ARTICLE

Speciation in the presence of gene flow: population genomics of closely related and diverging *Eucalyptus* species

Susan Rutherford[1,2] · Maurizio Rossetto[2] · Jason G. Bragg[2] · Hannah McPherson[2] · Doug Benson[2] ·
Stephen P. Bonser[1] · Peter G. Wilson[2]

Abstract

Speciation is a complex process that is fundamental to the origins of biological diversity. While there has been considerable progress in our understanding of speciation, there are still many unanswered questions, especially regarding barriers to gene flow in diverging populations. *Eucalyptus* is an appropriate system for investigating speciation mechanisms since it comprises species that are rapidly evolving across heterogeneous environments. We examined patterns of genetic variation within and among six closely related *Eucalyptus* species in subgenus *Eucalyptus* section *Eucalyptus* in south-eastern Australia (commonly known as the "green ashes"). We used reduced representation genome sequencing to genotype samples from populations across altitudinal and latitudinal gradients. We found one species, *Eucalyptus cunninghamii*, to be highly genetically differentiated from the others, and a population of mallees from Mount Banks to be genetically distinct and therefore likely to be a new undescribed species. Only modest levels of differentiation were found between all other species in the study. There was population structure within some species (e.g., *E. obstans*) corresponding to geographical factors, indicating that vicariance may have played a role in the evolution of the group. Overall, we found that lineages within the green ashes are differentiated to varying extents, from strongly diverged to much earlier stages of the speciation continuum. Furthermore, our results suggest the green ashes represent a group where a range of mechanisms (e.g., reticulate evolution and vicariance) have been operating in concert. These findings not only offer insights into recent speciation mechanisms in *Eucalyptus*, but also other species complexes.

Introduction

Speciation is responsible for the origin of biological diversity (Gavrilets 2003), and typically begins with a barrier to gene flow, which promotes further genetic and phenotypic divergence (Nosil and Feder 2012). Species are traditionally defined as a group of interbreeding populations that are reproductively isolated from other such groups (i.e., the biological species concept; Mayr 1963). However, more recently there has been a significant change in our understanding of the role of barriers to gene flow between diverging populations (Coyne and Orr 2004). It is now widely recognised that hybridisation between lineages can be common, leading to increased levels of genealogical discordance among loci within the genome (Lexer and Widmer 2008). In the genic view of speciation, species divergence occurs along a continuum of genetic differentiation (Wu 2001), with incipient species passing through a phase where they are only partly reproductively isolated (Kopp and Frank 2005). Evidence of partial speciation has been documented in many organisms, such as *Drosophila* (Legrand et al. 2011) and in the plant genus, *Nuphar* (Shiga and Kadono 2007). Lineages that are in the early stages of speciation offer important opportunities for understanding the mechanisms that drive species divergence (Schield et al. 2015). In this study, we examine patterns of diversification within *Eucalyptus*.

✉ Susan Rutherford
Susan.Rutherford@rbgsyd.nsw.gov.au

1 Evolution and Ecology Research Centre, School of Biological, Earth and Environmental Sciences, UNSW, Kensington, Sydney, Australia

2 National Herbarium of NSW, Royal Botanic Garden Sydney, Sydney, Australia
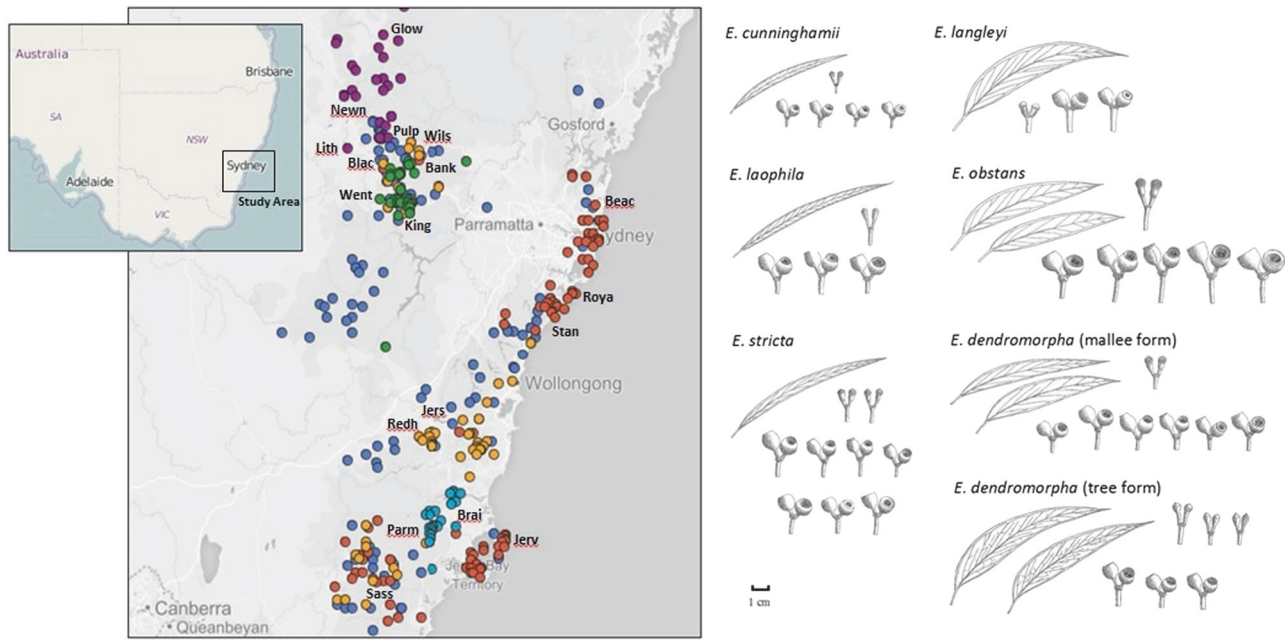
Fig. 1 Distribution of the study species showing: *Eucalyptus cunninghamii* (●), *E. laophila* (●), *E. stricta* (●), *E. langleyi* (●), *E. obstans* (●), and *E. dendromorpha* (●). Leaf morphology and variations in bud and capsule size of each species are also shown (Klaphake 2012, pp. 45–49). Locations sampled in the present study are shown. Details of species and populations sampled (and location codes) are presented in Table 1. Maps were generated using Australia's Virtual Herbarium (2015)

Eucalyptus is a large genus (comprising more than 700 species) in the family Myrtaceae that forms a significant component of the Australian biota (Smith et al. 2003). Although *Eucalyptus* is Gondwanan in origin (with fossils dated 51.9 Ma, Gandolfo et al. 2011), the present-day dominance of this genus across Australia is considered to be relatively recent (Crisp et al. 2004). Pollen evidence indicates that *Eucalyptus* became widespread in the Pleistocene (1.5–5 Ma; Martin 1982; Kershaw et al. 1994), and molecular studies suggest the occurrence of rapid radiations during the Quaternary (McKinnon et al. 2004). In southeastern Australia, eucalypts are particularly diverse (in terms of both morphological and lineage diversity), especially in the sub-coastal area of central New South Wales including Sydney and the Blue Mountains (Wardell-Johnson et al. 1997). The latter region was declared a World Heritage Area partly due to its high eucalypt diversity (c. 100 species; Hager and Benson 2010). One diverse group of eucalypts from this area is the green ashes (subgenus *Eucalyptus* section *Eucalyptus*, *sensu* Brooker 2000), which includes tall trees on fertile soils, as well as smaller trees and mallees (multi-stemmed plants, usually less than 10 m tall) on shallow soils (Ladiges et al. 2010). While the larger green ashes can be fire sensitive (Nicolle 2006), mallees grow from an underground lignotuber from which they can resprout following fire (Mullette 1978). Many green ash species are morphologically similar (Fig. 1), have overlapping distributions, and cases of hybridisation have been reported (Johnson and Blaxell 1972). As such, the green

ashes represent an appropriate group to study active evolutionary processes.

Speciation has influenced the diversification of present-day species (Barraclough and Nee 2001) and a comprehensive understanding of population genetic differentiation along environmental gradients is therefore important when investigating speciation mechanisms of taxa (Gaudeul et al. 2012). Genetic variation within a species often has a geographic basis since the processes of adaptation, gene flow, and genetic drift act differentially across heterogeneous landscapes and may be strongly influenced by the demography and spatial distributions of populations (Eckert et al. 2008). Previous genetic studies have demonstrated how a range of speciation mechanisms have played an important role in the evolution of different groups of eucalypts. For example, McGowen et al. (2001) investigated genetic differentiation (using nuclear microsatellites) in closely related eucalypts along steep mountains in Tasmania, and found that *Eucalyptus vernicosa* at higher altitudes had evolved in allopatry from *E. subcrenulata* from lower altitudes. Alternatively, Foster et al. (2007) found, using both nuclear and chloroplast (cp) DNA markers, that morphologically distinct tree and dwarf forms of *E. globulus* were maintained despite being geographically close, indicating that ecotypes, as well as species, could evolve in parapatry. Hybridisation and incomplete lineage sorting are thought to be important evolutionary processes in the diversification of *Eucalyptus*. For example, cpDNA variation in eucalypts from varying elevations in Victoria was found to be more

correlated with geography than species identity, suggesting past or current inter-specific hybridisation (Pollock et al. 2013). The poor morphological resolution of many present-day *Eucalyptus* species complexes suggests that they may be the result of recent and ongoing speciation in conjunction with past and current hybridisation (Griffin et al. 1988). The hypothesis of recent speciation in many eucalypt groups is supported by observations of inter-specific hybridisation in extant *Eucalyptus* species (e.g., Vaillancourt et al. 1994; McKinnon et al. 2001; Field et al. 2011).

