ABSTRACT

PAYN, KITT GARNET. Molecular Genetic Diversity and Population Genetic Structure of the Commercially Important Tropical Forest Tree Species *Eucalyptus urophylla*. (Under the direction of William S. Dvorak.)

Eucalyptus urophylla is an economically important plantation tree species in countries with tropical and subtropical climates. Its natural distribution is limited to a series of populations on seven islands of the Sunda archipelago in eastern Indonesia. The largest populations are found on the islands of Timor and Wetar, whereas smaller populations occur on the islands of Alor, Pantar, Lomblen, Adonara and Flores. Of concern is the depletion in the species' genetic resource, primarily on the latter five islands, as a consequence of land conversion to agriculture. Proficient management of the *E. urophylla* genetic resource, with respect to both conservation and breeding programs, requires a good understanding of the level of genetic diversity and population genetic structure across the species' native range. This may be achieved through the application of DNA marker techniques that measure genetic diversity at different levels (e.g. organellar vs. nuclear, or genome-wide vs. gene locus). Historical events such as range expansion and genetic bottlenecks can leave an imprint on contemporary levels of genetic diversity. Therefore, the geographical distribution of chloroplast DNA sequence variation in *E. urophylla* was studied to gain insight into its historical seed migration routes along the island archipelago. Chloroplast DNA sequence data were obtained from 198 plants broadly representing the native range of the species. Twenty haplotypes were identified. A moderate to high level of chloroplast genetic differentiation (G_{ST} = 0.581, N_{ST} = 0.724) and significant phylogeographic structure ($N_{ST} > G_{ST}$; $P < 0.01$) were observed, suggesting low

levels of recurrent seed-mediated gene flow among the islands. The highest levels of haplotype diversity were observed on the eastern islands of Wetar and Timor. The two most westerly islands, Flores and Lomblen, were fixed for the ancestral haplotype. Chloroplast haplotype diversity exhibited a decreasing trend from east to west in the species' range, consistent with an east-to-west colonisation route across the seven islands. Environmental factors that may have contributed to the contemporary spatial distribution of chloroplast DNA haplotypes include island paleogeology, ocean currents, fluctuations in sea levels and possible hybridization events with *Eucalyptus alba*. Twelve nuclear microsatellite markers were used to infer the geographical distribution of nuclear genetic diversity and determine whether populations are highly structured, or whether there appears to be sufficient pollen flow among populations so as to limit population divergence. High levels of microsatellite diversity were observed throughout 19 geographically defined *E. urophylla* populations (H_E = 0.703 to 0.776). The level of genetic differentiation among populations was low (F_{ST} = 0.031), but the amount of differentiation increased with geographic distance. A Bayesian clustering approach revealed a cryptic population structure comprising two genetically homogeneous groups, broadly structured according to geography. Pollen flow among the populations is likely responsible for the apparent weak influence of geographic insularity on nuclear population structure of the island species. In order to gain insight into the diversity and population structure of genes associated with economically important traits, we investigated the geographical distribution of single nucleotide polymorphism (SNP) haplotypes in two wood formation genes (*sucrose synthase1*, *EuSuSy1* and *cinnamyl alcohol dehydrogenase2*, *EuCAD2*). Three haplotype-tag SNPs at the 5' and 3' ends of each gene

were genotyped using a polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) approach. High levels of SNP haplotype diversity were observed throughout 19 geographically defined *E. urophylla* natural populations (H_d = 0.58 to 0.76). Low levels of population genetic structure were observed for the SNP haplotypes of each gene ($F_{ST} \leq 0.08$). These findings reflect those obtained in the microsatellite study based on the same *E. urophylla* samples. The results of this study provide improved direction for conservation and breeding strategies in *E. urophylla*.

Molecular Genetic Diversity and Population Genetic Structure of the Commercially Important Tropical Forest Tree Species *Eucalyptus urophylla*

by
Kitt Garnet Payn

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfilment of the
requirements for the Degree of
Doctor of Philosophy

Forestry

Raleigh, North Carolina

2008

APPROVED BY:

_____
Chair of Advisory Committee
William S. Dvorak

_____
Committee Member
Alexander A. Myburg

_____
Committee Member
Ronald R. Sederoff

_____
Committee Member
Gary R. Hodge

DEDICATION

This dissertation is dedicated to my beautiful wife, Carolyn.

BIOGRAPHY

Kitt Payn was born in Pietermaritzburg, KwaZulu-Natal, South Africa on the 28 April 1978. He attended Clifton Preparatory School, followed by Michaelhouse, matriculating in 1995. He commenced his tertiary education at the University of Natal (now the University of KwaZulu-Natal) where he obtained a B.Sc. Degree in 1998, majoring in Cell and Environmental Biology. He was actively involved as a student representative in the Department of Biology and Environmental Sciences.

