Population genetic analysis and phylogeny reconstruction in *Eucalyptus* (Myrtaceae) using high-throughput, genome-wide genotyping

Dorothy A. Steane[a,]*, Dean Nicolle[b], Carolina P. Sansaloni[c,d], César D. Petroli[c,d], Jason Carling[e], Andrzej Kilian[e], Alexander A. Myburg[f], Dario Grattapaglia[c,d,g], and René E. Vaillancourt[a]

[a]School of Plant Science and CRC for Forestry, University of Tasmania, Private Bag 55, Hobart, Tasmania 7001, Australia

[b]Currency Creek Arboretum, PO Box 808, Melrose Park, South Australia 5039, Australia

[c]Plant Genetics Laboratory, EMBRAPA Genetic Resources and Biotechnology, EPqB, 70770-910 Brasilia, Brazil

[d]Dep. Cell Biology, Universidade de Brasília, 70910-900 Brasília, DF, Brazil

[e]Diversity Arrays Technology Pty Ltd, 1 Wilf Crane Crescent, Yarralumla, ACT 2600, Australia

[f]Department of Genetics, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, 0002, South Africa

[g]Genomic Sciences Program - Universidade Católica de Brasília - SGAN, 916 modulo B, 70790-160 Brasília – DF, Brazil

*Corresponding author: Dorothy Steane, School of Plant Science, University of Tasmania, Private Bag 55, Hobart, Tasmania 7001 AUSTRALIA. Tel. -61-(0)3-62261828. Fax. -61-(0)3-62262698. Email: Dorothy.Steane@utas.edu.au.

Abstract

A set of over 8000 Diversity Arrays Technology (DArT) markers was tested for its utility in high-resolution population and phylogenetic studies across a range of *Eucalyptus* taxa. Small-scale population studies of *E. camaldulensis*, *E. cladocalyx*, *E. globulus*, *E. grandis*, *E. nitens*, *E. pilularis* and *E. urophylla* demonstrated the potential of genome-wide genotyping with DArT markers to differentiate species, to identify interspecific hybrids and to resolve biogeographic disjunctions within species. The population genetic studies resolved geographically partitioned clusters in *E. camaldulensis*, *E. cladocalyx*, *E. globulus* and *E. urophylla* that were congruent with previous molecular studies. A phylogenetic study of 94 eucalypt species provided results that were largely congruent with traditional taxonomy and ITS-based phylogenies, but provided more resolution within major clades than had been obtained previously. Ascertainment bias (the bias introduced in a phylogeny from using markers developed in a small sample of the taxa that are being studied) was not detected. DArT offers an unprecedented level of resolution for population genetic, phylogenetic and evolutionary studies across the full range of *Eucalyptus* species.

Key Words: Australia, Diversity Arrays Technology, DArT, molecular markers, networks, parsimony, plant systematics

1. Introduction

Eucalyptus is the dominant taxon of many Australian ecosystems from subalpine woodlands, through cool and warm temperate wet and dry forests to tropical savannah (Ladiges 1997). While the genus is easily recognised by characteristic leaf, floral and fruit morphologies, there is a huge range of quantitative variation and homoplasy (convergence/parallelism) in phenotypic characters, both among and within species (Pryor and Johnson 1971, 1981). To further complicate matters, there is incomplete reproductive isolation of morphological species that can produce interspecific hybrids, morphological clines and hybrid swarms (Pryor and Johnson 1971, 1981; Griffin *et al*. 1988), although clines can also be produced by primary differentiation (Holman *et al*. 2003). As a result of these factors, reconstructing the phylogenetic history of *Eucalyptus* species has been problematic for systematists, even with the application of molecular techniques. Eucalypt researchers have tested a range of molecular techniques (see below), but none has proven to be suitable for resolving relationships among closely related species within sections or between closely related sections. A marker system is needed that can resolve species-level relationships; that can be applied to a large number of samples across a broad taxonomic range; and that is relatively cheap. Diversity Arrays Technology (DArT; Jaccoud *et al*. 2001), a massively-parallel, array-based genotyping system, may provide the genome-wide coverage, resolution and throughput to meet these requirements.

Allozymes were the first source of molecular markers in eucalypts (Brown *et al*. 1975). They were used mainly to target population-level questions such as mating system, genetic diversity and population differentiation (reviewed by Moran (1992) and Potts and Wiltshire (1997)). The first phylogenetic study using allozymes in eucalypts was by Burgess and Bell (1983) who examined allozyme frequencies in the intergrading species *E. grandis* and *E.*

saligna (subg. *Symphyomyrtus*, sect. *Latoangulatae*, ser. *Transversae*). In this and other studies (e.g., Cook and Ladiges 1998; House and Bell 1994, 1996; Wright and Ladiges 1997), allozymes provided only low to moderate levels of variation within single species or between closely related species. Although allozymes are relatively cheap and simple, they require optimisation for each study and provide very few polymorphic loci per species and few alleles per locus and, therefore, are not suitable for high-resolution population genetic studies nor for large-scale phylogenetic studies.

DNA-based studies of eucalypts began in the early 1990s (Steane *et al*. 1992) with the then state-of-the-art technology of restriction fragment length polymorphism (RFLP) analysis of chloroplast DNA (cpDNA). Because of the expense, a paucity of markers and the large amount of labour involved, only small numbers of samples and markers were used, providing rather coarse resolution of phylogenetic relationships among higher eucalypt taxa (genera and subgenera; Sale *et al*. 1993, 1996). As DNA analytical methods progressed and became cheaper, fine-scale restriction site analysis of cpDNA was tested as a means to resolve relationships among closely related species within ser. *Viminales* (sect. *Maidenaria*, subgenus *Symphyomyrtus*) (Steane *et al*. 1998). Although this methodology provided improved resolution of clades, it became apparent that cpDNA haplotypes were not species-specific in *Eucalyptus* and hence not useful for phylogenetic resolution at that low taxonomic level (they were, however, useful for studies of phylogeography; e.g., Byrne and Hines 2004, Byrne and Macdonald 2000, Jackson *et al*. 1999, Wheeler and Byrne 2006).

In contrast, RFLP analysis of nuclear loci proved to be effective for many genetic studies within species or complexes of a few closely related species (e.g., Butcher *et al*. 2002; Byrne *et al*. 1998; Byrne 1999; Elliott and Byrne 2003; Elliott and Byrne 2004; Glaubitz *et al*. 2003;

Hines and Byrne 2001; Wheeler *et al*. 2003). Nuclear RFLPs have not, however, yielded any useful phylogenetic data at taxonomic levels higher than the species level within *Eucalyptus*, because of issues associated with homology assessment and increasing risk of character state homoplasy with increasing taxonomic distances. Futhermore, membrane-based RFLP techniques did not lend themselves to studies requiring high-throughput analysis of large numbers of individuals.

In the early 1990s, 5S ribosomal DNA sequence variation was tested for use in phylogenetic resolution of *Eucalyptus* taxa (Udovicic *et al*. 1995), but the approach was only informative at high taxonomic levels (genera and subgenera). In the mid-1990s sequencing technology improved rapidly and by the end of the decade, PCR-based sequencing and automated DNA analysers allowed the production of relatively large and informative sequence data sets, with cost being the main limiting factor to the size of a study. Steane *et al*. (1999; 2002; 2007) and Whittock *et al*. (2003) used sequence data from the internal transcribed spacer (ITS) of the nuclear ribosomal DNA region to explore phylogenetic relationships across all subgenera of *Eucalyptus* and related eucalypt genera (*Corymbia* and *Angophora*). They found that ITS data provided good resolution of sections and higher taxa, but did not contain enough polymorphism to resolve effectively species-level relationships between and within sections. Furthermore, some of the higher-level relationships between eucalypt genera depicted by ITS sequence data caused consternation among the taxonomic community. For example, ITS sequence data, cpDNA RFLPs and chloroplast restriction site data all suggested that *Corymbia* was paraphyletic; this assertion was countered by evidence from other sources such as the external transcribed spacer (ETS) of the nuclear ribosomal DNA region (Parra-O *et al*. 2006), microsatellites (Ochieng *et al*. 2007b) and a pseudogene of ITS (Ochieng *et al*. 2007a). One problem with using sequence data from functional regions of DNA (such as ITS and ETS) comes from the functional constraints imposed on cistrons that might prevent "neutral" change of nucleotides during evolution. Furthermore, there are many copies of ribosomal RNA genes in a genome and this introduces a risk of comparing paralogous loci (Bayly and Ladiges 2007). Despite the limitations of ribosomal and chloroplast DNA for resolution of species-level relationships, Gibbs *et al*. (2009) successfully used ITS, ETS and cpDNA sequence data in combination with morphological characters to resolve relationships among species within subgenus *Eudesmia*. Although none of the data sets in isolation produced a well-resolved phylogeny of the eudesmids there were elements of congruence in a combined analysis that provided the basis of a sound system of subdivision for that subgenus.

Because of complications associated with paralogy in multiple-copy regions of DNA (e.g., nuclear ribosomal DNA), researchers turned to low-copy number nuclear genes for phylogenetic and phylogeographic analyses. McKinnon *et al*. (2005) used the cinnamoyl-CoA reductase (*CCR*) gene to gain insights into the evolutionary history of *E. globulus*. Two highly divergent lineages of the *CCR* gene were identified within *E. globulus*, one of which was also found in 16 other species in subg. *Symphyomyrtus*, sect. *Maidenaria*. The other lineage was unique to *E. globulus* among the *Maidenaria* taxa, but showed homology to *CCR* in *E. saligna* (subg. *Symphyomyrtus,* sect. *Latoangulatae*), suggesting either incomplete lineage sorting or reticulate evolution. Poke *et al*. (2006) investigated this further and found more evidence of intersectional hybridisation in *Eucalyptus*. The authors concluded that using (single-copy, functional) nuclear genes for phylogeny reconstruction of eucalypt taxa would be problematic unless recombination was taken into account.

A genome-wide approach to phylogeny reconstruction, preferably using "neutral" loci (the evolution of which was unconstrained by functional requirements) that could be analysed with a combination of population genetic and phylogenetic approaches, could circumvent complications experienced with single locus analyses in *Eucalyptus*. The development of microsatellite primers for eucalypt taxa (Brondani *et al*. 1998, 2006; Byrne *et al*. 1996; Glaubitz *et al*. 2001; Jones *et al*. 2001; Ottewell *et al*. 2005; Shepherd *et al*. 2006; Steane *et al*. 2001; Thamarus *et al*. 2002) opened the door for reliable genome-wide genotyping of a relatively large number of samples. Microsatellite markers gave researchers the power to examine genetic relationships within and among populations of one (e.g., Butcher *et al*. 2009; Elliott and Byrne 2003; Jones *et al.* 2007; Payn *et al*. 2008; Rathbone *et al*. 2007; Steane *et al*. 2006; Walker *et al*. 2009; see also Byrne 2008 and references therein) or a few closely related species (e.g., Holman *et al*. 2003, Le *et al*. 2009, Shepherd *et al*. 2008, Stokoe *et al*. 2001). While microsatellites were developed initially for mapping and population genetic studies, Ochieng *et al*. (2007b) found them helpful for phylogenetic resolution of eucalypt genera. Microsatellite loci are selected by researchers to be highly polymorphic within species and their use for taxonomic purposes between closely related species is limited by the unreliable transferability of these markers across species boundaries (e.g., see Nevill *et al*. 2008) and by the risk of high levels of homoplasy that might be encountered (e.g., Barkley *et al*. 2009; Curtu *et al*. 2004). Hence, while microsatellites have the potential to provide phylogenetic resolution at high taxonomic levels (between genera) and are very useful for population-level studies within species, they are impractical for phylogenetic reconstruction between taxonomic extremes. Furthermore, combining datasets from different studies can be problematic; all samples need to be scored concurrently (or at least a subset of samples should be common to all studies) in order to ensure consistency of microsatellite bin sizes.

Arbitrarily amplified dominant (AAD) markers such as RAPD (Randomly Amplified Polymorphic DNA), ISSR (inter-simple sequence repeats) and AFLP (amplified fragment length polymorphism) have had limited use in population and phylogenetic studies of *Eucalyptus*. AAD markers have a high potential for homoplasy, so their application to phylogenetic analysis requires careful consideration of heritability, homology and homoplasy. Bussell *et al*. (2005) recommended AAD markers for phylogenetic and systematic studies of closely related species and non-reticulating, subspecific lineages, since below these taxonomic levels, population genetic effects (reticulation) may swamp hierarchical signal in the data, while at higher taxonomic levels homoplasy is likely to be significant. AAD markers have been used to examine genetic structure within and between populations of individual species of *Eucalyptus* (e.g., RAPD - Nesbitt *et al*. 1995; Gaiotto *et al*. 1997, Li 2000; ISSR - Okun *et al*. 2008; AFLP – Gaiotto *et al*. 1997, Poltri *et al*. 2003) but also to examine relationships among closely related (reticulating) species. In accordance with the findings of Bussell *et al*. (2005), McKinnon *et al*. (2008) could not separate closely related species within sect. *Maidenaria* (subg. *Symphyomyrtus*), with AFLP markers, but they were able to resolve series and subseries. However, the task of checking homology and repeatability of 930 AFLP markers across 84 samples was time-consuming (and hence, expensive). When analysing AFLP data, all the data for a particular study need to be scored and binned at the same time; it is not possible to score a number of data sets separately and then combine them, unless the data are checked manually. The transferability of AFLP markers across projects (and laboratories) is also problematic. Clearly, a more robust, high-throughput method for studies of closely related species would be preferable.

We recently developed a set of Diversity Arrays Technology (DArT) markers for *Eucalyptus* (Sansaloni *et al*. 2010) that has the potential to provide a rich source of phylogenetic information across a range of species at various taxonomic levels. DArT markers are highly variable genome-wide binary markers, the diversity of which is derived from restriction site polymorphism (Jaccoud *et al*. 2001). Polymorphism is detected by DNA-DNA hybridization on microarrays, allowing rapid analysis of large numbers of samples through a stream-lined automated production line (see http://www.diversityarrays.com/molecularprincip.html). We developed the markers from 65 species of *Eucalyptus* from across the taxonomic range (see Sansaloni *et al*. 2010) with a view to producing generic markers that would be useful in a large proportion of the 700+ species of the genus. Because DArT marker fragments are cloned (and most have been sequenced and mapped on genetic linkage maps; Petroli *et al*., in prep.), they do not suffer from the issues of homology assessment that exist in anonymous AAD markers. Furthermore, because of the genome-wide coverage of coding and non-coding regions (Petroli *et al*. in prep.), DArT has the potential to provide insights into the regions of the genome that are involved in adaptation, speciation and evolution in *Eucalyptus*.

In this study we test over 8000 DArT markers for their transferability across species and their utility in population genetics and phylogeny reconstruction. Only one other study has explored the utility of DArT markers for studies of evolution in wild populations (James *et al*. 2008), but that study focused on relationships between populations within two species of cryptogam (a fern, *Asplenium viride* and a moss, *Garovaglia elegans* ssp. *dietrichiae*) rather than inter-specific relationships within a genus. James *et al*. (2008) found that DArT markers could be highly informative about relationships among populations of cryptogam species. The present study is the first in which DArT markers have been designed for cross-species applications and applied to genus-wide studies of populations and phylogeny. Our aim in this study was to determine the degree to which DArT data can be used for: (1) studies of differentiation within and between species; (2) hybrid identification; and (3) phylogenetic reconstruction in wild populations of *Eucalyptus*.

2. Materials and Methods

2.1 Genotyping with DArT markers

Three plates (94 samples per plate) of *Eucalyptus* DNA were genotyped with DArT markers. DArT-genotyping was carried out by DArT P/L (http://www.diversityarrays.com) following the procedure described by Sansaloni *et al*. (2010). Genotypes were scored as presence/absence of DArT markers and were formatted as binary matrices. It should be noted that this study took place in 2008 during the development of the DArT marker array for *Eucalyptus* (Sansaloni *et al*. 2010) and before the operational *Eucalyptus* DArT array (comprising 7680 markers) had been designed and become publicly available (mid-2009). Hence, because each plate was involved in a different phase of the DArT marker development process, the suite of DArT markers with which each plate of DNA samples was genotyped differed to some degree (Table 1). Since then, more libraries have been made and screened for polymorphism. Some of the markers that were used in the present study have now been replaced on the array with new markers that decrease redundancy. Despite these changes, we anticipate that results from the final array would be comparable to those reported in this paper, because using subsets of the available markers in this study gave results that were comparable to the other subsets and to the full set of markers (see Results).

2.2 Plant material for species and population differentiation surveys

Two microtitre plates of 94 samples (i.e., 188 samples in total) were genotyped with DArT markers to assess the efficacy of the markers in differentiating (1) species of *Eucalyptus* and (2) populations within a species. Plate 1 included DNA from seven species of *Eucalyptus* and one species of *Corymbia*, a close relative (formerly a subgenus) of *Eucalyptus* and was screened with 7052 DArT markers (see Table 1). Plate 2 included a larger representation of four of the most valuable timber and pulp species, *E. grandis*, *E. urophylla s.l.* (following

Brooker (2000)), *E. globulus* and *E. nitens*; it was screened with 4684 DArT markers (Table 1). The samples came from as divergent a set of provenances for each species as could be obtained. Provenance details for these samples are provided in Appendices A and B, respectively.

2.3 Sampling for phylogenetic study

Ninety-four samples (Table 2) from across the taxonomic range of *Eucalyptus sensu stricto* (following the infrageneric classification of Brooker (2000), excluding *Corymbia* Hill and Johnson (see Hill and Johnson 1995) and *Angophora* Cav. and treating these as genera) were genotyped with 8354 DArT markers. Most DNA samples came from previous phylogenetic studies (McKinnon *et al*. 2008; Steane *et al*. 1999, 2002, 2007; Whittock *et al*. 2003) and the set of DNAs used in Plates 1 and 2 (Table 1), but several fresh leaf samples were collected from Currency Creek Arboretum (South Australia; http://www.dn.com.au/) and a *Eucalyptus* arboretum (SeedEnergy Pty. Ltd., Cambridge, Tasmania).

2.4 DNA extraction

At least 1 µg of DNA was extracted from fresh or frozen leaf tissue, using a CTAB extraction protocol (Doyle and Doyle 1990) with several modifications (McKinnon *et al.* 2004). DNA was resuspended in a low-EDTA TE buffer (0.1 mM EDTA, 10 mM Tris pH 8.0). DArT analysis requires high molecular weight, restrictable DNA and all samples in a plate need to be of a uniform concentration (a total of 500-1000 ng DNA at a concentration of 50-100 ng/µl). The DNA concentration of each sample was measured using a *Pico*fluor™ handheld fluorometer (Turner designs, CA, USA) and checked by running 1 µl of DNA on a 0.8% agarose gel alongside a series of standard concentrations (10, 25, 50, 75 and 100 ng) of undigested λ DNA. The gel was post-stained with GoldView (Guangzhou Geneshun Biotech

Ltd, China) and visualised using a Molecular Imager® GelDoc™ XR imaging system (BioRad Laboratories Inc.). The concentration of every sample was adjusted to approx. 75 ng/µl and re-checked on a 0.8% agarose gel. The quality of every DNA sample was tested by restriction of 2 µl DNA (ca. 150 ng) with a six-cutter enzyme, either *Eco* RV or *Hin*d III (New England Biolabs), according to the recommendations of the manufacturer. Digests were visualised on a 0.8% agarose gel, as described above. DNA preparations that did not digest properly were discarded and the samples were re-extracted.

