0.75, Table 1). Similarly, but at a finer scale, the mid Cann Valley Highway locality (locality 36) had lower expected heterozygosity than the localities at the northern (locality 31, *maidenii*) and southern (locality 37, *maidenii–pseudoglobulus* intergrade) ends of the highway (data not shown).

Population differentiation

Most of the genetic variation in the complex was between individuals within regions; however, there was a significant proportion of genetic variation maintained between the regions. Considering region as the subpopulation,
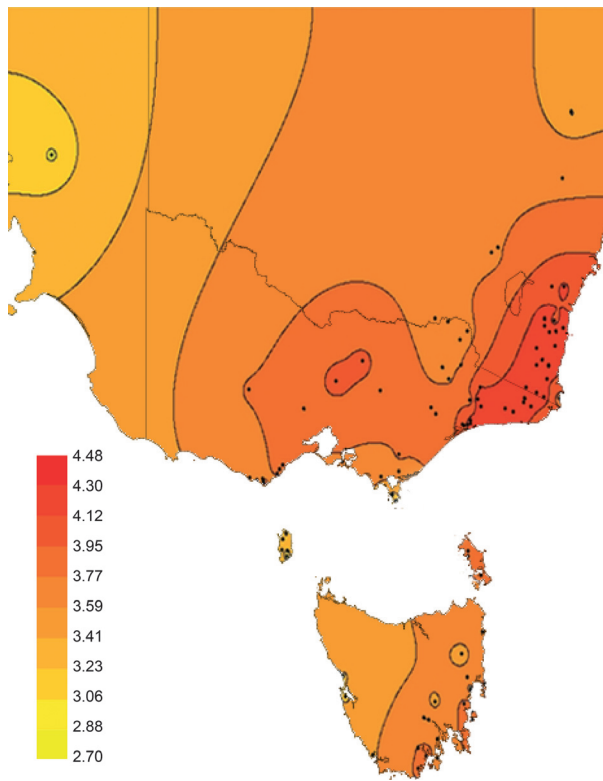
Figure 4. Allelic richness (R_t) for localities of *Eucalyptus globulus*, plotted as a contour map.

Figure 5. Neighbor-Joining radial tree with proportional branch lengths, based on a pairwise matrix of Nei's (1972) genetic distance between regions of the *Eucalyptus globulus* species complex. For definitions of region and taxon codes, see Table 1.

overall F_{ST} (i.e., inbreeding within a region relative to the entire species complex) was 0.10 (Table S2).

Differentiation was lowest among regions of *maidenii* ($F_{ST} = 0.011$, $n = 3$) and *pseudoglobulus* ($F_{ST} = 0.024$, $n = 2$) and highest among regions of *bicostata* ($F_{ST} = 0.131$, $n = 8$). Differentiation was moderate among regions of *globulus* ($F_{ST} = 0.071$, $n = 10$), increasing slightly when *globulus* intergrade regions were included ($F_{ST} = 0.081$, $n = 16$). Pairwise F_{ST} values among the four taxa revealed that *pseudoglobulus* was the least differentiated taxon (pairwise F_{ST} with other taxa 0.056–0.062, pairwise F_{ST} values among all taxa 0.056–0.112, data not shown).

Regions of *globulus* and its intergrades clustered tightly together in the Neighbor-Joining radial tree, distinct from the *maidenii* cluster and the *bicostata* cluster (Fig. 5). Populations from the 11-South East Forests region, including *pseudoglobulus*, *maidenii*, and their intergrades, were in the center of the tree (Regions 9–15, Fig. 5). Disjunct populations of *bicostata* (1-Mt. Bryan and 8-Nullo Mountain) were on long branches, as was the population at 24-Wilson's Promontory (Fig. 5). 7-Jenolan, a region classified as *bicostata*, was basal to the *maidenii* branch (Fig. 5).
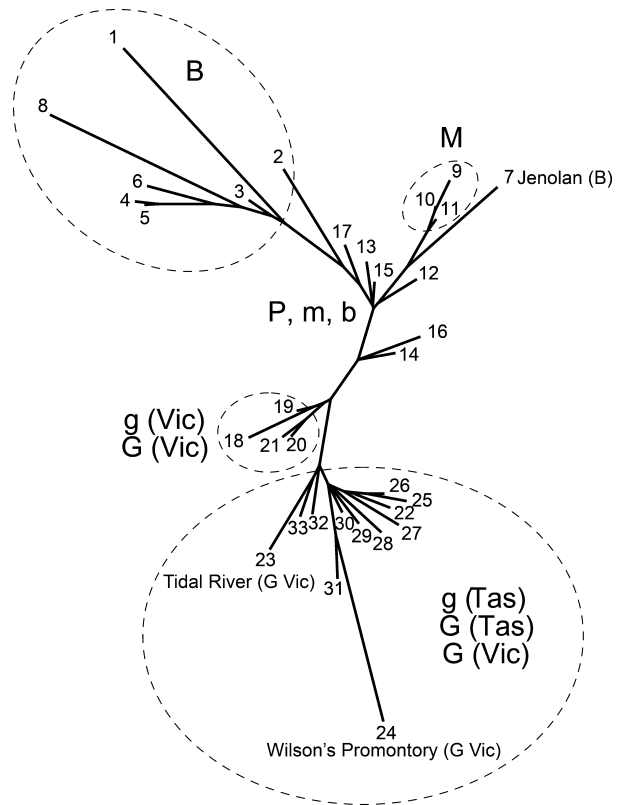
Individual affinities