The advent of next-generation sequencing (NGS) and associated technologies has enabled a much higher genomic resolution for the study of speciation mechanisms compared with traditional molecular methods (Keller et al. 2013). One technique that is increasingly being used in *Eucalyptus* is Diversity Arrays Technology sequencing (DArT, e.g., Steane et al. 2011; Rutherford et al. 2016). DArT is based on genome complexity reduction using restriction enzymes, followed by hybridisation to microarrays to simultaneously assay thousands of markers across the genome (Jaccoud et al. 2001). This method has recently been used in combination with NGS to develop DArTseq, which provides at least three times as many markers as the microarray DArT method, as well as an additional set of co-dominant single-nucleotide polymorphisms (SNPs, Sansaloni et al. 2011). With technological advances such as DArTseq, evolutionary processes at a genomic scale can now be investigated and genetic variation across landscapes can be examined in greater detail (Bragg et al. 2015).

The phylogeny of the green ashes was previously estimated using the DArT microarray method (Rutherford et al. 2016). It was found that while some of the lineage relationships were consistent with previous taxonomic classifications primarily based on morphology, other relationships were not. Recent speciation events may result in incongruence between gene trees, which can lead to difficulty in estimating relationships among taxa (Rieseberg and Brouillet 1994; Seehausen et al. 2014). Poor molecular resolution of taxa can also be due to hybridisation or incomplete lineage sorting (Jones et al. 2013). Reticulate evolution (hybridisation between divergent lineages) has been suggested to have played a role in the evolution of the green ashes (Hager and Benson 2010). However, "reticulate evolution" is poorly described by traditional evolutionary models (i.e., those based on the assumption that evolution is successfully captured by a bifurcating tree; Mindell 2013). Since a preliminary allozyme study by Prober et al. (1990) found low genetic differentiation between the green ashes, a population genomic study is likely to provide novel insights into the evolutionary history of these closely related lineages.

In the current study, we used genome-wide DArTseq markers to investigate the evolutionary origins and population genomics of six green ash species with overlapping distributions across altitudinal and latitudinal gradients in south-eastern Australia. Our objective was to improve our understanding of speciation mechanisms within the green ash group. In particular, based on previous phylogenetic analyses (Rutherford et al. 2016), we aimed to: (1) examine patterns of genetic differentiation among populations and taxa, (2) explore the role of inter-specific hybridisation, (3) investigate associations between genetic differentiation and geographic factors, and (4) better delineate species boundaries within the green ashes.

Materials and methods

Study system and sampling strategy

The green ashes have long been taxonomically challenging, with much disagreement over the number of recognised species. Many of the species recognised by Hill (2002) (e.g., *E. laophila* and *E. obstans*) were not recognised by Brooker (2000). To ensure that the species concepts of all major authorities were considered in our study, we followed the narrower species descriptions of Hill (2002). We selected six green ash species (Fig. 1), five of which were very closely related (i.e., in the same clade based on previous phylogenetic analyses; Rutherford et al. 2016). The sixth species, *E. cunninghamii*, are morphologically distinctive and in a previous phylogenetic study was found to be in a clade separate from the other study species (Rutherford et al. 2016). We chose to include *E. cunninghamii* in the present study so that we could compare taxa that are difficult to distinguish morphologically and genetically with a species that was morphologically and genetically differentiated, thereby providing a context for our findings. Sampling was designed to cover the distribution of each species across a range of latitudes and altitudes in Sydney, the South Coast, the Southern Highlands, and the Greater Blue Mountains World Heritage Area (GBMWHA) (Table 1). Locations of populations were obtained from the National Herbarium of New South Wales database, Benson and McDougall (1998) and Mills (2010); and are summarised as follows. Although the area sampled covers the full distribution of the rare and restricted species, the more widespread species, *E. stricta* and *E. dendromorpha*, extend further south. The southern extremity of *E. dendromorpha* is in Monga National Park (c. 270 km south of Sydney), while populations of *E. stricta* extend to Batemans Bay (280 km south of Sydney).

The study region is environmentally heterogeneous (0–1200 m a.s.l., 700–1800 mm annual rainfall). *Eucalyptus stricta* is widespread throughout this region, occurring in coastal, upland, and highland habitats. The other green ash

Table 1 Location details of species and populations of green ash eucalypts sampled for genomic DNA

Species	Location	Location code	Region	Growth form	Latitude (S)	Longitude (E)	Altitude (m)	Sample size	Sample code
E. cunninghamii	Mount Banks	Bank	GBMWHA	Mallee	−33.58	150.37	955–960	5	cunninghamii Bank
	Pulpit Rock	Pulp	GBMWHA	Mallee	−33.62	150.33	920–935	8	cunninghamii Pulp
	Wentworth Falls	Went	GBMWHA	Mallee	−33.73	150.37	810–820	3	cunninghamii Went
	Kings Tableland	King	GBMWHA	Mallee	−33.78	150.38	785–805	7	cunninghamii King
E. dendromorpha	Mount Wilson	Wils	GBMWHA	Mallee	−33.52	150.37	990–1010	8	dendromorpha Wils
	Mount Banks	Bank	GBMWHA	Mallee	−33.59	150.37	940–960	8	sp. Mt Banks
	Blackheath	Blac	GBMWHA	Mallee	−33.63	150.31	920–960	8	dendromorpha Blac
	Wentworth Falls	Went	GBMWHA	Mallee	−33.73	150.37	790–870	8	dendromorpha Went
	Redhills Road, Fitzroy Falls	Redh	Southern Highlands	Tree	−34.65	150.44	310–330	8	dendromorpha Redh
	Jersey Lookout, Fitzroy Falls	Jers	Southern Highlands	Tree	−34.65	150.48	290–310	8	dendromorpha Jers
E. laophila	Glow Worm Tunnel Road	Glow	GBMWHA	Mallee	−33.25	150.22	1115–1130	13	laophila Glow
	Lithgow	Lith	GBMWHA	Mallee	−33.50	150.17	900–935	5	laophila Lith
E. langleyi	Braidwood Road, Nowra	Brai	South Coast	Mallee	−34.97	150.50	195–265	8	langleyi Brai
	Parma Creek Firetrail, Nowra	Parm	South Coast	Mallee	−34.99	150.49	215–220	8	langleyi Parm
E. obstans	Beacon Hill	Beac	Sydney	Mallee	−33.74	151.26	135–145	7	obstans Beac
	Royal National Park	Roya	Sydney	Mallee	−34.12	151.08	110–125	8	obstans Roya
	Jervis Bay	Jerv	South Coast	Mallee	−35.01	150.83	20–30	7	obstans Jerv
	Stanwell Tops	Stan	Sydney	Mallee	−34.21	150.95	330	1	NSW908486
E. stricta	Newnes Plateau	Newn	GBMWHA	Mallee	−33.45	150.23	1170–1180	8	stricta Newn
	Mount Banks	Bank	GBMWHA	Mallee	−33.58	150.37	865–940	7	stricta Bank
	Blackheath	Blac	GBMWHA	Mallee	−33.63	150.31	910–925	7	stricta Blac
	Kings Tableland	King	GBMWHA	Mallee	−33.75	150.38	840–860	8	stricta King
	Stanwell Tops	Stan	Sydney	Mallee	−34.21	150.96	325–335	8	stricta Stan
	Sassafras	Sass	Southern Highlands	Mallee	−35.07	150.20	725–735	7	stricta Sass

GBMWHA Greater Blue Mountains World Heritage Area

Note: Latitude, longitude, and altitude of each sample were measured directly in the field at each site (GPS model: Garmin Rino 650)

species are rare, restricted, or localised. Both the mallee and tree forms of *E. dendromorpha* occur on sandstone substrates, generally in moist habitats (near waterfalls or creeks). *Eucalyptus obstans* occurs only in coastal habitats primarily from Sydney to Jervis Bay (200 km south of Sydney). *Eucalyptus laophila* is confined to higher elevations on sandstone ridges and pagodas (650–1100 m a.s.l.), while *E. cunninghamii* is restricted to escarpments (700–1000 m a.s.l). Many species are geographically proximate or have overlapping distributions. Other species are geographically disjunct and isolated. For example, *E. langleyi* is found only in the Nowra area (160 km south of Sydney).

Leaf material was collected from up to eight individuals (at least 10 m apart) per population. Mallees occur as stands (or clumps) that appear to grow from the same lignotuber, and as such, it can be difficult to discern a genetically distinct individual in the field (Rossetto et al. 1999). Therefore, a mallee stand was conservatively regarded as a genet and leaf material was collected from a single stem per stand. For some of the target populations (such as populations of *E. cunninghamii*), mallee stands were very close together and could not be sampled 10 m apart. Other populations (e.g., *E. cunninghamii* from Mount Banks) were small and it was not possible to sample eight individuals. In such cases, each mallee stand in the population was sampled. *Eucalyptus laophila* on the Glow Worm Tunnel Road (Wollemi National Park) was a scattered population, with some individuals occurring to the north and over the tunnel itself, while others were approximately 1.5 km south of the tunnel. For this population eight individuals at the tunnel itself and five individuals south of the tunnel were sampled. When sampling a population of *E. stricta* from Stanwell Tops, we found an individual that was morphologically differentiated from the others. We took a sample and voucher of this individual (NSW908486) and tentatively identified it as *E. obstans*. The geographic position (including elevation) of all samples were recorded (GPS model: Garmin Rino 650, Garmin Australasia, Sydney, NSW, Australia). Vouchers for each population were lodged at the National Herbarium of New South Wales. Leaf samples were freeze-dried and stored at –20 °C prior to DNA extraction.

