Molecular Phylogenetics and Evolution 105 (2016) 63–85

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

High density, genome-wide markers and intra-specific replication yield an unprecedented phylogenetic reconstruction of a globally significant, speciose lineage of *Eucalyptus*

Rebecca C. Jones [a,*], Dean Nicolle [b], Dorothy A. Steane [a], René E. Vaillancourt [a], Brad M. Potts [a]

[a] School of Biological Sciences, University of Tasmania, Private Bag 55, Hobart, Tasmania 7001, Australia
[b] Currency Creek Arboretum, PO Box 808, Melrose Park, South Australia 5039, Australia

ARTICLE INFO

Article history:
Received 16 May 2016
Revised 29 July 2016
Accepted 12 August 2016
Available online 12 August 2016

Keywords:
Eucalypt
Nuclear DNA
Australia
Diversity Arrays Technology
Reticulate evolution
Evolutionary history

ABSTRACT

We used genome-wide markers and an unprecedented scale of sampling to construct a phylogeny for a globally significant *Eucalyptus* lineage that has been impacted by hybridisation, recent radiation and morphological convergence. Our approach, using 3109 DArT markers distributed throughout the genome and 540 samples covering 185 terminal taxa in sections *Maidenaria*, *Exsertaria*, *Latoangulatae* and related smaller sections, with multiple geographically widespread samples per terminal taxon, produced a phylogeny that largely matched the morphological treatment of sections, though sections *Exsertaria* and *Latoangulatae* were polyphyletic. At lower levels there were numerous inconsistencies between the morphological treatment and the molecular phylogeny, and taxa within the three main sections were generally not monophyletic at the series (at least 62% polyphyly) or species (at least 52% polyphyly) level. Some of the discrepancies appear to be the result of morphological convergence or misclassifications, and we propose some taxonomic reassessments to address this. However, many inconsistencies appear to be the products of incomplete speciation and/or hybridisation. Our analysis represents a significant advance on previous phylogenies of these important eucalypt sections (which have mainly used single samples to represent each species), thus providing a robust phylogenetic framework for evolutionary and ecological studies.

1. Introduction

Accurate phylogenies are important for understanding evolution and for informing taxonomy. They are also increasingly being used for other practical purposes such as incorporating evolutionary diversity into biodiversity measures (i.e. phylogenetic diversity, Faith, 1992) for conservation and management applications, to give a phylogenetic perspective on ecological studies (ecophylogenetics, Mouquet et al., 2012), predicting the susceptibility of species to pests and pathogens (Gilbert et al., 2012), developing species-specific markers, identifying cryptic species, and assessing the risk and impact of hybridisation for biosecurity purposes. In the speciose Australian genus *Eucalyptus*, for example, molecular phylogenies have been used to evaluate how well existing nature reserves capture the evolutionary lineages of eucalypts (Pollock et al., 2015), to provide a framework to predict species responses to climate change (Senior et al., 2013) and the susceptibility of species to the recently introduced myrtle rust pathogen (Potts et al., 2016), and to assess the risk of exotic gene flow from eucalypt plantations and guide species selection for hybrid breeding programs, by predicting the hybridisation potential of species pairs (Larcombe et al., 2015). With their expanded applications, robust phylogenies are therefore in increasing demand in this ecologically and economically important genus.

Eucalypts are a quintessential feature of the Australian landscape, dominating forest and woodland ecosystems across most of the continent. Nearly 900 species (Slee et al., 2006) occur in a variety of habitats from semi-arid regions to the tropics, coastal environments to alpine regions, and are important foundation species in many of these ecosystems (Williams and Woinarski, 1997). As well as being ecologically significant, several species are of major economic importance: they are the main hardwoods grown in the industrial plantations of the world as sources of wood and fibre, with a total plantation area of 20 million hectares (Iglesias-Trabado and Wilstermann, 2009).