The ΔK method of Evanno et al. (2005) detected the major split in the entire *E. globulus* gene pool at $K = 2$ (Fig. S1), corresponding to a split of Tasmanian and mainland individuals, but with populations consisting of admixed individuals in inland central Victoria linking these two groups (Fig. 6). There was a secondary peak at $K = 4$ (Fig. S1), where three of the four morphological taxa were relatively well defined (Fig. 2). The fourth taxon, *pseudoglobulus*, emerged only at $K = 7$, but this solution was unstable, occurring in only three of the five highest likelihood runs (data not shown). At each increase in K, the group of regions that split into a new cluster were geographically contiguous, but at each given K, there were intermediate individuals that linked the K clusters (Fig. 6). At higher K values, solutions were unstable across runs, or the individuals split across the K clusters with few individuals strongly assigned to a single cluster, with no meaningful geographic pattern for the individuals that were strongly assigned (data not shown).
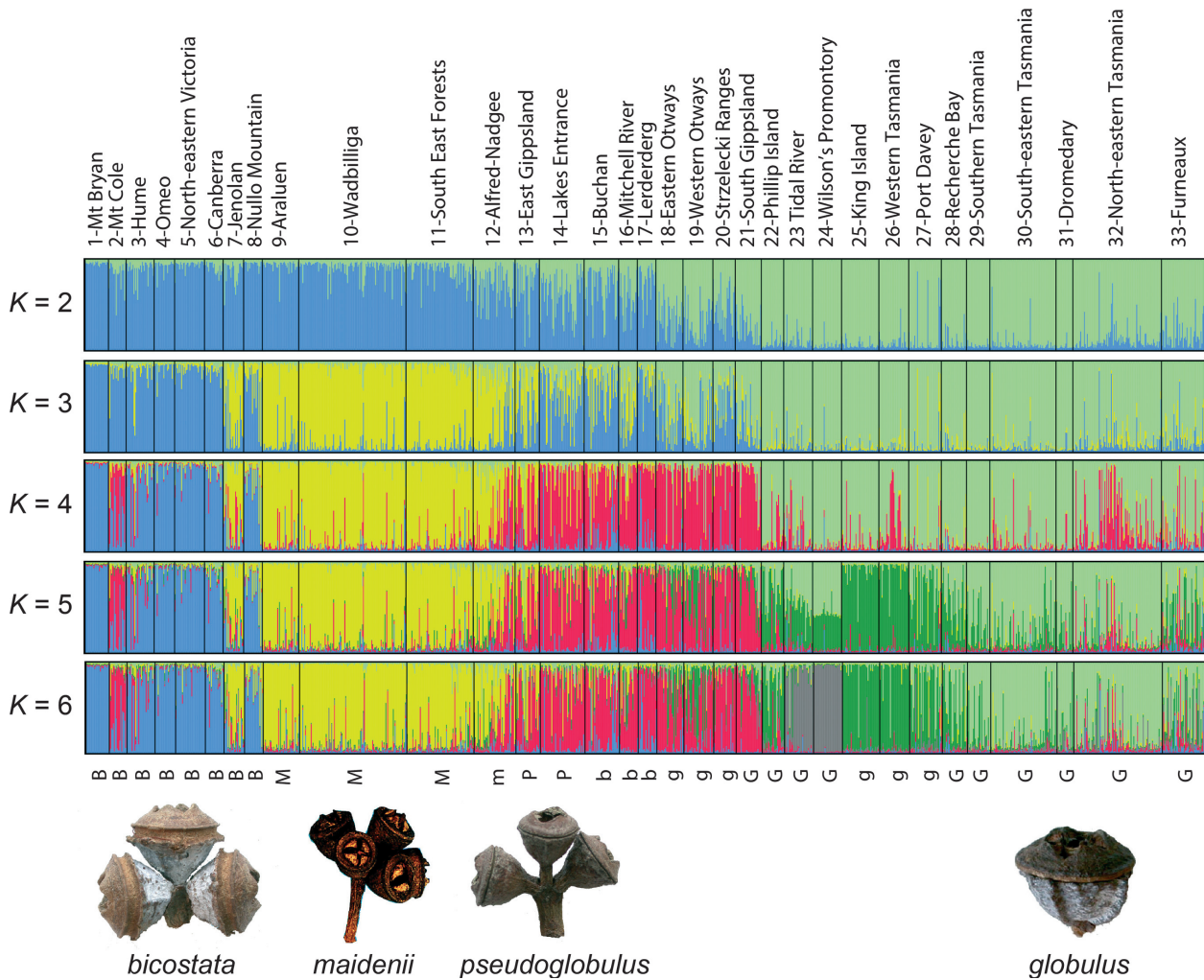
Figure 6. Proportion of membership for 1200 individuals of the *Eucalyptus globulus* species complex. Each individual is represented by a thin vertical line, partitioned into *K* colored segments representing the individual's estimated membership into the *K* genetic clusters. Individuals are grouped by region/race (indicated above the figure), and the overall capsule morphology of core regions is shown below the figure.

Morphological relationships

Capsule diameter, calycine ring height, number of capsules per umbel, and capsule height were the most important variables in differentiating the localities of the *E. globulus* complex (Fig. 3). Although the taxa of the *E. globulus* complex were differentiated, they were linked by a morphological continuum of intergrade populations. The localities that had been collected as core *pseudoglobulus* clustered together, but this group was not highly differentiated from the rest of the complex, unlike the core localities of *maidenii*, *bicostata*, and *globulus* that formed the extremities of the variation in the ordination (Fig. 3). Rather, the *pseudoglobulus* localities had morphological affinities to the intergrade localities that form part of the continuum between *bicostata* and *maidenii* (Fig. 3).

Relationship between geographic, morphological, and molecular affinities

There was a highly significant association between Nei's (1972) genetic distance and geographic distance ($r^2 = 0.370$, $P < 0.0001$) indicating that isolation-by-distance and genetic drift is an important process in the *E. globulus* species complex. The major outliers on this plot (not shown) were the pairwise comparisons of regions that were separated by a strong geographic barrier over a short geographic distance (e.g., Great Dividing Range). There was no association between Mahalanobis distance (calculated from the morphological data) and geographic distance ($r^2 = 0.007$, $P > 0.05$). Nei's (1972) genetic distance and Mahalanobis distance were weakly but significantly correlated ($r^2 = 0.042$, $P = 0.007$). Outliers on this plot

(not shown) were mainly pure Victorian *globulus* populations in South Gippsland compared with geographically proximal, but morphologically differentiated, intergrade populations, indicating that selection is acting over a steep gradient in this area.

As a recent hybrid zone would be expected to have correlated morphological and molecular affinities, the morphological and molecular affinities of individuals in regions of NSW and Victoria were tested for significant correlations in each region in which there was clear variation in morphological affinities. The only region that had a significant correlation between morphological and molecular affinities was 20-Strzelecki Ranges. In this region, there was a strong correlation between morphological and molecular affinity to Victorian *globulus* ($r = 0.64$, $P = 0.0007$, $n = 24$). Fourteen individuals had strong and similar morphological and molecular affinities to Victorian *globulus* in this region. No significant correlations involving the other taxa were observed and individuals with morphological and molecular affinities to Victorian *globulus* were not geographically clustered (data not shown).