DNA isolation and sequencing

Total genomic DNA was extracted from leaf samples following the protocol outlined in Rutherford et al. (2016). This was a CTAB protocol modified from Doyle and Doyle (1990). DNA concentrations were measured using a Qubit 2.0 Fluorometer (Invitrogen, Melbourne, VIC, Australia) and for each sample between 400 and 1000 ng of DNA (at a concentration of 50 ng mL^{-1}) was sent to Diversity Arrays Technology Pty Ltd. Samples were genotyped by Diversity Arrays Technology Pty Ltd using the DArTseq platform (Sansaloni et al. 2011).

Genotype data were generated from total genomic DNA using proprietary analytical pipelines (DArT Pty Ltd). These pipelines first remove poor-quality sequences from Illumina fastq files, leaving approximately 2,500,000 sequences per sample for marker calling. Identical sequences were collapsed (creating "fastqcoll" files), which were analysed using a proprietary algorithm (DArT Pty Ltd). This algorithm corrects a low-quality base in a singleton tag using a collapsed tag that is supported by multiple sequences as a reference. The corrected fastqcoll files were then used in a proprietary (DArT Pty Ltd) variant calling and filtering algorithm (version DArTsoft14). High-quality variant calling was ensured by sequencing to high mean depth of coverage (>30 reads) at each locus. Finally, the quality of marker calls was assayed by including technical replicates of multiple samples and scoring the consistency of genotype calls between the replicates (leading to a "reproducibility" score for each locus).

Relationships of green ash species and populations

DArTseq analysis produced a large dataset with a high proportion of high-quality markers. Of the 54 303 markers, 77.6% had a reproducibility of 100% (with 91.3% of markers having a reproducibility of ≥98%) and 50.2% of markers had a call rate of ≥90% (that is, genotype calls missing in <10% of samples). The proportion of missing data for samples ranged from 13.2 to 43.9%.

We examined the full dataset in *SplitsTree4* (Huson 1998; Huson and Bryant 2006). *SplitsTree4* infers the genetic relationships among a set of samples as a network, based on information from multiple loci. The network can represent evolutionary histories with substantial reticulation, arising from incomplete lineage sorting and hybridisation (Huson and Bryant 2006). In the present study, a relationship network of the total dataset was generated in *SplitsTree4* using the default settings of the software.

A reduced dataset of the higher-quality DArTseq markers was used for subsequent analyses. All SNPs with a reproducibility (proportion of replicate assay pairs for which the marker score is consistent) of less than 100% and which had more than 5% missing data were excluded from this dataset. To exclude the potential influence of linkage, one SNP was randomly selected from each of the restriction-associated loci that contained more than one SNP. This reduced dataset comprised 11,739 SNPs. To identify chloroplast data within the DArTseq dataset, we downloaded complete chloroplast genome sequences of 22 species of *Eucalyptus* (Table S1) from the NCBI (National Center for Biotechnology Information) database (https://www.ncbi.nlm.nih.gov/nuccore/?term=Eucalyptus,

accessed 23 August 2016). A BLAST search of the full DArTseq SNP dataset against the chloroplast genomes was performed in CLC Genomics Workbench using the default settings (version 8, www.clcbio.com). Eight DArTseq markers were identified as potential chloroplast sequences and checked to be homozygous before being submitted to a BLAST search of the whole genome of *E. grandis* on the NCBI database (https://blast.ncbi.nlm.nih.gov/Blast.cgi, accessed 24 August 2016). Results of the BLAST search were used to confirm hits to the chloroplast genome and five DArTseq markers were confirmed to be of chloroplast origin. These five DArTseq markers were excluded from the dataset for all subsequent analyses.

We used the R package adegenet 2.0.1 (Jombart 2008) to perform a principle coordinate analysis (PCoA) to explore genetic structure between species. To further investigate relationships among populations, analyses were performed using STRUCTURE version 2.3.4 (Pritchard et al. 2000; Falush et al. 2003). STRUCTURE uses a Bayesian model-based approach to group individuals by multi-locus genotypes, but does not impose taxonomic, population or geographic groups a priori (Pritchard et al. 2000). STRUCTURE was run using a subset of the highest quality markers (to reduce the run-time of the analysis). This subset excluded all markers with a reproducibility of less than 100%, an average count of minor alleles less than 10, as well as all markers with missing data at any locus (leaving 4783 markers). To determine whether this subset was representative of the larger dataset, a comparative PCoA was conducted based on Nei's unbiased genetic distances in the PCoA module of GenAlEx v6.501 (Peakall and Smouse 2006, 2012). A STRUCTURE analysis was performed using default parameters, K-values ranging from 1 to 25 (with five iterations for each K), 100,000 MCMC steps, and a burn-in period of 100,000 steps. The optimal value of genetically distinct clusters (K) was calculated based on the maximal mean posterior probability across replicates as well as the second rate of change (ΔK) (Evanno et al. 2005). We used STRUCTURE HARVESTER v0.6.94 (http://taylor0.biology.ucla.edu/structureHarvester, accessed 30 September 2016) to estimate the optimal value of K. Once the major sources of structure were identified, a hierarchical approach was used to examine the data further, with additional STRUCTURE analyses iterated for each subpopulation identified. Therefore, for each subpopulation identified we performed an analysis with K-values ranging from 1 to the total number of populations present and with the same parameters stated above (i.e., five iterations for each K, 100,000 MCMC steps, and a burn-in period of 100,000 steps).

Selecting the optimum value of K using STRUCTURE can be problematic as there is some degree of uncertainty as to what value of K best fits the data (Janes et al. 2017). The method of Evanno et al. (2005), which uses ΔK, is an ad hoc approximation that does not allow K to be 1 (Janes et al. 2017). Therefore, it is advised that users of STRUCTURE should not rely exclusively on ΔK, and to follow some recommendations when selecting the optimum value of K (Janes et al. 2017). So, in addition to using ΔK, we followed the recommendations of Janes et al. (2017), which involved inspecting all barplots produced by each STRUCTURE analysis at different values of K (from multiple runs and iterations), as well as including the Ln Pr(X/K) plots and STRUCTURE barplots for multiple values of K in the Supplementary Material when reporting results. Once the optimum value of K was identified for each STRUCTURE analysis, we used the *FullSearch* algorithm in *CLUMPP* v 1.1.2 (Jakobsson and Rosenberg 2007) to derive a single output from the five independent runs of K. We used Distruct v 1.1 (Rosenberg 2004) to visually display the probability of membership (Q) of individuals into the inferred groups.

Genetic diversity of species and populations

We calculated genetic diversity parameters from the highest quality markers (4783 SNPs). The expected heterozygosity (H_e), observed heterozygosity (H_o), number of observed alleles (N_a), genetic differentiation between regions (F_{ST}), and degree of inbreeding (F_{IS}) were determined using the Frequency Module of GenAlEx v6.501 (Peakall and Smouse 2006, 2012). Genetic diversity parameters were firstly calculated for the "taxonomic" species according to the species concepts of Hill (2002): *E. cunninghamii*, *E. dendromorpha* (including both the GBMWHA and Southern Highland populations), *E. stricta*, *E. laophila*, *E. obstans*, and *E. langleyi*. Genetic diversity parameters were also calculated for the groups found using the PCoA and STRUCTURE analyses described above, including *E. cunninghamii*, *E. dendromorpha* (GBMWHA populations only), *E. dendromorpha* (Southern Highland populations only), *E. obstans*, *E. langleyi*, as well as *E. stricta* and *E. laophila* combined.

We calculated pairwise F_{ST} values across all species and location combinations using the Frequency Module of GenAlEx v6.501. To better understand the contribution of species identity and geographical region to genetic differentiation, a hierarchical analysis of molecular variance (AMOVA) was used. AMOVA was performed using the PhiPT (Excoffier et al. 1992) analogue of F_{ST} in the AMOVA module of GenAlEx v6.501. The extent of genetic differentiation was determined between and within populations of each "taxonomic" species and for the groups found in the present study using the PCoA and STRUCTURE analyses. The extent of genetic differentiation was also determined between and within geographic regions
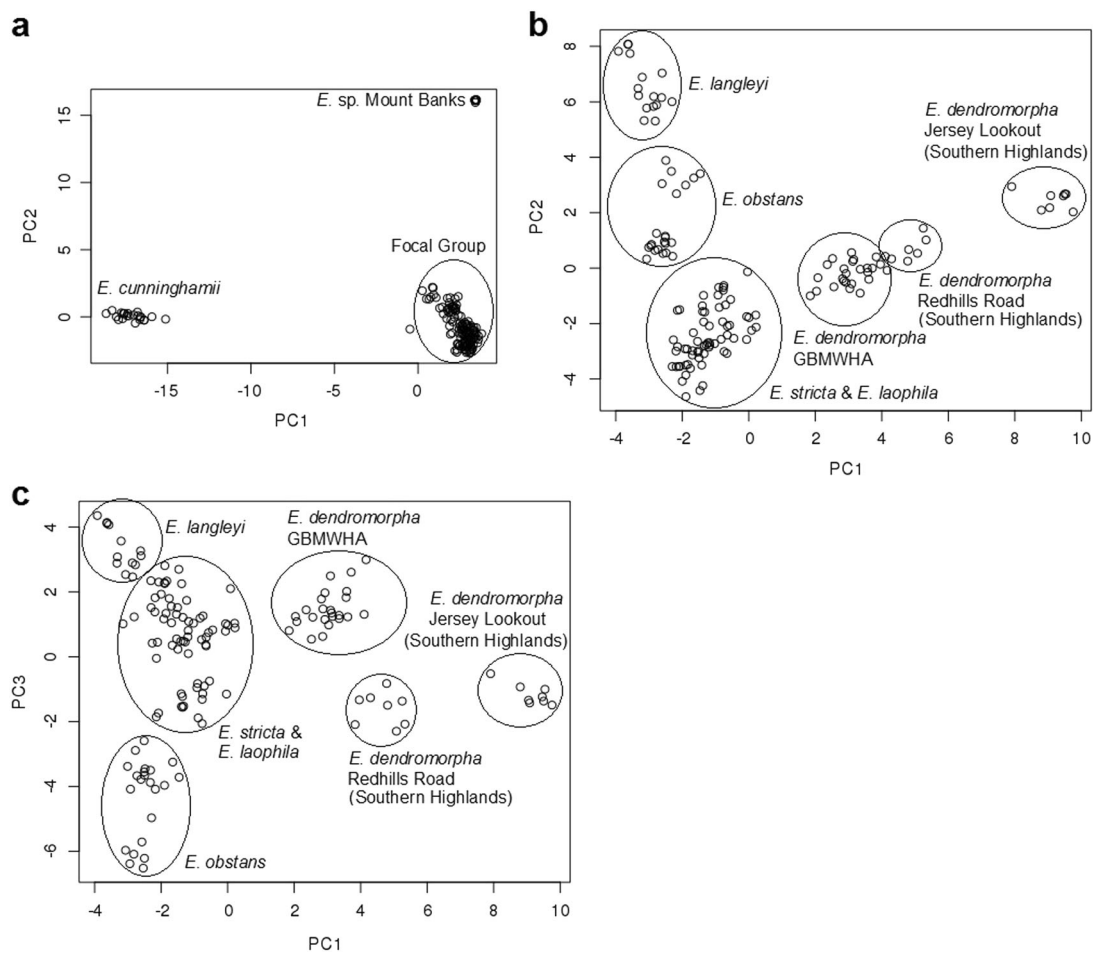
Fig. 2 Principle coordnate analysis (PCoA) derived from the DArTseq SNP dataset showing: **a** all study taxa (based on 11,739 DArTseq markers), **b** the focal group based on 10,868 DArTseq markers and including all taxa except for *Eucalyptus cunninghamii*, NSW908486 from Stanwell Tops and *E.* sp. Mount Banks (axis 1 and 2 shown), and **c** the focal group (axis 1 and 3 shown). GBMWHA, Greater Blue Mountains World Heritage Area

