Molecular Phylogenetics and Evolution 59 (2011) 206–224

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

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

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

[a] School of Plant Science, University of Tasmania, Private Bag 55, Hobart, Tasmania 7001, Australia
[b] CRC for Forestry, University of Tasmania, Private Bag 55, Hobart, Tasmania 7001, Australia
[c] Currency Creek Arboretum, PO Box 808, Melrose Park, South Australia 5039, Australia
[d] Plant Genetics Laboratory, EMBRAPA Genetic Resources and Biotechnology, EPqB, 70770-910 Brasilia, Brazil
[e] Department of Cell Biology, Universidade de Brasília, 70910-900 Brasília, DF, Brazil
[f] Diversity Arrays Technology Pty. Ltd., 1 Wilf Crane Crescent, Yarralumla ACT 2600, Australia
[g] Department of Genetics, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria 0002, South Africa
[h] Genomic Sciences Program, Universidade Católica de Brasília, SGAN, 916 modulo B, 70790-160 Brasília, DF, Brazil

ARTICLE INFO

Article history:
Received 17 June 2010
Revised 1 February 2011
Accepted 2 February 2011
Available online 16 February 2011

Keywords:
Australia
Diversity Arrays Technology
DArT
Molecular markers
Networks
Parsimony
Plant systematics

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 *Eucalyptus camaldulensis*, *Eucalyptus cladocalyx*, *Eucalyptus globulus*, *Eucalyptus grandis*, *Eucalyptus nitens*, *Eucalyptus pilularis* and *Eucalyptus 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.

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

* Corresponding author at: School of Plant Science, University of Tasmania, Private Bag 55, Hobart, Tasmania 7001, Australia. Fax: +61 (0)3 62262698.
E-mail address: Dorothy.Steane@utas.edu.au (D.A. Steane).

1055-7903/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2011.02.003

allozyme frequencies in the intergrading species *Eucalyptus grandis* and *Eucalyptus 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. Furthermore, 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 *Eucalyptus 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 se-