2.5 Analysis of genetic diversity within and between species

The output of a DArT genotyping comprises "presence/absence" data along with a range of statistics that provide insight into the information content of each marker and the reliability of the data derived from each marker in that particular analysis. The stringency of an analysis can be increased by excluding data on the basis of, for example, "Reproducibility" and/or "Call Rate". As replicated individuals should give identical results, replicated points are expected to fall into the same cluster (i.e., "presence" vs "absence"). "Reproducibility" is a measure of the consistency of scoring technical replicates (2-4 assays per sample per marker). It measures how often the replicates fall into the same cluster. This value tends to be kept above 98.9% in the data sets, but can be adjusted upwards to 100% by eliminating markers with low reproducibility. A complementary measure of reproducibility is "Discordance" and this latter measure is usually included in the data set that is provided by DArT P/L. Discordance expresses the overall variation of scores within the technical replicates (see above). Hence, maximising the Reproducibility or minimising the Discordance will have similar effects on a data set. The Call Rate value is an expression of reliability of the final scores for each marker. It represents the percentage of samples that could be scored as 0 or 1. For Plate 2 of this study, in which there were four species (and hybrids) being genotyped with markers derived predominantly from those same species (but also some from a few other species), the call rate always exceeded 80%. In the analysis of the data from Plates 1 and 2, increasing the stringency of the data by reducing Discordance to 0% and increasing the Call Rate to 95% did not greatly affect the overall results. Hence, all data were included in the analyses presented here.

In order to determine the efficacy of DArT markers in differentiating groups of eucalypts at different taxonomic levels, several types of analysis were undertaken. Analysis of Molecular Variance (AMOVA) and a distance-based Principal Coordinates Analysis (PCoA) were done on Plate 1 data using *GenAlex 6.1* (Peakall and Smouse 2006) to determine how the genetic diversity was partitioned among samples and whether DArT markers could be used to differentiate species, sections and subgenera. *Splitstree4* (Huson 1998; Huson and Bryant 2006) was used to generate relationship networks from Plate 1 and Plate 2 data sets, using the default settings of the software.

To check for potential bias in the number of polymorphic loci generated from DArT markers of different origin, the proportion of markers that gave polymorphic results on Plate 2 (two species from subg. *Symphyomyrtus*, sect. *Maidenaria* and two species from subg. *Symphyomyrtus*, sect. *Latoangulatae*) were calculated for each marker source (i.e., the taxon from which the markers were developed).

2.6 Comparison of ITS sequences of samples in the DArT phylogenetic study and samples in ITS-based phylogenies

ITS sequences were generated from 39 samples (GenBank HM596031-HM596069; Table 2) that had not been included in an ITS-based phylogenetic analysis previously (see Steane *et al*.

1999, 2002, 2007; Whittock *et al*. 2003). ITS sequences from all 94 samples were included in a phylogenetic analysis of all available ITS sequences (from the UTAS database) to confirm that the samples used in the DArT analysis were genetically comparable to other samples of the same species and/or section, and to provide a phylogeny that would be comparable to phylogenies derived from DArT data. Aligned ITS sequences were analysed as described previously (Steane *et al*. 2002), using both parsimony (PAUP*4.0b10; Swofford 1999) and Bayesian (MrBayes 3.1; Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003) methods using both desktop computers and the freely available University of Oslo Bioportal computer (www.bioportal.uio.no). The Bayesian analysis used the TrN+I+G model of nucleotide substitution, with parameters calculated by Modeltest Ver. 3.7 (Posada and Crandall 1998). Two runs of a Bayesian analysis were started from a random tree and were run simultaneously to convergence over 4×10^6 generations, using four incrementally heated Markov chains, employing the default heating values. The Markov chains were sampled each 100th generation, yielding 40001 trees, of which the first 10000 were discarded as "burnin". The remaining sample points were used to generate a consensus tree.

2.7 Phylogenetic analysis of DArT data

Bayesian and cladistic approaches were taken to the phylogenetic analysis of DArT data using both desktop computers and the freely available University of Oslo Bioportal computer (www.bioportal.uio.no).

The Bayesian analysis, using MrBayes 3.1 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003), adopted a restriction site (binary) model of evolution. An analysis that was conducted as described above for the ITS sequence data failed to reach convergence, so the analysis was modified so that the temperature was lowered from the default temperature of 0.2 to 0.1, and the number of swaps between chains was increased from the default value of 1 to 2 per swapping event. However, even with these modifications the parallel runs failed to converge. Bayesian analysis of DArT data was abandoned at this stage.

Maximum parsimony analysis of the DArT data set (using PAUP*4.0; Swofford 1999) comprised a heuristic search using 10 000 replicates of a random stepwise addition sequence, TBR branch swapping and steepest descent in effect. Bootstrapping comprised 1×10^6 replicates of the “fast step-wise” algorithm.

The DArT markers used in the phylogenetic study were derived from seven species, four sections and two subgenera of *Eucalyptus* (Table 3). The characters in the phylogenetic data set were divided into subsets of DArT markers that had been derived from particular taxa (e.g., subgenus *Symphyomyrtus* sections *Maidenaria* and *Latoangulatae*; subgenus *Symphyomyrtus* and subgenus *Eucalyptus*). The Incongruence Length Difference (ILD) test of Farris *et al*. (1994; as implemented in PAUP*4.0b10 (Swofford 1999) as the partition homogeneity test) was used to test whether each partition produced results that differed from results generated by random partitions of the data, and therefore whether such partitions would be expected to influence phylogenetic outcomes. One thousand replicates of the ILD test were conducted using a heuristic search strategy. The heuristic search used 10 replicates of random (stepwise) addition sequence and TBR branch swapping. The partition homogeneity test for the markers derived from subgenus *Eucalyptus* was carried out using the freely available University of Oslo Bioportal (www.bioportal.uio.no).

We also tested whether results changed when we excluded less stringent data (e.g., data that showed >0% discordance; data with call rates less than 90% or less than 95%) or when we genotyped the samples with markers derived from a particular taxon. The DArT characters were partitioned into nine subsets. Each subset was used in a *Splitstree4* analysis of the 94 species in the phylogenetic study:

1. No Discordance (i.e. excluding all characters with discordance value >0%) (7536 characters included)
2. Call rate ≥90% (i.e., excluding all characters with a call rate less than 90%) (6232 characters included)
3. Call rate ≥95% (i.e., excluding all characters with a call rate less than 95%) (4094 characters included)
4. No Discordance, Call rate ≥90% (5937 characters included)
5. No Discordance, Call rate ≥95% (3996 characters included)
6. Subg. *Symphyomyrtus* markers only (7899 characters included)
7. Subg. *Eucalyptus* markers only (455 characters included)
8. Subgenus *Symphyomyrtus*, section *Maidenaria* markers only (3615 characters included)
9. Subgenus *Symphyomyrtus*, section *Latoangulatae* markers only (2973 characters included)

Resulting cladograms were compared to the results derived from analysis of the full DArT data set (8354 characters) to detect differences in topologies. For subsets 8 and 9, special attention was paid to the relationships among taxa in sects. *Maidenaria* and *Latoangulatae*.

3. Results

3.1 Call rate, repeatability, polymorphism and missing data

The proportion of data missing from each sample in each of the three plates was quantified (Table 4). In Plate 1 greater than 95% of the markers could be scored unambiguously (i.e., there were fewer than 5% missing data) in 78% of the samples; all samples had greater than 90% scorable markers. In Plate 2, all samples had fewer than 15% missing data (i.e., greater than 85% unambiguously scored markers). In the phylogeny study, 95% of all samples had more than 90% unambiguously scored markers and 98% of the samples had more than 80% unambiguously scored markers. In the phylogeny study, one sample, *E. gamophylla*, had 24% missing data (reason not apparent); this might account for the anomalous position of *E. gamophylla* as sister to the rest of subgenus *Eudesmia* (Fig. 1; and see Gibbs *et al*. 2009). Overall the proportion of missing data for each plate was low, ranging from 4.4% in Plate 1 to 6.0% in Plate 2.

Table 5 shows the numbers of DArT markers from each marker source that were polymorphic in Plate 2 samples from subg. *Symphyomyrtus*, sect. *Maidenaria* (*E. globulus* and *E. nitens*) and compares them to the polymorphism observed in subg. *Symphyomyrtus*, sect. *Latoangulatae* (*E. grandis* and *E. urophylla*). There was a 10% difference overall in the proportion of markers that were polymorphic in the two sections, with the balance tipped in favour of sect. *Latoangulatae* (84% compared to 74% in sect. *Maidenaria*). However, the call rate (i.e., the proportion of samples for which the DArT markers provided scorable data) tended to be a little higher for sect. *Maidenaria* than for sect. *Latoangulatae*.

3.2 Differentiation within and between species

Early on in the development of the DArT markers it became clear that there was poor transferability of DArT markers between *Eucalyptus* and *Corymbia* and, because our focus was on *Eucalyptus*, we abandoned the development of DArT markers for *Corymbia* (Sansaloni *et al*., 2010). Hence, no results for *Corymbia* are reported.

An AMOVA analysis of the seven *Eucalyptus* species (up to 12 samples per species) in Plate 1 indicated that 36% of the DArT variation occurred within species and 64% occurred among species. The three principal coordinates in a PCoA of Plate 1 data (Fig. 2) explained 91.9% of the variation among samples and provided complete separation of each species, with *E. pilularis* (subgenus *Eucalyptus*, sect. *Pseudophloius*) being well-separated in PCo1 and PCo2 from a group of species from subgenus *Symphyomytus* (i.e., *E. globulus*, *E. nitens*, *E. grandis*, *E. urophylla*, *E. camaldulensis* and *E. cladocalyx*). The spatial distribution of the species from subgenus *Symphyomyrtus* is congruent with taxonomic relationships (see Steane *et al*. 2002, 2007): *E. camaldulensis* (sect. *Exsertaria*) clusters with *E. grandis* + *E. urophylla* (sect. *Latoangulatae*); *E. globulus* + *E. nitens* form a separate "*Maidenaria*" cluster.

Examination of single species in isolation demonstrated the potential of DArT to identify the provenance of individual samples. For example, DArT markers resolved population-level relationships in *E. globulus*. Figure 3 shows a *Splitstree4* network derived from Plate 2 data. Although networks generated by *Splitstree4* are somewhat complex to look at, they do impart an understanding of the complexity of DArT data and provide a useful summary of agreement and conflict among the data. Huson and Bryant (2006) provided a clear explanation of how to interpret network figures. Briefly, the parallel lines represent two-way splits in the data. If you cut the figure across the parallel lines you can visualise to which of two groups each sample belongs. The longer the line associated with a split, the more evidence there is to support that split. This method of depicting the data is an effective way of expressing the considerable homoplasy (character conflict) in DArT data sets. The network in Fig. 3 shows the geographic differentiation of Victorian, Furneaux, Eastern Tasmanian and “Western Link” provenances of *E. globulus*, results that are congruent with microsatellite studies (Steane *et al*. 2006). Also in agreement with other population genetic studies, geographic structuring could be seen in *E. camaldulensis*, *E. urophylla* and *E. cladocalyx* (Supplementary material Fig. S1). *Eucalyptus camaldulensis* samples from Queensland (probably mostly ssp. *acuta* but possibly also one or two samples of ssp. *simulata*; McDonald *et al*. 2009; David Lee, DEEDI, Queensland, pers. comm.) and Victoria (ssp. *camaldulensis*; McDonald *et al*. 2009) formed separate clusters (Supplementary material Fig. S1A), supporting the microsatellite results (Butcher *et al*. 2009) and subspecific taxonomy of the species (McDonald *et al.* 2009). Samples of *E. urophylla* clustered according to their Indonesian island of origin (Supplementary Material Fig. S1B), in agreement with the microsatellite study of Payn *et al*. (2008). Geographic partitioning of DArT variation in *E. cladocalyx* (Supplementary Material Fig. S1C), a species with several disjunct populations in South Australia, was similar to results obtained by McDonald *et al*. (2003) in an allozyme analysis of the species.

3.3 Hybrid identification

Results from the screening of the four commercially important species in Plate 2 (*E. globulus*, *E. nitens*, *E. grandis* and *E. urophylla*) demonstrated the potential of DArT markers to identify plants of hybrid origin. Figure 4 shows the intermediate position of *E. nitens* x *globulus* hybrids between the parent species. In contrast, hybrid progeny of *E. urophylla* x *grandis* emerged within the *E. urophylla* cluster, rather than between the two species clusters. Without detailed pedigree information about the parent plants, it is impossible to say whether this is a DArT artefact or whether the *E. urophylla* parent of these hybrids was, for example, an F1 or backcross hybrid, rather than pure *E. urophylla*. Such situations can occur in breeding populations. For example, one *E. grandis* sample (labelled on Fig. 4 as "putative hybrid, South Africa") was originally believed to be pure *E. grandis*. However, when the DArT results showed this tree to be an outlier relative to the other *E. grandis* samples, the tree was re-examined morphologically and atypical juvenile epicormic foliage demonstrated that the tree was, in fact, not pure *E. grandis* after all and probably originated from a *E. grandis* x *E. urophylla* hybrid. Similarly, a sample of *E. urophylla* clustered with this putative hybrid, raising concern that this sample may not be pure *E. urophylla*. Unfortunately, pedigree data for these samples were not available.

3.4 Phylogenetic analysis

3.4.1 ITS sequence data

ITS sequences were generated for all samples that had not been included in previous ITS-based phylogenetic analyses (Table 2). These sequences were added to the existing database, yielding a dataset comprising 140 operational taxonomic units and 680 characters (659 aligned nucleotide characters and 21 indels scored as presence/absence data). Maximum parsimony (MP; Supplementary material Fig. S2) and Bayesian analyses (not shown) showed that all new samples but one clustered in an appropriate clade, i.e., alongside samples of the same species or section. The ITS sequence from the sample of *E. grandis* (subg. *Symphyomyrtus*, sect. *Latoangulatae*) that had been included in the phylogenetic plate of the DArT study did not cluster with the existing three samples of that species (in Clade B; see Steane *et al*. 2002), but emerged in a clade with sect. *Maidenaria* (Clade C; Steane *et al*. 2002) (see Supplementary Material Fig. S2A). Despite this result, the sample of *E. grandis* (from a South African trial of Australian provenances) was retained in the DArT analysis for comparative purposes.

3.4.2 Checking the robustness of subsets of DArT data

The final DArT data set for the phylogeny analysis comprised 94 taxa and 8354 binary characters. Figure 1 shows the strict consensus of ten trees derived from cladistic analysis of the full data set. Each subset of DArT data of different technical robustness yielded similar results. Obviously, exclusion of characters resulted in shorter branch lengths in the trees (less support for some clades), but the relationships inferred from the different character sets remained reasonably stable. Exclusion of different sets did affect some of the finer details within clades, but the changes usually affected the same taxa in each instance (i.e., in subgenus *Symphyomyrtus*, *E. hallii* (sect. *Exsertaria*, monotypic ser. *Connexentes*), *E. dundasii* (sect. *Bisectae*, monotypic ser. *Dundasianae*), *E. cladocalyx* (monotypic sect. *Sejunctae*), *E. pumila* (monotypic sect. *Pumilo*) and species of sect. *Latoangulatae* relative to species of sect. *Exsertaria*; in subgenus *Eudesmia*, *E. gamophylla* (ULm); and in subgenus *Eucalyptus, E. nitida*).

The overall topology of the phylogeny did not change when only DArT markers of *Symphyomyrtus* origin were used in an analysis. Using markers derived from only sect. *Maidenaria* or sect. *Latoangulatae* did not affect the general results, and at the fine scale affected the positioning of the more “mobile” taxa only (e.g., the positioning of species of sect. *Latoangulatae* and *E. hallii* (SE) relative to sect. *Exsertaria*, *E. dundasii* (SB1) and *E. gamophylla* (ULm); see comment above). In contrast, using only the markers derived from subgenus *Eucalyptus* yielded a less well resolved phylogeny with some differences in topology when compared to the results from the full set of characters. However, considering that the number of characters was reduced to 5.4% of the full complement, the changes were relatively minor (results not shown). The critical influence here may be the sheer number of data that were excluded (i.e., 95% when only markers derived from subgenus *Eucalyptus* were included). The ILD test of Farris *et al*. (1994) was used to test the null hypothesis that a chosen partition (e.g., the markers derived from subgenus *Eucalyptus*) would not generate results that differed from any similar-sized random subset of the data. The alternative hypothesis in the test was that the partition was not random and that it generated results (in this case, phylogenies) that differed from results generated by similar-sized random subsets of the data. When the DArT markers that were derived from subgenus *Eucalyptus* were nominated as the "partition", the ILD test returned a P value of 0.43, indicating that this set of markers contained the same phylogenetic signal as a similar-sized random subset of markers. Hence, the different tree topologies that were generated by the "subg. *Eucalyptus* (monocalypt) markers only" subset of markers compared to those generated from using "*Symphyomyrtus* markers only" were probably a result of the size of the data partition rather than anything else.

Unlike the ILD test for the markers derived from subgenus *Eucalyptus* (above), ILD tests of sect. *Maidenaria* and sect. *Latoangulatae* did not yield non-significant results (Table 6). The results suggest that the source of the markers from these sections may affect the topology of inferred phylogenies. However, this result conflicts with results from such phylogenetic analyses (above), suggesting that the ILD test may be flawed (see Discussion).

3.4.3 Bayesian and maximum parsimony analyses of DArT data

Despite extensive searches that took up to two weeks (or up to 40 hours on the University of Oslo Bioportal facility), parallel runs of the Bayesian analyses failed to reach convergence, so this method was abandoned. This appears to be a common problem with Bayesian analysis of data sets comprising large numbers of taxa (see discussion at http://treethinkers.blogspot.com/2009/04/when-mrbayes-fails.html; viewed January 2011). A lack of convergence may also be a result of conflicting phylogenetic signals within a data set (see Mossel and Vigoda 2006) arising either (i) from different genes or genomic regions that have different phylogenetic histories or (ii) as a result of interspecific hybridization.

MP analysis (Fig. 1) of the complete DArT data set yielded ten equally most parsimonious trees (length = 74861; consistency index, CI=0.112; retention index, RI=0.624) with a strict consensus topology comparable to topologies derived from ITS sequence data in this (see Supplementary Material Fig. S2) and previous studies. Because CI is negatively correlated with the number of taxa and characters in an analysis (Forey *et al*. 1992), it is not surprising that the CI value was so low. This does not necessarily mean that homoplasy was particularly problematic, since the retention index (RI) indicated that 62% of the similarities on the tree were synapomorphic (Farris 1989). Although the resolution among taxa was high, bootstrap values revealed that many of the relationships depicted in the trees had low statistical support. Reweighting characters on the basis of CI or RI did not greatly alter the gross topology of the MP strict consensus trees derived from the DArT data, but resulted in a loss of resolution within and between the main clades (data not shown).

