NRC Research Press

ARTICLE

Microsatellite analysis of population structure in *Eucalyptus globulus*[1]

Joana Costa, René E. Vaillancourt, Dorothy A. Steane, Rebecca C. Jones, and Cristina Marques

Abstract: *Eucalyptus globulus* subsp. *globulus* Labill. (Tasmanian Blue Gum), native to southeast Australia, is a benchmark species for the pulp and paper industry. We genotyped 397 trees from 16 populations of *E. globulus* representing the native diversity in Australia using 24 microsatellite loci. Eight genetically distinct groups were detected, consistent with genetic groupings detected in previous quantitative and molecular studies. A sample of 29 Portuguese individuals was added to help clarify the origin of the Portuguese landrace. The results suggest a southern and eastern Tasmania origin for the Portuguese landrace. This genetic framework will enable researchers to investigate the provenance of individuals of unknown pedigree and assess the levels of representation of *E. globulus* natural variation in the Portuguese landrace.

Key words: Eucalyptus globulus breeding, simple sequence repeats, genetic diversity, Portuguese landrace.

Résumé : L'*Eucalyptus globulus* ssp. *globulus* Labill. (gommier bleu), indigène du sud-est de l'Australie, est une espèce de référence pour l'industrie des pâtes et papiers. Les auteurs ont génotypé 397 arbres, provenant de 16 populations de l'*E. globulus* et représentant la diversité présente en Australie, au moyen de 24 microsatellites. Huit groupes génétiques distincts ont été détectés, une observation conforme aux groupements génétiques établis suite à des études quantitatives et moléculaires antérieures. Un échantillon de 29 échantillons portugais a été ajouté afin de préciser l'origine d'une variété de pays portugaise. Les résultats suggèrent que cette variété de pays proviendrait du sud ou de l'est de la Tasmanie. Ce cadre génétique permettra aux chercheurs d'étudier la provenance d'individus de pedigree inconnu et de mesurer le degré de représentation de la variation naturelle de l'*E. globulus* au sein de la variété de pays portugaise. [Traduit par la Rédaction]

Mots-clés : amélioration génétique de l'*Eucalyptus globulus*, microsatellites, diversité génétique, variété de pays portugaise.

Introduction

A detailed knowledge of the population structure of a species is key to understanding its evolutionary history (Slatkin 1987). This knowledge has a direct impact on species management, ecology, and conservation in the implementation of genetic improvement programs (Grattapaglia and Kirst 2008) and in associating genes and traits (Pritchard et al. 2000).

Eucalyptus globulus subsp. *globulus* Labill. (Tasmanian Blue Gum) is native to southeast Australia, occurring naturally in Victoria and Tasmania, including the islands in the Bass Strait. The commercial value of the species prompted a seed collection in the late 1980s, which was considered representative of the species natural diversity throughout its area of distribution. Using this collection, and 35 quantitative traits such as flowering, growth, and leaf morphology, it was determined that *E. globulus* has significant population structure, consisting of 13 races and 20 subraces (Dutkowski and Potts 1999). However, because many quantitative traits may be important in adaptation, patterns of relationship between the *E. globulus* races inferred from quantitative data may not accurately reflect the true genetic relationships between races, owing to selection on the very quantitative traits used to infer relationships. Therefore, this structure was later studied using molecular markers, and the genetic affinities of races were further clarified (Steane et al. 2006; Yeoh et al. 2012). Using eight microsatellite loci, Steane et al. (2006) detected that the native populations of *E. globulus* in southeastern Australia were genetically diverse, and that the species had a highly structured genetic architecture. The genetic relationships among races of *E. globulus* correlated well with geographic location, in contrast to relationships previously inferred from quantitative genetic approaches. Using 16 microsatellite loci, Yeoh et al. (2012) detected five genetically distinct groups, consistent with the geographic distribution of the samples. Jones et al. (2013) expanded sampling to the related subspecies *E. globulus* subsp. *pseudoglobulus*, *E. globulus* subsp. *bicostata*, and *E. globulus* subsp. *maidenii* and found that some of the intergrade races included in the *E. globulus* breeding populations, such as Strzelecki Ranges, were part of a genetic lineage that included populations classified as *E. globulus* subsp. *pseudoglobulus*.

It is believed that the breeding of *E. globulus* began in Portugal in 1966, based on phenotypic selections from local populations (Potts et al. 2004). The Australian origin of *E. globulus* introductions is unrecorded and there are concerns that the landrace may have originated from a narrow genetic base (Eldridge et al. 1993; Potts et al. 2004). Previous studies suggest that the Portuguese landrace has affinities with the southeastern Tasmanian race (Lopez et al.

Received 3 December 2016. Accepted 26 March 2017.

Corresponding Editor: Andrew L. Hipp.

J. Costa and C. Marques. RAIZ, Instituto de Investigação da Floresta e do Papel, Herdade de Espirra, 2985-270 Pegões-Gare, Portugal.
R.E. Vaillancourt. School of Biological Sciences, University of Tasmania, Private Bag 55, Hobart, 7001, Tasmania, Australia; ARC Training Centre for Forest Value, University of Tasmania, Private Bag 55, Hobart, 7001, Tasmania, Australia.
D.A. Steane and R.C. Jones. School of Biological Sciences, University of Tasmania, Private Bag 55, Hobart, 7001, Tasmania, Australia.
Corresponding author: Joana Costa (email: joana.costa@thenavigatorcompany.com).

[1]This paper is part of a Special Issue entitled The Evolution of Tree Diversity.

Table 1. Information on the 45 native localities of *Eucalyptus globulus* and the number of samples collected within each population.