Discussion

This combined analysis of molecular and morphological data, sampled over a large scale and focusing on morphologically intermediate populations as well as "pure" populations, revealed that multiple evolutionary processes have driven the patterns of genetic differentiation in the *E. globulus* species complex. Across the broad geographic range, continuous molecular and morphological variation link the four recognized taxa. These continua appear to be driven by the interaction of the key evolutionary processes of genetic drift, selection, and, to a lesser extent, hybridization. These processes, occurring at multiple geographic locations and scales, result in a poor relationship between molecular and morphological affinities in many intergrade regions, which has implications for gene pool management and exploitation.

Genetic drift

The neutral molecular data revealed that drift is a strong process at both broad and fine geographic scales in *E. globulus*. Across the broad geographic range, there is a continuum in molecular genetic variation and a significant and positive association between genetic and geographic distance. While weak, the observed relationship with distance ($r^2 = 0.37$) is comparable to that reported in other eucalypts (e.g., *Eucalyptus camaldulensis* $r = 0.48$, Butcher et al. 2009). Such a pattern in neutral genetic markers is expected under the isolation-by-distance model (Wright 1943) and indicates that drift is a

key driver of the patterns of genetic variation. As in most eucalypts (Byrne 2008), gene flow in *E. globulus* is likely to be more a function of pollen than seed dispersal (Jones et al. 2007). Nevertheless, most pollen appears to be deposited within 100 m of the source (Mimura et al. 2009). Such distance is many orders of magnitude smaller than the geographic distance over which *E. globulus* is continuously distributed, providing the conditions for isolation-by-distance to occur. At a finer geographic scale, there is evidence to suggest that drift has been accentuated in small, geographically or ecologically isolated populations, resulting in strong divergence from the rest of the *E. globulus* gene pool.

Genetic drift and limited gene flow are key processes underlying the central-peripheral hypothesis, which predicts that populations at the margins of a species' range should exhibit lower genetic diversity and greater genetic differentiation than in the central part of the range (reviewed by Eckert et al. 2008). This hypothesis appears to hold at least in *bicostata*, with highest diversity in the continuous, central part of the range and lower diversity at the edge and in disjunct regions in the north and west of the range. For the species complex as a whole, though, high levels of genetic diversity were detected in the southeastern rather than the central part of the range. High genetic diversity is typically used as evidence for a center of origin (Vavilov 1992), suggesting that the southeastern region is the origin of the species complex. High genetic diversity in northern populations of the *Acacia saligna* species complex in Western Australia was evidence for a center of diversity or refugium (Millar et al. 2011), and in *E. camaldulensis*, 40% of the variation in allelic richness could be explained by latitude, with lower diversity in the south (Butcher et al. 2009), supporting the suggestion that the southern populations originated from the northern red gums (Hope 1994). A strong relationship between latitude and mean number of alleles per allozyme locus was also shown in populations of *Eucalyptus obtusiflora* ($r = 0.91$, Kennington and James 1998), a widespread species with a linear distribution pattern around 500-km long by 100-km wide. The non-linear/non-radial migration patterns and evolutionary history of *E. globulus* (Freeman et al. 2001) and the geographic barriers to dispersal that divide its distribution (the Murray Darling Depression, the highlands of the Great Dividing Range and Bass Strait, see Fig. 2) could account for the fact that the relationship between distance from South East Forests and heterozygosity/allelic richness was not strong. Two lines of evidence give further strength to the hypothesis that the South East Forests region is the origin of the species complex: (1) the Neighbour-Joining tree, in which the South East Forests populations cluster centrally; and (2) the capsule morphology

in this region, which is assumed to be ancestral (Jordan et al. 1993), as seven-fruited umbels occur in other members of the subseries *Globulinae,* but single-fruited umbels, as in *globulus*, do not (Slee et al. 2006).

Hybridization

There is good evidence that hybridization with species outside the *E. globulus* complex has affected the *E. globulus* gene pool, although nuclear DNA marker studies (McKinnon et al. 2010) indicate that this may be less extensive than was initially predicted from chloroplast markers (Jackson et al. 1999; McKinnon et al. 2001, 2004b). Similarly, this study provides evidence to suggest that hybridization and introgression among differentiated components of the *E. globulus* complex may have occurred in parts of the geographic range (e.g., 20-Strzelecki Ranges, see below), but most of the intergrade populations appear to be a result of primary differentiation as suggested by Kirkpatrick (1975). This is inferred from the absence of a peak in genetic diversity and the general lack of correlation between morphological and molecular affinities in intergrade zones. However, it is difficult to distinguish primary differentiation from ancient secondary contact (Endler 1977). It is sometimes achieved, using a combination of nuclear DNA markers and uniparentally inherited cytoplasmic DNA evidence (e.g., McKinnon et al. 2010). However, this has not been possible in *E. globulus* intergrade populations because of the confounding effects of potentially extensive chloroplast capture from species outside the *E. globulus* complex that can result in chloroplast DNA variation among species being correlated with geography rather than species boundaries (Steane et al. 1998; McKinnon et al. 1999, 2001, 2004a).

Although primary differentiation and ancient secondary intergradation may be difficult to distinguish, there is evidence to support the occurrence of recent secondary intergradations in *E. globulus*. A zone of recent secondary intergradation would be expected to display increased microsatellite diversity (due to gene flow from differentiated populations; Rieseberg and Wendel 1993) and a correlation between morphological and molecular affinities, as was the case in the *Q. crassifolia–Q. crassipes* hybrid zone in Mexico (Tovar-Sanchez and Oyama 2004). Of all the intergrade regions that were analyzed, only the 20-Strzelecki Ranges exhibited a significant correlation between morphological and molecular affinities, but showed no increase in genetic diversity compared with neighboring regions. A previous study of this region did not detect any microsatellite differentiation between individuals with *globulus* and *bicostata* morphological affinity and concluded that there was no evidence of recent sec-

ondary contact and introgression (Jones et al. 2002). This study, using a more sensitive analysis and greater sampling of the subspecies cores, provides evidence for recent gene flow from the adjacent coastal *globulus* into the Strzelecki Ranges population.