* Corresponding author.
E-mail addresses: Rebecca.Jones@utas.edu.au (R.C. Jones), dn@dn.com.au (D. Nicolle), Dorothy.Steane@utas.edu.au (D.A. Steane), Rene.Vaillancourt@utas.edu.au (R.E. Vaillancourt), B.M.Potts@utas.edu.au (B.M. Potts).

http://dx.doi.org/10.1016/j.ympev.2016.08.009
1055-7903/

Despite the importance of eucalypts, there are still many phylogenetic questions and taxonomic issues that remain unresolved. At a high taxonomic level, *Eucalyptus* was subdivided into 13 subgenera by Brooker (2000) (the polytypic subgenera *Angophora*, *Corymbia*, *Blakella*, *Eudesmia*, *Symphyomyrtus*, *Minutifructa*, *Eucalyptus*, and the monotypic subgenera *Acerosa*, *Cruciformes*, *Alveolata*, *Cuboidea*, *Idiogenes* and *Primitiva*) but the most recent informal classification recognises *Angophora* and *Corymbia* (which includes *Blakella*) as genera (Nicolle, 2015), following molecular evidence (Bayly et al., 2013; Parra-O et al., 2006; Steane et al., 2002). The largest subgenus, *Symphyomyrtus*, consists of ~470 species belonging to 15 sections (Brooker, 2000). The sections are mostly well separated using molecular markers (Pollock et al., 2015; Steane et al., 2002, 2011) though there are some discrepancies with morphology-based classifications. Resolving lower level taxonomy (i.e. below the sectional level) of eucalypts is particularly complex, with morphological convergence of characters, recent divergence of taxa, and hybridisation and introgression often blurring boundaries between species. Molecular phylogenetic studies at this lower level are few (e.g. McKinnon et al., 2008; Rutherford et al., 2016; Steane et al., 2011) due to difficulties with finding appropriate markers to resolve species (Poke et al., 2006). This problem has now been largely overcome since an extensive set of genome-wide Diversity Array Technology (DArT) markers developed for *Eucalyptus* (Sansaloni et al., 2010) has proven useful for phylogenetic analyses at various taxonomic levels (Rutherford et al., 2016; Steane et al., 2011).

Many DNA-based phylogenetic studies use a single sample to represent a species, assuming that the species are monophyletic and that each sample is a good representative of the species. However, non-monophyly of species as a result of hybridisation or recent speciation is common in plants and a major challenge in phylogenetic studies (Rieseberg and Brouillet, 1994; Syring et al., 2007). Hence, phylogenetic studies using single representatives of each species may not reflect true relationships among taxa (Spinks et al., 2013; Syring et al., 2007) and, furthermore, will not allow the detection of misidentified samples or technical errors. Although eucalypts are known for their weak reproductive barriers between species (demonstrated in Griffin et al., 1988; Larcombe et al., 2015), most eucalypt phylogenies to date have been based on single representatives of each species (for exceptions see McKinnon et al., 2008; Rutherford et al., 2016). The reasons for this are historical: DNA-based studies were time-consuming and expensive, analyses of data sets containing large numbers of samples were problematic, and the logistics and cost of collecting multiple samples of each species to represent the genetic diversity across the species' range were prohibitive. The recent revolution in DNA technologies has overcome some of these issues, in particular the cost and time involved with obtaining DNA data from large numbers of samples. The rate of advance of computing capacity and the development of efficient phylogenetic algorithms have lagged behind the generation of genomic data, but are now sufficiently advanced to allow effective analysis of large phylogenetic data sets. Currency Creek Arboretum in South Australia, which has a collection of over 900 eucalypt species and subspecies planted over the last two decades, now facilitates the collection of multiple samples of many species from across the species' ranges. Hence, it is now possible to design very robust studies of species-level DNA-based phylogenies of *Eucalyptus* that incorporate multiple samples of a large number of species and data from across the whole genome. Such experimental design will produce more robust eucalypt phylogenies that will help to resolve some of the long standing evolutionary and taxonomic issues in the genus.