	Locality	Region (race)	No. of individuals	Latitude	Longitude
1	Cannan Spur	Western Otways	8	−38.7639	143.5461
2	Lavers Hill		7	−38.7461	143.2833
3	Otway State Forest		7	−38.7745	143.4239
4	Parker Spur		8	−38.8435	143.5410
5	Cape Patton	Eastern Otways	7	−38.6731	143.8582
6	Jamieson Creek		8	−38.6021	143.9179
7	Lorne		7	−38.5327	143.9700
8	Phillip Island	Phillip Island	13	−38.4836	145.2618
9	Unnamed	Strzelecki Ranges	2	−38.44714	145.84682
10	Bowden-Carrajung		1	−38.3663	146.7012
11	Jeeralang		15	−43.1918	146.8960
12	Alberton West-Welshpool-Hedley	South Gippsland	9	−38.6662	146.4407
13	Fish Creek		6	−38.6990	146.1047
14	Toora		8	−38.6722	146.2803
15	Tidal River	Tidal River	20	−39.0212	146.3356
16	Wilson's Promontory Lighthouse	Wilson's Promontory	23	−39.1269	146.4246
17	Cape Barren Island	Furneaux	25	−40.3408	148.2334
18	Flinders Island central		14	−40.0605	148.0562
19	Flinders Island north		1	−39.8449	147.8362
20	Flinders Island south		3	−40.2097	148.2388
21	Cape Tourville Dwarf	Northeastern Tasmania	1	−42.1231	148.3422
22	Cape Tourville Tall		7	−42.1214	148.3371
23	Humbug Hill		9	−41.1912	148.2424
24	Jericho		6	−42.4268	147.2618
25	Mayfield north		7	−42.0173	147.9681
26	Mayfield south		5	−42.2115	148.0375
27	Pepper Hill		8	−41.6376	147.8450
28	Maria Island north	Southeastern Tasmania	5	−42.5861	148.0638
29	Maria Island south		10	−42.7171	148.0721
30	Tinderbox		8	−43.0338	147.3202
31	Triabunna		8	−42.4572	147.8932
32	Mt. Dromedary (Tas)	Dromedary	10	−42.6988	147.1237
33	Platform Peak		2	−42.6915	147.0300
34	Dover	Southern Tasmania	5	−43.3002	147.0772
35	Geeveston		8	−43.1918	146.8960
36	South Bruny Island		8	−43.1918	146.8960
37	Recherche Bay	Recherche Bay	25	−43.5283	146.8959
38	Port Davey	Port Davey	19	−43.2918	145.9199
39	Badgers Creek	Western Tasmania	7	−42.0017	145.2761
40	Little Henty River		6	−41.9416	145.1980
41	Macquarie Harbour		16	−42.3439	145.3359
42	Central King Island	King Island	10	−39.9429	144.0278
43	Central North King Island		4	−39.7579	143.9792
44	North King Island		6	−39.6499	144.0249
45	South King Island		5	−39.8729	143.9949
		Total	397		

Note: Each region corresponds to the 16 populations used for analysis, reflecting the 13 races determined by Dutkowski and Potts (1999) and some new populations (Tidal River, Phillip Island, and Port Davey).

2001; Freeman et al. 2007). *Eucalyptus globulus* landrace material is now a major component of the breeding and deployment populations in Portugal (Eldridge et al. 1993; Araújo et al. 1997; Costa e Silva et al. 2005), together with introduced germplasm of known Australian origin (Eldridge et al. 1993). The area of origin within the native gene pool is an important issue as *E. globulus* is highly variable, and races have different traits of interest (Apiolaza et al. 2005).

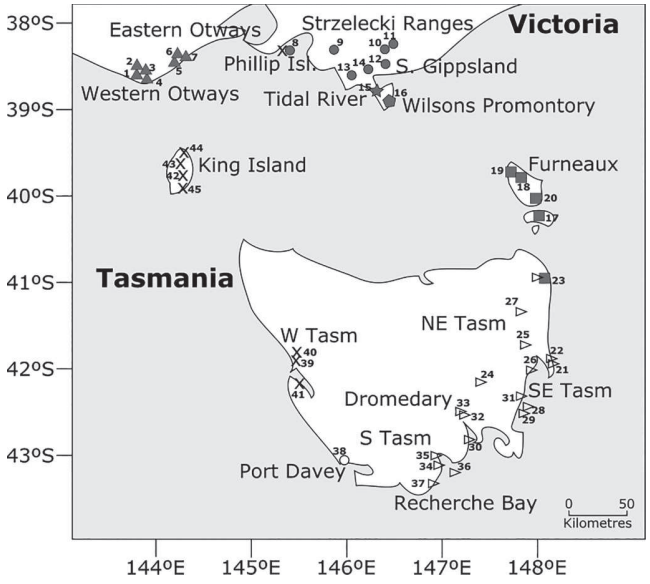
The aim of this study was to study the genetic structure of *E. globulus*, sampling populations from throughout its entire natural distribution, using SSR markers. Based on the results, the origin of individuals collected in Portugal and introduced in genetic improvement programs can be investigated, relative to the diversity and structure of the Australian population.

Materials and methods

Plant material sampling

The sample size of 397 individuals of Australian origin aims to be representative of *E. globulus* natural diversity and its races. Original localities (1–45, Table 1 and Fig. 1) were grouped into 16 populations reflecting the 13 races determined by Dutkowski and Potts (1999), as well as some new populations (Tidal River, Phillip Island, and Port Davey, Fig. 1). Sample locations and DNA extraction protocols are given in Steane et al. (2006) and Jones et al. (2013). A sample of 29 individuals collected throughout Portugal was added for later analysis, representing a 17th population of unknown Australian origin. Four samples of known Australian origin (Furneaux region) were included in the Portuguese sample and treated as blind controls.

Fig. 1. Sampling locations across the *Eucalyptus globulus* species' range (grouped in the 16 populations considered) and the eight clusters resolved from STRUCTURE and GENELAND analyses. Each symbol matches a colour in Fig. 2, as follows: ○ refers to the colour grey (Port Davey), ● refers to red (eastern Victoria), ■ refers to yellow (Furneaux), ★ refers to blue (Tidal River), ▲ refers to green (western Victoria), ✕ refers to pink (western Tasmania and King Island), ▷ refers to orange (southern and eastern Tasmania), and ⬟ to the colour purple (Wilson's Promontory).



Twenty-four microsatellite (SSR) markers were selected to cover the linkage groups in the *Eucalyptus* genome (Table 2) and amplified to provide genotypes for the 397 samples using the conditions described in Glaubitz et al. (2001), Steane et al. (2001), and Brondani et al. (2006).

Amplified DNA fragments were separated by electrophoresis on 0.25 mm 65% polyacrylamide gels and visualized using a DNA sequencer (Gene Readir 4200; LI-COR). RFLP Scan analysis software (Scanalytic, Fairfax, Va., USA) was used for measuring SSR allele size. Manual editing of allele size was performed for increased accuracy.

Data analysis

Population-level genetic parameters were estimated. Allelic frequencies were generated using GENALEX v6.5 (Peakall and Smouse 2012). The number of alleles per loci (A), the number of effective alleles per loci (A_e), and allelic richness (A_r) were calculated using FSTAT v2.9.3 (Goudet 1995, 2001). The observed (H_o) and expected (H_e) heterozygosity, the inbreeding coefficient (F_{IS}), and the Hardy-Weinberg equilibrium (HWE) were calculated using the software FSTAT and GENEPOP v4 (Rousset 2008). The polymorphism information content (PIC) of SSR markers was calculated by POWERMARKER v3.25 (Liu and Muse 2005). The null allele frequencies were estimated following the Brookfield method, implemented in MICRO-CHECKER v2.2.3 package (Van Oosterhout et al. 2004). The population differentiation indices F_{ST} (fixation index) and D_{EST} (Jost's D) were also calculated. The global F_{ST} was calculated using GENEPOP and pairwise F_{ST} measurements (Weir and Cockerham 1984) for the 16 populations were estimated using ARLEQUIN v3.5 (Excoffier and Lischer 2010, 10 000 permutations), as well as the significance levels for the registered F_{ST} values. SMOGD v1.2.5 (Crawford 2010) was used to measure the actual differentiation among populations (D_{EST}) (Jost 2008; Meirmans and Hedrick 2011).