There is also the possibility of secondary gene flow across the eastern Bass Strait, between the Tasmanian *globulus* and the *pseudoglobulus* genetic groups. Tasmania and the mainland would have been linked by a land bridge formed following the lowering of sea levels during the Pleistocene glacial periods (McKinnon et al. 2004a). There are shared chloroplast DNA haplotypes (Freeman et al. 2001, 2007) and nuclear microsatellite (Steane et al. 2006; Yeoh et al. 2012; present study) affinities between western Victoria, western Tasmania, and King Island that support a western seed-dispersal link across this land bridge. However, although eastern mainland and Tasmanian choroplast DNA haplotypes occur on the Furneaux Group of islands, these haplotypes do not span the Bass Strait, thereby arguing for an eastern barrier to seed dispersal (Freeman et al. 2001, 2007). Nevertheless, pollen-mediated gene flow between eastern Victoria, Furneaux Group and northeast Tasmanian populations is suggested by the sharing of a rare haplotype of the nuclear *CCR* gene (McKinnon et al. 2005), as well as the occasional occurrence of mainland leaf chemotypes (Wallis et al. 2011) and the *pseudoglobulus* microsatellite group in the Furneaux Group and Northeastern Tasmania ($K = 6$ in Fig. 4). The far northern populations in the Furneaux group have morphological affinities to *pseudoglobulus*, and Jordan et al. (1993) have hypothesized that the Furneaux Group may be a zone of secondary contact between Tasmanian *globulus* and the *pseudoglobulus* genetic groups. Indeed, Wallis et al. (2011) hypothesized that on-going pollen-mediated gene flow between Tasmania and the mainland occurs along this eastern route, facilitated by the migratory swift parrot (*L. discolor*), a major pollinator of *globulus* (Hingston et al. 2004). Nevertheless, this study shows no evidence of increased nuclear microsatellite marker variability in the Furneaux Group expected from recent contact between well-differentiated gene pools.

Natural selection

At various geographic locations across the range of the *E. globulus* complex, there is clear evidence for selection on capsule morphology characters, which are the key taxonomic traits. The lack of association between morphological and geographic distance is evidence for selection on capsule morphology traits, and while the association between genetic distance and morphological distance among regions was significant, it was small and there were a number of key correlation breakers where there was a discrepancy between morphology and neutral

markers. For example, close morphological affinities but large genetic distances between 1-Mt. Bryan and the other *bicostata* regions are evidence not only for genetic drift but also for selection maintaining the *bicostata* capsule morphology despite migration over large geographic distances. Conversely, the 21-South Gippsland region of core *globulus* morphology had molecular affinities to geographically close, but morphologically distant regions, such as 20-Strzelecki Ranges, 15-Buchan, and the *pseudoglobulus* regions. The presence of these correlation breakers suggests that selection for large capsules over a steep environmental gradient has occurred in the Gippsland area (Jones et al. 2002). There is a positive correlation between seed size and capsule size at the population level within *E. globulus* (McGowen et al. 2004). Seed size influences seedling growth and bigger/heavier seed has been shown to result in faster growth of *E. globulus* during seedling establishment (Lopez et al. 2003). Selection for large capsules could, thus, have occurred in dry coastal areas as, for example, a mechanism to protect seeds during intense fire (Murray and Gill 2001) or to produce larger seed required for establishment in drought-prone areas (Hallett et al. 2011). Selection could, thus, explain discrepancies between morphology and microsatellite affinities, and it is noteworthy that several key discrepancies at the population level involve deviation of the molecular lineage in the direction of larger capsules and fewer capsules per inflorescence. This trend is evident, for example, in the *maidenii* molecular lineage at 7-Jenolan and in the *pseudoglobulus* molecular lineages at 21-South Gippsland and 2-Mt. Cole.

With its broad native range, the *E. globulus* complex spans diverse climatic zones in Australia ranging from summer (*maidenii*) to winter (*bicostata*, *globulus*) rainfall zones (Jordan et al. 1993). *Eucalyptus globulus* provenance trials established within and outside the native range of the complex in Australia have shown adaptive divergence at various geographic scales, from subspecies and regions (races) to local population level, indicating that natural selection has also played a significant role in shaping variation patterns in the *E. globulus* complex. At the subspecies level, core *bicostata* and *maidenii* are poorly adapted to the cool temperate climate of Tasmania (Volker and Orme 1988; Kube et al. 1995), whereas *globulus* is clearly more susceptible to leaf disease than *maidenii* (Carnegie et al. 1994; Hood et al. 2002), a factor contributing to the mal-adaptation of *globulus* to summer rainfall zones (Carnegie et al. 1994).

Of the four subspecies, *globulus* and its intergrades are the most intensively studied (because of their commercial importance; Dutkowski and Potts 1999; Stackpole et al. 2011) and there is plentiful evidence of broad-scale differential adaptation among races in many quantitative traits.

By comparing differentiation of putatively neutral molecular (F_{ST}) and quantitative (Q_{ST}) traits, Steane et al. (2006) identified a number of phenotypic traits that had been influenced by diversifying selection ($Q_{ST} > F_{ST}$). Subsequent studies have revealed other traits where $Q_{ST} > F_{ST}$ (Hamilton et al. 2011; Stackpole et al. 2011; Dutkowski and Potts 2012). Many of these traits show further evidence of adaptation, as their variation is associated with broad-scale (Stackpole et al. 2011; Dutkowski and Potts 2012) or local (Jordan et al. 2000; Foster et al. 2007; Dutkowski and Potts 2012) environmental gradients.

Practical implications

The molecular perspective on the *E. globulus* species complex is consistent with the taxonomic and morphological treatment in many respects. For example, while there is some admixture in intergrade populations, the cores of three of the four recognized taxa (*globulus*, *bicostata,* and *maidenii*) are relatively well-defined genetic entities, consistent with their geographic separation. The key problem is the lack of morphological definition of a large fraction of the gene pool that includes *pseudoglobulus* and the populations south of the Great Dividing Range in Victoria that comprise morphological intergrades between the cores (Jordan et al. 1993). In this region, there is little relationship between morphological and molecular affinities; yet at the molecular level, populations in this region form a relatively well-defined group that is genetically and geographically intermediate between the other three genetic groups. The taxonomic treatment of the populations from this "cryptic genetic entity" has been historically variable, which has caused confusion for conservation and exploitation of this gene pool. For example, different breeding programs have allocated the intergrade 20-Strzelecki region to different taxa (Jones et al. 2002). On the other hand, some populations classified morphologically as *globulus* in southern Gippsland belong to this cryptic genetic entity. From a practical perspective, the simplest higher level summary of the genetic diversity in this complex would be to expand the definition of *pseudoglobulus* to fully encompass populations in the cryptic genetic entity. The definition of taxa based on molecular data may be taxonomically challenging and difficult to apply. However, in this case, the genetic entity can be well defined geographically within Victoria, with the exception of a few peripheral populations (e.g. 2-Mt. Cole, 13-East Gippsland). Thus, geographic location would be the most pragmatic approach to classify germplasm in this region into taxonomic units for breeding and conservation purposes.