Because of the limited sampling in this analysis, we were hesitant to draw many inferences from the fine topological details of the MP strict consensus cladogram. In general, the DArT analyses (of the whole data set or of subsets of taxa) produced trees that were congruent with existing phylogenies from DNA sequence data (Steane *et al*. 1999, 2002, 2007; Gibbs *et al*. 2009), SSR-based population studies (e.g., *E. globulus* species complex; Jones 2009), AFLP studies (McKinnon *et al*. 2008) and morphology-based classifications (e.g., Brooker 2000). The four main clades (A-D) of subgenus *Symphyomyrtus* that are always found in ITS analyses were apparent. The main difference was the lack of differentiation of Clades B and

C and the sister relationship of Clades A and D in the DArT-based analysis (compare Fig. 1 with Supplementary Material Fig. S2). Even with the exclusion from the MP analysis of the aberrant sample of *E. grandis* (sect. *Latoangulatae*), the position of *E. urophylla* (sect. *Latoangulatae*) relative to sect. *Maidenaria* (Clade C) and the other representative of sect. *Latoangulatae* (*E. biterranea*) that clustered with sect. *Exsertaria* (Clade B) remained unresolved (results not shown). Also in contrast to ITS-based analyses, *E. cladocalyx* (monotypic sect. *Sejunctae*) formed the sister group to the rest of Clade A, rather than being embedded within Clade A.

DArT data provided more resolution than ITS sequence data within subgenus *Eucalyptus*. Some closely related species formed clades (e.g., *E. regnans* and *E. obliqua* from sect. *Eucalyptus*; *E. delegatensis* and *E. pauciflora* from sect. *Cineraceae; E. pulchella* and *E. risdonii* from sect. *Aromatica*, ser. *Insulanae*), but many of Brooker's (2000) sections were not found to be monophyletic. The positions of subgenus *Idiogenes* as sister to subgenus *Eucalyptus* and of subgenus *Primitiva* embedded within subgenus *Eucalyptus* are congruent with previous studies based on nuclear ribosomal DNA sequence data (Steane *et al*. 1999, 2002; Ladiges *et al*. 2010).

Subgenus *Eudesmia* formed a clade that was well-supported apart from the positioning of *E. gamophylla* at the base of the clade (Fig. 1), a position that was incongruent with other data (Gibbs *et al*. 2009) and probably due to a high level of missing data for that species. Monotypic subgenus *Alveolata* (*E. microcorys*) emerged as sister to subgenera *Symphyomyrtus* and *Minutifructus*; monotypic subgenus *Cuboidea* (*E. tenuipes*) was sister to all other eucalypts except for *E. curtisii* (monotypic subgenus *Acerosae*) which was used as the outgroup in this analysis (Fig. 1).

4. Discussion

4.1 Application of DArT to studies of genetic differentiation within and between species

Rigorous statistical analysis of populations requires a large number of individuals and/or large numbers of characters that sample genetic diversity throughout the genome. Although attempts have been made to develop high-throughput marker systems, it has been difficult to generate large numbers of polymorphic markers that can be applied efficiently to large numbers of samples using techniques such as RFLP (e.g., Byrne *et al*. 1998; Butcher *et al*. 2002), RAPD (Nesbitt *et al*. 1995) and even microsatellites (Steane *et al*. 2001; Brondani *et al*. 1998, 2006). AFLP markers can be scaled up to produce a high-throughput system (Myburg *et al*. 2001), but even with these relatively abundant markers, developing large numbers of reliable, high-quality polymorphic markers is laborious and expensive due to the requirement for gel or capillary electrophoresis. The development of an automated DArT marker system for use across many species of *Eucalyptus* allows rapid, high-throughput whole-genome analysis of numerous individuals across a wide range of species.

In the development of the DArT markers, although we focussed on species of commercial importance, we aimed to produce an array that would provide polymorphic markers for use in all species of *Eucalyptus*. Most of the markers were developed from four commercially important species in two sections of subgenus *Symphyomyrtus*: *E. grandis* and *E. urophylla* from sect. *Latoangulatae*, and *E. globulus* and *E. nitens* from sect. *Maidenaria*. In addition, substantial numbers of markers were developed from *Corymbia variegata, E. camaldulensis* (subgenus *Symphyomyrtus*, sect. *Exsertaria*) and *E. pilularis* (subgenus *Eucalyptus*). Libraries were also created from 64 DNA samples from the phylogeny plate, so there is a small set of (ca. 700) markers derived from a mixture of DNAs representing the full taxonomic range of *Eucalyptus* (see Sansaloni *et al*., 2010). We explored the possibility that markers derived from one taxon would be more polymorphic in that taxon than in another (the so-called "ascertainment bias" (Clark *et al*. 2005)). We found a very slight bias towards higher polymorphism in sect. *Latoangulatae* than in sect. *Maidenaria* when using markers derived from sect. *Latoangulatae*, but since a slight bias was also observed in sect. *Latoangulatae* with markers derived from other sections (e.g., sect. *Exsertaria* and a range of taxa from phylogeny plate; Table 5) this may have been an artefact of how the markers were selected (e.g., a slight bias towards selecting markers that were polymorphic in section *Latoangulatae*) rather than an intrinsic taxonomic bias. Thus, there was no evidence of strong ascertainment bias in the eucalypt DArT array. The level of DArT variation within populations and the application of DArT to population-level studies were surveyed in just a small number of species from which most of the markers were developed (*E. cladocalyx*, *E. globulus, E. nitens, E. grandis, E. urophylla* and *E. pilularis*). Although we feel that the success of these studies is a good indication of their wider applicability, further studies involving larger numbers of individuals per population and species will be required to determine the utility of the markers for fine-scale population studies in other taxa.

Each individual tree in our study had a unique genotype, as would be expected from a genome-wide fingerprint comprising thousands of markers. The fact that polymorphism of DArT markers relies primarily on restriction site polymorphism (in most cases polymorphisms come from single nucleotide mutations in restriction sites) means that DArT markers are highly heritable and can be traced readily in pedigrees (Sansaloni *et al*., 2010). The sample sizes for each species in this study were small, but where known geographic partitioning existed among samples of a species, DArT markers identified this partitioning. For example, DArT markers resolved population-level relationships in *E. globulus* that were consistent with results from previous molecular analyses: the DArT-based clusters of Victorian, Furneaux, Eastern Tasmanian and “Western Link” provenances were supported by both nuclear microsatellite data (Steane *et al*. 2006) and chloroplast DNA data (Freeman *et al*. 2001). Geographic structuring was also observed in *E. camaldulensis*, *E. urophylla* and *E. cladocalyx* (Supplementary material, Fig. S1). *Eucalyptus camaldulensis* is widespread across mainland Australia and has significant geographical variation; McDonald *et al*. (2009) recognised seven infraspecific taxa, but sampling for this study included samples only from Victoria (ssp. *camaldulensis*) and Queensland (probably all ssp. *acuta*) and these formed discrete geographic clusters in *Splitstree4* analyses. This result is congruent with results from microsatellite data that showed distinct geographic clustering of *E. camaldulensis* populations across the full distribution of the species (Butcher *et al*. 2009). Timor mountain gum, *Eucalyptus urophylla s.l.,* is a tropical species comprising a limited number of disjunct populations located on volcanic soils on seven of the Lesser Sunda Islands in eastern Indonesia. DArT results derived from Plate 2 showed distinct geographic clustering of samples from several islands (i.e., Flores, Wetar and Timor), although a few samples from Lembata (previously Lomblen) did not cluster tightly. These results are congruent with those derived from microsatellite data by Payn *et al*. (2008), who detected subtle island-based geographic structuring of *E. urophylla* within a highly homogeneous gene pool. The sugar gum, *E. cladocalyx*, grows naturally in three disparate regions in South Australia (Kangaroo Island, southern Eyre Peninsula and the southern Flinders Ranges) and displays significant partitioning of allozyme (McDonald *et al*. 2003) and DArT variation. In other species where our sampling was more-or-less continuous across part or all of the species' distribution, geographic partitioning was not observed (i.e., *E. grandis*, *E. nitens* and *E. pilularis*). This seems surprising in the case of *E. nitens*, where geographic partitioning of genetic variation has been reported previously (Byrne *et al*. 1998), but our sampling of *E. nitens* in this study (only one or two samples per locality; see Appendices A and B) may have been insufficient to detect such diversity partitioning. A larger sample size with multiple samples from more localities might yield more definitive results. *Eucalyptus grandis* provides an interesting counterpoise. In this species that has a long and, in places, disjunct latitudinal distribution from far northern Queensland to mid-coast northern New South Wales, one might expect evidence of geographic partitioning of molecular genetic variation. However, such partitioning was not detected either with allozymes (Burgess and Bell 1983) or with cpDNA sequence data (Jones *et al*. 2006), lending support to the otherwise perplexing (negative) results of the DArT study. Hence, DArT markers have the potential to be a powerful tool for detecting the geographic substructuring of genetic variation within *Eucalyptus* species. However, caution should be exercised if using DArT for estimating genetic diversity and inbreeding parameters (e.g., *F* statistics). Dominant markers are generally not considered ideal for such studies, but there are algorithms - such as those in AFLP-SURV (Vekemans 2002) - that allow for calculations of these statistics from dominant markers.

In contrast to other molecular marker systems (including chloroplast and ribosomal DNA sequence data), DArT markers were useful for differentiating closely related species and, to some degree, closely related sections. Hence, DArT could play a role in species identification, especially if taxon-specific markers or suites of markers were identified and incorporated onto a single "taxonomy" array. However, DArT would not be a practical tool to use for "DNA barcoding" (see Kress and Erickson 2008) of plants generally, since DArT arrays would need to be developed for all plant groups (species or genera) and this would be prohibitively expensive and time-consuming.

4.2 Application of DArT to hybrid identification and studies of introgression

Due to the genome-wide sampling of thousands of marker loci, DArT has the potential to identify interspecific hybrid material. We observed that known *E. nitens* x *globulus* hybrids fell between clusters of the two parent species when genotyped with DArT markers. However, more rigorous studies of hybrids are required to test further the behaviour of DArT markers in hybrids. Progenies of a *E. urophylla* x *grandis* cross that were genotyped in this study did not yield DArT profiles that were intermediate between the two parents, contrary to what might be expected. This might be a result of using non-pure parental material of *E. urophylla* (e.g., the parent might have been a hybrid between *E. urophylla* and a species other than *E. grandis*), or there may be directional segregation distortion of DArT markers occurring in the hybrid progeny that favours *E. urophylla* alleles over *E. grandis* alleles. Large, targeted pedigree studies are required to determine how DArT markers behave in such pedigrees.

Problems with species identification may sometimes arise when historical hybridisation has left traces of one species in morphologically pure material of another species. DArT could be used as a tool to investigate such ‘reticulate evolution’ among closely related species (e.g., *E. cordata* and *E. globulus*; McKinnon *et al.* 2004, 2008). It may also be used to examine historical introgression among less-closely related taxa, such as taxa in different series or sections. Previous research using the cinnamoyl CoA reductase (*CCR*) gene (Poke *et al.* 2006) demonstrated historic recombination among the genomes of sects. *Latoangulatae*, *Exsertaria* and *Maidenaria* (subg. *Symphyomyrtus*). The sample of *E. grandis* included in the phylogenetic analysis in this study appeared to be intermediate in its DArT profile between sect. *Maidenaria* and other representatives of sect. *Latoangulatae*. Hence, this could be an example of historic genetic recombination among closely related sections, a topic that will be examined in a future study.

4.3 Application of DArT to phylogeny reconstruction

Traditionally DNA-based phylogenetic analyses of plants have utilised sequence data from one or several small region(s) of the genome, for example, chloroplast DNA, the ITS and/or ETS region(s) of nuclear ribosomal DNA, or sundry single copy nuclear genes. More often now, with increasingly economical DNA sequencing technologies, several regions of the genome are combined into a single analysis, but these still represent just a small proportion of the whole genome. There has been considerable debate about how well single-gene phylogenies reflect species phylogenies (see Liu *et al*. 2009 and references therein) and researchers have often lamented the lack of an efficient method of whole-genome phylogeny reconstruction. To overcome this issue, genome-wide marker systems such as microsatellites (e.g., Ochieng *et al*. 2007b; Eggert *et al*. 2009) and AFLP (e.g., McKinnon *et al*. 2008, Perrie and Shepherd 2009) have been used for phylogenetic reconstruction (usually in studies of quite closely related species), but these systems have their limitations in terms of labour, numbers of markers, cost of marker development, transferability of markers between laboratories, the ease with which different (linked) data sets can be combined, time-consuming analysis and hierarchical level at which they are effective.

This study was the first to use DArT data to examine phylogenetic relationships among species from across a large and diverse taxonomic group. Two issues concerning the use of the markers for phylogeny reconstruction were the transferability of the markers (i.e. the

DArT polymorphisms) across species and the potential for biased results from markers that were developed from one taxon and applied to phylogeny reconstruction in another (ascertainment bias).

Transferability of the markers among species of *Eucalyptus s.s*. was generally good, with polymorphic markers being available for all species; transferability between closely related taxa *Eucalyptus* and *Corymbia* was poor.

We tested a number of subsets of DArT markers (that varied either in technical reproducibility or taxonomic origin) to see whether the results obtained in different groups were markedly different; on the whole, the phylogenies did not change much. We also used the ILD test (Farris *et al*. 1994) to determine whether a subset of markers from a particular taxon (e.g., all markers from sect. *Maidenaria*) would yield phylogenetic patterns that differed from patterns generated by similar-sized random subsets of the full complement of markers in the data set. While the subsets of markers from subgenus *Eucalyptus* (the monocalypts) and sect. *Exsertaria* did not appear to be biased, we found positive suggestions of bias for sects. *Latoangulatae* and *Maidenaria* even though actual phylogenetic analyses showed no such effect. Ramirez (2006) reviewed numerous problems associated with the ILD test, including the fact that highly significant ILD values can be obtained when there is homoplasy in one of the data partitions and there are characters that are irrelevant to the groups-in-conflict in the other data partition. This may well be the case with DArT data where the proportion of DArT markers (derived from a particular taxon) that provide phylogenetically useful information might decrease as the taxonomic distance (between the source of the DArT marker and the taxon being genotyped) increases. For example, markers from sect. *Latoangulatae* might be more likely to be phylogenetically informative within sect. *Latoangulatae* than outside that group. However, when subsets of the markers (e.g., markers derived from sect. *Latoangulatae* or sect. *Maidenaria*) were used in phylogenetic analyses, the overall topology of the phylogenetic trees did not change greatly, suggesting that the DArT markers are informative across the full taxonomic range of *Eucalyptus* and not just in taxa that are close to the source of the markers.

The results of the DArT analyses of higher-level taxa (subgenera, sections, clades A-D within subgenus *Symphyomyrtus*) were largely concordant with those generated from ITS sequence data, regardless of which analytical method was used for the DArT data. For example, the position of subgenus *Minutifructus* within subgenus *Symphyomyrtus* (Whittock *et al*. 2003) was supported by the DArT data, as were the close relationships among subgenera *Eucalyptus*, *Idiogenes* and *Primitiva* (Fig. 1). At the species level, there were a few "mobile" samples (see Section 3.4.2) that tended to move around depending on which partition of data or which analytical method was used, but these were mostly taxa whose positions were unresolved in ITS analyses as well. Because of the sparse sampling across the genus for this study, opportunities for direct comparisons of species-level phylogenies derived from the DArT analysis and other studies are limited. There is one published study of AFLP variation across endemic Tasmanian species from subg. *Symphyomyrtus* sect. *Maidenaria* (McKinnon *et al*. 2008) that can be compared to the DArT data set. McKinnon *et al*. (2008) assayed 84 samples across 21 species and found that within subsect. *Euryotae*, the *E. globulus* complex (*E. globulus*, *E. bicostata*, *E. pseudoglobulus* and *E. maidenii*; ser. *Globulares*) formed a distinct group and its putative sister-species, *E. nitens* (ser. *Globulares*), was an outlier; *E. perriniana* (ser. *Orbiculares*) clustered with most species from ser. *Viminales* (e.g., *E. rubida* and *E. viminalis*); and the boundary between other species from ser. *Orbiculares* (e.g., Tasmanian endemics *E. cordata*, *E. morrisbyi* and *E. gunnii*) and subsect. *Triangulares* ser.

Foveolatae (e.g., *E. ovata*) was blurred. The DArT study included a small subset (13) of the samples used by McKinnon *et al*. (2008), as well as some additional samples from sect. *Maidenaria* from mainland Australia representing ser. *Orbiculares* (*E. glaucescens*) and ser. *Bridgesiana* (*E. dunnii*) from subsect. *Euryote,* and ser. *Microcarpae* from subsect. *Triangulares*. The DArT study reinforced the observation (McKinnon *et al*. 2008) that *E. nitens* is not closely related to other species in sect. *Maidenaria* ser. *Globulares* (Fig. 1). Furthermore, both the AFLP study and the DArT study found that the Tasmanian species belonging to ser. *Orbiculares* (e.g., *E morrisbyi*, *E. gunnii* and *E. cordata*) formed a cluster that was distinct from mainland representatives of ser. *Orbiculares* (e.g., *E. glaucescens*, *E. pulverulenta*) and the latter tended to cluster with Tasmanian species from ser. *Viminales* (Fig. 1). *Eucalyptus perrininana* (ser. *Orbiculares*) grows on both the island of Tasmania and mainland Australia. The sample of *E. perriniana* in this study was from mainland Australia and grouped with the other mainland samples of ser. *Orbiculares*. However, when the other mainland samples of ser. *Orbiculares* were omitted from the analysis, *E. perrininana* clustered with the Tasmanian species from that series. The AFLP and the DArT studies both suggested that the boundary between series within section *Maidenaria* are blurred. This may be a biological phenomenon or could reflect a lack of resolution provided by the two marker systems. Much denser sampling within sect. *Maidenaria* would be required for the DArT results to be convincing.

The phylogenetic analysis presented in this study represented only about 12% of all the species of *Eucalyptus* from across the species range. Analysis of the 94 species together gave fine-scale (species-level) topologies that differed slightly from those obtained when different subsets of the taxa were analysed independently (e.g., subg. *Symphyomyrtus* sect. *Maidenaria*; results not shown), most likely because of a reduction in the level of homoplasy among the characters across the taxa. Future phylogenetic studies based on DArT will employ high-density sampling of species within small taxonomic units, for example, subgenus *Eudesmia* (18 species), the genetic clusters identified in this and other studies (e.g., Clades A-D) and individual sections (e.g., sect. *Maidenaria*). The inclusion of multiple samples of closely related species in a small taxonomic unit may help to increase the accuracy of a DArT-based interpretation of species-level evolution.