The genetic structure was investigated using a Bayesian model-based clustering method implemented in the software STRUCTURE v2.3.4 (Pritchard et al. 2000). All the analyses were conducted using a discarded burnin of 300 000 steps, followed by 800 000 Markov Chain Monte Carlo (MCMC) steps. Ten independent runs were performed for each cluster K (K = 1 to K = 16). All the analyses were based on the "admixture" and "correlated allele frequencies" models and the outputs were consistent in all iterations. STRUCTURE HARVESTER v0.6.92 (Earl and vonHoldt 2012) was used to extract the results. The optimum number of genetically distinct clusters was determined based on the maximum value of mean posterior probability across replicates (Pritchard et al. 2000) and also on the Evanno test (Evanno et al. 2005). Software CLUMPP v1.1.2 (Jakobsson and Rosenberg 2007) was used to obtain the average of the permuted individuals and Q matrices, throughout the several replicates for each K. Final results were visualized using DISTRUCT v1.1 (Rosenberg 2004), used to obtain the bar plots where each individual is represented as a segment, divided in K colours. Population analyses were repeated with the same parameters for 17 populations after adding the Portuguese samples to the data matrix. The POPFLAG = 1 parameter was also explored with the same result.

The degree of population subdivision was also explored using the R-package GENELAND v4.0.2 (Guillot et al. 2005*b*). This program, as well as STRUCTURE, is based on Bayesian clustering model, executed in a MCMC scheme, attributing individuals to subpopulations based on its genetic information. However, GENELAND also uses the spatial coordinates as prior information at an initial stage of the simulation, which may provide a better estimation of the number of clusters (Guillot et al. 2005*a*). Ten replicates were performed, with the following parameters: the number of subpopulations (clusters, K) varied from 1 to 16 with 1 500 000 MCMC iterations and a thining of 100. The Dirichlet model was used as the alternative model, and it was verified to have a tendency to overestimate the number of clusters (Guillot et al. 2005*a*). The 10 replicates were ordered by the values of mean logarithm of posterior probability and the three best analyses were processed and visually verified for consistency.

Nei et al. (1983) DA distance implemented in POWERMARKER was calculated for pairwise populations/races. Bootstrapping over loci with 1000 replications was carried out to assess the strength of the evidence for the branching patterns in the resulting neighbor-joining tree. A consensus tree with bootstrap values was reconstructed by the consensus program of PHYLIP (Felsenstein 2006) and displayed by FigTree v1.3.1 (Rambaut 2014).

To estimate the proportion of genetic variability explained by the subpopulations inferred by GENELAND and STRUCTURE, an analysis of molecular variance (AMOVA) was conducted in ARLEQUIN. The analysis examines the hierarchic proportion of total genetic variability among groups (F_{CT}), among populations within groups (F_{SC}), among individuals within populations (F_{IS}), and within individuals (F_{IT}). The robustness of the frontiers between clusters was evaluated by comparing the proportion of variance among groups.

Results

Diversity analysis

The 24 SSR markers were tested for several coefficients (Table 2) in the Australian populations. The PIC values ranged from 0.6347 (*EMBRA19*) to 0.9378 (*EMBRA37*), with an average of 0.8456. All loci were considered very informative (PIC > 0.5; Botstein et al. 1980). The number of alleles (A) varied between 10 (*EMBRA213*) and 42 (*EMBRA42*), with an average of 22.5 alleles/locus. The effective number of alleles (A_e) was considerably lower than A (Table 2), suggesting many alleles were relatively rare. The observed heterozygosity (H_o) was high in most loci (Table 2), with an average value of 0.76. The expected heterozygosity (H_e) showed an average value of 0.79. The native Australian populations in this study did

Table 2. Information of the 24 microsatellite markers used in this study, with repeat motif, number of alleles observed, size range of amplification product, and linkage group.

Locus	Repeat	Linkage group	*A*	Size range	A_e	A_r	PIC	H_o	H_e	F_{IS}	F_{ST}	D_{EST}	Null
EMBRA10	$(CCT)_3(AG)$	10	22	115–165	6.31	11.65	0.91	0.84	0.85	0.01	0.08	0.49	0.02
EMBRA11	$(AG)_4GG(AG)_{13}$	1	28	93–165	5.34	9.86	0.85	0.82	0.81	0.00	0.07	0.31	0.00
EMBRA12	$(AG)_{22}$	1	20	108–174	6.90	11.98	0.91	0.86	0.86	−0.01	0.07	0.47	0.00
EMBRA18	$(AG)_3GG(AG)_{19}$	9	29	97–167	6.42	12.38	0.92	0.88	0.85	−0.03	0.09	0.57	0.02
EMBRA19	$(AG)_{23}$	4	11	146–176	2.70	4.89	0.64	0.36	0.62	0.44***	0.08	0.15	0.22
EMBRA23	$(AG)_{16}$	8	24	96–168	7.02	11.90	0.91	0.83	0.87	0.05*	0.05	0.35	0.02
EMBRA30	$(AG)_{22}$	8	30	98–166	6.49	10.88	0.89	0.85	0.82	−0.02	0.08	0.46	0.00
EMBRA37	$(AG)_{16}$	5	31	114–188	9.47	13.88	0.94	0.93	0.89	−0.03	0.05	0.47	0.00
EMBRA38	$(AG)_{11}(TA)_4$	10	25	100–158	6.08	10.87	0.88	0.81	0.81	0.02	0.09	0.45	0.02
EMBRA42	$(AG)_{15}$	7	42	107–197	7.44	13.67	0.92	0.86	0.87	0.02	0.06	0.44	0.00
EMBRA63	$(AG)_{21}$	2	22	171–233	3.74	7.03	0.74	0.71	0.73	0.03	0.06	0.19	0.00
EMBRA120	$(TC)_{17}$	5	25	118–168	7.17	12.84	0.93	0.86	0.87	0.03	0.07	0.49	0.00
EMBRA210	$(TC)_{25}$	9	25	189–249	6.94	12.04	0.92	0.89	0.86	−0.05	0.07	0.48	0.00
EMBRA213	$(CT)_{17}$	4	10	204–224	2.74	4.65	0.64	0.56	0.60	0.09*	0.16	0.29	0.01
EMBRA362	$(TC)_3TT(TC)_{18}(AC)_{13}$	6	19	108–166	4.77	8.29	0.82	0.80	0.79	−0.01**	0.06	0.27	0.00
EMBRA645	—	4	28	182–238	4.95	10.49	0.87	0.65	0.80	0.16***	0.10	0.46	0.07
EMBRA712	—	11	18	160–226	4.00	8.27	0.79	0.71	0.74	0.03	0.09	0.29	0.00
EMBRA747	—	11	21	242–288	4.91	9.57	0.85	0.76	0.79	0.04*	0.09	0.40	0.00
EMCRC6	$(CT)_{11}(CA)_{23}$	8	20	147–203	4.56	9.46	0.81	0.70	0.74	0.05*	0.09	0.30	0.00
EMCRC7	$(TG)_{15}(AG)_8$	3	19	265–317	3.83	8.37	0.79	0.66	0.71	0.08*	0.10	0.33	0.01
EMCRC8	$(CT)_{13}(CA)_{24}$	2	25	223–279	6.26	12.34	0.92	0.79	0.82	0.03	0.12	0.61	0.00
EMCRC11	$(TC)_{10}(AC)_{10}$	6	18	223–267	5.38	9.32	0.85	0.78	0.80	0.03	0.09	0.39	0.06
Eg61	$GAA_9(GAT)_9$(1 T to A)	7	16	300–360	3.46	7.74	0.75	0.65	0.70	0.08*	0.10	0.29	0.02
Eg131	$TTC_{11}(CTC)_2$	3	13	108–147	5.22	8.34	0.86	0.74	0.82	0.10***	0.07	0.34	0.04
Average			22.54			10.03±2.51	0.85	0.76	0.79		0.08***	0.36***	

