Molecular Phylogenetics and Evolution 69 (2013) 704–716

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

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

Chloroplast genome analysis of Australian eucalypts – *Eucalyptus*, *Corymbia*, *Angophora*, *Allosyncarpia* and *Stockwellia* (Myrtaceae)

Michael J. Bayly [a,1], Philippe Rigault [b,1], Antanas Spokevicius [c], Pauline Y. Ladiges [a], Peter K. Ades [c], Charlotte Anderson [d], Gerd Bossinger [c], Andrew Merchant [e], Frank Udovicic [f], Ian E. Woodrow [a], Josquin Tibbits [d,*,1]

[a] School of Botany, The University of Melbourne, Vic. 3010, Australia
[b] GYDLE, 1363 Av. Maguire, Suite 301, Québec, QC G1T 1Z2, Canada
[c] Department of Forest and Ecosystem Science, The University of Melbourne, Vic. 3010, Australia
[d] Victorian Department of Environment and Primary industries, AgriBiosciences Center, La Trobe University, 5 Ring Road, Bundoora, Vic. 3083, Australia
[e] Faculty of Agriculture and Environment, University of Sydney, NSW 2006, Australia
[f] National Herbarium of Victoria, Royal Botanic Gardens Melbourne, Private Bag 2000, South Yarra, Vic. 3141, Australia

ARTICLE INFO

Article history:
Received 29 April 2013
Revised 28 June 2013
Accepted 8 July 2013
Available online 19 July 2013

Keywords:
Chloroplast genome
Eucalypts
Evolution
Illumina sequencing
Phylogeny

ABSTRACT

We present a phylogenetic analysis and comparison of structural features of chloroplast genomes for 39 species of the eucalypt group (genera *Eucalyptus*, *Corymbia*, *Angophora*, and outgroups *Allosyncarpia* and *Stockwellia*). We use 41 complete chloroplast genome sequences, adding 39 finished-quality chloroplast genomes to two previously published genomes. Maximum parsimony and Bayesian analyses, based on >7000 variable nucleotide positions, produced one fully resolved phylogenetic tree (35 supported nodes, 27 with 100% bootstrap support). *Eucalyptus* and its sister lineage *Angophora* + *Corymbia* show a deep divergence. Within *Eucalyptus*, three lineages are resolved: the 'eudesmid', 'symphyomyrt' and 'monocalypt' groups. *Corymbia* is paraphyletic with respect to *Angophora*. Gene content and order do not vary among eucalypt chloroplasts; length mutations, especially frame shifts, are uncommon in protein-coding genes. Some non-synonymous mutations are highly incongruent with the overall phylogenetic signal, notably in *rbc*L, and may be adaptive. Application of custom informatics pipelines (GYDLE Inc.) enabled direct chloroplast genome assembly, resolving each genome to finished-quality with no need for PCR gap-filling or contig order resolution. Analysis of whole chloroplast genomes resolved major eucalypt clades and revealed variable regions of the genome that will be useful in lower-level genetic studies (including phylogeography and geneflow).

1. Introduction

1.1. Eucalypt phylogeny

Eucalypts are iconic hardwood trees that characterise the Australian landscape. They are mostly endemic to Australia and are significant ecologically and economically (e.g., Williams and Woinarski, 1997), with many species used internationally in industrial forestry plantations in both temperate and tropical regions of the world (Grattapaglia and Kirst, 2008). The largest genus *Eucalyptus* L'Hér., with c. 800 species, includes the tallest angiosperm in the world, *E. regnans*, and important plantation species such as *E. globulus*, *E. grandis*, *E. nitens*, and *E. deglupta*.

Eucalypt phylogeny has been of interest for many years as the basis for classification, evolutionary and ecological research. Considerable progress in resolving relationships of major clades has been made using morphological characters and sequence data of nuclear and chloroplast DNA, but determining genetic relationships at the level of sections, series and species within clades has been challenging (e.g., Bayly and Ladiges, 2007; Bayly et al., 2008; Ochieng et al., 2007a,b; Steane et al., 1999, 2002; Udovicic and Ladiges, 2000). Recent papers have assessed the suitability of low copy nuclear genes (Poke et al., 2003, 2006), AFLPs (McKinnon et al., 2008), microsatellites (Ochieng et al., 2007a; Steane et al.,2005), and Diversity Array Technology (DArT) markers (Steane et al., 2011). However, a greater knowledge of eucalypt genomes is required for identifying variable and more informative regions for phylogenetic analysis. Studies of eucalypt genomes will also allow discovery and interpretation of functional elements encoded within sequences, providing a basis for understanding key

* Corresponding author.
E-mail address: Josquin.Tibbits@depi.vic.gov.au (J. Tibbits).
[1] These authors contributed equally to this work.