5. Conclusion

Our studies have shown that the DArT markers developed for *Eucalyptus* have great potential for studies of (1) genetic differentiation within and among species; (2) hybridisation and introgression; and (3) phylogeny reconstruction at many taxonomic levels within *Eucalyptus*. One of the most appealing aspects of DArT markers is that they are cloned, which means that issues of homology assessment are negligible. Many of the markers have been sequenced and mapped onto linkage maps (e.g., Sansaloni *et al.*, 2010) and soon will be traced onto the completed *Eucalyptus* genome sequence (http://eucalyptusdb.bi.up.ac.za/). Searches on GenBank have indicated that at least 30% of the markers come from coding regions of the genome. Hence, in future studies we will be able to divide our data sets into markers from coding and non-coding regions, allowing us to compare phylogenies derived from regions of the genome that are under selection and regions of the genome that are (presumably) selectively neutral. We also hope to be able to identify mutations in regions of the genome that are diagnostic for particular taxa and that exhibit segregation distortion in F2 hybrid progeny, that might provide insight into genomic regions that are linked to speciation and postzygotic isolation. DArT markers in combination with the complete *Eucalyptus* genome sequence hold much promise for breakthroughs in the understanding of evolution and speciation in this complex genus.

Acknowledgements

We would like to thank the following individuals and organisations for their assistance with providing plant material or DNA: Dean Williams (Forestry Tasmania), Chris Harwood (CSIRO Sustainable Ecosystems), Peter Gore and David Boomsma (SeedEnergy), Merv Shepherd and Tim Sexton (Southern Cross University, NSW), Luke McManus (University of Melbourne), Minique de Castro (University of Pretoria, South Africa), David Lee (Department of Primary Industry, Queensland), Kelsey Joyce (Gunns Ltd.), Christophe Plomion and Jean-Marc Gion (INRA, France), Nigel England and Keiji Tomita (Albany Forestry Research Centre, Oji Paper Co., Ltd.), Cristina Marques (RAIZ, Portugal), Rebeca Sanhueza Herrera (Forestal Mininco, Chile), Kitt Payn (Mondi, South Africa) and Geoff Galloway (Sappi, South Africa). Thanks also to the following people (University of Tasmania) for technical assistance with sample collection, processing and DNA sequence analysis: Scott Nicolls, Robert Barbour, James Marthick, Jules Freeman, James Worth. We are very grateful to Gay McKinnon (formerly of University of Tasmania) for her very constructive comments on the manuscript. This research was funded with a grant from the Australian Research Council (DP0770506) and assistance from the Cooperative Research Centre for Forestry (Australia).

References

Barkley, N.A., Krueger, R.R., Federici, C.T., Roose, M.L., 2009. What phylogeny and gene genealogy analyses reveal about homoplasy in citrus microsatellite alleles. Plant Syst. Evol. 282, 71-86.

Bayly MJ and Ladiges PY (2007) Divergent paralogues of ribosomal DNA in eucalypts (Myrtaceae). Molecular Phylogenetics and Evolution 44, 346-356.

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

Brondani, R.P.V., Williams, E.R., Brondani, C., Grattapaglia, D., 2006. A microsatellite-based consensus linkage map for species of *Eucalyptus* and a novel set of 230 microsatellite markers from the genus. BMC Plant Biol. 6, 20.

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

Brown, A.H.D., Matheson, A.C., Eldridge, K.G., 1975. Estimation of mating system of *Eucalyptus obliqua* L'Herit. by using allozyme polymorphisms. Aust. J. Bot. 23, 931-949.

Burgess, I., Bell, J.C., 1983. Comparative morphology and allozyme frequencies of *Eucalyptus grandis* Hill ex Maiden and *Eucalyptus saligna* SM. Aust. For. Res. 13, 133-149.

Bussell, J.D., Waycott, M., Chappill, J.A. 2005. Arbitrarily amplified DNA markers as characters for phylogenetic inference. Perspect. Plant Ecol. Evol. Syst. 7, 3-26.

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

Butcher, P.A., Otero, A., McDonald, M.W., Moran, G.F., 2002. Nuclear RFLP variation in *Eucalyptus camaldulensis* Dehnh. from northern Australia. Heredity 88, 402-412.

Byrne, M., 1999. High genetic identities between three oil mallee taxa, *Eucalyptus kochii* ssp. *kochii,* ssp. *plenissima* and *E. horistes,* based on nuclear RFLP analysis. Heredity 82, 205-211.

Byrne, M., 2008. Phylogeny, diversity and evolution of eucalypts, in: Sharma, A.K., Sharma, A. (Eds.), Plant Genome Biodiversity and Evolution Vol. 1, Part E. Phanerogams – Angiosperms. Science Publishers, pp 303-346.

Byrne, M., Hines, B., 2004. Phylogeographical analysis of cpDNA variation in *Eucalyptus loxophleba* (Myrtaceae). Aust. J. Bot. 52, 459-470.

Byrne, M., Marques-Garcia, M.I., Uren, T., Smith, D.S., Moran, G.F., 1996. Conservation and genetic diversity of microsatellite loci in the genus *Eucalyptus*. Aust. J. Bot. 44, 331–341.

Byrne, M., Parrish, T.L., Moran, G.F., 1998. Nuclear RFLP diversity in *Eucalyptus nitens*. Heredity 81, 225-233.

Clark, A.G., Hubisz, M.J., Bustamante, C.D., Williamson, S.H., Nielsen, R., 2005. Ascertainment bias in studies of human genome-wide polymorphism. Genome Res. 15, 1496-1502.

Cook, I.O., Ladiges, P.Y., 1998. Isozyme variation in *Eucalyptus nitens* and *E. denticulata*. Aust. J. Bot. 46, 35-44.

Curtu, A.L., Finkeldey, R., Gailing, O., 2004. Comparative sequencing of a microsatellite locus reveals size homoplasy within and between European oak species (*Quercus* spp.). Plant Mol. Bio. Reporter 22, 339-346.

Doyle, J.J., Doyle, J.L., 1990. Isolation of plant DNA from fresh tissue. Focus 12, 13-15.

Drinnan, A.N., Ladiges, P.Y., 1991. Floral development and systematic position of *Eucalyptus curtisii*. Aust. Syst. Bot. 4, 539-552.

Eggert, L,S,, Beadell, J.S., McClung, A., McIntosh, C.E., Fleischer, R.C., 2009. Evolution of microsatellite loci in the adaptive radiation of Hawaiian honeycreepers. J. Hered. 100, 137-147.

Elliott, C., Byrne, M., 2003. Genetic diversity within and between natural populations of *Eucalyptus occidentalis* (Myrtaceae). Silvae Genet. 52, 169-173.

Elliott, C.P., Byrne, M., 2004. Phylogenetics and the conservation of rare taxa in the *Eucalyptus angustissima* complex in Western Australia. Conserv. Genet. 5, 39-47.

Farris, J.S., 1989. The retention index and the rescaled consistency index. Cladistics 5, 417-419.

Farris, J.S., Källersjö, M., Kluge, A.G., Bult, C., 1994. Testing significance of incongruence. Cladistics 10, 315-319.

Forey, P.L., Humphries, C.J., Kitching, I.J., Scotland, R.W., Siebert, D.J., Williams, D.M., 1992. Cladistics. A practical course in systematics. The Sytematics Association Publication No. 10. Oxford University Press, New York.

Freeman, J.S., Jackson, H.D., Steane, D.A., McKinnon, G.E., Dutkowski G.W., Potts, B.M., Vaillancourt R.E., 2001. Chloroplast DNA phylogeography of Eucalyptus globulus. Aust. J. Bot. 49, 585-596.

Gaiotto, F.A., Bramucci, M., Grattapaglia, D. 1997. Estimation of outcrossing rate in a breeding program of *Eucalyptus urophylla* with dominant RAPD and AFLP markers. Theor. Appl. Genet. 95, 842-849.

Gibbs, A.K., Udovicic, F., Drinnan, A.N., Ladiges, P.Y., 2009. Phylogeny and classification of *Eucalyptus* subgenus *Eudesmia* (Myrtaceae) based on nuclear ribosomal DNA, chloroplast DNA and morphology. Aust. Syst. Bot. 22, 158-179.

Glaubitz, J., Emebiri, L., Moran, G. 2001. Dinucleotide microsatellites from *Eucalyptus sieberi*: inheritance, diversity, and improved scoring of single-base differences. Genome 44, 1041–1045.

Glaubitz, J.C., Murrell, J.C., Moran, G.F. 2003. Effects of native forest regeneration practices on genetic diversity in *Eucalyptus consideniana*. Theor. Appl. Genet. 107, 422-431.

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

Hill, K.D., Johnson L.A.S., 1995. Systematic studies in the eucalypts. 7. A revision of the bloodwoods, genus *Corymbia* (Myrtaceae). Telopea 6, 185-504.

Hines, B., Byrne, M., 2001. Genetic differentiation between mallee and tree forms in the *Eucalyptus loxophleba* complex. Heredity 87, 566-572.

Holman, J.E., Hughes, J.M., Fensham, R.J., 2003. A morphological cline in *Eucalyptus*: a genetic perspective. Mol. Ecol. 12, 3013-3025.

House, A.P.N., Bell, J.C., 1994. Isozyme variation and mating system in *Eucalyptus urophylla* ST Blake. Silvae Genet. 43, 167-179.

House, A.P.N., Bell, J.C., 1996. Genetic diversity, mating system and systematic relationships in two red mahoganies, *Eucalyptus pellita* and *E. scias*. Aust. J. Bot. 44, 157-174.

Huelsenbeck, J.P., Ronquist, F.R., 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17, 754-755.

Huson, D.H., 1998. SplitsTree: A program for analyzing and visualizing evolutionary data. Bioinformatics 14, 68-73.

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254-267.

Jaccoud, D., Peng, K., Feinstein, D., Kilian, A., 2001. Diversity Arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Res. 29: e25.

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

James, K.E., Schneider, H., Ansell, S.W., Evers, M., Robba, L., Uszynski, G., Pedersen, N., Newton, A.E., Russell, S.J., Vogel, J.C., Kilian, A., 2008. Diversity Arrays Technology (DArT) for pan-genomic evolutionary studies of non-model organisms. PLoS ONE **3**, e1682 [doi:10.1371/journal.pone.0001682].

Jones, M.E., Shepherd, M., Henry, R.J., Delves, A., 2006. Chloroplast DNA variation and population structure in the widespread forest tree, Eucalyptus grandis. Conservation Genetics 7, 691-703.

Jones, M.E., Stokoe, R.L., Cross, M.J., Scott, L.J., Maguire, T.L., Shepherd, M., 2001. Isolation of microsatellite loci from spotted gum (*Corymbia variegata*) and cross-species amplification in *Corymbia* and *Eucalyptus*. Mol. Ecol. Notes 1, 276-278.

Jones, R.C., 2009. Molecular evolution and genetic control of flowering in the *Eucalyptus globulus* species complex. PhD thesis, University of Tasmania.

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

Kress, W.J., Erickson, D.L. (2008) DNA barcodes: Genes, genomics and bioinformatics. Proc. Natl. Acad. Sci. USA 105: 2761-2762.

Ladiges, P.Y., 1997. Phylogenetic history and classification of eucalypts, in: Williams, J.E., Woinarski, J.C.Z. (Eds.), Eucalypt ecology: individuals to ecosystems. Cambridge University Press, Cambridge, pp.16-29.

Ladiges, P.Y., Bayly, M.J., Nelson, G.J., 2010. East-west continental vicariance in *Eucalyptus* subgenus *Eucalyptus*, in: Williams, D.M., Knapp, S. (Eds.), Beyond Cladistics: The Branching of a Paradigm. University of California, pp. 267-302.

Le, S., Nock, C., Henson, M., Shepherd, M., 2009. Genetic differentiation among and within three red mahoganies (ser. *Annulares*), *Eucalyptus pellita*, *E. resinifera* and *E. scias* (Myrtaceae). Aust. Syst. Bot. 22, 332-343.

Li, C.Y., 2000. RAPD analysis of genetic variation in *Eucalyptus microtheca* F. Muell. populations. Hereditas 132, 151-156.

Liu, L., Yu, L., Kubatko, L., Pearl, D.K., Edwards, S.V. 2009. Coalescent methods for estimating phylogenetic trees. Mol. Phylog. Evol. 53, 320-328.

McDonald M.W., Rawlings, M., Butcher, P.A., Bell, J.C., 2003. Regional divergence and inbreeding in *Eucalyptus cladocalyx* (Myrtaceae). Aust. J. Bot. 51, 393-403.

McDonald, M.W., Brooker, M.I.H., Butcher, P.A. 2009., A taxonomic revision of *Eucalyptus camaldulensis* (Myrtaceae). Aust. Syst. Bot. 22: 257-285.

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

McKinnon, G.E., Vaillancourt, R.E., Steane, D.A., Potts, B.M., 2004. 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., Vaillancourt, R.E., Steane, D.A., Potts, B.M., 2008. An AFLP marker approach to lower-level systematics in *Eucalyptus* (Myrtaceae). Am. J. Bot. 95, 368-380.

Moran, G.F., 1992. Patterns of genetic diversity in Australian tree species. New Forests 6, 49-66.

Mossel, E., Vigoda, E., 2006. Limitations of Markov Chain Monte Carlo algorithms for Bayesian inference of phylogeny. Ann. Appl. Probab. 16, 2215-2234.

Myburg, A.A., Remington, D.L., O'Malley, D.M., Sederoff, R.R., Whetten, R.W., 2001. High-throughput AFLP analysis using infrared dye-labeled primers and an automated DNA sequencer. Biotechniques 30, 348-357.

Nesbitt, K.A., Potts, B.M., Vaillancourt, R.E., West, A.K., Reid, J.B., 1995. Partitioning and distribution of RAPD variation in a forest tree species, *Eucalyptus globulus*. Heredity 74, 628-637.

Nevill, P.G., Reed, A., Bossinger, G., Vaillancourt, R.E., Larcombe, M., Ades, P.K., 2008. Cross-species amplification of *Eucalyptus* microsatellite loci. Mol. Ecol. Resources 8, 1277-1280.

Ochieng, J.W., Henry, R.J., Baverstock, P.R., Steane, D.A., Shepherd, M., 2007a. Nuclear ribosomal pseudogenes resolve a corroborated monophyly of the eucalypt genus *Corymbia* despite misleading hypotheses at functional ITS paralogs. Mol. Phylog. Evol. 44, 752-764.

Ochieng, J.W., Steane, D.A., Ladiges, P.Y., Baverstock, P.R., Henry, R.J., Shepherd, M., 2007b. The genus *Corymbia* (Myrtaceae) is monophyletic at microsatellite loci. Genet. Mol. Biol. 30, 1125-1134.

Okun, D.O., Kenya, E.U., Oballa, P.O., Odee, D.W., Muluvi, G.M., 2008. Analysis of genetic diversity in *Eucalyptus grandis* (Hill ex Maiden) seed sources using inter simple sequence repeats (ISSR) molecular markers. Afr. J. Biotech. 7, 2119-2123.

Ottewell, K.M., Donnellan, S.C., Moran, G.F., Paton, D.C., 2005. Multiplexed microsatellite markers for the genetic analysis of *Eucalyptus leucoxylon* (Myrtaceae) and their utility for ecological and breeding studies in other *Eucalyptus* species. J. Hered. 96, 445–451.

Parra-O, C., Bayly, M., Udovicic, F., Ladiges, P., 2006. ETS sequences support the monophyly of the eucalypt genus *Corymbia* (Myrtaceae). Taxon 55, 653-663.

Payn, K.G., Dvorak, W.S., Janse, B.J.H., Myburg, A.A., 2008, Microsatellite diversity and genetic structure of the commercially important tropical tree species *Eucalyptus urophylla*, endemic to seven islands in eastern Indonesia. Tree Genet. Genomes 4, 519-530.

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

Perrie, L.R., Shepherd, L.D. 2009. Reconstructing the species phylogeny of *Pseudopanax* (Araliaceae), a genus of hybridizing trees. Mol. Phylog. Evol. 52, 774-783.

Petroli, C.D., Sansaloni, C.P., Carling, J., Hudson, C., Steane, D.A., Myburg, A.A., Vaillancourt, R.E., Kilian, A., Grattapaglia, D. in prep. A high-density sub-centi-Morgan DArT/microsatellite genetic linkage map for species of Eucalyptus based on 2,450 markers. BMC Plant Biology.

Poke, F.S., Martin, D.P., Steane, D.A., Vaillancourt, R.E., Reid, J.B., 2006. The impact of intragenic recombination on phylogenetic reconstruction at the sectional level in *Eucalyptus* when using a single copy nuclear gene (cinnamoyl CoA reductase). Mol. Phylog. Evol. 39, 160-170.

Poltri, S.N.M., Zelener, N., Traverso, J.R., Gelid, P., Hopp, H.E., 2003. Selection of a seed orchard of *Eucalyptus dunnii* based on genetic diversity criteria calculated using molecular markers. Tree Physiol. 23, 625-632.

Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817-818.

Potts, B.M., Wiltshire, R.J.E., 1997. Eucalypt genetics and genecology, in: Williams, J. and Woinarski, J. (Eds.), Eucalypt ecology: individuals to ecosystems. Cambridge University Press, Cambridge, pp. 56-91.

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

Pryor, L.D., Johnson, L.A.S., 1981. *Eucalyptus*, the universal Australian, in: Keast, A. (Ed.), Ecological Biogeography of Australia. Dr W. Junk, The Hague, pp. 499-536.

Ramirez, M.J., 2006. Further problems with the incongruence length difference test: “hypercongruence” effect and multiple comparisons. Cladistics 22, 289-295.

Rathbone, D.A., McKinnon, G.E., Vaillancourt, R.E., Steane, D.A., Potts, B.M., 2007. Microsatellite and cpDNA variation in island and continental populations of a regionally rare eucalypt, *Eucalyptus perriniana* (Myrtaceae). Aust. J. Bot. 55, 513-520.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572-1574.

Sale, M.M., Potts, B.M., West, A.K., Reid, J.B., 1993. Relationships within *Eucalyptus* using chloroplast DNA. Aust. Syst. Bot. 6, 127-138.

Sale, M.M., Potts, B.M., West, A.K., Reid, J.B., 1996. Relationships within *Eucalyptus* (Myrtaceae) using PCR-amplification and southern hybridization of chloroplast DNA. Aust. Syst. Bot. 9, 273-282.

Sansaloni, C.P., Petroli, C.D., Carling, J., Hudson, C., Steane, D.A., Myburg, A.A., Grattapaglia, D., Vaillancourt, R.E., Kilian, A., 2010. A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in *Eucalyptus*. Plant Methods.

Shepherd, M., Kasem, S., Ablett, G., Ochieng, J., Crawford, A., 2008. Genetic structuring in the spotted gum complex (genus *Corymbia*, section *Politaria*). Aust. Syst. Bot. 21, 15-25.

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

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

Steane, D.A., Conod, N., Jones, R.C., Vaillancourt, R.E., 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.

Steane, D.A., McKinnon, G.E., Vaillancourt, R.E., Potts, B.M., 1999. ITS sequence data resolve higher-level relationships among the eucalypts. Mol. Phylog. Evol. 12, 215-233.