Here, we present a phylogenetic analysis of taxa from three closely related (Steane et al., 2011) sections of the subgenus *Symphyomyrtus*: *Exsertaria*, *Latoangulatae* and *Maidenaria*. Along with a number of smaller related sections, *Exsertaria*, *Latoangulatae* and *Maidenaria* are part of a single genetic lineage within subgenus *Symphyomyrtus* (Steane et al., 2011), the largest subgenus in the genus *Eucalyptus*. This lineage includes all of the "Big Nine" species that account for 90–95% of the world's planted eucalypts (*E. tereticornis* and *E. camaldulensis* in section *Exsertaria*; *E. grandis*, *E. saligna*, *E. pellita* and *E. urophylla* in section *Latoangulatae*; *E. dunnii*, *E. globulus*, *E. nitens* in section *Maidenaria*; Harwood, 2011), including Australia's most naturally widespread eucalypt, *E. camaldulensis*, and the first eucalypt to have its genome sequenced, *E. grandis* (Myburg et al., 2014). We also include the related smaller sections *Racemus*, *Inclusae*, *Similares*, *Incognitae*, *Liberivalvae*, *Platysperma* and *Pumilio*, hypothesised to be sister sections to, or embedded within, *Exsertaria*, *Latoangulatae* and *Maidenaria* (Brooker, 2000; Nicolle, 2015; Steane et al., 2007) to further elucidate the genetic affinities of these anomalous taxa. We use a genome-wide approach (3109 DArT markers distributed throughout the genome, Petroli et al., 2012) and an unprecedented scale of sampling (540 samples covering 185 of the 193 terminal taxa - as recognised by Nicolle, 2015 - in these sections), with multiple geographically widespread samples per terminal taxon (where possible) to account for intraspecific genetic diversity. We aim to contribute genetic data to inform taxonomic revisions in these important sections and provide robust phylogenies that can be used for a range of practical and fundamental purposes.

2. Materials and methods

2.1. Sample collection

Leaf tissue for DNA extraction and a herbarium specimen were sampled from 542 trees in either natural populations or growing at Currency Creek Arboretum (South Australia, http://www.dn.com.au/Currency_Creek_Arboretum.html). Samples were obtained from 185 of the 193 terminal taxa (153 of the 160 species) in sections *Maidenaria*, *Latoangulatae*, *Exsertaria* and the related smaller sections *Racemus* (monotypic, *E. michaeliana*), *Inclusae* (monotypic, *E. diversicolor*), *Similares* (monotypic, *E. longifolia*), *Incognitae* (two species, *E. cosmophylla* and *E. paludicola*), *Liberivalvae* (seven terminal taxa), *Platysperma* (eight terminal taxa) and *Pumilio* (monotypic, *E. pumila*) (Supplementary Table 1, summarised in Table 1). Where possible, 3 or 4 samples were collected per taxon, each sample representing a different population; populations were geographically well-spaced across the distribution of each taxon (Supplementary Table 1). Some additional sampling was conducted for widespread, polymorphic or taxonomically contentious taxa. Of the 185 terminal taxa (species and subspecies) sampled, 161 were represented by more than one sample, and of the 153 species sampled, 136 were represented by more than one sample. One sample each of *Eucalyptus cornuta* and *E. wandoo* (subgenus *Symphyomyrtus*, section *Bisectae*) were included as outgroups.

2.2. DNA extraction and DArT genotyping

DNA was extracted from fresh, frozen or dried leaf tissue following the method of Doyle and Doyle (1990) with several modifications (McKinnon et al., 2004). Due to quarantine restrictions on imports of eucalypt leaf material into Tasmania, DNA for some samples was extracted at the Australian Genome Research Facility (Adelaide) using a NucleoSpin® 96 Plant II kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany). DNA quality was tested by cutting with the *Hin*dIII restriction enzyme, following the methods of Steane et al. (2011). DNA samples that did not digest properly were either re-extracted or cleaned by ethanol precipitation in the presence of 2M NaCl. DArT genotyping was carried out at DArT Pty Ltd