Note: Diversity indices are presented as *A*, number of alleles; A_e, effective number of alleles; A_r, allelic richness; PIC, polymorphism information content; H_o/H_e, observed/expected heterozygosity; F_{IS}, inbreeding coefficient. ***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$ indicate significant deficit or excess of heterozygosity in comparison with the expected under Hardy-Weinberg equilibrium. F_{ST} (fixation index) and Jost's *D* (D_{EST}) values for each locus and overall are highly significant ($p < 0.001$). Null refers to null alleles frequency estimates.

not deviate significantly from Hardy–Weinberg equilibrium (HWE), with most of the loci in balance, or just slightly deviated from the expected values. The most significant deviations from HWE were in markers *EMBRA19*, *EMBRA645*, and *EG131*, and all presented positive values for the inbreeding coefficient (F_{IS}). F_{IS} positive values may be a sign of null alleles, but only *EMBRA19* tested positive for null alleles (Table 2).

Overall F_{ST} (fixation index, 0.08 with 95% CI equal to 0.074–0.092) and Jost's *D* (D_{EST}, 0.36) values indicated significant ($p < 0.001$) population differentiation. The F_{ST} values in the native populations, although relatively low to moderate (0.02–0.2), were highly significant for all pairs of populations ($p < 0.001$) (supplementary data, Table S1[2]). The D_{EST} values were, in most cases, higher than the F_{ST} values (0.1–0.5, Table S1[2].

Population structure

STRUCTURE outputs analysed using the Δ*K* method of Evanno et al. (2005) indicated that there were five main genetic clusters ($K = 5$, Fig. S1[2]) in the Australian populations. The clusters/groups compositions were as follows: Group A (green) is composed of populations from eastern Otways, Phillip Island, south Gippsland, Strzelecki Ranges, and western Otways; Group B (blue) comprised genetic material from Tidal River and Wilson's Promontory; Group C (yellow) is formed by individuals from Furneaux; Group D (orange) composed of genetic materials from Dromedary, northeast Tasmania, Port Davey, Recherche Bay, southeast Tasmania, and southern Tasmania; and Group E (pink) clustered individuals from King Island and western Tasmania (Fig. 2).

GENELAND analysis consistently revealed $K = 8$ for all replicates. Although this was a higher number of clusters, these eight groups were congruent with the structure found with $K = 5$ in STRUCTURE and were also nearly identical to the STRUCTURE solution at $K = 8$. The structure of eight populational clusters was as follows (Figs. 1 and 2): Group A (green) composed of individuals from eastern Otways and western Otways; Group B (red) constituted by genetic materials from south Gippsland, Strzelecki Ranges, and Phillip Island; Group C (blue) composed of individuals from Tidal River; Group D (purple) formed by genetic materials from Wilson's Promontory; Group E (yellow) formed by individuals from Furneaux; Group F (orange) is composed by Dromedary, northeast Tasmania, Recherche Bay, southeast Tasmania, and southern Tasmania; Group G (grey) constituted by genetic materials from Port Davey; and Group H (pink) by King Island and western Tasmania. The most important difference between the STRUCTURE $K = 8$ and $K = 5$ solutions was that the group containing most of the Victorian races was split into eastern Victoria (south Gippsland and Strzelecki Ranges) and western Victoria (eastern and western Otways) groups. This splitting was also apparent in the dendrogram (Fig. 3) with 65% bootstrap support. One other difference was the splitting of the Port Davey race from the group that contained most of the Tasmanian races. The last difference was the splitting of the group joining the adjacent Tidal River and Wilson's Promontory races. In general and in both the $K = 8$ and $K = 5$ solutions, the attribution of populations to clusters coincided with geographic distribution (Figs. 1 and 2).

Most clusters were very homogeneous and showed little admixture apart from northeast Tasmania and Phillip Island (Fig. 2). Some individuals from Humbug Hill (locality 23, Table 1 and Fig. 1), the northernmost locality in the northeastern Tasmania race, presented a strong influence (10%–40%) of Group E (Furneaux), or even a complete genetic constitution equivalent to that of

[2]Supplementary data are available with the article through the journal Web site at http://nrcresearchpress.com/doi/suppl/10.1139/gen-2016-0218.

Fig. 2. STRUCTURE clustering results obtained for the *Eucalyptus globulus* Australian populations. Each individual is represented by a thin vertical bar. Each colour represents an estimated population group (cluster, *K*). The colours in each individual represent the estimated proportion of its presence in each of the *K* groups. (*a*) *K* = 5; (*b*) *K* = 8, which was similar to the GENELAND analysis.