Kitt was introduced to forest research during his 1998 summer vacation. This comprised voluntary work in the nursery and laboratories of Mondi in Hilton, KwaZulu-Natal. Kitt was subsequently awarded a bursary from Mondi, receiving financial support for his Honours and Masters degrees. During this time, he was based at the Forestry and Forest Products Research Center at the Councl for Scientific and Industrial Research (CSIR), Durban, where he focused primarily on the inheritance of wood property traits.

In January 2002, Kitt joined Mondi's Tree Improvement Research Department in Pietermaritzburg as a Research Scientist. Kitt married Carolyn Mitchell in June 2003. They moved to Raleigh, North Carolina in August of that same year, where Kitt enrolled in a Doctoral Program in Forestry at NC State University. Kitt worked part-time for Camcore in accordance with the Camcore stipend. Following the completion of the course work and preliminary examination, Kitt returned to South Africa in August 2005 to commence the research

component of the degree in the laboratories of Professor Myburg at the University of Pretoria. Upon completion of his PhD., Kitt will return to KwaZulu-Natal, where he will take up the position of Program Leader for Softwood Breeding in Mondi.

In addition to being a keen fly-fisherman, Kitt enjoys hiking, rugby and overseas travel.

ACKNOWLEDGEMENTS

I would like to thank my entire committee for their support and encouragement during my studies at North Carolina State University. I am grateful to Bill Dvorak and the Camcore staff for welcoming me into the team during my time in Raleigh. I would like to thank Zander Myburg and the students of the Myburg laboratories at the University of Pretoria for creating an enjoyable and intellectually stimulating environment in which to conduct my research. I would also like to extend my thanks to Mondi's Tree Improvement Department for assisting me with my fieldwork.

I would like to thank Camcore for awarding me the Camcore stipend, and I hope that future students will continue to benefit from the program. Finally, I am particularly grateful to Mondi for their continued financial support, without which this study would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

PREFACE

The genus *Eucalyptus* was named in the late eighteenth century by the French botanist, Charles Louis L'He´ritier de Brutelle. At that time there was little appreciation of the potential of the genus to become a major source of forest products. The wood was difficult to saw and season, and it was often considered to have value only as firewood. In Brazil and South Africa, eucalypts were initially planted along railways to provide fuel for wood-burning locomotives. More recently, eucalypts have been planted to meet the growing demand for short fibre pulp. Today, *Eucalyptus* is the most widely planted hardwood genus in the world due to its wide adaptability, extremely fast growth rate, good form and excellent fiber and strength properties.

Eucalyptus urophylla is one of the most commercially important *Eucalyptus* species in tropical and subtropical regions of Africa, South America and Asia, primarily as a hybrid parent. Unfortunately, the conservation status of the majority of the populations that naturally occur on seven islands in eastern Indonesia ranges from vulnerable to critically endangered. The accumulation and long-term conservation of genetic material from different *E. urophylla* populations are fundamental to growing healthy, productive plantations and remaining competitive in an industry where forest products, markets, and environmental conditions continually change. Therefore, *E. urophylla* seed collections have been carried out by several international organizations over the last 30 years to provide a broad genetic resource necessary for effective breeding and conservation programs.

An understanding of the genetic diversity of natural *E. urophylla* populations is important because knowledge of their genetic structure can be applied to better conserve and utilize the evolutionary potential of the species. Genetic diversity and population genetic structure are subject to evolutionary forces including, selection, migration and genetic drift. Natural selection favoring adaptation to environmental gradients or discontinuous habitat variation, for example may produce genetic differentiation among populations. This genetic structure is evident in *ex situ E. urophylla* provenance trials. Factors such as small population size and inbreeding, which reduce within population genetic variation via genetic drift, further contribute to the development of population genetic structure. In contrast to these potentially disruptive forces, historical and contemporary gene flow among populations reduce population differentiation. Evolutionary forces such as migration and drift are difficult to quantify based on phenotypic provenance trials. Molecular marker techniques are more suitable in this regard.

This dissertation describes the results of a detailed molecular marker study of the genetic diversity and population genetic structure of *E. urophylla*. Plants originating from an extensive series of *E. urophylla* seed collections, conducted by Camcore, North Carolina State University, were genotyped at the chloroplast and nuclear genome levels. This molecular approach served to achieve the following objectives:

1) Gained insight into the evolution of *E. urophylla* and, more specifically, inferred the historical seed migration routes of the species across the seven islands in its natural range in the context of historical and contemporary environmental factors.
2) Described the level of genome-wide genetic variation of *E. urophylla*, and how the variation is distributed within populations, among populations, and among islands.
3) Determined the level of gene flow among the island populations.
4) Described the level of gene-specific genetic variation of two wood formation genes in *E. urophylla*, and the distribution of this variation within populations, among populations, and among islands.
6) Provided recommendations for *E. urophylla* conservation and breeding strategies.