(GBMWHA, Sydney, Southern Highlands and South Coast). Statistical significance was determined by comparison to 999 random permutations of the data.
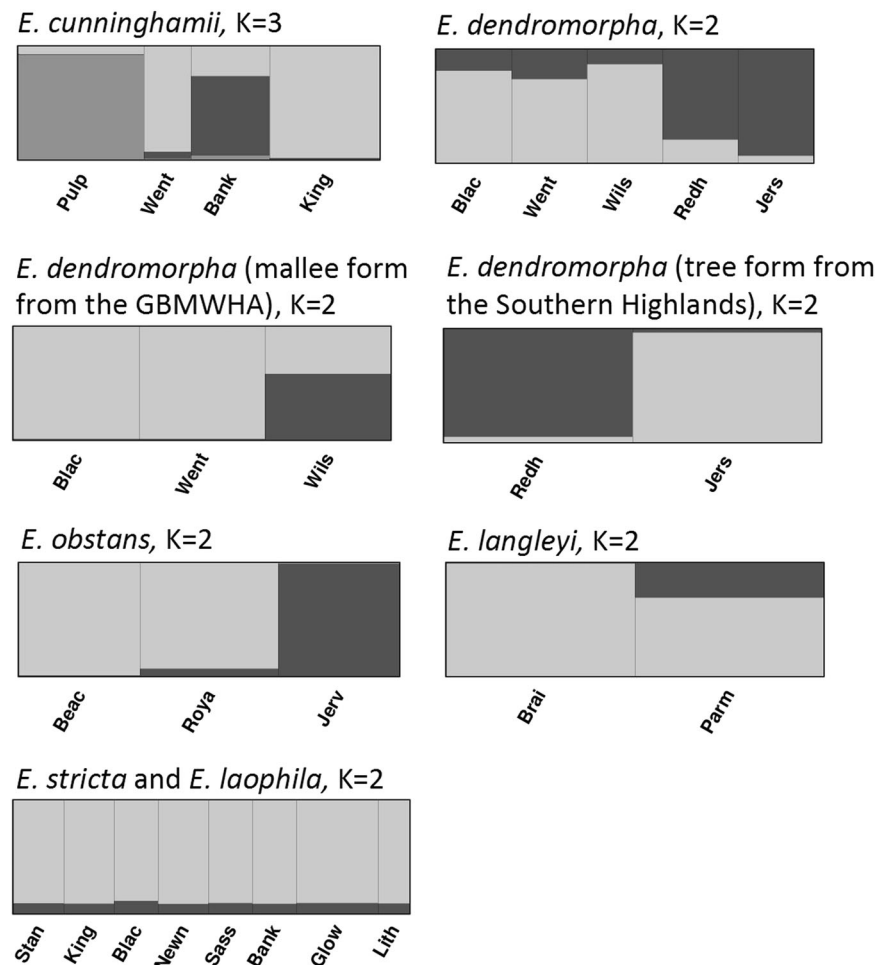
Analysis of connectivity and between-species gene flow

Patterns of connectivity were examined using network analysis implemented in EDENetworks version 2.18 (Kivelä et al. 2015). Network analysis can illustrate gene evolution by taking into account uncertainties in mutational pathways or reticulate events such as recombination, lateral transfer, and hybridisation (Posada and Crandall 2001). This approach is free of many of the "a priori" assumptions that usually underlie other methods of interpreting population molecular datasets (e.g., geographic clustering) and provide a graphical approach of viewing multidimensional data (Kivelä et al. 2015). In EDENetworks, networks are constructed from genetic distances that are calculated using the F_{ST}-based distance of Reynolds (Reynolds et al. 1983). In the present study, population-level and individual centred networks were produced from the dataset comprising the highest quality markers (4783 SNPs). To investigate connectivity across the landscape, geographic locations of each sample were used to overlay the network onto a map of the study region.

We applied TreeMix (version 1.13, Pickrell and Pritchard 2012) to test for and visualise gene flow and admixture between species using the reduced DArTseq dataset (11,739 SNPs) that was used to generate the PCoA (described above). TreeMix uses allele frequency data and a Gaussian approximation for genetic drift among taxa and populations to estimate a maximum likelihood tree (Pickrell and Pritchard 2012). Admixture between branches of the tree is determined in a stepwise likelihood procedure, where the tree is searched for the optimal placement of each migration event (Pickrell and Pritchard 2012). The proportion and directionality of gene flow is displayed on the tree (this is

Results

Relationships of green ash species and populations

A relationship network derived from the full dataset was generated in *SplitsTree 4* (Supplementary Fig. S1). In a relationship network, the parallel lines indicate splits in the data, with the longer lines suggesting more support for that particular split (Huson and Bryant 2006). *Eucalyptus cunninghamii* formed a distinct cluster, as did one population of *E. dendromorpha* (from Mount Banks). A sample from the Stanwell Tops site (NSW908486) was also on a longer branch and isolated from all other species. The *SplitsTree* analysis highlighted the complexity of the full DArTseq dataset.

The higher-level relationships of the green ashes was visualised in a PCoA based on the reduced dataset (11,739 SNPs, Fig. 2a), which identified four main groups: *E. cunninghamii*, *E. dendromorpha* from Mount Banks (referred to as *E.* sp. Mount Banks hereafter), a single sample from Stanwell Tops (NSW908486) and the remainder of the green ashes. A PCoA of this latter group allowed the resolution of further divisions among the remainder of the green ashes (Fig. 2b). *Eucalyptus langleyi* and *E. obstans* formed species-specific groups, although the population of *E. obstans* from the South Coast (Jervis Bay) was positioned between *E. langleyi* and the Sydney populations of *E. obstans*. The GBMWHA populations of *E. dendromorpha* and *E. dendromorpha* from Jersey Lookout (Southern Highlands) formed separate groups. However, when axis 1 was plotted against axis 3 in the PCoA (Fig. 2c), differentiation was detected between *E. dendromorpha* from the GBMWHA, *E. dendromorpha* from Jersey Lookout, as well as *E. dendromorpha* from Redhills Road

estimated from the asymmetries in the relationships inferred by the tree, Martin et al. 2015). For our analysis, we estimated a maximum likelihood tree with *E. cunninghamii* (all populations) selected as the outgroup taxon (based on the phylogenetic analysis of Rutherford et al. 2016). We inferred a topology without admixture, as well as allowing for between one and six migration events.

Table 2 Summary of genetic diversity parameters for the green ashes

Species	N	N_A (SE)	H_o (SE)	H_e (SE)	F_{IS} (SE)	F_{ST} (SE)
E. cunninghamii	23	1.621 (0.006)	0.234 (0.003)	0.206 (0.002)	−0.101 (0.008)	0.127 (0.003)
E. langleyi	16	1.767 (0.008)	0.246 (0.004)	0.220 (0.003)	−0.090 (0.007)	0.054 (0.001)
E. obstans	22	1.631 (0.006)	0.202 (0.003)	0.183 (0.002)	−0.082 (0.006)	0.094 (0.002)
E. dendromorpha (excluding *E.* sp. Mount Banks)	40	1.521 (0.004)	0.153 (0.002)	0.143 (0.001)	−0.045 (0.004)	0.087 (0.001)
E. dendromorpha (GBMWHA only)	24	1.642 (0.006)	0.185 (0.002)	0.174 (0.002)	−0.050 (0.005)	0.063 (0.001)
E. dendromorpha (Southern Highlands only)	16	1.770 (0.007)	0.232 (0.004)	0.217 (0.003)	−0.054 (0.006)	0.053 (0.002)
E. stricta	45	1.507 (0.004)	0.148 (0.002)	0.138 (0.001)	−0.052 (0.004)	0.079 (0.001)
E. laophila	18	1.727 (0.007)	0.204 (0.003)	0.197 (0.003)	−0.027 (0.007)	0.055 (0.001)
E. stricta and *E. laophila*	63	1.458 (0.003)	0.133 (0.001)	0.125 (0.001)	−0.042 (0.004)	0.086 (0.001)

Note: Values are mean and standard errors (SE) across loci

N_A observed number of alleles, F_{IS} degree of inbreeding in each individual relative to its local population, F_{ST} inbreeding in local populations relative to the total sample, GBMWHA Greater Blue Mountains World Heritage Area, H_e expected heterozygosity, H_o observed heterozygosity, N sample size

(Southern Highlands). All samples of *E. stricta* and *E. laophila* clustered together.