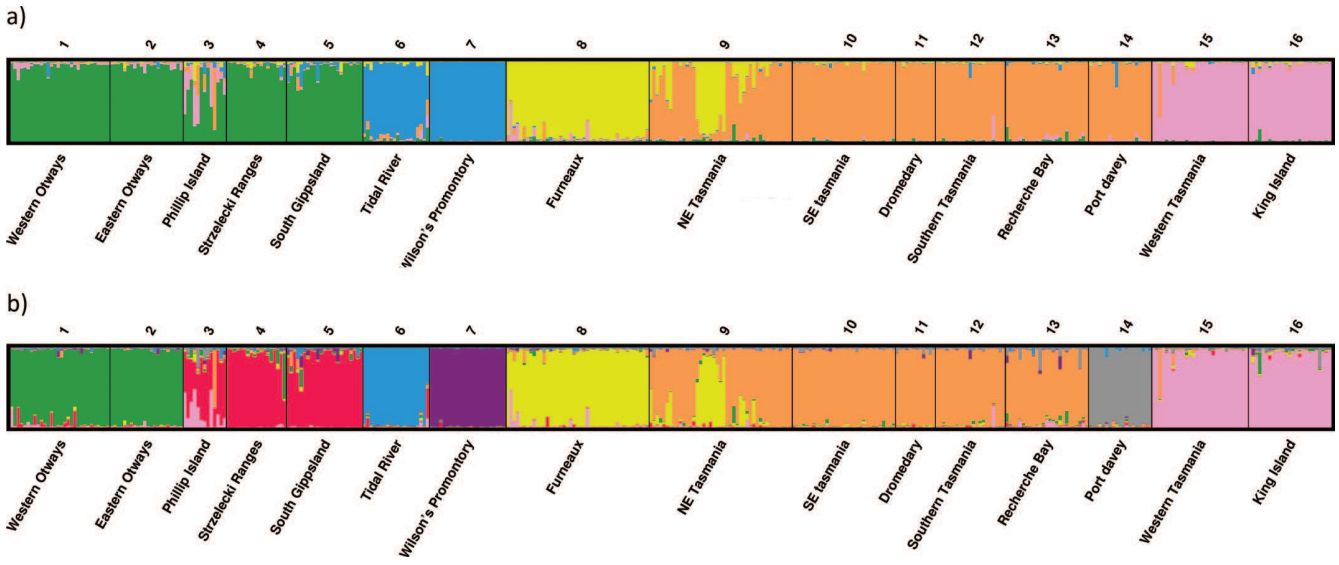
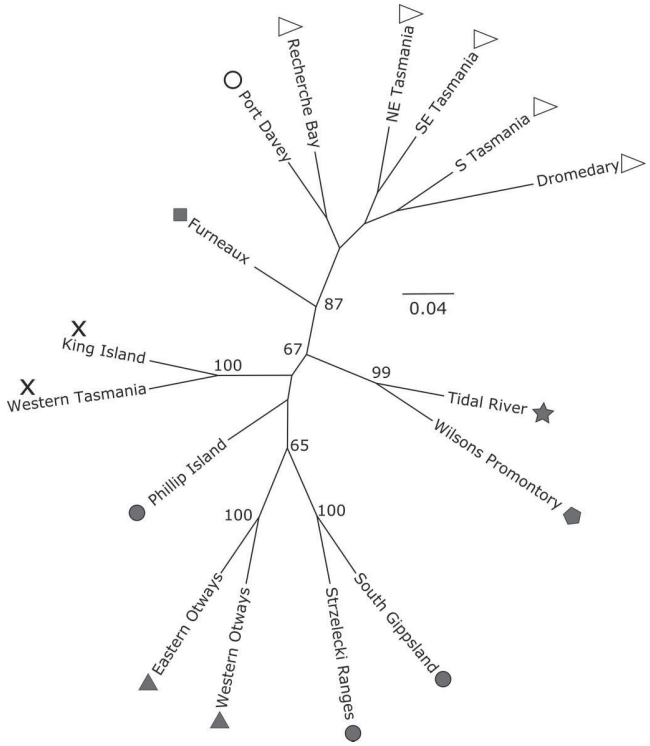


Fig. 3. Neighbor-joining tree of 16 Australian populations of *Eucalyptus globulus*, based on Nei's genetic distance (DA, Nei et al. 1983). The symbols represent the different groups inferred by the STRUCTURE and GENELAND analysis represented in Fig. 2*b*, as follows: ○ refers to the colour grey (Port Davey), ● refers to red (eastern Victoria), ■ refers to yellow (Furneaux), ★ refers to blue (Tidal River), ▲ refers to green (western Victoria), ✕ refers pink (western Tasmania and King Island), ▷ refers to orange (southern and eastern Tasmania), and ⬟ to the colour purple (Wilson's Promontory). Only bootstrap values above 60% are shown.



Group E (Fig. 2). The Phillip Island population appeared as a region of genetic admixture, with most individuals belonging to Group B (south Gippsland and Strzelecki Ranges), but a strong genetic influence from other clusters, particularly from Group H (King Island and western Tasmania).

AMOVA results were used to study the partitioning of the genetic variation at both values of *K* obtained (*K* = 5 and *K* = 8, Table 3). The analysis showed that most of the diversity was found within individuals (F_{IT} = 87% in both *K*). All other partitions of the genetic variances were small. There was also no difference in the proportion of the variation among individuals within populations (F_{IS} = 3.6%) between *K* = 8 and *K* = 5. However, there were small differences between *K* = 8 and *K* = 5 in the proportions of the variance groups (F_{CT}) and among populations within groups (F_{SC}). The major proportion of difference between groups (F_{CT}) was obtained when the populations were classified in eight groups (5.94% vs. 3.63%).

The neighbor-joining phenogram illustrates the relationship among the 16 *E. globulus* populations based on Nei et al.'s (1983) genetic distance with bootstrap values (Fig. 3). The genetic distance-based cluster analysis also grouped populations according to geographic locations that agreed closely with the clustering of populations based on the Bayesian method. All populations from the Otway Ranges formed a group with bootstrap support of 100%. The closest cluster to the Otway group consisted of populations from south Gippsland and Strzelecki Ranges (100%), with a bootstrap support of 65%. Populations from western Tasmania and King Island formed a cluster that links to the previous populations (and Phillip Island). The next clear group is between Wilson's Promontory and Tidal River (100%). Data suggests that the Furneaux population is an outgroup of the Gippsland/Strzelecki, Otways, King Island/west Tasmania, and Tidal River/Promontory groups, with a strong bootstrap support (87%). All of the populations in southern and eastern Tasmania, together with Dromedary, Recherche Bay, and Port Davey grouped together but with poor bootstrap support. Recherche Bay tended to group with Port Davey, in contrast to the results from the Bayesian analysis.

Portuguese landrace

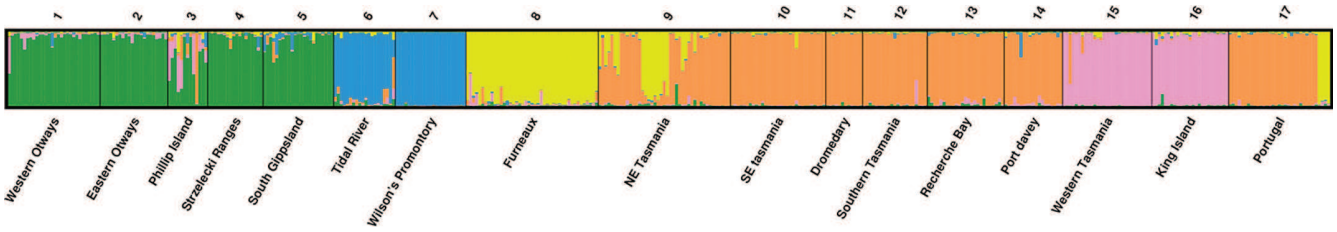
The Portuguese population was highly polymorphic, with a mean of 28.92 alleles per locus, a value within the range of the observed for the Australian populations (average *A* = 22.54). Sim-

Table 3. AMOVA design and results considering the 16 populations defined and the degree of population subdivision detected (K = 5 and K = 8).