Steane, D.A., Nicolle, D., McKinnon, G.E., Vaillancourt, R.E., Potts, B.M., 2002. Higher level relationships among the eucalypts are resolved by ITS-sequence data. Aust. Syst. Bot. 15, 49-62.

Steane, D.A., Nicolle, D., Potts, B.M., 2007. Phylogenetic positioning of anomalous eucalypts by using ITS sequence data. Aust. Syst. Bot. 20, 402-408.

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

Steane, D.A., West, A.K., Potts, B.M., Ovenden, J.R., Reid, J.B., 1992. Restriction fragment length polymorphisms in chloroplast DNA from six species of *Eucalyptus*. Aust. J. Bot. 39: 399-414.

Stokoe, R.L., Shepherd, M., Lee, D.J., Nikles, D.G., Henry, R.J., 2001. Natural inter-subgeneric hybridisation between *Eucalyptus acmenoides* Schauer and *Eucalyptus cloeziana* F. Muell (Myrtaceae) in southeast Queensland. Ann. Bot. 88, 563-570.

Swofford, D.L., 1999. ‘PAUP*. Phylogenetic analysis using parsimony (*and other methods), Version 4. Sinauer, Sunderland, MA.

Thamarus, K.A., Groom, K., Murrel, J., Byrne, M., Moran, G.F., 2002. A genetic linkage map for *Eucalyptus globulus* with candidate loci for wood, fibre and floral traits. Theor. Appl. Genet. 104, 379-387.

Udovicic, F., McFadden, G., Ladiges, P.Y., 1995. Phylogeny of *Eucalyptus* and *Angophora* based on 5S rDNA spacer sequence DNA. Mol. Phylog. Evol. 4, 247-256.

Vekemans, X., 2002. AFLP-SURV version 1.0. Distibuted by the author, Laboratoire de Genetique et Ecologie Vegetale, Universite Libre de Bruxelles, Belgium.

Walker, E., Byrne, M., Macdonald, B., Nicolle, D., McComb, J., 2009. Clonality and hybrid origin of the rare *Eucalyptus bennettiae* (Myrtaceae) in Western Australia. Aust. J. Bot. 57, 180-188.

Wheeler, M.A., Byrne, M., 2006. Congruence between phylogeographic patterns in cpDNA variation in *Eucalyptus marginata* (Myrtaceae) and geomorphology of the Darling Plateau, south-west of Western Australia. Aust. J. Bot. 54, 17-26.

Wheeler, M.A., Byrne, M., McComb, J.A., 2003. Little genetic differentiation within the dominant forest tree, *Eucalyptus marginata* (Myrtaceae) of South-Western Australia. Silvae Genet. 52, 254-259.

Whittock, S., Steane, D.A., Vaillancourt, R.E., Potts, B.M., 2003. Molecular evidence shows that the tropical boxes (*Eucalyptus* subgenus *Minutifructus*) are over-ranked. Trans. Roy. Soc. S. Austr. 127, 27-32.

Wright, I.J., Ladiges, P.Y., 1997. Geographic variation in *Eucalyptus diversifolia* (Myrtaceae) and the recognition of new subspecies *E. diversifolia* subsp. *hesperia* and *E. diversifolia* subsp. *megacarpa*. Aust. Syst. Bot. 10, 651-680.

Figure legends:

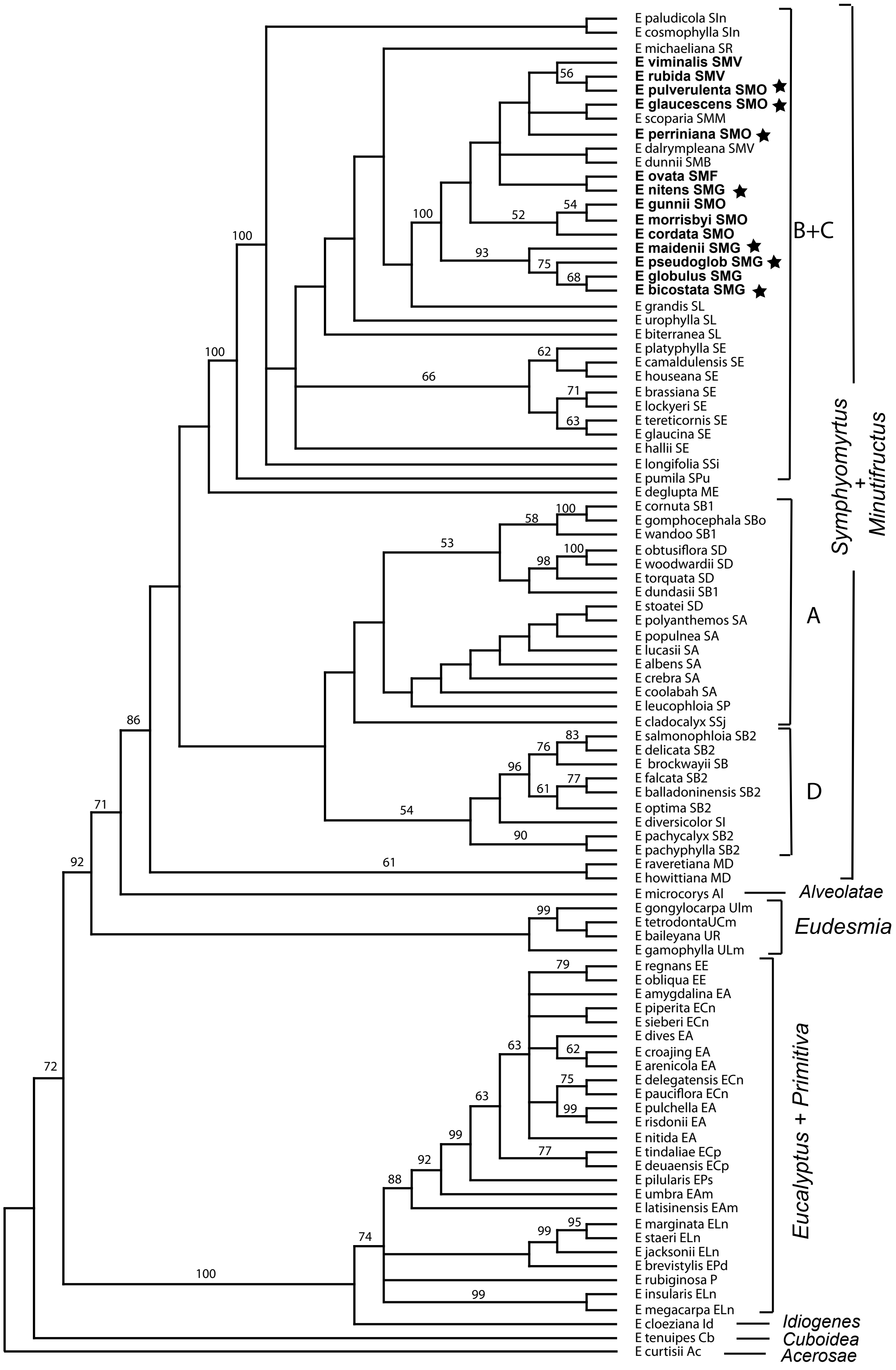
Fig.1. Strict consensus of 10 trees derived from DArT data using PAUP*4.0b10 (length (including ten autapomorphies) = 74861; CI (excl. autapomorphies) = 0.111; RI = 0.624). The cladograms were rooted on *E. curtisii* (subg. *Aceroceae*) on the basis of previous studies (Drinnan and Ladiges 1991; Steane *et al*. 2002). Numbers above branches represent bootstrap values greater than 50%. Although resolution within subgenus *Symphyomyrtus* is good, many nodes have poor bootstrap support. Clades A, B, C and D refer to clades within subgenus *Symphyomyrtus* that were identified in phylogenies based on ITS sequence data (Steane *et al*. 1999, 2002, 2007). Refer to Table 2 of main text for taxon abbreviations (subgenus, section) after each species name. Series within section *Maidenaria* (SM) are shown: B – *Bridgesianae*; F – *Foveolatae*; G – *Globulares*; M – *Microcarpae*; O – *Orbiculares*; V – *Viminales*. Species relating to the AFLP study of McKinnon *et al*. (2008) are shown in bold, with species from mainland Australia marked with a star (see Discussion).

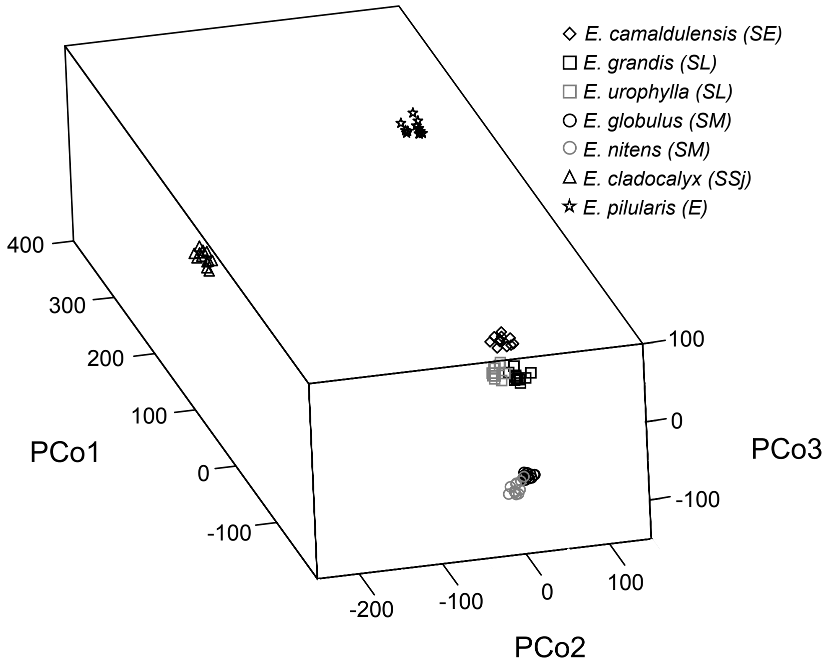
Fig. 2. Three dimensional representation of the first three Principal Components for DArT variation in seven species of *Eucalyptus*, calculated using *GenAlex* (Peakall and Smouse 2006). The first three coordinates explained 91.9% (68.9%, 16.1% and 6.9% for axes 1, 2 and 3, respectively) of the variation among samples. Subgenera and sections are given in parentheses after species name. Subgenus *Symphyomyrtus* sections are as follows: SE = *Exsertaria*, SL = *Latoangulatae*, SM = *Maidenaria* and SSj = *Sejunctae*. Subgenus *Eucalyptus* is represented by E.

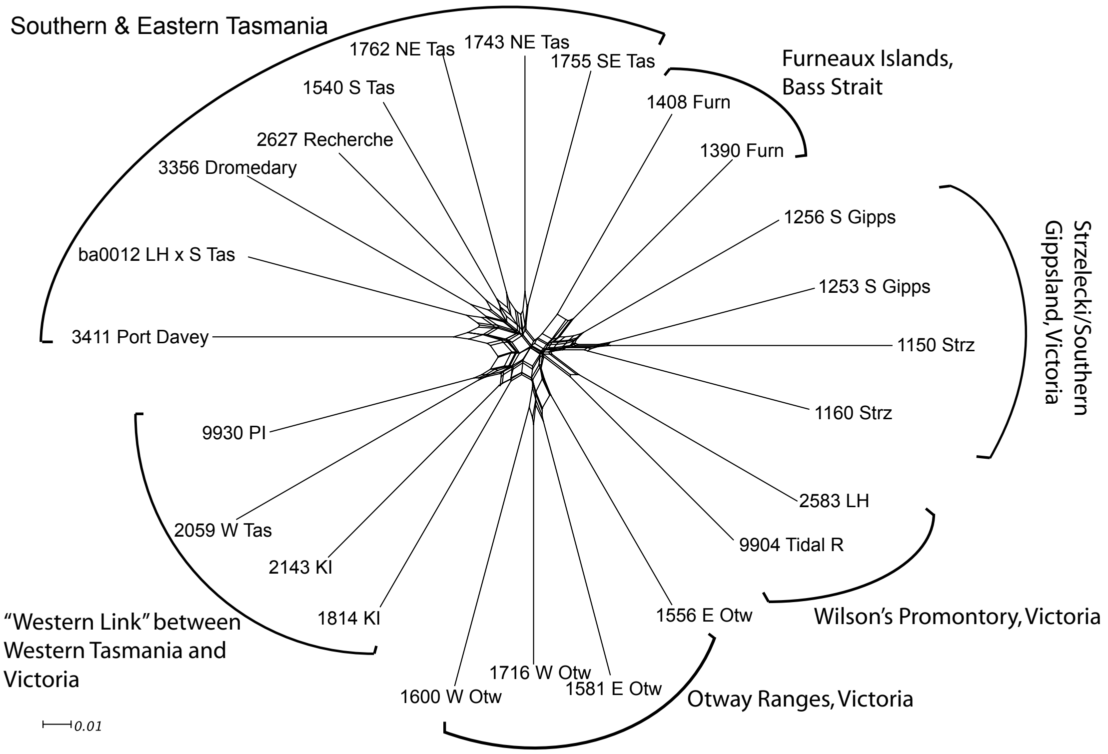
Fig. 3. Network generated by *Splitstree4* showing relationships among races of *Eucalyptus globulus* (based on DArT results from Plate 2). Abbreviations: E Otw – eastern Otway Ranges (Victoria); Furn – Furneaux group of islands, eastern Bass Strait; KI – King Island,

western Bass Strait; LH – Wilson’s Promotory Lighthouse (Victoria); NE Tas – northeastern Tasmania; PI – Phillip Island (Victoria); S Gipps – Southern Gippsland (Victoria); S Tas – southern Tasmania; SE Tas – southeastern Tasmania; Strz – Strzelecki Ranges (Victoria); Tidal R – Tidal River, Wilson’s Promontory (Victoria); W Otw – western Otway Ranges (Victoria); W Tas – western Tasmania. Scale bar shows *Uncorrected P* genetic distance equivalent to 0.01.

Fig. 4. *Splitstree4* network showing the positions of eucalypt hybrids relative to their parent species. While the *E. nitens* x *globulus* hybrids appear to be intermediate between the two parent species, the *E. urophylla* x *grandis* hybrids that were included in this study appear to be more closely related to *E. urophylla* than to *E. grandis* at the genome-wide level.


E paludicola SIn
E cosmophylla SIn
E michaeliana SR
E viminalis SMV
E rubida SMV
E pulverulenta SMO ★
E glaucescens SMO ★
E scoparia SMM
E perriniana SMO ★
E dalrympleana SMV
E dunnii SMB
E ovata SMF
E nitens SMG ★
E gunnii SMO
E morrisbyi SMO
E cordata SMO
E maidenii SMG ★
E pseudoglob SMG ★
E globulus SMG
E bicostata SMG ★
E grandis SL
E urophylla SL
E biterranea SL
E platyphylla SE
E camaldulensis SE
E houseana SE
E brassiana SE
E lockyeri SE
E tereticornis SE
E glaucina SE
E hallii SE
E longifolia SSi
E pumila SPu
E deglupta ME
E cornuta SB1
E gomphocephala SBo
E wandoo SB1
E obtusiflora SD
E woodwardii SD
E torquata SD
E dundasii SB1
E stoatei SD
E polyanthemos SA
E populnea SA
E lucasii SA
E albens SA
E crebra SA
E coolabah SA
E leucophloia SP
E cladocalyx SSj
E salmonophloia SB2
E delicata SB2
E brockwayii SB
E falcata SB2
E balladoninensis SB2
E optima SB2
E diversicolor SI
E pachycalyx SB2
E pachyphylla SB2
E raveretiana MD
E howittiana MD
E microcorys Al
E gongylocarpa Ulm
E tetrodontaUCm
E baileyana UR
E gamophylla ULm
E regnans EE
E obliqua EE
E amygdalina EA
E piperita ECn
E sieberi ECn
E dives EA
E croajing EA
E arenicola EA
E delegatensis ECn
E pauciflora ECn
E pulchella EA
E risdonii EA
E nitida EA
E tindaliae ECp
E deuaensis ECp
E pilularis EPs
E umbra EAm
E latisinensis EAm
E marginata ELn
E staeri ELn
E jacksonii ELn
E brevistylis EPd
E rubiginosa P
E insularis ELn
E megacarpa ELn
E cloeziana Id
E tenuipes Cb
E curtisii Ac
B+C
A
D
Symphyomyrtus + Minutifructus
Alveolatae
Eudesmia
Eucalyptus + Primitiva
Idiogenes
Cuboidea
Acerosae
56
100
54
52
100
93
75
68
100
62
66
71
63
100
58
53
100
98
83
76
96
77
61
86
54
90
61
71
92
99
79
63
62
75
63
99
77
72
99
92
88
99
95
74
100
99


◇ E. camaldulensis (SE)
□ E. grandis (SL)
□ E. urophylla (SL)
○ E. globulus (SM)
○ E. nitens (SM)
△ E. cladocalyx (SSj)
☆ E. pilularis (E)
PCo1
400
300
200
100
0
-100
PCo2
-200
-100
0
100
PCo3
100
0
-100


Southern & Eastern Tasmania
1762 NE Tas
1743 NE Tas
1755 SE Tas
1540 S Tas
2627 Recherche
3356 Dromedary
ba0012 LH x S Tas
3411 Port Davey
Furneaux Islands, Bass Strait
1408 Furn
1390 Furn
Strzelecki/Southern Gippsland, Victoria
1256 S Gipps
1253 S Gipps
1150 Strz
1160 Strz
2583 LH
9904 Tidal R
Wilson's Promontory, Victoria
9930 PI
2059 W Tas
2143 KI
1814 KI
"Western Link" between Western Tasmania and Victoria
1556 E Otw
1716 W Otw
1581 E Otw
1600 W Otw
Otway Ranges, Victoria
0.01

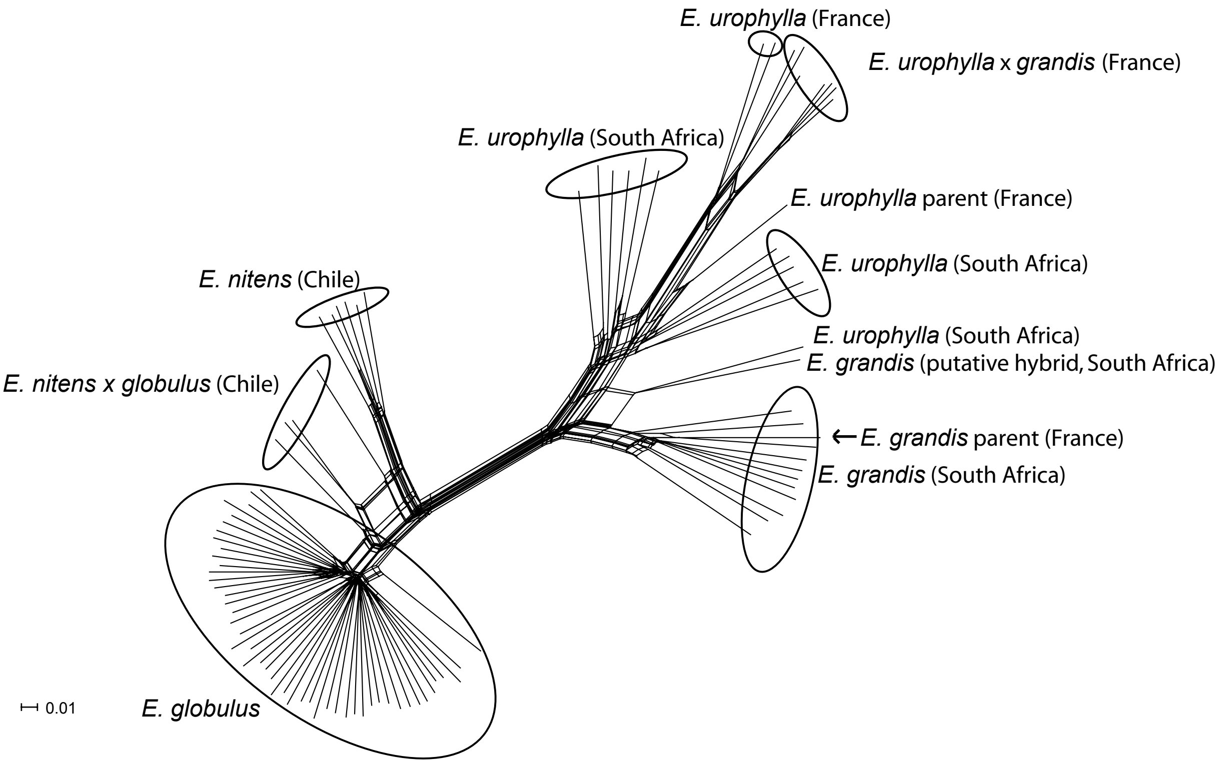

E. urophylla (France)
E. urophylla x *grandis* (France)
E. urophylla (South Africa)
E. urophylla parent (France)
E. urophylla (South Africa)
E. nitens (Chile)
E. urophylla (South Africa)
E. grandis (putative hybrid, South Africa)
E. nitens x globulus (Chile)
← *E. grandis* parent (France)
E. grandis (South Africa)
0.01
E. globulus

Table 1 Summary of taxa represented on each plate (number of samples given in parentheses) and the number of DArT markers used in the screening of each plate. See Appendices A and B and Table 2 for details.