A comparative PCoA of the high-quality dataset (4783 loci) used for STRUCTURE analyses displayed similar groupings of species and populations (Supplementary Fig. S2). In the STRUCTURE analyses, coefficient of membership values indicated that all populations of *E. cunninghamii* were assigned to one group ($Q > 0.9$), while the remainder of species, with the exception of NSW908486 from Stanwell Tops, were assigned to a second group ($Q > 0.9$, Supplementary Table S2). However, when all populations of *E. cunninghamii* were excluded from the dataset, only *E.* sp. Mount Banks registered a high coefficient of membership to one group ($Q > 0.9$, Supplementary Table S3). All populations of other species had lower coefficients of membership to a second group ($Q < 0.8$), with small coefficients of membership assigned to one or more of another five groups (Supplementary Table S3). STRUCTURE analyses identified distinct genetic clusters within many species (Fig. 3, Supplementary Fig. S4). In *E. cunninghamii*, the Pulpit Rock and Mount Banks populations formed separate groups to the other two populations, which were geographically closer together (Wentworth Falls and Kings Tableland). The mallee form (from the GBMWHA) and the tree form (from the Southern Highlands) of *E. dendromorpha* formed two groups (Fig. 3). STRUCTURE analyses of the mallee form of *E. dendromorpha* from the GBMWHA revealed that there was admixture between the populations. The two populations of the tree form of *E. dendromorpha* from the Southern Highlands (Redhills Road and Jersey Lookout) formed separate groups. *Eucalyptus obstans* from the South Coast (Jervis Bay) was in a separate cluster to the two Sydney populations (Beacon Hill and Royal National Park). We found ongoing admixture between *E. stricta* and *E. laophila* despite the geographic location of the populations. Although the results from

STRUCTURE HARVESTER suggested that the value of K for the STRUCTURE analysis of *E. stricta* and *E. laophila* was 2, when we inspected the barplots from the analysis we found that there was no genetic differentiation between the two species (Fig. 3). Ongoing gene flow was detected between the two populations of *E. langleyi*.

Genetic diversity of populations and species

Eucalyptus cunninghamii had the highest F_{ST} among and the lowest F_{IS} within populations (0.127 and −0.101, respectively; Table 2). *Eucalyptus laophila* had the highest F_{IS} (−0.027), followed by *E. stricta* and *E. laophila* together (−0.042). *Eucalyptus langleyi* had the lowest F_{ST} (0.054) and the second lowest F_{IS} (−0.090). *Eucalyptus langleyi* had the highest N_A, H_e, and H_o (1.767, 0.246, and 0.220), while *E. stricta* and *E. laophila* together had the lowest N_A, H_e, and H_o (1.458, 0.133, and 0.125). Pairwise F_{ST} values between all populations revealed that there was relatively strong genetic differentiation between *E. cunninghamii* and the other study species (Supplementary Table S4). Similarly, high pairwise F_{ST} values were found between *E.* sp. Mount Banks and all the other study species. In contrast, the magnitude of inter-specific pairwise F_{ST} values was not greater than that of the intra-specific (between-population) F_{ST} for *E. stricta*, *E. laophila*, *E. obstans*, *E. langleyi*, and *E. dendromorpha*.

AMOVA revealed that for each taxonomic species (that were also supported by the groupings in the PCoA and STRUCTURE analyses), more than 80% of the variation could be attributed to within population variation (Supplementary Table S5). For example, *Eucalyptus obstans* had the highest between-population variation (19%), followed by *E. cunninghamii* (18%). *Eucalyptus stricta* and *E. laophila* (which were consistently grouped together in all other analyses) had the lowest between-population variation, as
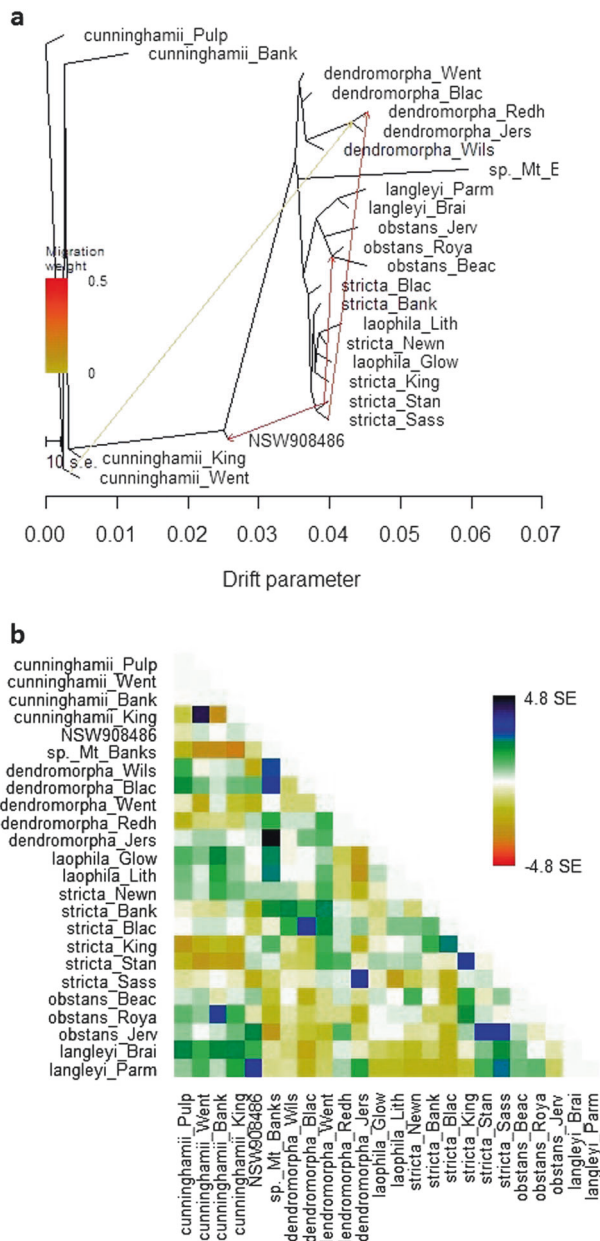
Fig. 4 TreeMix analysis of the green ash eucalypts (based on 11,739 SNPs) showing the **a** inferred maximum likelihood phylogeny showing four migration events and **b** residual fit plotted from the maximum likelihood tree in (**a**). In (**a**) the directionality of gene flow is indicated by arrows and coloured (yellow to red) according to their weight (0–50%). In (**b**), the colour bar to the right of the matrix indicates degree of relatedness between populations, with residuals above zero indicating populations that are more closely related to each other in the data than in the best-fit tree (i.e., bluer shades indicate population pairs that are candidates for admixture events). Sample codes correspond to those presented in Table 1. TreeMix analyses under varying numbers of admixture events (0–3 events) are shown in Supplementary Fig. S7

four times greater than variation among geographic regions (28 and 6%, respectively).

Analysis of connectivity and between-species gene flow

In networks produced by EDENetworks, network nodes (or vertices) represent populations as defined by sampling sites, and links (or edges) represent their relationships and interactions (Kivelä et al. 2015). In the current study, the population network (minimum-spanning tree) grouped most populations according to species (Supplementary Fig. S5). The population network of species and sampling sites overlain on a topographic map of the study region revealed a high degree of connectivity between many species across latitudes and along altitudinal gradients (Supplementary Fig. S6). There was a relatively high degree of genetic similarity between the population of *E. obstans* from Royal National Park and the other two populations of *E. obstans*, a population of *E. stricta* from Stanwell Tops, *E. langleyi* and NSW908486. In contrast, there was relatively low genetic similarity between populations of *E. dendromorpha* and *E. stricta* from Blackheath in the GBMWHA and samples from Stanwell Tops.

To test whether the low levels of genetic differentiation between species found in our other analyses were due to introgression, we used TreeMix on the reduced dataset (11,739 SNPs) to construct a maximum likelihood tree featuring varying numbers of admixture events. In our TreeMix analyses, we found that with the addition of each migration event, the log-likelihood (LL) of the analysis increased by a value of 36–132 until the addition of the sixth migration event (where the LL decreased by a value of 6). However, we found that TreeMix analyses with greater than four admixture events only showed minor improvements in LL. Furthermore, the TreeMix analysis with four migration events was consistent with all our other analyses. Therefore, the outputs from the TreeMix analysis featuring four migration events are presented in Fig. 4.