The study has provided a molecular framework for the management of the genetic resources of the *E. globulus* complex from several perspectives. First, the genetic entities

defined at various levels with the STRUCTURE analysis can guide translocation of *E. globulus* germplasm within its native range for forestry, agro-forestry, and restoration purposes. The level of molecular differentiation between a local and a non-local seed source can provide an early predictor of the genetic impact of introgression between translocated and native populations (Potts et al. 2003). This is because the more the populations differ, the more likely new genes are to be introgressed into the native population. Assessing the long-term consequences of introgression from tree plantings is extremely difficult given the long generation times and such information is one component of assessing genetic risk associated with translocation (Laikre et al. 2010; Byrne and Stone 2011; Byrne et al. 2011). Second, *globulus*, together with some Victorian intergrade populations, which are here classified with *pseudoglobulus*, forms the base populations of *E. globulus* breeding programs in many countries (Potts et al. 2004). As there are close molecular relationships among most Victorian populations, and breeding programs already include populations from the cryptic genetic entity identified here (Jones et al. 2006), a broader range of Victorian populations could readily be infused into the current breeding programs. Finally, molecular relationships among populations can be used to improve quantitative genetic evaluation of germplasm in breeding programs, because usual practice assumes unrelated genetic groups of founder individuals. In the Australian *E. globulus* breeding program, the subrace classification of Dutkowski and Potts (1999) is used to define genetic groups in these evaluations (McRae et al. 2004). However, molecular studies (Steane et al. 2006; Yeoh et al. 2012; this study) suggest that this grouping is conservative and, as discussed by Steane et al. (2006), some subraces are more closely related than others, which could be accounted for in quantitative genetic evaluations.

Acknowledgments

Conflict of Interest

None declared.

References

Aldrich, P. R., G. R. Parker, C. H. Michler, and J. Romero-Severson. 2003. Whole-tree silvic identifications and the microsatellite genetic structure of a red oak species complex in an Indiana old-growth forest. Can. J. For. Res. 33:2228–2237.

Barbour, R. C., B. M. Potts, and R. E. Vaillancourt. 2005. Gene flow between introduced and native *Eucalyptus* species: crossability of native Tasmanian species with exotic *E. nitens*. Aust. J. Bot. 53:465–477.

Barbour, R. C., Y. Otahal, R. E. Vaillancourt, and B. M. Potts. 2008. Assessing the risk of pollen-mediated gene flow from exotic *Eucalyptus globulus* plantations into native eucalypt populations of Australia. Biol. Conserv. 141:896–907.

Barbour, R. C., J. M. O'Reilly-Wapstra, D. W. de Little, G. J. Jordan, D. A. Steane, J. R. Humphreys, et al. 2009. A geographic mosaic of genetic variation within a foundation tree species and its community-level consequence. Ecology 90:1762–1772.

Beckstrom-Sternberg, S. M., L. H. Rieseberg, and K. Doan. 1991. Gene lineage analysis in populations of *Helianthus niveus* and *H. petiolaris* (Asteraceae). Plant Syst. Evol. 175:125–138.

Brondani, R. P. V., C. Brondani, R. Tarchini, and D. Grattapaglia. 1998. Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. Theor. Appl. Genet. 97:816–827.

Brondani, R. P. V., C. Brondani, and D. Grattapaglia. 2002. Towards a genus-wide reference linkage map for *Eucalyptus* based exclusively on highly informative microsatellite markers. Mol. Genet. Genomics 267:338–347.

Brooker, M. I. H. 2000. A new classification of the genus *Eucalyptus* L'Her (Myrtaceae). Aust. Syst. Bot. 13:79–148.

Brooker, M. I. H., and D. A. Kleinig. 2006. Field guide to eucalypts. Volume 1. South-eastern Australia. 3rd ed. Bloomings Books, Hawthorn, Vic.

Butcher, P. A., M. W. McDonald, and J. C. Bell. 2009. Congruence between environmental parameters, morphology and genetic structure in Australia's most widely distributed eucalypt, *Eucalyptus camaldulensis*. Tree Genet. Genomes 5:189–210.

Byrne, M. 2008. Phylogeny, diversity and evolution of eucalypts. Pp. 303–346 *in* A. K. Sharma and A. Sharma, eds. Plant genome: biodiversity and evolution. Volume 1, Part E: Phanerogams – Angiosperm. Science Publishers, Enfield.

Byrne, M., and L. Stone. 2011. The need for 'duty of care' when introducing new crops for sustainable agriculture. Curr. Opin. Environ. Sustain. 3:50–54.

Byrne, M., B. MacDonald, and D. Coates. 2002. Phylogeographical patterns in chloroplast DNA variation

within the *Acacia acuminata* (Leguminosae: Mimosoideae) complex in Western Australia. J. Evol. Biol. 15:576–587.

Byrne, M., L. Stone, and M. A. Millar. 2011. Assessing genetic risk in revegetation. J. Appl. Ecol. 48:1365–1373.

Caetano, S., L. Nusboumer, and Y. Naciri. 2008. Chloroplast and microsatellite markers in *Astronium urundeuva* (Allemao) Engl. and close species of Anacardiaceae: toward the definition of a species complex? Candollea 63:115–130.

Carnegie, A. J., P. J. Keane, P. K. Ades, and I. W. Smith. 1994. Variation in susceptibility of *Eucalyptus globulus* provenances to *Mycosphaerella* leaf disease. Can. J. For. Res. 24:1751–1757.

Chybicki, I. J., and J. Burczyk. 2009. Simultaneous estimation of null alleles and inbreeding coefficients. J. Hered. 100:106–113.

Cook, L. G., D. C. Morris, R. D. Edwards, and M. D. Crisp. 2008. Reticulate evolution in the natural range of the invasive wetland tree species *Melaleuca quinquenervia*. Mol. Phylogenet. Evol. 47:506–522.

Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer Associates, Inc., Sunderland, MA.

Doyle, J. J., and J. L. Doyle. 1990. Extraction of plant DNA from fresh tissue. Focus 12:13–15.

Dressler, R. L., and C. H. Dodson. 1960. Classification and phylogeny in the Orchidaceae. Ann. Mo. Bot. Gard. 47:25–68.

Duminil, J., and M. Di Michele. 2009. Plant species delimitation: a comparison of morphological and molecular markers. Plant Biosyst. 143:528–542.