	Among groups		Among populations within groups		Among individuals within populations		Within individuals	
	%	F_{CT}	%	F_{SC}	%	F_{IS}	%	F_{IT}
Five clusters (K = 5)	3.63	0.03627***	5.28	0.05477***	3.63	0.03986***	87.46	0.12537***
Eight clusters (K = 8)	5.94	0.05939***	2.97	0.03153***	3.63	0.03986***	87.46	0.12536***

Note: ***, $p < 0.001$.

Fig. 4. STRUCTURE clustering results obtained for the *Eucalyptus globulus* Australian populations and the Portuguese landrace. Each individual is represented by a thin vertical bar. Each colour represents an estimated population group (cluster, K). The colours in each individual represent the estimated proportion of its presence in each of the K groups. The last four samples in the Portuguese population represent the blind control, samples with an Australian origin in the Furneaux region (yellow cluster).



ilar to that observed in the Australian populations, the mean effective number of alleles per locus was considerably lower than the observed number of alleles per locus (A_e = 10.88). The high number of alleles was reflected in the high observed and expected heterozygosity (H_o = 0.83 and H_e = 0.84, respectively), and the values were similar to, but higher than, those found in the Australian populations (H_o = 0.76 and H_e = 0.79).

The addition of the Portuguese samples to the population structure analysis revealed a similar result with five main genetic clusters (K = 5). The Portuguese samples merged with samples from southern and eastern Tasmania (Fig. 4). The four blind control samples demonstrated the power of the analysis as they grouped with the Australian race of origin (i.e., Furneaux).

Discussion

The study investigated the genetic diversity and population structure of native *E. globulus* using 24 SSR markers. This is significantly more SSR markers than previously used in this species (8 in Steane et al. (2006); 16 in Yeoh et al. (2012); and 9 in Jones et al. (2013)). More power in analysing the architecture of genetic diversity is expected from using more markers (Luikart et al. 2003). The level of genetic diversity, as measured using heterozygosity statistics (H_o and H_e) as well as those measuring allele diversity was high, which agrees with similar results reported previously in *E. globulus* (Steane et al. 2006; Ribeiro et al. 2011), but also in other eucalypt species (Arumugasundaram et al. 2011; Faria et al. 2011). However, genetic diversity was not as high as in eucalypt species with more continuous (connected) distribution, such as *E. obliqua* and *E. pauciflora* (Bloomfield et al. 2011; Gauli et al. 2014).

Meirmans and Hedrick (2011) defended the use of measures other than F_{ST} when highly polymorphic markers such as SSR are used. In a meta-analysis of 34 published studies, Heller and Siegismund (2009) determined that D_{EST} seemed to be more appropriate to estimate populational differentiation and was a better measure for allelic differentiation between populations. However, F_{ST} is still useful as a fixation index and is largely used, so it is presented as a comparative measure, as the values obtained are similar to those presented by Steane et al. (2006) and Yeoh et al. (2012) for *E. globulus*. Nevertheless, *E. globulus* populations showed more genetic differentiation (overall F_{ST} = 0.08) than reported for other eucalypt species, such as *E. pauciflora* (F_{ST} = 0.03, Gauli et al. 2014) and *E. obliqua* populations (F_{ST} = 0.02, Bloomfield et al. 2011).

Although most of the variation in *E. globulus* is found within individuals, significant but small proportions of the variation can be ascribed to various levels of tree groupings (locality, race, or STRUCTURE groups). Similar results were found using different types of markers (i.e., RAPD, Nesbitt et al. 1995) and other SSR studies of this species (Steane et al. 2006; Yeoh et al. 2012; Jones et al. 2013). Species with large population and that are more continuous in their distribution have smaller proportions of their variation between individuals, populations, and groups (Bloomfield et al. 2011; Gauli et al. 2014).

The genetic structure of *E. globulus* uncovered in this study is congruent with that found in previous studies using SSRs (Steane et al. 2006; Yeoh et al. 2012; Jones et al. 2013) and consistent with geography. Differences between studies appear to result from differences in sampling. For example, the populations at Wilson's Promontory, Phillip Island, and Port Davey were not included in Yeoh et al. (2012), and Steane et al. (2006) had a small number of trees from Phillip Island and Port Davey and none from Wilson's Promontory. All analyses agree that the more or less continuous populations on the east coast of Tasmania, although sub-divided into a number of races based on quantitative traits, form a cohesive genetic group. Yeoh et al. (2012) pointed out that the populations at the northeastern extreme of Tasmania have an influence from the Furneaux Islands, but not vice-versa. This is consistent with unidirectional pollen dispersal through migrating birds, while a barrier to seed dispersal was discovered between the Furneaux Island and Tasmania through studies of cpDNA (Freeman et al. 2001). The small remnant population of *E. globulus* on the west coast of Tasmania (Port Davey, western Tasmania) has affinities with southern Tasmanian east coast populations (Recherche Bay) at their southern extreme and King Island (in Bass Strait) at their northern extreme. Interestingly, King Island and Phillip Island in eastern Victoria appear to have closer genetic affinities than would be expected considering their geographic separation. The population on Phillip Island may be a remnant of a larger population, which inhabited the Bassian plains at times of low sea level during past ice ages, when Tasmania was connected to the mainland (Mckinnon et al. 2004). The populations of *E. globulus* in western Victoria (the Otways) are clearly connected to those in eastern Victoria (south Gippsland and Strzelecki Ranges), as shown in the K = 5 results. However, the populations in eastern and western Victoria form separate groups in the K = 8 analysis, consistent with their geographic separation. Wilson's Promontory

and Tidal River each have their own group in $K = 8$, but they are grouped together in $K = 5$. The separation of the disjunct and small population at Wilson's Promontory may be due to drift as evidenced by its low genetic diversity (Jones et al. 2013), which may also be the case for Port Davey. The studies of cpDNA (Freeman et al. 2001, 2007) have shown strong connections, through sharing of haplotypes between the populations in Bass Strait (King Island and Furneaux) and those on the mainland.