Our TreeMix analysis recovered a maximum likelihood tree with similar groupings of taxa to those found in the PCoA (Fig. 4). All populations of *E. dendromorpha* formed a clade as did the three populations of *E. obstans* and the two populations of *E. langleyi* (Fig. 4a). All populations of *E. stricta* and *E. laophila* were grouped together, while *E. sp.* Mount Banks and NSW908486 each occupied isolated positions in the phylogeny. Weak admixture (<10%) was detected between *E. cunninghamii* from Wentworth Falls and a clade comprising the populations of *E. dendromorpha* from the Southern Highlands (Jersey Lookout and Redhills Road). Stronger admixture (>40%) was detected between *E. stricta* from Stanwell Tops (Sydney) and *E. obstans* (from Beacon Hill and Royal National Park, Sydney), between *E.*

did populations of *E. dendromorpha* only from the GBMWHA (8%). We found that variation was better explained by "species" identity rather than by geographic partitioning. The variation among species was more than

stricta and NSW908486 (both from Stanwell Tops), and between *E. stricta* and *E. dendromorpha* from the Southern Highlands (Sassafras and Redhills Road, respectively). The population of *E. stricta* from Stanwell Tops is 70 km from the population of *E. obstans* from Beacon Hill, while the population of *E. stricta* from Sassafras is 50 km from the Redhills Road population of *E. dendromorpha*. A number of populations were identified as likely candidates for recent admixture events (Fig. 4b). For example, *E. stricta* from the Southern Highlands (Sassafras) and *E. obstans* from the South Coast (Jervis Bay), and populations of *E. stricta* and *E. dendromorpha* from Blackheath (GBMWHA).

Discussion

A set of SNP genotype markers provided useful insights into the evolutionary diversification of a group of closely related species within the green ash group. Patterns of gene flow across the study species and associations between genetic variation and geographic factors suggest that a range of speciation mechanisms are likely to have been operating within the one species complex. DarTseq markers also provided some insights regarding species boundaries within a group where there has been much uncertainty regarding the divergence and differentiation of taxa. Our findings have implications not only for speciation mechanisms in *Eucalyptus*, but also for other groups comprising taxa that have recently diverged.

Speciation by ecological isolation mechanisms

We found *E. cunninghamii* to be highly genetically differentiated compared to all the other study species. This was consistent with the phylogenetic analysis of Rutherford et al. (2016), where *E. cunninghamii* was in a clade with species found in northern New South Wales and southern Queensland (*E. approximans*, *E. codonocarpa*, and *E. microcodon*) that was sister to a clade that included the green ashes from the Sydney region and GBMWHA. *Eucalyptus cunninghamii* is morphologically distinct from all other green ash species (in that it is usually less than 1 m tall with thin, soft, silvery green leaves; Hill 2002) and occupies a narrow environmental niche on exposed cliff edges (Benson and McDougall 1998). While the TreeMix analysis suggested some introgression between one population of *E. cunninghamii* (Wentworth Falls) and *E. dendromorpha* from the Southern Highlands, this admixture was relatively weak (<10%, Fig. 4a). Furthermore, although *E. cunninghamii* is sympatric with *E. stricta* and *E. dendromorpha* from the GBMWHA, it has remained genetically differentiated (with high inter-specific pairwise F_{ST} values relative to intra-specific pairwise values,

Supplementary Table S4) while maintaining high levels of outcrossing (its populations have the lowest F_{IS}, Table 2). The comparatively high genetic differentiation of *E. cunninghamii* may be due to differences in phenology between species, as *E. cunninghamii* is thought to be an autumn-winter flowering species, whereas *E. stricta* and *E. dendromorpha* flower during the summer months (based on voucher specimens from the National Herbarium of New South Wales; Benson and McDougall 1998). Differences in flowering times may limit interbreeding opportunities among closely related sympatric species (Mohler 1990; Cavender-Bares and Pahlich 2009) and additional research on the phenology of green ash species would further clarify the role of temporal isolation in facilitating species co-occurrence without admixture.

Specimens of *E.* sp. Mount Banks in the National Herbarium of New South Wales were previously identified as *E. dendromorpha*, or in one case, as a possible hybrid between *E. dendromorpha* and *E. cunninghamii*. However, our results strongly suggest that this population is differentiated from other populations of *E. dendromorpha*, as well as all other green ash species in the study (Fig. 2a). In Rutherford et al. (2016), *E.* sp. Mount Banks (referred to as *E. dendromorpha* from Mount Banks) was in a clade with *E. stricta* and *E. laophila*. *Eucalyptus* sp. Mount Banks has morphologically distinct characters, in that its leaves have a bluish tinge and are shorter than *E. dendromorpha* and wider than *E. stricta*. While it is possible that *E.* sp. Mount Banks is a hybrid between a green ash and another non-green ash *Eucalyptus* species, the results of the STRUCTURE analysis suggest that there was very little admixture between this population and the other green ash species (in Supplementary Table S3 *E.* sp. Mount Banks was strongly assigned to Group 2 with a *Q*-value of 0.9982). It is therefore more likely that *E.* sp. Mount Banks is an undescribed species existing in sympatry with related species.

Between-species admixture and reticulate evolution

Relatively low genetic differentiation was detected among all other species. This was consistent with the findings of previous studies (Prober et al. 1990; Rutherford et al. 2016) and could be interpreted as a signature of recent radiation and reticulate evolution. Recent molecular dating of eucalypts places the radiation of subgenus *Eucalyptus* in the last 10 Ma (Crisp et al. 2011), while the earliest known fossils of subgenus *Eucalyptus* are from Late Miocene deposits (5–10 Ma; Blazey 1994). However, despite low between-species genetic differentiation many species in the present study maintain distinct morphological characters. For example, *E. dendromorpha* and *E. stricta* in the GBMWHA are often sympatric but can be distinguished by habit (*E. dendromorpha* is taller than *E. stricta*; Hill 2002) and leaf

size (leaves are significantly narrower in *E. stricta*; Hill 2002). Low genetic differentiation between sympatric species that can be distinguished morphologically has been documented in previous studies of *Eucalyptus* (e.g., Shepherd and Raymond 2010; Pollock et al. 2013), as well as studies of other plant genera, such as *Quercus* (e.g., Petit et al. 1997) and the Australian and South American genus *Lomatia* (McIntosh et al. 2014). This phenomenon could be due to strong selection of morphological characters resulting in taxa that are phenotypically different despite ongoing gene flow (Latta 2004). Alternatively, taxa could still be in the early stages of speciation and have developed morphological differences even though they are not yet strongly differentiated (Wu 2001).

Between-species admixture and hybridisation in the current study was supported by our TreeMix analysis, which showed relatively strong admixture between many sympatric and geographically close populations of many species (Fig. 4a). This is concordant with the findings of inter-specific hybridisation in other eucalypt species (e.g., McKinnon et al. 2001; Nevill et al. 2014; Jones et al. 2016), as well as species of *Lomatia* (Milner et al. 2012), *Quercus* (e.g., Petit et al. 1997), and *Pinus* (e.g., Matos and Schaal 2000). Between-species gene flow found for many species in the present study is consistent with historic and/or present hybridisation, as well as incomplete lineage sorting. Natural inter-specific hybridisation and introgression is likely to have played a significant role in plant evolution and speciation (Mallet 2005). While gene flow has been considered to be a countervailing process to speciation, more recently hybridisation is thought to have resulted in the evolution of new and stable evolutionary lineages (Rieseberg 1997; Abbott et al. 2013). Hybridisation can be a source of adaptive variation, functional novelty, and new species (Seehausen 2004). Incomplete barriers to gene flow can result in the introgression of selectively favoured alleles from one population into another (Abbott et al. 2013). For example, in Darwin's finches, it was estimated that the genetic variation introduced to populations through hybridisation was 2–3 times greater than that introduced by mutation (Grant and Grant 1994). Hybridisation can also act as a gene dispersal mechanism by extending a population's gene pool (and hence ecological range) through introgression (Potts and Reid 1988). Gene flow via hybridisation and introgression could be very important in species with small, fragmented, and isolated populations (Pollock et al. 2015), such as many of the species within the green ashes, and could contribute to the high outcrossing rates measured for most populations (Table 2). We note that individuals of *E. cunninghamii* at the Mount Banks site had strong assignments to different ancestral populations, as did individual samples of *E. langleyi* from Parma Creek Firetrail (as indicated by the STRUCTURE groups, Fig. 3). Also, the

TreeMix analysis identified populations of *E. langleyi* from Parma Creek Firetrail and *E. cunninghamii* from Mount Banks as likely candidates for admixture with other populations (Fig. 4b). In a previous study, evidence of hybridisation was found between *E. langleyi* and a co-occurring species, *E. consideniana* (which is also in subgenus *Eucalyptus*) (Rutherford et al. 2016). A detailed field study focusing on seed and pollen dispersal is needed to better understand hybridisation in the green ashes. Such a study could include progeny trials and genetic analysis of seeds from proximate and geographically isolated populations, as well as from stands of potential hybrid zones.

Speciation by geographic isolation

We found that patterns of genetic differentiation across some species could be associated with geographic factors. For example, genetic differentiation between *E. langleyi* and the other study species was higher than between *E. stricta*, *E. laophila*, and *E. obstans* (as indicated by the coefficient membership values, Supplementary Table S3). *Eucalyptus langleyi* is one of the most morphologically distinct species within the green ashes (with angular buds; Klaphake 2012) and is geographically restricted and isolated (occurring within a 17×7 km area near Nowra; Mills 2010). The nearest green ash populations examined here were more than 20 km away (*E. obstans* from Jervis Bay and *E. stricta* from Sassafras). Consequently, drift and divergence in allopatry are likely mechanisms impacting the evolution of *E. langleyi*.