Duminil, J., H. Caron, I. Scotti, S. O. Cazal, and R. J. Petit. 2006. Blind population genetics survey of tropical rainforest trees. Mol. Ecol. 15:3505–3513.

Dutkowski, G. W., and B. M. Potts. 1999. Geographic patterns of genetic variation in *Eucalyptus globulus* ssp. *globulus* and a revised racial classification. Aust. J. Bot. 47:237–263.

Dutkowski, G. W., and B. M. Potts. 2012. Genetic variation in the susceptibility of *Eucalyptus globulus* to drought damage. Tree Genet. Genomes 8:757–773.

Eckert, C. G., K. E. Samis, and S. C. Lougheed. 2008. Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. Mol. Ecol. 17:1170–1188.

El Mousadik, A., and R. J. Petit. 1996. High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. Theor. Appl. Genet. 92:832–839.

Eldridge, K., J. Davidson, C. Harwood, and G. van Wyk. 1993. Eucalypt domestication and breeding. Clarendon Press, Oxford.

Endler, J. A. 1977. Geographic variation, speciation and clines. Princeton Univ. Press, Princeton, NJ.

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

Foster, S. A., G. E. McKinnon, D. A. Steane, B. M. Potts, and R. E. Vaillancourt. 2007. Parallel evolution of dwarf ecotypes in the forest tree *Eucalyptus globulus*. New Phytol. 175:370–380.

Frankel, O. H., and M. E. Soule. 1981. Conservation and evolution. Cambridge Univ. Press, Cambridge, U.K.

Freeman, J. S., H. D. Jackson, D. A. Steane, G. E. McKinnon, G. W. Dutkowski, B. M. Potts, et al. 2001. Chloroplast DNA phylogeography of *Eucalyptus globulus*. Aust. J. Bot. 49:585–596.

Freeman, J. S., C. M. P. Marques, V. Carocha, N. Borralho, B. M. Potts, and R. E. Vaillancourt. 2007. Origins and diversity of the Portuguese landrace of *Eucalyptus globulus*. Ann. For. Sci. 64:639–647.

Galouchko, V. 2007. 3DField Version 2.9.6. Available via http://field.hypermart.net/.

Gomory, D., I. Yakovlev, P. Zhelev, J. Jedinakova, and L. Paule. 2001. Genetic differentiation of oak populations within the *Quercus robur/Quercus petraea* complex in Central and Eastern Europe. Heredity 86:557–563.

Gomory, D., L. Paule, and J. Vysny. 2007. Patterns of allozyme variation in western Eurasian *Fagus*. Bot. J. Linn. Soc. 154:165–174.

Goudet, J. 2002. FSTAT, a program to estimate and test gene diversities and fixation indices Version 2.9.3.2. Available via http://www2.unil.ch/popgen/softwares/fstat.htm.

Grattapaglia, D., R. E. Vaillancourt, M. Shepherd, B. R. Thumma, W. J. Foley, C. Kulheim, et al. 2012. Progress in Myrtaceae genomics: *Eucalyptus* as the pivotal genus. Tree Genet. Genomes (special Compendium issue) 8:463–508.

Griffin, A. R., I. P. Burgess, and L. Wolf. 1988. Patterns of natural and manipulated hybridization in the genus *Eucalyptus* L'Herit – a review. Aust. J. Bot. 36:41–66.

Hallett, L. M., R. J. Standish, and R. J. Hobbs. 2011. Seed mass and summer drought survival in a Mediterranean-climate ecosystem. Plant Ecol. 212:1479–1489.

Hamilton, M. G., P. A. Tilyard, D. R. Williams, R. E. Vaillancourt, T. J. Wardlaw, and B. M. Potts. 2011. The genetic variation in the timing of heteroblastic transition in *Eucalyptus globulus* is stable across environments. Aust. J. Bot. 59:170–175.

Hess, W. J., and N. A. Stoynoff. 1998. Taxonomic status of *Quercus acerifolia* (Fagaceae) and a morphological comparison of four members of the *Quercus shumardii* complex. Syst. Bot. 23:89–100.

Hingston, A. B., B. M. Potts, and P. B. McQuillan. 2004. The swift parrot, *Lathamus discolor* (Psittacidae), social bees (Apidae) and native insects as pollinators of *Eucalyptus globulus* ssp. *globulus* (Myrtaceae). Aust. J. Bot. 52:371–379.

Hokanson, S. C., J. G. Isebrands, R. J. Jensen, and J. F. Hancock. 1993. Isozyme variation in oaks of the Apostle Islands in Wisconsin – genetic structure and levels of inbreeding in *Quercus rubra* and *Q. ellipsoidalis* (Fagaceae). Am. J. Bot. 80:1349–1357.

Hood, I. A., J. F. Gardner, M. O. Kimberley, and K. Molony. 2002. Variation among eucalypt species in early susceptibility to the leaf spot fungi *Phaeophleospora eucalypti* and *Mycosphaerella* spp. NZ J. Forest. Sci. 32:235–255.

Hope, G. S. 1994. Quarternary vegetation. Pp. 368–389 *in* R. S. Hill, ed. History of the Australian vegetation: cretaceous to recent. Cambridge Univ. Press, Cambridge.

Jackson, H. D., D. A. Steane, B. M. Potts, and R. E. Vaillancourt. 1999. Chloroplast DNA evidence for reticulate evolution in *Eucalyptus* (Myrtaceae). Mol. Ecol. 8:739–751.

Jakobsson, M., and N. A. Rosenberg. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806.

Johnson, L. A. S. 1976. Problems of species and genera in *Eucalyptus* (Myrtaceae). Plant Syst. Evol. 125:155–167.

Jones, R. C., D. A. Steane, B. M. Potts, and R. E. Vaillancourt. 2002. Microsatellite and morphological analysis of *Eucalyptus globulus* populations. Can. J. For. Res. 32:59–66.

Jones, T. H., D. A. Steane, R. C. Jones, D. Pilbeam, R. E. Vaillancourt, and B. M. Potts. 2006. Effects of domestication on genetic diversity in *Eucalyptus globulus*. For. Ecol. Manage. 234:78–84.

Jones, T. H., R. E. Vaillancourt, and B. M. Potts. 2007. Detection and visualization of spatial genetic structure in continuous *Eucalyptus globulus* forest. Mol. Ecol. 16:697–707.

Jones, R. C., V. F. G. Hecht, B. M. Potts, R. E. Vaillancourt, and J. L. Weller. 2011. Expression of a *FLOWERING LOCUS T* homologue is temporally associated with annual flower bud initiation in *Eucalyptus globulus* subsp. *globulus* (Myrtaceae). Aust. J. Bot. 59:756–769.