1055-7903/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ympev.2013.07.006

evolutionary changes that correlate with the diversification and adaptation of clades.

1.2. Diversity and distribution of the eucalypt group

The monophyletic eucalypt group (Myrtaceae, comprising tribe Eucalypteae *sensu* Wilson et al., 2005) includes seven genera: *Allosyncarpia* S.T. Blake, *Stockwellia* Carr, Carr & Hyland, *Eucalyptopsis* C.T. White, *Arillastrum* Pancher ex Baill., *Eucalyptus* L'Hér. *sensu strict*., *Angophora* Cav. and *Corymbia* K.D. Hill & L.A.S. Johnson. The first four are rainforest genera of five species, while the other three are sclerophylls comprising c. 900 species. Four of the rainforest species form a clade, including *Allosyncarpia ternata* endemic to the Australian monsoon tropics, *Stockwellia quadrifida* of the Queensland wet tropics, and *Eucalyptopsis papuana* and *E. alauda* in New Guinea and adjacent islands. The fifth rainforest species, *Arillastrum gummiferum*, is endemic to New Caledonia and hypothesised as a basal lineage in the eucalypt group (Bohte and Drinnan, 2005; Udovicic and Ladiges, 2000).

Groups within the largest genus, *Eucalyptus*, are treated taxonomically as subgenera (Brooker, 2000), including three main subgenera, *Eucalyptus*, *Symphyomyrtus* and *Eudesmia*, and seven small subgenera, *Acerosae* (*E. curtisii*), *Cuboidea* (*E. tenuipes*), *Idiogenes* (*E. cloeziana*), *Alveolata* (*E. microcorys*), *Cruciformes* (*E. guilfoylei*), *Minutifructus* (four species of tropical boxes, nested within *Symphyomyrtus*, Whittock et al., 2003) and *Primitiva* (*E. rubiginosa*but nested within subgenus *Eucalyptus*, Ladiges et al., 2010). The lineage of *Eucalyptus* is sister to the lineage of *Angophora* Cav. (c. 13 species) + *Corymbia* K.D. Hill & L.A.S. Johnson (c. 100 species of bloodwoods, spotted gums and ghost gums).

A number of species within *Eucalyptus* occur outside, or extend beyond, the Australian continent, with four species endemic to New Guinea and/or Malesia (to the southern Philippines). Macrofossils of eucalypts extend this range historically to South America (Early Eocene; Gandolfo et al., 2011) and New Zealand (Early Miocene; Pole, 1989). Evidence from fossils, biogeography and molecular dating indicates that the eucalypt group has origins in the Late Cretaceous (Ladiges et al., 2003).

Since the morphological phylogenetic analysis of Hill and Johnson (1995), molecular phylogenetic studies have been contradictory as to whether *Corymbia* is monophyletic or paraphyletic (Udovicic et al., 1995; Steane et al., 1999, 2002; Udovicic and Ladiges, 2000; Whittock et al., 2003; Parra-O. et al., 2006). The strongest molecular evidence so far for the monophyly of *Corymbia* and its sister relationship to *Angophora* comes from analysis of nuclear ribosomal ETS sequences (Parra-O. et al., 2006, 2009), microsatellites (Ochieng et al., 2007a) and ITS pseudogenes (riboform H2 sequences of Ochieng et al.,2007b). The small genus *Angophora* is unquestionably monophyletic (e.g., Thiele and Ladiges, 1988; Parra-O. et al., 2009).

We have commenced a comparative genomic study, sampling broadly across the phylogeny of the Australian eucalypts, to maximise divergence between major lineages. Here we report 39 new chloroplast genomes of finished quality for 38 species, representing all five Australian genera, combined with published data for *E. globulus* (Steane, 2005) and *E. grandis* (Paiva et al., 2011).