Plate name	Species (No. samples) on plate	No. DArT markers
Plate 1	*Eucalyptus globulus (12), E. nitens (11), E. grandis (12), E. urophylla (12), E. camaldulensis (11), E. cladocalyx (12), E. pilularis (12), Corymbia variegata (12)*	7052
Plate 2	*E. globulus (49), E. nitens (6), E. nitens* x *globulus (4), E. grandis (13), E. urophylla (15), E. grandis* x *urophylla (7)*	4684
Phylogeny	94 species	8354

Table 2. Samples used in phylogeny trial of DArT markers. Samples with superscripts were included in previous studies (ITS-based phylogenies). “Code” gives an abbreviation of the subgenus and section (where applicable) names. CCA – Currency Creek Arboretum; NSW – New South Wales; NT – Northern Territory; Qld – Queensland; SA – South Australia; Tas – Tasmania; Vic – Victoria; WA – Western Australia.

Species	Subgenus	Section	Code	Provenance	Origin or Herbarium number (ITS GenBank accession No.)
E. albens	*Symphyomyrtus*	*Adnataria*	SA	West of Wagga Wagga, NSW	DN 2898 (HM596031)
E. amygdalina	*Eucalyptus*	*Aromatica*	EA	Kingston, SE Tas	Bridport Pole 33 (HM596032)
E. arenicola[a,h]	*Eucalyptus*	*Aromatica*	EA	Holey Plains, Gippland, SE Vic	CCA 32,13 (AF058499)
E. baileyana	*Eudesmia*	*Reticulatae*	UR	B/n Grafton & Baryulgil, NSW	DN 665 (HM596033)
E. balladoniensis ssp. *balladoniensis*	*Symphyomyrtus*	*Bisectae (II)*	SB2	Nr Mt Ney, WA	DN 3602 (HM596034)
E. bicostata	*Symphyomyrtus*	*Maidenaria*	SM	Bruthen, E Vic	UTAS 4075 (HM596035)
E. biterranea[f]	*Symphyomyrtus*	*Latoangulatae*	SL	Iron Range, Cape York Peninsula, Qld	DN 2518 (HM596036)
E. brassiana	*Symphyomyrtus*	*Exsertaria*	SE	W of Cooktown, Qld	DN 1316 (HM596037)
E. brevistylis[b]	*Eucalyptus*	*Pedaria*	EPd	E of Mt Frankland, WA	CCA 76,24: seedling from DN 1141 (AF390527)
E. brockwayi[b]	*Symphyomyrtus*	*Bisectae (II)*	SB2	NW of Norseman, WA	CCA 15,15: seedling from DN 136 (AF390505)

E. camaldulensis[g]	*Symphyomyrtus*	*Exsertaria*	SE	Palmer, Qld.	B10626 (From DArT Phase 1, Plate 1) (HM596038)
E. cladocalyx[c]	*Symphyomyrtus*	*Sejunctae*	SSj	Port Lincoln, Eyre Peninsula, SA	DN 4134 (progeny of DN 3182)(EF488228)
E. cloeziana	*Idiogenes*		Id	Isla Gorge NP, Qld	DN696 (no sequence available)
E. coolabah	*Symphyomyrtus*	*Adnataria*	SA	SE of Wilcannia, NSW	DN 2957 (HM596039)
E. cordata ssp. *cordata*	*Symphyomyrtus*	*Maidenaria*	SM	Cape Queen Elizabeth	UTAS 1475 (HM596040)
E. cornuta	*Symphyomyrtus*	*Bisectae (I)*	SB1	Nr Bremer Bay, WA	DN 3748 (HM596041)
E. cosmophylla[c]	*Symphyomyrtus*	*Incognitae*	SIn	Kingscote, Kangaroo Is., SA	CCA: DN 819 (EF488226)
E. crebra[b]	*Symphyomyrtus*	*Adnataria*	SA	NE of Tara, Qld	CCA 51, 01: seedling from DN 680 (AF390503)
E. croajingolensis[a]	*Eucalyptus*	*Aromatica*	EA	Holey Plains, Gippsland, SE Vic	TU: 29/16 (AF058497)
E. curtisii	*Acerosae*		Ac	Nr Beerwah, Qld	DN 2108 (HM596042)
E. dalrympleana[a]	*Symphyomyrtus*	*Maidenaria*	SM	Central Plateau, Tas	TU: 458/DA1 (AF058466)
E. deglupta	*Minutifructus*	*Equatoria*	ME	Philippines	Flecker BG, Cairns (HM596043)
E. delegatensis ssp. *tasmaniensis*[a]	*Eucalyptus*	*Cineraceae*	ECn	Mt Wellington, SE Tas	TU: 636 (AF058480)

E. delicata	*Symphyomyrtus*	*Bisectae (II)*	SB2	Peak Charles, WA	DN 2262 (HM596044)
E. deuaensis	*Eucalyptus*	*Capillulus*	ECp	Deua Nat Park, Southern Tablelands, NSW	DN 1769 (HM596045)
E. diversicolor[b]	*Symphyomyrtus*	*Inclusae*	SI	Walpole, WA	CCA 76, 18: seedling from DN 1142 (AY039754)
E. dives[a]	*Eucalyptus*	*Aromatica*	EA	Gembrook, S Vic.	TU: GEM4 (AF058503)
E. dundasii[b]	*Symphyomyrtus*	*Bisectae (I)*	SB1	Nr Fraser Range, WA	CCA 11, 15: seedling from DN 129 (AF390501)
E. dunnii[b]	*Symphyomyrtus*	*Maidenaria*	SM	Nr Legume, NSW	CCA 89, 31: seedling from DN 1257 (AF390510)
E. falcata[b]	*Symphyomyrtus*	*Bisectae (II)*	SB2	NE of Hopetoun, WA	CCA 17, 30: seedling from DN 198 ((AF390506)
E. gamophylla[a]	*Eudesmia*	*Limbatae*	ULm	Road to Kings Canyon, NT	CCA 09, 11 (HM596046)
E. glaucescens	*Symphyomyrtus*	*Maidenaria*	SM	St Guinear, NSW	St Guinear (HM596047)
E. glaucina	*Symphyomyrtus*	*Exsertaria*	SE	Nr Paterson, NSW	DN 2085 (HM596048)
E. globulus	*Symphyomyrtus*	*Maidenaria*	SM	Strzelecki	UTAS 1160 (HM596049)
E. gomphocephala[c]	*Symphyomyrtus*	*Bolites*	SBo	Bunbury, West Coast, WA	CCA: DN 1148 (EF488231)
E. gongylocarpa[b]	*Eudesmia*	*Limbatae*	ULm	Great Victoria Desert, WA	CCA 40, 25: seedling from DN 519 (AF390466)

E. grandis	*Symphyomyrtus*	*Latoangulatae*	SL	South Africa	#17 Zander Myburg, S. Afr. (HM596050)
E. gunnii ssp. *gunnii*[a]	*Symphyomyrtus*	*Maidenaria*	SM	Snug, SE Tas	TU: 460 (AF058464)
E. hallii[b]	*Symphyomyrtus*	*Exsertaria*	SE	Nr Goodwood, Qld	CCA 58, 06: seedling from DN 716 (AF390512)
E. houseana[b]	*Symphyomyrtus*	*Exsertaria*	SE	March Fly Glen, Kimberley, WA	CCA 134, 22: seedling from DN 1911 (AF390487)
E. howittiana[d]	*Minutifructus*	*Domesticae*	MD	Greenvale, Qld	CCA: DN 2526 (EF694709)
E. insularis	*Eucalyptus*	*Longistylus*	ELn	Mt LeGrand, S. Coast, WA	DN 1637 (HM596051)
E. jacksonii[b]	*Eucalyptus*	*Longistylus*	ELn	Valley of the Giants, WA	CCA 76, 10: seedling from DN 1140 (AF390529)
E. latisinensis[b]	*Eucalyptus*	*Amentum*	EAm	Goodwood, Qld	CCA 58, 33: seedling from DN 715 (AF390532)
E. leucophloia ssp. *leucophloia*	*Symphyomyrtus*	*Platysperma*	SP	Round Hill, W of Capricorn Roadhouse, WA	DN 539 (HM596052)
E. lockyeri ssp. *lockyeri*[b]	*Symphyomyrtus*	*Exsertaria*	SE	NW of Ravenshoe, Qld	CCA: seedling from DN 1323 (AF390488)
E. longifolia[c]	*Symphyomyrtus*	*Similares*	SSi	Eden, South Coast, NSW	CCA: DN 1750 (EF488224)
E. lucasii[b]	*Symphyomyrtus*	*Adnataria*	SA	W of Wiluna, WA	CCA 39, 12: seedling from DN 545 (AF390494)
E. maidenii	*Symphyomyrtus*	*Maidenaria*	SM	Mt Myrtle, NSW	UTAS 3125 (HM596053)

E. marginata ssp. *thalassica*[b]	*Eucalyptus*	*Longistyla*	ELn	Gingin, WA	CCA 26, 29: seedling from DN 246 (AF390530)
E. megacarpa[b]	*Eucalyptus*	*Longistylus*	ELn	Two people's Bay, WA	CCA 70, 37: seedling from DN 1137 (AF390528)
E. michaeliana[b]	*Symphyomyrtus*	*Racemus*	SR	Nr Hillgrove, NSW	CCA: seedling from DN 843 (AF390484)
E. microcorys[d]	*Alveolata*		Al	Johns River, N NSW	CCA: DN 1238 (EF694714)
E. morrisbyi	*Symphyomyrtus*	*Maidenaria*	SM	Calverts Hill, Tas	UTAS 2307 (HM596054)
E. nitens	*Symphyomyrtus*	*Maidenaria*	SM	Heyfield, Vic.	8-185 (From Phase I, Plate 1 of DArT) (HM596055)
E. nitida[a]	*Eucalyptus*	*Aromatica*	EA	Flinders Is. NE Tas	TU trial 90/1; TU N42 (AF058481)
E. obliqua[a]	*Eucalyptus*	*Eucalyptus*	EE	Mt Nelson, SE Tas	TU: 634 (AF058484)
E. obtusiflora	*Symphyomyrtus*	*Dumaria*	SD	S of Shark Bay, WA	DN 1173 (HM596056)
E. optima	*Symphyomyrtus*	*Bisectae (II)*	SB2	B/n Balladonia and Norseman, WA	DN 2154 (HM596057)
E. ovata var. *ovata*[b]	*Symphyomyrtus*	*Maidenaria*	SM	Kingston, Tas	TU 204 (AF390480)
E. pachycalyx	*Symphyomyrtus*	*Bisectae (II)*	SB2	B/n Herberton and Irvinebank, Qld	DN 1307 (HM596058)
E. pachyphylla[b]	*Symphyomyrtus*	*Bisectae (II)*	SB2	E of Mt. Webb, Gibson Desert, WA	CCA 73, 28: seedling of DN 1203 (AF390473)

E. paludicola[c]	*Symphyomyrtus*	*Incognitae*	SIn	Ashbourne, Fleurieu Peninsula, SA	CCA: DN 69 (EF488227)
E. pauciflora ssp. *pauciflora*[a]	*Eucalyptus*	*Cineraceae*	ECn	Tomahawk, NE Tas	TU: 638 (AF058489)
E. perriniana[b]	*Symphyomyrtus*	*Maidenaria*	SM	Kosciusko, NSW	UTAS 688 (AF390476)
E. pilularis[b]	*Eucalyptus*	*Pseudophloia*	EPs	Domain, Sydney, NSW (RBGS 19739)	NA (AF390533)
E. piperita ssp. urceolaris[a]	*Eucalyptus*	*Cineraceae*	ECn	Nowra, SE NSW	DN 610 (AF058485)
E. aff. platyphylla[b,e]	*Symphyomyrtus*	*Exsertaria*	SE	E of Kupiano, Papua New Guinea	CSIRO 13400B (AF390485)
E. polyanthemos ssp. polyanthemos[b]	*Symphyomyrtus*	*Adnataria*	SA	Nr Rylston, NSW	CCA 46, 36: seedling from DN 742 (AF390513)
E. populnea ssp. *populnea*	*Symphyomyrtus*	*Adnataria*	SA	SE of Kogan, Qld	DN 679 (HM596059)
E. pseudoglobulus	*Symphyomyrtus*	*Maidenaria*	SM	Wiben's Hill, Vic	UTAS 627 (HM596060)
E. pulchella[a]	*Eucalyptus*	*Aromatica*	EA	Mt Nelson, SE Tas	TU: 633 (AF058487)
E. pulverulenta	*Symphyomyrtus*	*Maidenaria*	SM	Cult. Uni of Tas (Regent St)	UTAS 6065 (HM596061)
E. pumila[c]	*Symphyomyrtus*	*Pumilio*	SPu	Broken Back Range, NSW	CCA: DN 636 (EF488232)

E. raveretiana[a]	*Minutifructus*	*Domesticae*	MD	Oaky Ck, NW of Mingela, Qld	DN 1297 (HM596062)
E. regnans	*Eucalyptus*	*Eucalyptus*	EE	Leslie Vale, SE Tas	Cambridge Arboretum, Rep 1, Row F, Col. 6, Serpentine (HM596063)
E. risdonii[a]	*Eucalyptus*	*Aromatica*	EA	Meehan Range, SE Tas	TU: MRTHC1 (AF058493)
E. rubida ssp. *rubida*	*Symphyomyrtus*	*Maidenaria*	SM	Fingal/South Esk, N Tas	UTAS 693 (HM596064)
E. rubiginosa	*Primitiva*		P	Isla Gorge, Qld	DN 2114 (HM596065)
E. salmonophloia[b]	*Symphyomyrtus*	*Bisectae (II)*	SB2	Great Victoria Desert, WA	CCA 42, 14: seedling from DN 341 (AF390509)
E. scoparia[b]	*Symphyomyrtus*	*Maidenaria*	SM	Mt Norman, Qld	CCA 72, 34: seedling from DN 672 (AF390479)
E. sieberi[a]	*Eucalyptus*	*Cineraceae*	ECn	Freycinet Peninsula, E. Tas. (TU trial 90/1)	TU SIBFBR (AF058495)
E. staeri[b]	*Eucalyptus*	*Longistylus*	ELn	Wellstead, WA	CCA 70, 34: seedling from DN 1133 (AF3905531)
E. stoatei[b]	*Symphyomyrtus*	*Dumaria*	SD	SE Raventhorpe, WA	CCA 41, 34: seedling from DN 181 (AF390498)
E. tenuipes[b]	*Cuboidea*		Cb	Auburn Rd, 44km N of Warrego Hwy	RBGS 842705 (AF390523)
E. tereticornis ssp. *tereticornis*	*Symphyomyrtus*	*Exsertaria*	SE	B/n Helidon and Crows Nest, Qld	DN 2937 (HM596066)

E. tetrodonta	*Eudesmia*	*Complanatae*	UCm	Darwin, NT	DN 5157 (HM596067)
E. tindaliae[b]	*Eucalyptus*	*Capillulus*	ECp	SE of Grafton, NSW	CCA 138, 3: seedling from DN 1243 (AF390534)
E. torquata[b]	*Symphyomyrtus*	*Dumaria*	SD	NW of Norseman, WA	CCA 43, 3: seedling from DN135 (AF390499)
E. umbra[a,h]	*Eucalyptus*	*Amenta*	EAm	NSW/Qld	Waite Arboretum, #1537 (AF058505)
E. urophylla	*Symphyomyrtus*	*Latoangulatae*	SL	Domesticated, South Africa	#15 Zander Myburg, S. Afr. (HM596068)
E. viminalis ssp. *viminalis*	*Symphyomyrtus*	*Maidenaria*	SM	Leith, NW Tas	UTAS 923 (HM596069)
E. wandoo ssp. *wandoo*[b]	*Symphyomyrtus*	*Bisectae (I)*	SB1	Stirling Range NP, WA	CCA 23, 31: seedling from DN 230 (AF390497)
E. woodwardii[a]	*Symphyomyrtus*	*Dumaria*	SD	Southern WA	Waite Arboretum, #136 (AF058479)

Samples with superscripts were included in previous studies (ITS-based phylogenies): [a]Steane *et al*. 1999; [b]Steane *et al*. 2002; [c]Steane *et al*. 2007; [d]Whittock *et al*. 2003.[e]This sample was listed as *E. alba* by Steane *et al*. (2002), a close relative of *E. platyphylla* (same series) but not found in Papua New Guinea. [f]Brooker (2000) includes *E. biterranea* in the more well-known *E. pellita*. [g]*E. camaldulensis* ssp. *acuta* or ssp. *simulata* - both subspecies occur in this region but not enough morphological information is available to determine the classification of this sample. [h]*E. arenicola* was listed as *E. willisii* ssp. *willisii* by Steane *et al*. (1999).

Table 3 Sources of DArT markers that were used to genotype 94 species of *Eucalyptus* for phylogenetic analysis.