Jordan, G. J., B. M. Potts, J. B. Kirkpatrick, and C. Gardiner. 1993. Variation in the *Eucalyptus globulus* complex revisited. Aust. J. Bot. 41:763–785.

Jordan, G. J., B. M. Potts, P. Chalmers, and R. J. E. Wiltshire. 2000. Quantitative genetic evidence that the timing of vegetative phase change in *Eucalyptus globulus* ssp. *globulus* is an adaptive trait. Aust. J. Bot. 48:561–567.

Kashani, N., and R. S. Dodd. 2002. Genetic differentiation of two California red oak species, *Quercus parvula* var. *shreveii* and *Q. wislizeni*, based on AFLP genetic markers. General Technical Report – Pacific Southwest Research Station, USDA Forest Service:417–426.

Kennington, W. J., and S. H. James. 1998. Allozyme and morphometric variation in two closely related mallee species from Western Australia, *Eucalyptus argutifolia* and *E. obtusiflora* (Myrtaceae). Aust. J. Bot. 46:173–186.

Kirkpatrick, J. 1975. Geographical variation in *Eucalyptus globulus*. Department of Agriculture, Forestry and Timber Bureau, Canberra, Australia.

Kube, P. D., A. C. Matheson, C. A. Raymond, and D. B. Boomsma. 1995. Provenance variation in *Eucalyptus globulus*. Eucalypt Plantations: Improving Fibre Yield and Quality, 19–24 February, at Hobart, Tas.

Kulheim, C., S. H. Yeoh, I. R. Wallis, S. Laffan, G. F. Moran, and W. J. Foley. 2011. The molecular basis of quantitative variation in foliar secondary metabolites in *Eucalyptus globulus*. New Phytol. 191:1041–1053.

Laikre, L., M. K. Schwartz, R. S. Waples, and N. Ryman. 2010. Compromising genetic diversity in the wild: unmonitored large-scale release of plants and animals. Trends Ecol. Evol. 25:520–529.

Lewis, P. O., and D. Zaykin. 2001. Genetic data analysis: computer program for the analysis of allelic data. Version 1.1. Available via http://hydrodictyon.eeb.uconn.edu/people/plewis/software.php.

Lopez, G. A., B. M. Potts, R. E. Vaillancourt, and L. A. Apiolaza. 2003. Maternal and carryover effects on early growth of *Eucalyptus globulus*. Can. J. For. Res. 33:2108–2115.

Makarenkov, V. 2001. T-REX: reconstructing and visualizing phylogenetic trees and reticulation networks. Bioinformatics 17:664–668.

Manos, P. S., and D. E. Fairbrothers. 1987. Allozyme variation in populations of six Northeastern American red oaks (Fagaceae, *Quercus* Subg *Erythrobalanus*). Syst. Bot. 12:365–373.

McGowen, M. H., B. M. Potts, R. E. Vaillancourt, P. L. Gore, D. R. Williams, and D. J. Pilbeam. 2004. The genetic control of sexual reproduction in *Eucalyptus globulus*. *Eucalyptus* in a Changing World, 11–15 October, at Aveiro, Portugal.

McKinnon, G. E., D. A. Steane, B. M. Potts, and R. E. Vaillancourt. 1999. Incongruence between chloroplast and species phylogenies in *Eucalyptus* subgenus *Monocalyptus* (Myrtaceae). Am. J. Bot. 86:1038–1046.

McKinnon, G. E., R. E. Vaillancourt, H. D. Jackson, and B. M. Potts. 2001. Chloroplast sharing in the Tasmanian eucalypts. Evolution 55:703–711.

McKinnon, G. E., G. J. Jordan, R. E. Vaillancourt, D. A. Steane, and B. M. Potts. 2004a. Glacial refugia and reticulate evolution: the case of the Tasmanian eucalypts. Philos. Trans. R. Soc. Lond. B Biol. Sci. 359:275–284.

McKinnon, G. E., R. E. Vaillancourt, D. A. Steane, and B. M. Potts. 2004b. The rare silver gum, *Eucalyptus cordata*, is leaving its trace in the organellar gene pool of *Eucalyptus globulus*. Mol. Ecol. 13:3751–3762.

McKinnon, G. E., B. M. Potts, D. A. Steane, and R. E. Vaillancourt. 2005. Population and phylogenetic analysis of the cinnamoyl coA reductase gene in *Eucalyptus globulus* (Myrtaceae). Aust. J. Bot. 53:827–838.

McKinnon, G. E., R. E. Vaillancourt, D. A. Steane, and B. M. Potts. 2008. An AFLP marker approach to lower-level systematics in *Eucalyptus* (Myrtaceae). Am. J. Bot. 95:368–380.

McKinnon, G. E., J. J. Smith, and B. M. Potts. 2010. Recurrent nuclear DNA introgression accompanies chloroplast DNA exchange between two eucalypt species. Mol. Ecol. 19:1367–1380.

McRae, T. A., D. J. Pilbeam, M. Powell, G. W. Dutkowski, K. Joyce, and B. Tier. 2004. Genetic evaluation in Eucalypt breeding programs. *Eucalyptus* in a changing world, 11–15 October, at Aveiro, Portugal.

Mergeay, J., and L. Santamaria. 2012. Evolution and Biodiversity: the evolutionary basis of biodiversity and its potential for adaptation to global change. Evol. Appl. 5:103–106.

Millar, M. A., M. Byrne, and W. O'Sullivan. 2011. Defining entities in the *Acacia saligna* (Fabaceae) species complex using a population genetics approach. Aust. J. Bot. 59:137–148.

Mimura, M., R. C. Barbour, B. M. Potts, R. E. Vaillancourt, and K. N. Watanabe. 2009. Comparison of contemporary mating patterns in continuous and fragmented *Eucalyptus globulus* native forests. Mol. Ecol. 18:4180–4192.

Murray, B. R., and A. M. Gill. 2001. A comparative study of interspecific variation in fruit size among Australian eucalypts. Ecography 24:651–658.

Nei, M. 1972. Genetic distance between populations. Am. Nat. 106:283–292.

Oline, D. K. 2008. Geographic variation in chloroplast haplotypes in the California red fir-noble fir species complex and the status of Shasta red fir. Can. J. For. Res. 38:2705–2710.

Peakall, R., and P. E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6:288–295.