2. Material and Methods

2.1. Taxon sampling

Table 1 lists species and accessions sampled. In total, 38 species were selected to represent the major clades: major sections and series within *Eucalyptus* and allied genera. Sampling included 12 *Eucalyptus* subg. *Eucalyptus* species representing six sections, 13 species from subg. *Symphyomyrtus* representing seven sections, and one species from subg. *Eudesmia* (*E. erythrocorys*). We sampled from five of the seven small subgenera, with one species each representing *Acerosae* (*E. curtisii*), *Idiogenes* (*E. cloeziana*), *Alveolata* (*E. microcorys*), *Cruciformes* (*E. guilfoylei*)and *Minutifructus* (*E. deglupta*). We also sampled two species of *Angophora*, four species of *Corymbia* representing two subgenera and four sections, and two outgroup species, *Allosyncarpia ternata* and *Stockwellia quadrifida*.

2.2. DNA isolation from leaves and cambial tissue

Nuclear and plastid DNA was isolated from fresh leaf material or cambial tissue. Cambial extraction methods followed Tibbits et al. (2006). For leaf samples, nucleus enrichment, adapted from Peterson et al. (1997), was followed by DNA extraction based on Tibbits et al. (2006). Nucleus enrichment was carried out to reduce the proportion of plastid DNA in leaf extracts (much higher than in cambial tissue) in order to achieve greater sequencing coverage of nuclear genomes for use in subsequent studies. Extracted DNA was quantitated using a NanoDrop 2000 (NanoDrop Products).

For nucleus enrichment from leaf tissue, fast growing, new leaf material was collected into an ice/TE (10:1, pH 8.0) slurry, and kept at or below 4 °C. Between 10 and 30 g of leaf material was washed in ice-cold diethyl ether and then washed twice in ice-cold TE (10:1, pH 8.0). Washed leaves were homogenised using a Bamix blender in 300 mL ice-cold MEB buffer (1.0 M 2-methyl-2,4-pentanediol, 10 mM PIPES buffer, 10 mM magnesium chloride $6H_2O$, 2% w/v polyvinylpyrrolidone (PVP-10), 10 mM sodium metabisulphite, 5 mM 2-mercaptoethanol, 0.5% w/v sodium diethyldithiocarbamate) and then filtered three times – first through a 50 mL syringe, with no filter to remove the largest fragments, then through 16 layers of cheesecloth, and finally through 32 layers of cheesecloth. To lyse chloroplasts andmitochondria, Triton-X 100 (Sigma–Aldrich) was added to a final concentration of 0.5% (v/v) and samples were gently mixed at 4 °C for 20–30 min (until the Triton-X 100 was completely dissolved). Samples were centrifuged at 800*g* for 20 min at 4 °C and the supernatants discarded. Pellets were gently re-suspended in 25 mL MEB/0.5% Triton-X 100 solution. This second lysis significantly improved the enrichment over the single lysis step in the parent protocol. Samples were again centrifuged at 800*g* for 20 min at 4 °C and the supernatants discarded. Pellets were re-suspended in 2 mL MPDB buffer (0.5 M 2-methyl-2,4-pentanediol, 10 mM PIPES buffer, 10 mM magnesium chloride $6H_2O$, 10 mM sodium metabisulphite,5 mM 2-mercaptoethanol), layered onto a 37.5% Percoll bed and centrifuged at 650*g* for 60 min at 4 °C. The supernatant was discarded leaving a pellet of enriched nuclei, but including some residual chloroplast and mitochondrial DNA.

For DNA extraction from leaf material, following enrichment for nuclei, 20 mL of preheated (65 °C) DNA extraction buffer (2% w/v cetyltrimethylammonium bromide (CTAB), 2% polyvinylpyrrolidone 40,000 (PVP-40), 1.4 M sodium chloride, 20 mM EDTA, 100 mM Tris–HCl pH 8.0) was added to the nuclei pellet. Samples were gently mixed to lyse the enriched nuclei and, once completely re-suspended and lysed, 2.75 mL NaCl/BSA (10:1; 10 M NaCl, 4% BSA) (v/v) was added. Two 1:1 extractions with Chloroform:Isoamyl alcohol (24:1) were performed with 15 min of gentle shaking followed by centrifugation at 10,000*g* for 10 min at room temperature. Approximately 15 mL of supernatant was collected and 2/3 volume of 100% iso-propanol added to precipitate the nucleic acids. Nucleic acids werepelleted by centrifugation at 10,000*g* for 10 min and pellets were washed with 70% ethanol. Washed pellets were carefully transferred to 1.75 mL microcentrifuge tubes and excess ethanol was aspirated.

For clean-up of DNA extracted from either leaf or cambial tissue, DNA pellets were resuspended in 400 μL 100 mM EDTA and