Source species	Subgenus [Section]	No. clones
E. globulus	*Symphyomyrtus* [*Maidenaria*]	2754
E. nitens	*Symphyomyrtus* [*Maidenaria*]	861
[Subtotal]	[*Maidenaria*]	[3615]
E. grandis	*Symphyomyrtus* [*Latoangulatae*]	2517
E. urophylla	*Symphyomyrtus* [*Latoangulatae*]	882
E. grandis x *urophylla*	*Symphyomyrtus* [*Latoangulatae*]	456
[Subtotal]	[*Latoangulatae*]	[3855]
E. camaldulensis	*Symphyomyrtus* [*Exsertaria*]	429
[Subtotal]	[*Symphyomyrtus*]	7899
E. pilularis	*Eucalyptus* [*Pseudophloius*]	455
[Subtotal]	[*Eucalyptus*]	455
Total		8354

Table 4. Proportion (percentage) of samples on each plate for which ≥95%, ≥90%, ≥85% and ≥80% of DArT markers were scorable (i.e., hybridisation between the sample being genotyped and the DArT marker could be scored unambiguously as either “present” or “absent”). For example, in Plate 1, 100% of samples had scorable binary data (i.e., 0, 1) for at least 90% of the markers used in the screening of that plate; in other words, 100% of samples had less than 10% missing data.

	No. Taxa	No. markers	Percentage scorable data				Total missing data per plate
			≥95%	≥90%	≥85%	≥80%	
Plate 1	7	7052	78%	100%			4.4%
Plate 2	4	4684	33%	86%	100%		6.0%
Phylogeny plate	94	8354	50%	95%		99%*	5.6%

* *E. gamophylla* had 24% missing data.

Table 5 Comparison of levels of polymorphism within sections *Maidenaria* and *Latoangulatae* (Plate 2) in DArT markers from different sources. The call rate is the percentage of samples that could be scored as "0" or "1" (i.e., not missing).

Marker source	Total number of markers from source used to screen Plate 2	No. (%) of markers polymorphic in *Maidenaria*	No. markers (%) with a call rate of 100% across *Maidenaria* samples	No. (%) of markers polymorphic in *Latoangulatae*	No. markers (%) with call rate of 100% across *Latoangulatae* samples
Section *Maidenaria*	1293	1006 (78%)	451 (35%)	1026 (79%)	370 (29%)
Section *Latoangulatae*	2425	1708 (70%)	981 (41%)	2107 (86%)	802 (33%)
Section *Exsertaria*	132	99 (75%)	35 (27%)	110 (83%)	38 (29%)
Subgenus *Eucalyptus*	109	89 (82%)	24 (22%)	93 (85%)	30 (28%)
Phylogeny plate	669	489 (73%)	245 (37%)	588 (88%)	204 (31%)
Corymbia	56	40 (71%)	29 (52%)	47 (84%)	26 (46%)
Mean percentage		74%	36%	84%	33%

Table 6 Probability values for partition homogeneity tests. *A significant P value suggests that a data partition contains phylogenetic signal that is different from that generated by random partitions of the data set.

Data partition	No. clones in partition	P value from ILD test
Subg. *Eucalyptus*	455	0.423
Sect. *Exsertaria*	429	0.115
Sect. *Latoangulatae*	3855	0.003*
Sect. *Maidenaria*	3615	0.013 *

Appendix A. *Eucalyptus* samples in Plate 1 of DArT study. *Eucalyptus grandis* BRASUZ1 is the tree from which the USA Department of Energy, Joint Genome Institute produced the first complete genome sequence for *Eucalyptus.* Abbreviations: NSW - New South Wales; SC Vic - Southern Central Victoria; SA- South Australia. N/A - not applicable because *Corymbia* does not have subgenera.

Species	**Subgenus, Section**	**Identifier**	**Provenance**
E. globulus	*Symphyomyrtus, Maidenaria*	1138	Strzelecki, Victoria
E. globulus	*Symphyomyrtus, Maidenaria*	1279	Sth Gippsland, Victoria
E. globulus	*Symphyomyrtus, Maidenaria*	1393	Furneaux, eastern Bass Strait
E. globulus	*Symphyomyrtus, Maidenaria*	1424	Furneaux, eastern Bass Strait
E. globulus	*Symphyomyrtus, Maidenaria*	1542	Southern Tas
E. globulus	*Symphyomyrtus, Maidenaria*	1573	Eastern Otways, Victoria
E. globulus	*Symphyomyrtus, Maidenaria*	1704	Western Otways, Victoria
E. globulus	*Symphyomyrtus, Maidenaria*	1707	Western Otways, Victoria
E. globulus	*Symphyomyrtus, Maidenaria*	1723	Northeastern Tasmania
E. globulus	*Symphyomyrtus, Maidenaria*	1768	Southeastern Tasmania
E. globulus	*Symphyomyrtus, Maidenaria*	2071	Western Tasmania
E. globulus	*Symphyomyrtus, Maidenaria*	ba0010	Light House (Vic) X Southern Tasmania
E. nitens	*Symphyomyrtus, Maidenaria*	8-151	Toolangi, SC Victoria
E. nitens	*Symphyomyrtus, Maidenaria*	8-185	Mt Wellington, Victoria
E. nitens	*Symphyomyrtus, Maidenaria*	7-206	Toorongo, SC Victoria
E. nitens	*Symphyomyrtus, Maidenaria*	8-155	Toolangi, SC Victoria
E. nitens	*Symphyomyrtus, Maidenaria*	5-201	Toorongo, SC Victoria
E. nitens	*Symphyomyrtus, Maidenaria*	4-173	Mt Erica, Thomson Valley, SC Victoria
E. nitens	*Symphyomyrtus, Maidenaria*	17-2,5	N. NSW
E. nitens	*Symphyomyrtus, Maidenaria*	8-207	N Toorongo, SC Victoria
E. nitens	*Symphyomyrtus, Maidenaria*	17-9,16	S. NSW
E. nitens	*Symphyomyrtus, Maidenaria*	CP20(1)	Connors Plains, Victoria
E. nitens	*Symphyomyrtus, Maidenaria*	CP186(3)	Connors Plains, Victoria
E. grandis	*Symphyomyrtus, Latoangulatae*	1	Baldy State Forest, Queensland
E. grandis	*Symphyomyrtus, Latoangulatae*	2	Mareeba, Queensland
E. grandis	*Symphyomyrtus, Latoangulatae*	3	Ravenshoe, Queensland
E. grandis	*Symphyomyrtus, Latoangulatae*	4	Townsville, Queensland
E. grandis	*Symphyomyrtus, Latoangulatae*	5	Kenilworth, Queensland
E. grandis	*Symphyomyrtus, Latoangulatae*	6	Veteran Gympie, Queensland
E. grandis	*Symphyomyrtus, Latoangulatae*	8	Toonumba, NSW
E. grandis	*Symphyomyrtus, Latoangulatae*	9	Lake Cathie, NSW
E. grandis	*Symphyomyrtus, Latoangulatae*	11	Mt George, Taree, NSW
E. grandis	*Symphyomyrtus, Latoangulatae*	12	Coffs Harbour, NSW
E. grandis	*Symphyomyrtus, Latoangulatae*	13	Parent of mapping population
E. grandis	*Symphyomyrtus, Latoangulatae*	BRASUZ1	DOE-JGI target genome
E. urophylla	*Symphyomyrtus, Latoangulatae*	2	Lere-Baukrenget, Indonesia
E. urophylla	*Symphyomyrtus, Latoangulatae*	3	Ile Nggele, Indonesia
E. urophylla	*Symphyomyrtus, Latoangulatae*	4	Jontona, Indonesia
E. urophylla	*Symphyomyrtus, Latoangulatae*	5	Labalekan, Indonesia
E. urophylla	*Symphyomyrtus, Latoangulatae*	6	Beangonong, Indonesia
E. urophylla	*Symphyomyrtus, Latoangulatae*	7	Delaki, Indonesia
E. urophylla	*Symphyomyrtus, Latoangulatae*	8	Bonleu, Indonesia
E. urophylla	*Symphyomyrtus, Latoangulatae*	9	Mollo, Indonesia
E. urophylla	*Symphyomyrtus, Latoangulatae*	10	Pintu Mas, Indonesia
E. urophylla	*Symphyomyrtus, Latoangulatae*	11	Apui, Indonesia
E. urophylla	*Symphyomyrtus, Latoangulatae*	12	Nesunhuhun, Indonesia
E. urophylla	*Symphyomyrtus, Latoangulatae*	13	Parent of mapping population
E. camaldulensis	*Symphyomyrtus, Exsertaria*	3R33B	Lake Albacutya E, Victoria
E. camaldulensis	*Symphyomyrtus, Exsertaria*	3R32A	Kororoit Ck, Melton, Victoria
E. camaldulensis	*Symphyomyrtus, Exsertaria*	2R32D	Edenhope, Victoria
E. camaldulensis	*Symphyomyrtus, Exsertaria*	2R34B	Edenhope, Victoria
E. camaldulensis	*Symphyomyrtus, Exsertaria*	PET32	Petford, Queensland
E. camaldulensis	*Symphyomyrtus, Exsertaria*	PET116	Petford, Queensland
E. camaldulensis	*Symphyomyrtus, Exsertaria*	CST01009	Lake Albacutya, Victoria
E. camaldulensis	*Symphyomyrtus, Exsertaria*	B10530	Mitchell, Queensland

E. camaldulensis	*Symphyomyrtus, Exsertaria*	B10540	Morehead, Queensland
E. camaldulensis	*Symphyomyrtus, Exsertaria*	B10531	Laura, Queensland
E. camaldulensis	*Symphyomyrtus, Exsertaria*	B10626	Palmer, Queensland
E. cladocalyx	*Symphyomyrtus, Sejunctae*	DN 2569	Horrocks Pass, Flinders Ranges, SA
E. cladocalyx	*Symphyomyrtus, Sejunctae*	DN 3182	Port Lincoln, Eyre Peninsula, SA
E. cladocalyx	*Symphyomyrtus, Sejunctae*	K006	Kangaroo Is, SA
E. cladocalyx	*Symphyomyrtus, Sejunctae*	K046	Kangaroo Is, SA
E. cladocalyx	*Symphyomyrtus, Sejunctae*	K047	Kangaroo Is, SA
E. cladocalyx	*Symphyomyrtus, Sejunctae*	K048	Kangaroo Is, SA
E. cladocalyx	*Symphyomyrtus, Sejunctae*	K051	Kangaroo Is, SA
E. cladocalyx	*Symphyomyrtus, Sejunctae*	W002	Wirrabara, Flinders Ranges, SA
E. cladocalyx	*Symphyomyrtus, Sejunctae*	W004	Wirrabara, Flinders Ranges, SA
E. cladocalyx	*Symphyomyrtus, Sejunctae*	W012	Wirrabara, Flinders Ranges, SA
E. cladocalyx	*Symphyomyrtus, Sejunctae*	W025	Wirrabara, Flinders Ranges, SA
E. cladocalyx	*Symphyomyrtus, Sejunctae*	W027	Wirrabara, Flinders Ranges, SA
E. pilularis	*Eucalyptus, Pseudophloia*	9025	Goonengerry, NSW
E. pilularis	*Eucalyptus, Pseudophloia*	9468	Tamban, NSW
E. pilularis	*Eucalyptus, Pseudophloia*	9479	Gallangowan, Queensland
E. pilularis	*Eucalyptus, Pseudophloia*	9659	Olney, NSW
E. pilularis	*Eucalyptus, Pseudophloia*	9720	Kiwarrak, NSW
E. pilularis	*Eucalyptus, Pseudophloia*	9742	Bulga, NSW
E. pilularis	*Eucalyptus, Pseudophloia*	9789	Clouds Creek, NSW
E. pilularis	*Eucalyptus, Pseudophloia*	9910	Whain Whain, NSW
E. pilularis	*Eucalyptus, Pseudophloia*	9375	Orara East, NSW
E. pilularis	*Eucalyptus, Pseudophloia*	9415	Coopernook, NSW
E. pilularis	*Eucalyptus, Pseudophloia*	9464	Kerewong, NSW
E. pilularis	*Eucalyptus, Pseudophloia*	9519	Newry, NSW
Corymbia variegata	*N/A, Politaria*	4893	Wedding Bells, NSW
Corymbia variegata	*N/A, Politaria*	7782	Richmond Ranges, NSW
Corymbia variegata	*N/A, Politaria*	7785	Presho, SW of Yeppoon, Queensland
Corymbia variegata	*N/A, Politaria*	7797	Brisbane Forest Park, Queensland
Corymbia variegata	*N/A, Politaria*	7989	Woondum, Queensland
Corymbia variegata	*N/A, Politaria*	7991	Woondum, Queensland
Corymbia variegata	*N/A, Politaria*	10178	Cherry Tree, NSW
Corymbia variegata	*N/A, Politaria*	10394	Brooyar, W of Gympie, Queensland
Corymbia variegata	*N/A, Politaria*	10406	Ewingar, NSW
Corymbia variegata	*N/A, Politaria*	10262	Cherry Tree, NSW
Corymbia variegata	*N/A, Politaria*	10286	Woondum, Queensland
Corymbia variegata	*N/A, Politaria*	10348	Candole, NSW

Appendix B. *Eucalyptus* samples used in Plate 2 of DArT study. Abbreviations: WPLH - Wilson's Promontory Lighthouse, Victoria; NSW - New South Wales.

	Species	**Identifier**	**Race**
1	*E. globulus*	1150	Strzelecki, Victoria
2	*E. globulus*	1160	Strzelecki, Victoria
3	*E. globulus*	1253	Southern Gippsland, Victoria
4	*E. globulus*	1256	Southern Gippsland, Victoria
5	*E. globulus*	1390	Furneaux, eastern Bass Strait
6	*E. globulus*	1408	Furneaux, eastern Bass Strait
7	*E. globulus*	1540	Southern Tasmania
8	*E. globulus*	1556	Eastern Otways, Victoria
9	*E. globulus*	1581	Eastern Otways, Victoria
10	*E. globulus*	1600	Western Otways, Victoria
11	*E. globulus*	1716	Western Otways, Victoria
12	*E. globulus*	1743	Northeastern Tasmania
13	*E. globulus*	1751	Southeastern Tasmania
14	*E. globulus*	1762	Northeastern Tasmania
15	*E. globulus*	1814	King Island, western Bass Strait
16	*E. globulus*	2059	Western Tasmania
17	*E. globulus*	2143	King Island, western Bass Strait
18	*E. globulus*	2583	Wilson's Promontory Lighthouse, Victoria
19	*E. globulus*	2627	Recherche Bay, southern Tasmania
20	*E. globulus*	3356	Dromedary, southeastern Tasmania
21	*E. globulus*	3411	Port Davey, Western Tasmania
22	*E. globulus*	9904	Tidal River, Wilson's Promontory, Victoria
23	*E. globulus*	9930	Phillip Island, Victoria
24	*E. globulus*	ba0012	WPLH X Southern Tasmania
25	*E. globulus*	OP1	unknown
26	*E. globulus*	OP2	unknown
27	*E. globulus*	OP3	unknown
28	*E. globulus*	OP4	unknown
29	*E. globulus*	OP5	unknown
30	*E. globulus*	OP6	unknown
31	*E. globulus*	OP7	unknown
32	*E. globulus*	OP8	unknown
33	*E. globulus*	OP9	unknown
34	*E. globulus*	OP10	unknown
35	*E. globulus*	OP11	unknown
36	*E. globulus*	OP12	unknown
37	*E. globulus*	Port1	Portugal
38	*E. globulus*	Port2	Portugal
39	*E. globulus*	Ch1	Chivilingo, Chile
40	*E. globulus*	Ch2	Chivilingo, Chile

41	*E. globulus*	Ch3	Manzano Miramar, Chile
42	*E. globulus*	Ch4	Cerro Alto, Chile
43	*E. globulus*	Ch5	Araneda, Chile
44	*E. globulus*	Ch6	Araneda, Chile
45	*E. globulus*	Ch7	Manzano Miramar, Chile
46	*E. globulus*	Ch8	Maquehua, Chile
47	*E. globulus*	Ch9	Araneda, Chile
48	*E. globulus*	Ch10	Cerro Alto, Chile
49	*E. globulus*	Ch11	Araneda, Chile
50	*E. nitens*	Ch12	Tallaganda State Forest, NSW
51	*E. nitens*	Ch13	Badja State Forest, NSW
52	*E. nitens*	Ch14	Tallaganda State Forest, NSW
53	*E. nitens*	Ch15	Tallaganda State Forest, NSW
54	*E. nitens*	Ch16	Tallaganda State Forest, NSW
55	*E. nitens*	Ch17	Thomson Valley, Victoria
56	*E. nitens x globulus*	Ch18	Cuatro del recorte, Chile
57	*E. nitens x globulus*	Ch19	Cuatro del recorte, Chile
58	*E. nitens x globulus*	Ch20	Cuatro del recorte, Chile
59	*E. nitens x globulus*	Ch21	La Huina, Chile
60	*E. grandis*	grand13	Mareeba, Queensland
61	*E. grandis*	grand14	Townsville, Queensland
62	*E. grandis*	grand15	Baldy State Forest, Queensland
63	*E. grandis*	grand16	Woondum/Gympie, Queensland
64	*E. grandis*	grand17	Kenilworth, Queensland
65	*E. grandis*	grand18	Belthorpe, Queensland
66	*E. grandis*	grand19	Wauchope, NSW
67	*E. grandis*	grand20	Lake Cathie, NSW
68	*E. grandis*	grand21	Mt George Taree, NSW
69	*E. grandis*	grand22	Taree, NSW
70	*E. grandis*	grand23	Bulahdelah, NSW
71	*E. grandis*	grand24	Wauchope, NSW
72	*E. urophylla*	uro13	Lasinisir, Indonesia
73	*E. urophylla*	uro14	Rotus, Indonesia
74	*E. urophylla*	uro15	Lasinisir, Indonesia
75	*E. urophylla*	uro16	Lembata, Indonesia
76	*E. urophylla*	uro17	Padekluwa Lembata, Indonesia
77	*E. urophylla*	uro18	Lembata, Indonesia
78	*E. urophylla*	uro19	Padekluwa Lembata, Indonesia
79	*E. urophylla*	uro20	Lelobatan, Timor, Indonesia
80	*E. urophylla*	uro21	Leloboko, Timor, Indonesia
81	*E. urophylla*	uro22	Kilawair Ille, Wodong, Flores, Indonesia
82	*E. urophylla*	uro23	Hokeng, Flores, Indonesia
83	*E. urophylla*	uro24	Leloboko, Timor, Indonesia
84	*E. urophylla x grandis*	Fr1	France

85	*E. urophylla x grandis*	Fr3	France
86	*E. urophylla x grandis*	Fr5	France
87	*E. urophylla x grandis*	Fr7	France
88	*E. urophylla x grandis*	Fr11	France
89	*E. urophylla x grandis*	Fr13	France
90	*E. urophylla x grandis*	Fr15	France
91	*E. urophylla*	Fr17	France
92	*E. grandis*	Fr24	France
93	*E. urophylla*	Fr20	France
94	*E. urophylla*	Fr21	France