A relatively high level of genetic diversity was found in the Portuguese landrace. Molecular evidence suggests that the Portuguese landrace is dominated by genetic material from southern and eastern Tasmania. The high levels of genetic diversity contradict the suggestions that the Portuguese landrace may have been derived from a very narrow original collection (Eldridge et al. 1993; Potts et al. 2004), and this is supported by a previous molecular study (Freeman et al. 2007) that used seven nuclear SSR and one cpDNA marker. Freeman et al. (2007) found that the Portuguese landrace had a level of genetic diversity intermediate between that of an average race and that found in the species as a whole. In this study, the samples of the Portuguese landrace displayed a higher level of diversity than found overall in the species. The difference between studies may be due to differences in sampling of the landrace. The predominantly southern and eastern Tasmanian origin of the Portuguese landrace is consistent with evidence from quantitative data (Lopez et al. 2001) and previous molecular study (Freeman et al. 2007). However, Freeman et al. (2007) found that a few samples of the Portuguese landrace had affinities to other races, such as the Furneaux Islands, King Island, and mainland Australian races (e.g., Otways and south Gippsland). This discrepancy may again be due to differences in sampling or because Freeman et al.'s (2007) use of chloroplast markers increased their ability to detect genetic contributions from other origins. Typically, selections for pulpwood breeding purposes derived from *E. globulus* native populations have preferred germplasm from Strzelecki Ranges, Otways, and Furneaux (Potts et al. 2004), which are races that appear to have a diminished representation in the Portuguese landrace and might have to be imported by breeding programs.

In conclusion, several studies have demonstrated that traits of economic value, such as wood density or disease resistance, vary between *E. globulus* Australian races (Dutkowski and Potts 1999; Borralho et al. 2007; Stackpole et al. 2010), and better understanding of the genetic structure of the species will help quantitative genetic analyses in breeding programs as well as association genetic analyses (Thavamanikumar et al. 2014). Understanding the genetic structure of ecologically important species such as *E. globulus* is crucial as it informs management of the native populations, helps provide evidence of natural selection in the study of quantitative traits (e.g., F_{ST} vs. Q_{ST} comparisons; Steane et al. 2006), and provides a native database to allow discovering the origin of unpedigreed material including landraces from countries such as Argentina, Chile, Ecuador, India, Peru, and Spain (Potts et al. 2004).

Acknowledgements

We acknowledge support from Australian Research Council Discovery grants DP130104220 and DP160101650. We thank Sascha Wise and Fátima Cunha for help in preparing the DNA samples and José Araújo for providing the Portuguese samples.

References

Apiolaza, L., Raymond, C., and Yeo, B.J. 2005. Genetic variation of physical and chemical wood properties of *Eucalyptus globulus*. Silvae Genet. **54**(4–5): 160–166.

Araújo, J.A., Lemos, L., Ramos, A., Ferreira, J.G., and Borralho, N.M.G. 1997. The RAIZ *Eucalyptus globulus* breeding program: a BLUP rolling front strategy with a mixed clonal and seedling deployment scheme. *In* Proceedings IUFRO Conference on Silviculture and Improvements of Eucalypts, Salvador, Brazil. pp. 371–376.

Arumugasundaram, S., Ghosh, M., Veerasamy, S., and Ramasamy, Y. 2011. Species discrimination, population structure and linkage disequilibrium in *Eucalyptus camaldulensis* and *Eucalyptus tereticornis* using SSR markers. PLoS ONE, **6**(12): e0028252. doi:10.1371/journal.pone.0028252.

Bloomfield, J.A., Nevill, P., Potts, B.M., Vaillancourt, R.E., and Steane, D.A. 2011. Molecular genetic variation in a widespread forest tree species *Eucalyptus obliqua* (Myrtaceae) on the island of Tasmania. Aust. J. Bot. **59**(3): 226–237. doi:10.1071/BT10315.

Borralho, N.M.G., Almeida, H., and Potts, B.M. 2007. O Melhoramento do Eucalipto em Portugal. *In* Eucaliptal em Portugal: Impactes Ambientais e Investigação Científica. *Edited by* A.M. Alves, J.S. Pereira, and J.M.N. Silva. ISAPress, Lisbon. pp. 61–110.

Botstein, D., White, R.L., Skolnick, M., and Davis, R.W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. **32**(3): 314–31. PMID:6247908.

Brondani, R.P., Williams, E.R., Brondani, C., and Grattapaglia, D. 2006. A microsatellite-based consensus linkage map for species of *Eucalyptus* and a novel set of 230 microsatellite markers for the genus. BMC Plant Biol. **6**: 20. doi:10.1186/1471-2229-6-20. PMID:16995939.

Costa, e, Silva, J., Dutkowski, G., and Borralho, N.M.G. 2005. Across-site heterogeneity of genetic and environmental variances in the genetic evaluation of *Eucalyptus globulus* trials for height growth. Silvae Genet. **62**: 183–191. doi:10.1051/forest:2005010.

Crawford, N.G. 2010. Smogd: software for the measurement of genetic diversity. Mol. Ecol. Resour. **10**(3): 556–557. doi:10.1111/j.1755-0998.2009.02801.x. PMID:21565057.

Dutkowski, G.W., and Potts, B.M. 1999. Geographic patterns of genetic variation in *Eucalyptus globulus* ssp. *globulus* and a revised racial classification. Aust. J. Bot. **47**(2): 237–263. doi:10.1071/BT97114.

Earl, D.A., and vonHoldt, B.M. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. **4**(2): 359–361. doi:10.1007/s12686-011-9548-7.

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

Evanno, G., Regnaut, S., and Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. **14**(8): 2611–2620. doi:10.1111/j.1365-294X.2005.02553.x. PMID:15969739.

Excoffier, L., and Lischer, H.E.L. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. **10**(3): 564–567. doi:10.1111/j.1755-0998.2010.02847.x. PMID:21565059.

Faria, D.A., Mamani, E.M.C., Pappas, G.J., and Grattapaglia, D. 2011. Genotyping systems for *Eucalyptus* based on tetra-, penta-, and hexanucleotide repeat EST microsatellites and their use for individual fingerprinting and assignment tests. Tree Genet. Genomes, **7**(1): 63–77. doi:10.1007/s11295-010-0315-9.

Felsenstein, J. 2006. PHYLIP (Phylogeny Inference Package) v3.66. Department of Genome Sciences, University of Washington; Seattle, Wash., USA.

Freeman, J.S., Jackson, H.D., Steane, D.A., McKinnon, G.E., Dutkowski, G.W., Potts, B.M., and Vaillancourt, R.E. 2001. Chloroplast DNA phylogeography of *Eucalyptus globulus*. Aust. J. Bot. **49**(5): 585–596. doi:10.1071/BT00094.

Freeman, J.S., Marques, C.M.P., Carocha, V., Borralho, N., Potts, B.M., and Vaillancourt, R.E. 2007. Origins and diversity of the Portuguese Landrace of *Eucalyptus globulus*. Ann. For. Sci. **64**(6): 639–647. doi:10.1051/forest:2007042.

Gauli, A., Steane, D.A., Vaillancourt, R.E., and Potts, B.M. 2014. Molecular genetic diversity and population structure in *Eucalyptus pauciflora* subsp. *pauciflora* (Myrtaceae) on the island of Tasmania. Aust. J. Bot. **62**(3): 175–188. doi:10.1071/BT14036.

Glaubitz, J.C., Emebiri, L.C., and Moran, G.F. 2001. Dinucleotide microsatellites from *Eucalyptus sieberi*: inheritance, diversity, and improved scoring of single-base differences. Genome, **44**(6): 1041–1045. doi:10.1139/g01-106. PMID:11768207.