Potts, B. M., R. C. Barbour, A. B. Hingston, and R. E. Vaillancourt. 2003. Genetic pollution of native eucalypt gene pools – identifying the risks. Aust. J. Bot. 51:1–25.

Potts, B. M., R. E. Vaillancourt, G. J. Jordan, G. W. Dutkowski, J. Costa e Silva, G. E. McKinnon, et al. 2004. Exploration of the *Eucalyptus globulus* gene pool. *Eucalyptus* in a changing world, 11–15 October, at Aveiro, Portugal.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Pryor, L. D., and L. A. S. Johnson. 1971. A classification of the eucalypts. Australian National Univ. Press, Canberra, ACT.

Rieseberg, L. H., and J. F. Wendel. 1993. Introgression and its consequences in plants. Pp. 70–114 *in* R. G. Harrison, ed. Hybrid zones and the evolutionary process. Oxford Univ. Press, New York.

Rieseberg, L. H., S. M. Beckstrom-Sternberg, A. Liston, and D. M. Arias. 1991. Phylogenetic and systematic inferences from chloroplast DNA and isozyme variation in *Helianthus* sect. *Helianthus* (Asteraceae). Syst. Bot. 16:50–76.

Rosenberg, N. A. 2004. DISTRUCT: a program for the graphical display of population structure. Mol. Ecol. Notes 4:137–138.

Saitou, N., and M. Nei. 1987. The Neighbor-Joining method – a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406–425.

Sgrò, C. M., A. J. Lowe, and A. A. Hoffmann. 2010. Building evolutionary resilience for conserving biodiversity under climate change. Evol. Appl. 4:326–337.

Skabo, S., R. E. Vaillancourt, and B. M. Potts. 1998. Fine-scale genetic structure of *Eucalyptus globulus* ssp. *globulus* forest revealed by RAPDs. Aust. J. Bot. 46:583–594.

Slee, A. V., M. I. H. Brooker, S. M. Duffy, and J. G. West. 2006. EUCLID: eucalypts of Australia. 3rd ed. CSIRO Publishing, Collingwood, Vic.

Stackpole, D. J., R. E. Vaillancourt, A. Alves, J. Rodrigues, and B. M. Potts. 2011. Genetic variation in the chemical components of *Eucalyptus globulus* wood. G3 (Bethesda) 1:151–159.

Steane, D. A., M. Byrne, R. E. Vaillancourt, and B. M. Potts. 1998. Chloroplast DNA polymorphism signals complex interspecific interactions in *Eucalyptus* (Myrtaceae). Aust. Syst. Bot. 11:25–40.

Steane, D. A., R. E. Vaillancourt, J. Russell, W. Powell, D. Marshall, and B. M. Potts. 2001. Development and characterisation of microsatellite loci in *Eucalyptus globulus* (Myrtaceae). Silvae Genet. 50:89–91.

Steane, D. A., N. Conod, R. C. Jones, R. E. Vaillancourt, and B. M. Potts. 2006. A comparative analysis of population structure of a forest tree, *Eucalyptus globulus* (Myrtaceae), using microsatellite markers and quantitative traits. Tree Genet. Genomes 2:30–38.

Steane, D. A., D. Nicolle, C. P. Sansaloni, C. D. Petroli, J. Carling, A. Kilian, et al. 2011. Population genetic analysis and phylogeny reconstruction in *Eucalyptus* (Myrtaceae) using high-throughput, genome-wide genotyping. Mol. Phylogenet. Evol. 59:206–224.

Stebbins, G. L. 1989. Plant speciation and the founder principle. Pp. 113–125 *in* L. Giddings, K. Kaneshiro and W. Anderson, eds. Genetics, speciation and the founder principle. Oxford Univ. Press, New York, NY.

Tovar-Sanchez, E., and K. Oyama. 2004. Natural hybridization and hybrid zones between *Quercus crassifolia* and *Quercus crassipes* (Fagaceae) in Mexico: morphological and molecular evidence. Am. J. Bot. 91:1352–1363.

Vavilov, N. I. 1992. Origin and geography of cultivated plants. Cambridge Univ. Press, Cambridge, U.K.

Volker, P. W., and R. K. Orme. 1988. Provenance trials of *Eucalyptus globulus* and related species in Tasmania. Aust. For. 51:257–265.

Wallis, I. R., A. Keszei, M. L. Henery, G. F. Moran, R. Forrester, J. Maintz, et al. 2011. A chemical perspective on the evolution of variation in *Eucalyptus globulus*. Perspect. Plant Ecol. Evol. Syst. 13:305–318.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370.

Whitham, T. G., G. D. Martinsen, K. D. Floate, H. S. Dungey, B. M. Potts, and P. Keim. 1999. Plant hybrid zones affect biodiversity: tools for a genetic-based understanding of community structure. Ecology 80:416–428.

Whitham, T. G., C. A. Gehring, L. M. Evans, C. J. LeRoy, R. K. Bangert, J. A. Schweitzer, et al. 2010. A community and ecosystem genetics approach to conservation biology and management. Pp. 50–73 *in* A. B. DeWoody, J. W. Bickham, C. Michler, K. M. Nichols, O. E. Rhodes Jr and K. Woeste,

eds. Molecular approaches in natural resource conservation and management. Cambridge Univ. Press, Cambridge, U.K.

Wright, S. 1943. Isolation by distance. Genetics 28:114–138.

Yatabe, Y., N. C. Kane, C. Scotti-Saintagne, and L. H. Rieseberg. 2007. Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, *Helianthus annuus* and *H. petiolaris*. Genetics 175:1883–1893.

Yeoh, S. H., J. C. Bell, W. J. Foley, I. R. Wallis, and G. F. Moran. 2012. Estimating population boundaries using regional and local-scale spatial genetic structure: an example in *Eucalyptus globulus*. Tree Genet. Genomes 8:695–708. Online only. doi:10.1007/s11295-011-0457-4.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web site:

Table S1. Sample sizes (*n*) of *Eucalyptus globulus* localities in this study, after excluding individuals that were missing data from three or more loci.

Table S2. Overall genetic diversity parameters for the microsatellite loci used in the analysis of the *Eucalyptus globulus* species complex.

Figure S1. Calculation of ΔK, following Evanno et al. (2005).

Data S1. Description of the repeatability of scoring for the microsatellite loci used in the analysis of the *Eucalyptus globulus* species complex.

Data S2. Description of the overall genetic diversity parameters for the microsatellite loci used in the analysis of the *Eucalyptus globulus* species complex.