Text for Supplementary Material

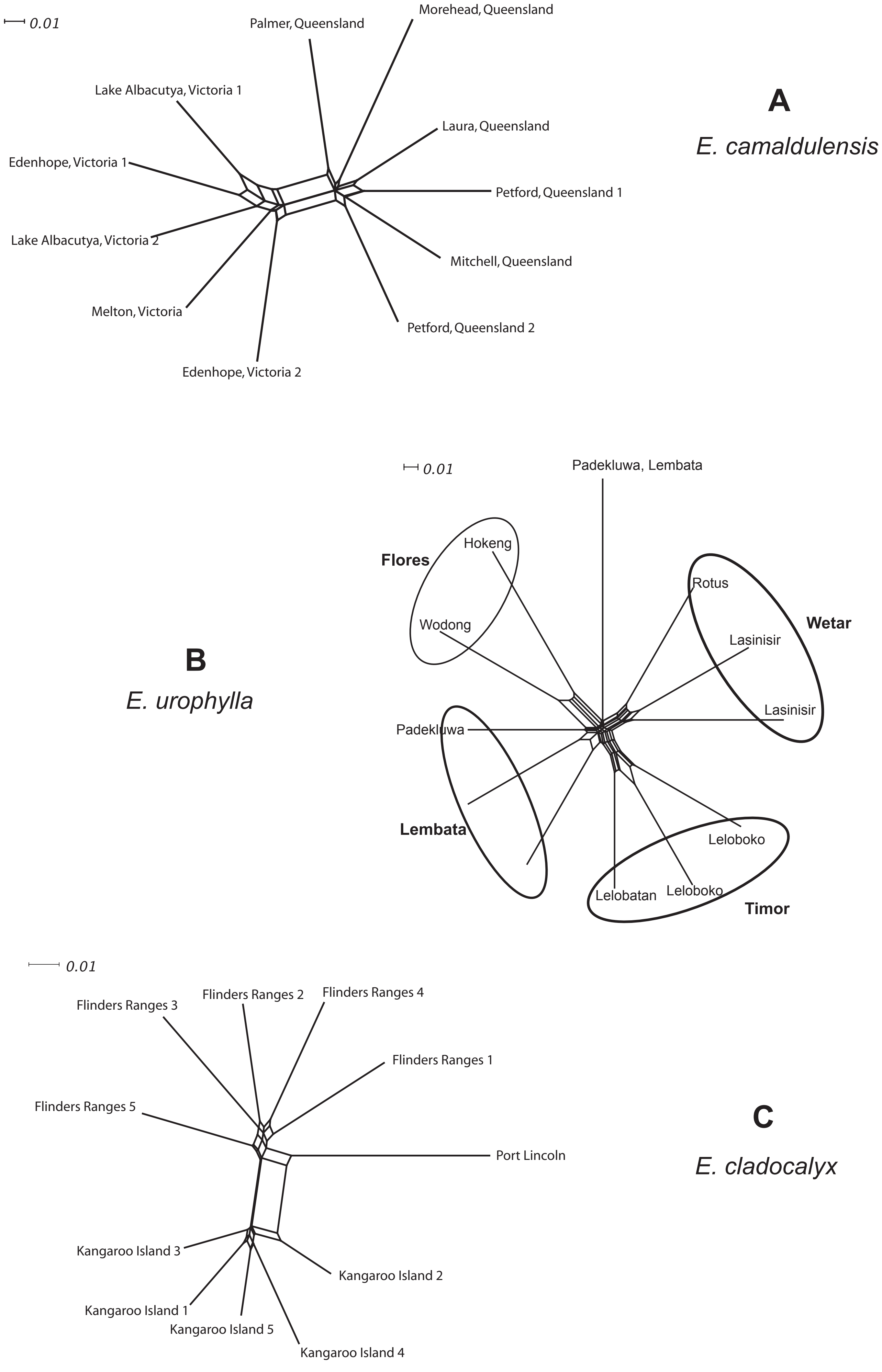
Supplementary Material Figure S1

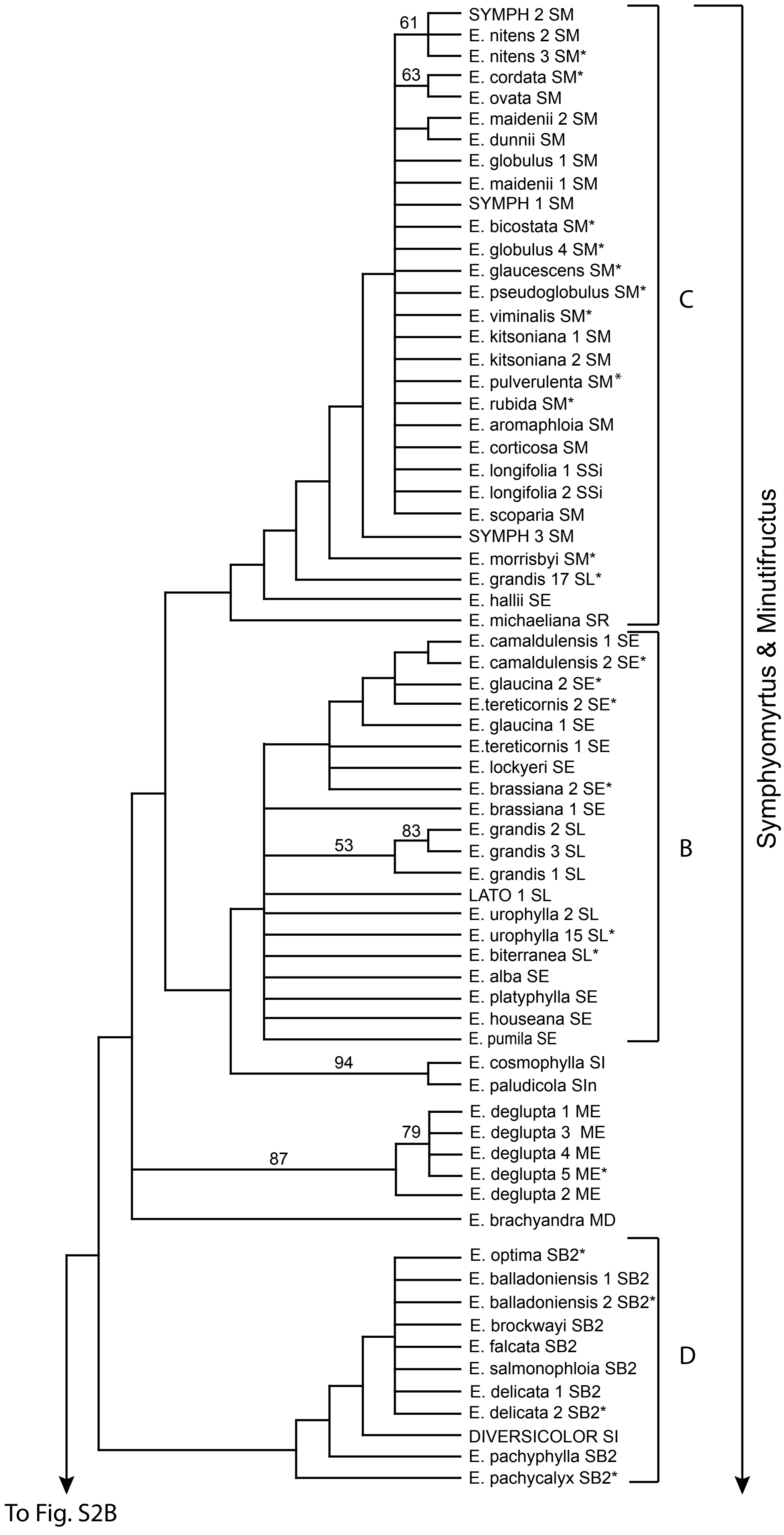
Splitstree 4 networks showing geographic partitioning of samples of (A) *E. camaldulensis* (Queensland and Victoria), (B) *E. urophylla* (Indonesia) and (C) *E. cladocalyx* (South Australia). Data for A and B came from Plate 1; data for C came from Plate 2 (commercial species). Precise locality information was not available for two of the samples from the island of Lembata.

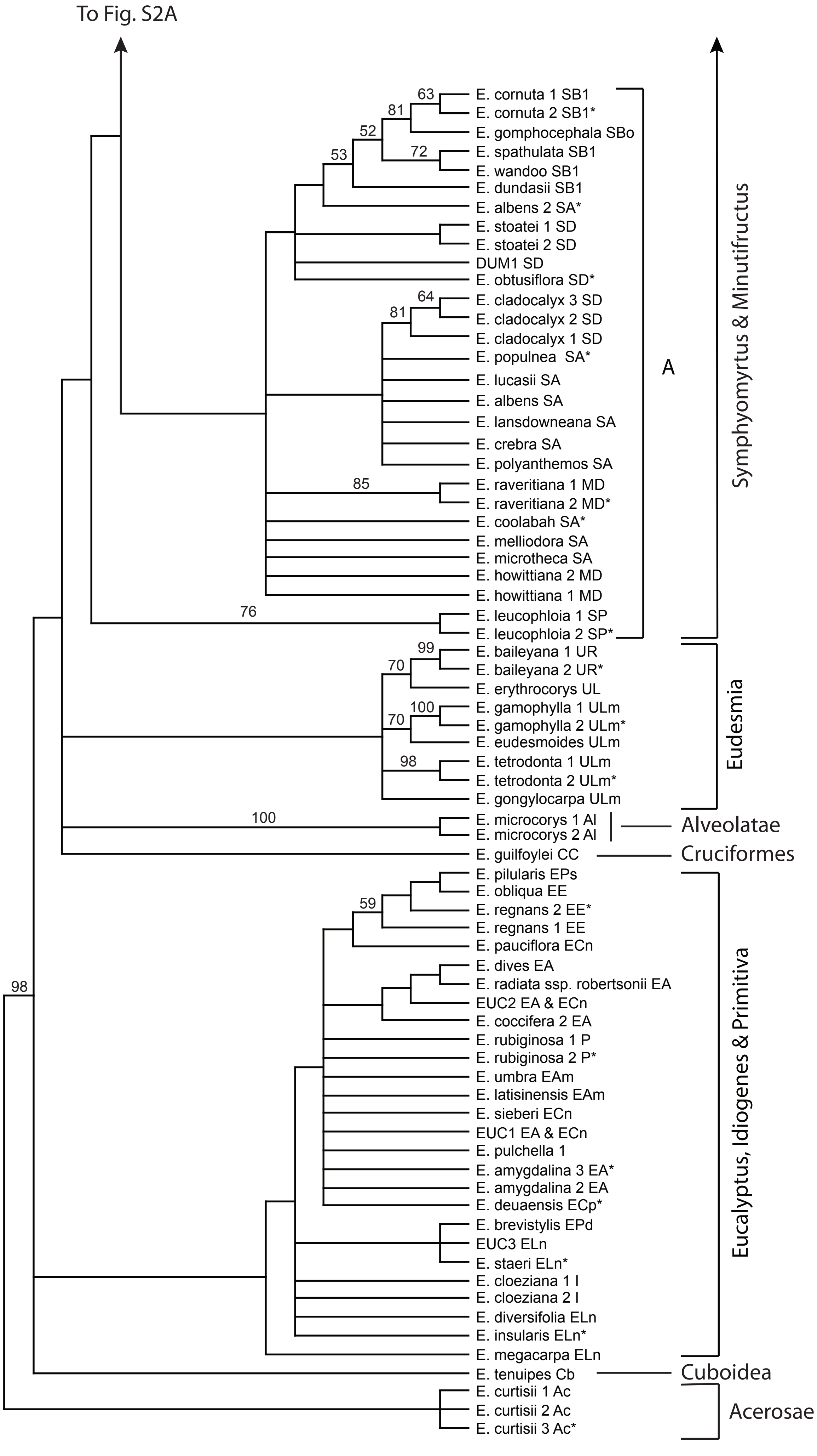
Supplementary Material Figure S2 (A and B).

Strict consensus of 39,600 equally most parsimonious trees (length = 473 steps; CI = 0.524; RI = 0.862) derived from MP analysis of ITS DNA sequence data. The 39 samples that are labelled with asterisks are the samples for which ITS sequences were acquired for this study. Bootstrap values greater than 50% are shown above branches. Subgenera are indicated to the right of the cladogram. Clades A-D are groups within subgenus *Symphyomyrtus* that have been discussed previously (see Steane *et al*. 1999, 2002, 2007; Whittock *et al*. 2003). Only one sample, *E. grandis* 17 SL (Clade C), did not come out in a position congruent with closely related samples (Clade B). Refer to Table 2 of main text for full species details and explanations of taxon abbreviations (subgenus, section) after each species name. Groups of species from previous studies (Steane *et al*. 1999, 2002) that had identical ITS sequences were clustered into single operational taxonomic units: EUC1 – *E. amygdalina* 1, *E. coccifera* 1, *E. coccifera* 3, *E. croajingolensis*, *E. elata*, *E. piperita*, *E. tindaliae*, *E. pulchella* 2, *E. risdonii*, *E. tenuiramis* 1 and 2, *E. willisii* ssp. *falciformis*, *E. arenicola* (formerly *E. willisii* ssp. *willisii*); EUC2 – *E. delegatensis*, *E. nitida*; EUC3 – *E. jacksonii*, *E. staeri*, *E. marginata*. DUM1 – *E. torquata*, *E. obtusiflora*; LATO1 – *E. urophylla* 1, *E. urophylla* 3, *E. pellita*, *E. wetarensis*; SYMPH1 – *E. gunnii* 1, *E. gunnii* 2, *E. perenniana* 1, *E. perenniana* 2, *E.*

dalrympleana; SYMPH2 – *E. nitens* 1, *E. nitens* 3, *E. pseudoglobulus, E. globulus* 2; SYMPH3 – *E. globulus* 3, *E. bicostata*; DIVERSICOLOR – *E. diversicolor* 1, 2, 3.


0.01
Palmer, Queensland
Morehead, Queensland
A
E. camaldulensis
Lake Albacutya, Victoria 1
Laura, Queensland
Edenhope, Victoria 1
Petford, Queensland 1
Lake Albacutya, Victoria 2
Mitchell, Queensland
Melton, Victoria
Petford, Queensland 2
Edenhope, Victoria 2
0.01
Padekluwa, Lembata
Flores
Hokeng
Wodong
Rotus
Wetar
Lasinisir
Lasinisir
B
E. urophylla
Padekluwa
Lembata
Leloboko
Lelobatan
Leloboko
Timor
0.01
Flinders Ranges 3
Flinders Ranges 2
Flinders Ranges 4
Flinders Ranges 1
Flinders Ranges 5
C
Port Lincoln
E. cladocalyx
Kangaroo Island 3
Kangaroo Island 2
Kangaroo Island 1
Kangaroo Island 5
Kangaroo Island 4


SYMPH 2 SM
E. nitens 2 SM
E. nitens 3 SM*
E. cordata SM*
E. ovata SM
E. maidenii 2 SM
E. dunnii SM
E. globulus 1 SM
E. maidenii 1 SM
SYMPH 1 SM
E. bicostata SM*
E. globulus 4 SM*
E. glaucescens SM*
E. pseudoglobulus SM*
E. viminalis SM*
E. kitsoniana 1 SM
E. kitsoniana 2 SM
E. pulverulenta SM*
E. rubida SM*
E. aromaphloia SM
E. corticosa SM
E. longifolia 1 SSi
E. longifolia 2 SSi
E. scoparia SM
SYMPH 3 SM
E. morrisbyi SM*
E. grandis 17 SL*
E. hallii SE
E. michaeliana SR
E. camaldulensis 1 SE
E. camaldulensis 2 SE*
E. glaucina 2 SE*
E.tereticornis 2 SE*
E. glaucina 1 SE
E.tereticornis 1 SE
E. lockyeri SE
E. brassiana 2 SE*
E. brassiana 1 SE
E. grandis 2 SL
E. grandis 3 SL
E. grandis 1 SL
LATO 1 SL
E. urophylla 2 SL
E. urophylla 15 SL*
E. biterranea SL*
E. alba SE
E. platyphylla SE
E. houseana SE
E. pumila SE
E. cosmophylla SI
E. paludicola SIn
E. deglupta 1 ME
E. deglupta 3 ME
E. deglupta 4 ME
E. deglupta 5 ME*
E. deglupta 2 ME
E. brachyandra MD
E. optima SB2*
E. balladoniensis 1 SB2
E. balladoniensis 2 SB2*
E. brockwayi SB2
E. falcata SB2
E. salmonophloia SB2
E. delicata 1 SB2
E. delicata 2 SB2*
DIVERSICOLOR SI
E. pachyphylla SB2
E. pachycalyx SB2*
61
63
83
53
94
79
87
C
B
D
Symphyomyrtus & Minutifructus
To Fig. S2B


To Fig. S2A
63
81
52
72
53
E. cornuta 1 SB1
E. cornuta 2 SB1*
E. gomphocephala SBo
E. spathulata SB1
E. wandoo SB1
E. dundasii SB1
E. albens 2 SA*
E. stoatei 1 SD
E. stoatei 2 SD
DUM1 SD
E. obtusiflora SD*
64
81
E. cladocalyx 3 SD
E. cladocalyx 2 SD
E. cladocalyx 1 SD
E. populnea SA*
E. lucasii SA
E. albens SA
E. lansdowneana SA
E. crebra SA
E. polyanthemos SA
85
E. raveritiana 1 MD
E. raveritiana 2 MD*
E. coolabah SA*
E. melliodora SA
E. microtheca SA
E. howittiana 2 MD
E. howittiana 1 MD
76
E. leucophloia 1 SP
E. leucophloia 2 SP*
A
Symphyomyrtus & Minutifructus
99
70
E. baileyana 1 UR
E. baileyana 2 UR*
E. erythrocorys UL
100
70
E. gamophylla 1 ULm
E. gamophylla 2 ULm*
E. eudesmoides ULm
98
E. tetrodonta 1 ULm
E. tetrodonta 2 ULm*
E. gongylocarpa ULm
Eudesmia
100
E. microcorys 1 Al
E. microcorys 2 Al
Alveolatae
E. guilfoylei CC
Cruciformes
59
E. pilularis EPs
E. obliqua EE
E. regnans 2 EE*
E. regnans 1 EE
E. pauciflora ECn
E. dives EA
E. radiata ssp. robertsonii EA
EUC2 EA & ECn
E. coccifera 2 EA
E. rubiginosa 1 P
E. rubiginosa 2 P*
E. umbra EAm
E. latisinensis EAm
E. sieberi ECn
EUC1 EA & ECn
E. pulchella 1
E. amygdalina 3 EA*
E. amygdalina 2 EA
E. deuaensis ECp*
E. brevistylis EPd
EUC3 ELn
E. staeri ELn*
E. cloeziana 1 I
E. cloeziana 2 I
E. diversifolia ELn
E. insularis ELn*
E. megacarpa ELn
Eucalyptus, Idiogenes & Primitiva
98
E. tenuipes Cb
Cuboidea
E. curtisii 1 Ac
E. curtisii 2 Ac
E. curtisii 3 Ac*
Acerosae