Our findings also indicated geographic factors to be potential drivers of population genetic structure within species (apart from *E. stricta* and *E. laophila*). For example, in *E. cunninghamii*, the Mount Banks and Pulpit Rock populations were genetically differentiated from the others (these populations are higher in altitude and are on the other side of the ridge from the Wentworth Falls and Kedumba Valley populations, Fig. 3). A similar pattern was found along a latitudinal gradient for *E. obstans*, where the South Coast population was genetically more distinct than the Sydney populations (Fig. 3). These findings suggest that there can be strong associations between genetic variation and geographic factors, which is consistent with previous studies (e.g., Eckert et al. 2008), and indicate that vicariance and/or local adaptation may have had a role in the genetic differentiation and broader evolution of the green ashes. While dispersal could offer an explanation for the patterns observed here, the majority of studies suggest that eucalypts have very limited seed dispersal capabilities (Booth 2017). Earlier studies found that the bulk of seeds in most eucalypt species fall within a distance approximately equal to the height of the tree (e.g., Cremer 1966). More recently, Ruthrof et al. (2003) suggested that seed dispersal rates in

many eucalypt species is 1–2 m per year over a 70-year period. While pollen in many eucalypt species may travel as far as 1 km, most of the pollen has been found to be distributed within 200 m of the parent plant (Byrne et al. 2008; Broadhurst 2013).

Implications for species boundaries within rapidly differentiating lineages

We found no genetic differentiation between *E. stricta* and *E. laophila*. These two taxa are difficult to distinguish on the basis of morphology alone (both are mallees with overlapping leaf, bud, and fruit dimensions) and are distinguished on the basis of geographic location (Hill 2002). Our results together with the current morphological evidence suggest that *E. laophila* has been incorrectly assigned specific rank and is, rather, likely to be an ecotype of *E. stricta*. We found that *E.* sp. Mount Banks is likely to be an undescribed species. However, morphological data will be required to investigate this hypothesis further. With the exception of *E. cunninghamii* and *E.* sp. Mount Banks, we found moderate levels of connectivity between all the studied green ash species. However, combining all taxa (excluding *E. cunninghamii* and *E.* sp. Mount Banks) into one species would ignore observed morphological differences and may underestimate the diversity of the green ashes. Our findings suggest that the green ashes fit into the genic view of speciation (Wu 2001), with taxa being at varying stages along the speciation continuum. Despite occurrences of inter-specific hybridisation, many of the species can be recognised based on morphological discontinuities (e.g., *E. obstans*, *E. dendromorpha*, *E. stricta*, and *E. langleyi*). Although there is some geographic and ecological differentiation, most taxa in the GBMWHA have overlapping ranges. It is notable that the geographically restricted *E. langleyi* is one of the most morphologically distinct species and this is reflected in the higher level of genetic differentiation between it and the GBMWHA taxa. Nevertheless, there is still a degree of gene flow between it and nearby populations of *E. obstans*. We found that populations currently recognised as *E. dendromorpha* did not all form one "group". While most authorities, including Brooker (2000) and Hill (2002), regard all populations of *E. dendromorpha* to be the one species, Klaphake (2012) considers the tree form found in the Southern Highlands to be a taxon separate from the mallee form in the GBMWHA. Our findings supported genetic differentiation between the GBMWHA and Southern Highlands populations. Previous studies suggest that difference in growth habit in eucalypts and other species (such as *Nothofagus*) is likely to be due to environmental-induced variation, including soil, wind exposure, and fire regimes (e.g., Byrne et al. 2016; Barrera et al. 2000). Further investigation is required to determine

whether this is the case for the mallee and tree forms of *E. dendromorpha*. Although *E. laophila* has been shown to be inadequately differentiated, presently accepted concepts of other species remain as hypotheses that can be tested further in future. A detailed morphometric analysis of the study species would complement this molecular dataset and could be used to resolve these taxonomic issues. A large-scale population genetic analysis (>10 populations) of species may provide further insights into species boundaries.

Finally, a sample from Stanwell Tops (NSW908486) should be investigated in greater detail. It was genetically distinct from the other species in all of our analyses and was morphologically distinct from the *E. stricta* population found at Stanwell Tops with which it co-occurred (in that it had broader leaves). However, genomic DNA from more individuals from this site will be needed to investigate this further.

Conclusions

Speciation in the presence of gene flow was historically considered to be problematic because gene flow constrains population differentiation, thereby preventing the evolution of reproductive isolation (Mayr 1963). However, increasing evidence of speciation in the presence of gene flow is emerging (e.g., Nosil 2008; Pinho and Hey 2010), and in such cases, concepts that recognise speciation with gene flow are regarded as more appropriate in understanding species divergence (Schield et al. 2015). The clade of green ash eucalypts contains lineages at a range of different points along a "continuum" of speciation. In aggregate, our findings indicate that many mechanisms (e.g., vicariance, ecological speciation, and reticulate evolution) can or have been operating in tandem within the one species complex. This interplay of mechanisms has implications for speciation in other recently radiated taxa.

Data archiving

Genotype data is available from the Dryad Digital Repository (https://doi.org/10.5061/dryad.gd8gp17).

research was undertaken. We thank three anonymous reviewers, whose comments allowed us to significantly improve the manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N et al. (2013) Hybridization and speciation. J Evol Biol 26:229–246

Australia's Virtual Herbarium (2015) Occurrence records. Available at http://avh.ala.org.au/occurrences/search?q=qid:1521606130498#tab_mapView. Accessed 1 August, 2015

Barrera MD, Frangi JL, Ricther LL, Perdomo MH, Pinedo LB (2000) Structural and functional changes in Nothofagus pumilio forests along an altitudinal gradient in Tierra del Fuego. Argent J Veg Sci 11:179–188

Barraclough TG, Nee S (2001) Phylogenetics and speciation. Trends Ecol Evol 16:391–399

Benson D, McDougall L (1998) Ecology of Sydney plant species part 6: dicotyledon family Myrtaceae. Cunninghamia 5:808–987

Blazey J (1994) A Late Miocene flora from Bacchus Marsh. Honours Thesis, School of Botany, University of Melbourne

Booth TH (2017) Going nowhere fast: a review of seed dispersal in eucalypts. Aust J Bot. https://doi.org/10.1071/BT17019

Bragg JG, Supple MA, Andrew RL, Borevitz JO (2015) Genomic variation across landscapes: insights and applications. New Phytol 207:953–967

Broadhurst LM (2013) A genetic analysis of scattered Yellow Box trees (Eucalyptus melliodora A.Cunn. ex Schauer, Myrtaceae) and their restored cohorts. Biol Conserv 161:48–57

Brooker MIH (2000) A new classification of the genus Eucalyptus L'Hér (Myrtaceae). Aust Syst Bot 13:79–148

Byrne M, Elliott CP, Yates CJ, Coates DJ (2008) Maintenance of high pollen dispersal in Eucalyptus wandoo, a dominant tree of the fragmented agricultural region in Western Australia. Conserv Genet 9:97–105

Byrne M, Koenders A, Rogerson K, Sampson J, van Etten EJB (2016) Genetic and morphological analysis of multi-stemmed plants of tuart (Eucalyptus gomphocephala). Aust J Bot 64:704–714

Cavender-Bares J, Pahlich A (2009) Molecular, morphological and ecological niche differentiation of sympatric sister oak species, Quercus virginiana and Q. geminata (Fagaceae). Am J Bot 96:1690–1702

Coyne JA, Orr HA (2004) Speciation. Sinauer Associates, Sunderland, MA

Cremer KW (1966) Dissemination of seed from Eucalyptus regnans. Aust For 30:33–37

Crisp M, Cook L, Steane D (2004) Radiation of the Australian flora: what can comparisons of molecular phylogenies across multiple taxa tell us about the evolution of diversity in present-day communities? Philos Trans R Soc B Biol Sci 359:1551–1571

Crisp MD, Burrows GE, Cook LG, Thornhill AH, Bowman DMJS (2011) Flammable biomes dominated by eucalypts originated at the Cretaceous–Palaeogene boundary. Nat Commun 2:193

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15

Eckert CG, Samis KE, Lougheed SC (2008) Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. Mol Ecol 17:1170–1188

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

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587

Field DL, Ayre DJ, Whelan RJ, Young AG (2011) Patterns of hybridization and asymmetrical gene flow in hybrid zones of the rare Eucalyptus aggregata and common E. rubida. Heredity 106:841–853

Foster SA, McKinnon GE, Steane DA, Potts BM, Vaillancourt RE (2007) Parallel evolution of dwarf ecotypes in the forest tree Eucalyptus globulus. New Phytol 175:370–380

Gandolfo MA, Hermsen EJ, Zamaloa MC, Nixon KC, Gonzalez CC, Wilf P et al. (2011) Oldest known Eucalyptus macrofossils are from South America. PLoS ONE 6:e21084

Gaudeul M, Rouhan G, Gardner MF, Hollingsworth PM (2012) AFLP markers provide insights into the evolutionary relationships and diversification of the New Caledonian Araucaria species (Araucariaceae). Am J Bot 99:68–81

Gavrilets S (2003) Models of speciation: what have we learned in 40 years? Evolution 57:2197–2215

Grant PR, Grant BR (1994) Phenotypic and genetic effects of hybridization in Darwin's Finches. Evolution 48:297–316

Griffin AR, Burgess IP, Wolf L (1988) Patterns of natural and manipulated hybridization in the genus Eucalyptus L'Herit—a review. Aust J Bot 36:41–66

Hager T, Benson D (2010) The eucalypts of the Greater Blue Mountains World Heritage Area: distribution, classification and habitats of the species of Eucalyptus, Angophora and Corymbia (family Myrtaceae) recorded in its eight conservation reserves. Cunninghamia 11:425–444

Hill KD (2002) Myrtaceae: Eucalyptus. In: Harden GJ (ed) Flora of New South Wales. Revised edition. Vol. 2. University of New South Wales Press, Sydney, pp 96–164

Huson DH (1998) SplitsTree: analysing and visualizing evolutionary data. Bioinformatics 14:68–73

Huson DH, 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

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

Janes JK, Miller JM, Dupuis JR, Malenfant RM, Gorrell JC, Cullingham CI et al. (2017) The K=2 conundrum. Mol Ecol 26:3594–3602