Goudet, J. 1995. FSTAT. Version 1.2. A computer program to calculate F-statistics. J. Hered. **86**(6): 485–486.

Goudet, J. 2001. FSTAT, a program to estimate and test gene diversities and fixation indices. Version 2.9.3.

Grattapaglia, D., and Kirst, M. 2008. *Eucalyptus* applied genomics: from gene sequences to breeding tools. New Phytol. **179**(4): 911–929. doi:10.1111/j.1469-8137.2008.02503.x. PMID:18537893.

Guillot, G., Estoup, A., Mortier, F., and Cosson, J.F. 2005*a*. A spatial statistical model for landscape genetics. Genetics, **170**(3): 1261–1280. doi:10.1534/genetics.104.033803. PMID:15520263.

Guillot, G., Mortier, F., and Estoup, A. 2005*b*. Geneland: a computer package for landscape genetics. Mol. Ecol. Notes, **5**(3): 712–715. doi:10.1111/j.1471-8286.2005.01031.x.

Heller, R., and Siegismund, H.R. 2009. Relationship between three measures of genetic differentiation G_{ST}, D_{EST} and G'_{ST}: How wrong have we been? Mol Ecol. **18**: 2080–2083. doi:10.1111/j.1365-294X.2009.04185.x. PMID:19645078.

Jakobsson, M., and Rosenberg, N.A. 2007. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics, **23**(14): 1801–1806. doi:10.1093/bioinformatics/btm233. PMID:17485429.

Jones, R.C., Steane, D.A., Lavery, M., Vaillancourt, R.E., and Potts, B.M. 2013. Multiple evolutionary processes drive the patterns of genetic differentiation in a forest tree species complex. Ecol. Evol. **3**(1): 1–17. doi:10.1002/ece3.421. PMID:23403692.

Jost, L. 2008. G_{ST} and its relatives do not measure differentiation. Mol. Ecol. **17**(18): 4015–4026. doi:10.1111/j.1365-294X.2008.03887.x. PMID:19238703.

Liu, K., and Muse, S.V. 2005. PowerMarker: an integrated analysis environment for genetic maker analysis. Bioinformatics, **21**(9): 2128–2129. doi:10.1093/bioinformatics/bti282. PMID:15705655.

Lopez, G.A., Potts, B.M., Dutkowski, G.W., and Rodriguez Traverso, J.M. 2001. Quantitative genetics of *Eucalyptus globulus*: affinities of land race and native stand localities. Silvae Genet. **50**(5–6): 244–252.

Luikart, G., England, P.R., Tallmon, D., Jordan, S., and Taberlet, P. 2003. The power and promise of population genomics: from genotyping to genome typing. Nat. Rev. Genet. **4**(12): 981–994. doi:10.1038/nrg1226. PMID:14631358.

Mckinnon, G.E., Jordan, G.J., Vaillancourt, R.E., Steane, D.A., and Potts, B.M. 2004. Glacial refugia and reticulate evolution: the case of the Tasmanian eucalypts. Philos. Trans. R. Soc. B Biol. Sci. **359**: 275–284. doi:10.1098/rstb.2003.1391.

Meirmans, P.G., and Hedrick, P.W. 2011. Assessing population structure: F_{ST} and related measures. Mol. Ecol. Resour. **11**(1): 5–18. doi:10.1111/j.1755-0998.2010.02927.x. PMID:21429096.

Nei, M., Tajima, F., and Tateno, Y. 1983. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. J. Mol. Evol. **19**(2): 153–170. doi:10.1007/BF02300753. PMID:6571220.

Nesbitt, K.A., Potts, B.M., Vaillancourt, R.E., Wesf, A.K., and Reid, J.B. 1995. Partitioning and distribution of RAPD variation in a forest tree species, *Eucalyptus globulus* (Myrtaceae). Heredity (Edinb.), **74**: 628–637. doi:10.1038/hdy.1995.86.

Peakall, R., and Smouse, P.E. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research — an update. Bioinformatics, **28**: 2537–2539. PMID:22820204.

Potts, B.M., Vaillancourt, R.E., Jordan, G., Dutkowski, G., Costa, J., Mckinnon, G., et al. 2004. Exploration of the *Eucalyptus globulus* gene pool. *In Eucalyptus* in a changing world. Proceedings of the IUFRO Conference. *Edited by* N.M.G. Barralho, J.S. Pereira, C. Marques, J. Coutinho, M. Madeira, and M. Tomé. Aveiro, Portugal 11–15 October 2004.

Pritchard, J.K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics, **155**(2): 945–59. PMID:10835412.

Rambaut, A. 2014. FigTree v1.4.2. Available from http://tree.bio.ed.ac.uk/software/figtree/ [accessed 18 February 2016].

Ribeiro, M.M., Sanchez, L., Ribeiro, C., Cunha, F., Araújo, J., Borralho, N.M.G., and Marques, C. 2011. A case study of *Eucalyptus globulus* fingerprinting for breeding. Ann. For. Sci. **68**: 701–714. doi:10.1007/s13595-011-0087-x.

Rosenberg, N.A. 2004. DISTRUCT: a program for the graphical display of population structure. Mol. Ecol. Notes, **4**(1): 137–138. doi:10.1046/j.1471-8286.2003.00566.x.

Rousset, F. 2008. GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. Mol. Ecol. Resour. **8**(1): 103–106. doi:10.1111/j.1471-8286.2007.01931.x. PMID:21585727.

Slatkin, M. 1987. Gene flow and the geographic structure of natural populations. Science, **236**(4803): 787–792. doi:10.1126/science.3576198. PMID:3576198.

Stackpole, D.J., Vaillancourt, R.E., Downes, G.M., Harwood, C.E., and Potts, B.M. 2010. Genetic control of kraft pulp yield in *Eucalyptus globulus*. Can. J. For. Res. **40**(5): 917–927. doi:10.1139/X10-035.

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

Steane, D.A., Conod, N., Jones, R.C., Vaillancourt, R.E., and Potts, B.M. 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. doi:10.1007/s11295-005-0028-7.

Thavamanikumar, S., McManus, L.J., Ades, P.K., Bossinger, G., Stackpole, D.J., Kerr, R., et al. 2014. Association mapping for wood quality and growth traits in *Eucalyptus globulus* ssp. *globulus* Labill identifies nine stable marker-trait associations for seven traits. Tree Genet. Genomes, **10**(6): 1661–1678. doi:10.1007/s11295-014-0787-0.

van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M., and Shipley, P. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol. Ecol. Notes, **4**(3): 535–538. doi:10.1111/j.1471-8286.2004.00684.x.

Weir, B., and Cockerham, C. 1984. Estimating F-statistics for the analysis of population structure. Evolution, **38**(6): 1358–1370. doi:10.2307/2408641. [Published by Society for the Study of Evolution.]

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