Johnson LAS, Blaxell DF (1972) New taxa and combinations in Eucalyptus – I. Contrib New South Wales Natl Herb 4:287

Jombart T (2008) Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405

Jones RC, Steane DA, Lavery M, Vaillancourt RE, Potts BM (2013) Multiple evolutionary processes drive the patterns of genetic differentiation in a forest tree species complex. Ecol Evol 3:1–17

Jones RC, Nicolle D, Steane DA, Vaillancourt RE, Potts BM (2016) High density, genome-wide markers and intra-specific replication yield an unprecedented phylogenetic reconstruction of a globally significant, speciose lineage of Eucalyptus. Mol Phylogenet Evol 105:63–85

Keller I, Wagner CE, Greuter L, Mwaiko S, Selz OM, Sivasundar A et al. (2013) Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. Mol Ecol 22:2848–2863

Kershaw AP, Martin HA, McEwen Mason JRC (1994) The Neogene: a period of transition. In: Hill RS (ed) History of the Australian vegetation: cretaceous to recent. Cambridge University Press, Cambridge, pp 299–327

Kivelä M, Arnaud-Haond S, Saramäki J (2015) EDENetworks: a user-friendly software to build and analyse networks in biogeography, ecology and population genetics. Mol Ecol Resour 15:117–122

Klaphake V (2012) Eucalypts of the Sydney region, 3rd edn. Van Klaphake, Byabarra, NSW

Kopp A, Frank AK (2005) Speciation in progress? A continuum of reproductive isolation in Drosophila bipectinata. Genetica 125:55–68

Ladiges PY, Bayly MJ, Nelson GJ (2010) East–west continental vicariance in Eucalyptus subgenus Eucalyptus. In: Williams DM, Knapp S (eds) Beyond Cladistics: the branching of a paradigm. University of California Press, Los Angeles, CA, pp 267–301

Latta RG (2004) Gene flow, adaptive population divergence and comparative population structure across loci. New Phytol 161:51–58

Legrand D, Chenel T, Campagne C, Lachaise D, Cariou ML (2011) Inter-island divergence within Drosophila mauritiana, a species of the D. simulans complex: past history and/or speciation in progress? Mol Ecol 20:2787–2804

Lexer C, Widmer A (2008) The genic view of plant speciation: recent progress and emerging questions. Philos Trans R Soc B Biol Sci 363:3023–3036

Mallet J (2005) Hybridization as an invasion of the genome. Trends Ecol Evol 20:229–237

Martin CH, Cutler JS, Friel JP, Touokong CD, Coop G, Wainwright PC (2015) Complex histories of repeated gene flow in Cameroon crater lake cichlids cast doubt on one of the clearest examples of sympatric speciation. Evolution 69:1406–1422

Martin HA (1982) Changing Cenozoic barriers and the Australian Paleobotanical record. Ann Mo Bot Gard 69:625–667

Matos JA, Schaal BA (2000) Chloroplast evolution in the Pinus montezumae complex: a coalescent approach to hybridization. Evolution 54:1218–1233

Mayr E (1963) Animal species and evolution. Harvard University Press, Cambridge, MA

McGowen MH, Wiltshire RJE, Potts BM, Vaillancourt RE (2001) The origin of Eucalyptus vernicosa, a unique shrub eucalypt. Biol J Linn Soc Lond 74:397–405

McIntosh EJ, Rossetto M, Weston PH, Wardle GM (2014) Maintenance of strong morphological differentiation despite ongoing natural hybridization between sympatric species of Lomatia (Proteaceae). Ann Bot 113:861–872

McKinnon GE, Vaillancourt RE, Jackson HD, Potts BM (2001) Chloroplast sharing in the Tasmanian eucalypts. Evolution 55:703–711

McKinnon GE, Jordan GJ, Vaillancourt RE, Potts BM (2004) Glacial refugia and reticulate evolution: the case of the Tasmanian eucalypts. Philos Trans R Soc B Biol Sci 359:275–284

Mills K (2010) Rare plant species in the Illawarra 4: Eucalyptus langleyi (Myrtaceae). Illawarra vegetation studies (20). Coachwood Publishing, Jamberoo, NSW

Milner ML, Rossetto M, Crisp MD, Weston PH (2012) The impact of multiple biogeographic barriers and hybridization on species-level differentiation. Am J Bot 99:2045–2057

Mindell DP (2013) The tree of life: metaphor, model and heuristic device. Syst Biol 62:479–489

Mohler CL (1990) Co-occurrence of oak subgenera: implications for niche differentiation. Bull Torre Bot Club 117:247–255

Mullette KJ (1978) Studies of the lignotubers of Eucalyptus gummifera (Gaertn. and Hochr.). I: the nature of the lignotuber. Aust J Bot 26:9–13

Nevill PG, Després T, Bayly MJ, Bossinger G, Ades PK (2014) Shared phylogeographic patterns and widespread chloroplast haplotype sharing in Eucalyptus species with different ecological tolerances. Tree Genet Genomes 10:1079–1092

Nicolle D (2006) A classification and census of regenerative strategies in the eucalypts (Angophora, Corymbia and Eucalyptus—Myrtaceae), with special reference to the obligate seeders. Aust J Bot 54:391–407

Nosil P (2008) Speciation with gene flow could be common. Mol Ecol 17:2103–2106

Nosil P, Feder JL (2012) Genomic divergence during speciation: causes and consequences. Philos Trans R Soc B Biol Sci 367:332–342

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

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics 28:2537–2539

Petit RJ, Pineau E, Demesure B, Bacilieri R, Ducousso A, Kremer A (1997) Chloroplast DNA footprints of postglacial recolonization by oaks. Proc Natl Acad Sci USA 94:9996–10001

Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet 8: e1002967

Pinho C, Hey J (2010) Divergence with gene flow: models and data. Annu Rev Ecol Evol Syst 41:215–230

Pollock LJ, Bayly MJ, Nevill PG, Vesk PA (2013) Chloroplast DNA diversity associated with protected slopes and valleys for hybridizing Eucalyptus species on isolated ranges in south-eastern Australia. J Biogeogr 40:155–167

Pollock LJ, Bayly MJ, Vesk PA (2015) The roles of ecological and evolutionary processes in plant community assembly: the environment, hybridization, and introgression influence co-occurrence of Eucalyptus. Am Nat 185:784–796

Posada D, Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks. Trends Ecol Evol 16:37–45

Potts BM, Reid JB (1988) Hybridization as a dispersal mechanism. Evolution 42:1245–1255

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

Prober S, Bell JC, Moran G (1990) A phylogenetic and allozyme approach to understanding rarity in three "green ash" eucalypts (Myrtaceae). Plant Syst Evol 172:99–118

Reynolds J, Weir BS, Cockerham CC (1983) Estimation of the coancestry coefficient: basis of a short-term genetic distance. Genetics 105:767–779

Rieseberg LH (1997) Hybrid origins of plant species. Annu Rev Ecol Syst 28:359–389

Rieseberg LH, Brouillet L (1994) Are many plant species paraphyletic? Taxon 43:21–32

Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. Mol Ecol Notes 4:137–138

Rossetto M, Jezierski G, Hopper SD, Dixon KW (1999) Conservation genetics and clonality in two critically endangered eucalypts from the highly endemic south-western Australian flora. Biol Conserv 88:321–331

Ruthrof KX, Loneragan WA, Yates CJ (2003) Comparative population dynamics of Eucalyptus cladocalyx in its native habitat and as an invasive species in an urban bushland in south-western Australia. Divers Distrib 9:469–484

Rutherford S, Wilson PG, Rossetto M, Bonser SP (2016) Phyloge-nomics of the green ash eucalypts (Myrtaceae): a tale of reticulate evolution and misidentification. Aust Syst Bot 28:326–354

Sansaloni C, Petroli C, Jaccoud D, Carling J, Detering F, Grattapaglia D et al. (2011) Diversity Arrays Technology (DArT) and next generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of *Eucalyptus*. BMC Proc 5:54

Schield DR, Card DC, Adams RH, Jezkova T, Reyes-Velasco J, Nicole Proctor F et al. (2015) Incipient speciation with biased gene flow between two lineages of the Western Diamondback Rattlesnake (*Crotalus atrox*). Mol Phylogenet Evol 83:213–223

Seehausen O (2004) Hybridization and adaptive radiation. Trends Ecol Evol 19:198–207

Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA et al. (2014) Genomics and the origin of species. Nat Rev 15:176–192

Shepherd M, Raymond C (2010) Species differentiation and gene flow in the Blackbutts (genus *Eucalyptus* subgenus *Eucalyptus* section *Pseudophloius*). Conserv Genet 11:1965–1978

Shiga T, Kadono Y (2007) Natural hybridization of the two *Nuphar* species in northern Japan: homoploid hybrid speciation in pro-gress? Aquat Bot 86:123–131

Smith S, Hughes J, Wardell-Johnson G (2003) High population dif-ferentiation and extensive clonality in a rare mallee eucalypt: *Eucalyptus curtisii*. Conserv Genet 4:289–300

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

Vaillancourt RE, Potts BM, Manson A, Reid JB (1994) Detection of QTLs in a *Eucalyptus gunni×E. globulus* F2 using a RAPD linkage map. Proc Int Wood Biotechnol Symp 31:63–70

Wardell-Johnson GW, Williams JE, Hill KD, Cumming R (1997) Evolutionary biogeography and contemporary distribution of eucalypts. In: Williams J, Woinarski J (eds) Eucalypt ecology: individuals to ecosystems. Cambridge University Press, Cam-bridge, pp 92–128

Wu C-I (2001) The genic view of the process of speciation. J Evol Biol 14:851–865