Chapter 1 of the dissertation introduces three topics. The first topic is the importance of *Eucalyptus* in the international forestry industry. The second topic is the natural distribution of *E. urophylla* and conservation status of the species. The third topic is the importance of understanding the levels of genetic diversity and population genetic structure of *E. urophylla*, and the three molecular marker techniques used in this study to gain such understanding.

Chapter 2 comprises the phylogeographic analysis of *E. urophylla* based on chloroplast DNA sequence variation, and provides insight into the historical seed migration routes of the species along the island archipelago. Environmental factors likely contributing to the contemporary spatial distribution of chloroplast DNA haplotypes are discussed.

In Chapter 3, microsatellite markers are used to evaluate the structure of genome-wide genetic variation of *E. urophylla.* Gene flow among populations is quantified, and its effect on the apportionment of total genetic variation within populations, among populations, and among islands is discussed. SNP haplotype diversity and population structure of two wood formation genes (*EuSuSy1* and *EuCAD2*) are investigated in Chapter 4 to gain insight into allelic variation of candidate genes in *E. urophylla* populations. Finally, major conclusions of the research project in the context of managing an *ex situ E. urophylla* genetic resource population for conservation and breeding purposes are summarized in Chapter 5.

The findings presented in this PhD. dissertation represent the outcomes of a study undertaken from August 2003 to July 2008 in the Department of Forestry and Environmental Resources, North Carolina State University, under the supervision of William S. Dvorak, Alexander A. Myburg, Gary R. Hodge and Ronald R. Sederoff. Chapters 2, 3 and 4 have been prepared in the format of independent manuscripts for the publication in refereed journals as indicated at the beginning of each chapter. A certain degree of redundancy may therefore exist between the introductory sections of these chapters and Chapter 1. In addition, the following presentations and publications were produced during the course of this research project:

PAYN KG, Maleka FM, Dvorak WS, Janse BJH, Myburg AA (2006) Phylogeography of *Eucalyptus urophylla* based on the chloroplast J_{LA} region. 6th Symposium of the Southern African Plant Breeders' Association 13-15 March. Club Mykonos, Langebaan, South Africa. (oral presentation)

Maleka FM, PAYN KG, Bloomer P, Janse BJH, Dvorak WS, Myburg AA (2006) Allelic diversity and Linkage Disequilibrium in wood and fibre genes of *Eucalyptus urophylla*. South African Genetics Society (SAGS) Conference, 2-4 April. Bain's Game Lodge, Bloemfontein, South Africa. (oral presentation)

PAYN KG, Dvorak WS, Myburg AA (2007) Chloroplast Chloroplast DNA phylogeography reveals the island colonisation history of *Eucalyptus urophylla* (Myrtaceae). IUFRO Tree Biotechnology, 3-8 June. Ponta Delgada, Azores, Portugal. (oral presentation)

PAYN KG, Dvorak WS, Janse BJH, Myburg AA (2007) Genetic diversity and population structure of *Eucalyptus urophylla (Myrtaceae)* based on microsatellite markers. IUFRO WP 2.08.03 Improvement and Culture of Eucalypts, "Eucalypts and diversity: balancing productivity and sustainability", 22-26 October. Durban, South Africa. (oral presentation)

Dvorak WS, Hodge GR, PAYN KG (2007). The conservation and breeding of *Eucalyptus urophylla*: A case study to better protect important populations and improve productivity. IUFRO WP 2.08.03 Improvement and Culture of Eucalypts, "Eucalypts and diversity: balancing productivity and sustainability", 22-26 October. Durban, South Africa. (oral presentation)

PAYN KG, Dvorak WS, Myburg AA (2007) Chloroplast DNA phylogeography reveals the island colonisation route of *Eucalyptus urophylla* (Myrtaceae). *Australian Journal of Botany* **55**, 673-683

PAYN KG, Maleka FM, Dvorak WS, Janse BJH Myburg AA (2008) Population genetic analysis of *Eucalyptus urophylla* based on chloroplast DNA, microsatellite and single nucleotide polymorphism (SNP) diversity. South African Genetics Society Congress, 27-29 March. Pretoria, South Africa. (oral presentation)

PAYN KG, Dvorak WS, Janse BJH, Myburg AA (2008) Microsatellite diversity and genetic structure of the commercially important tropical tree species *Eucalyptus urophylla*, endemic to seven islands in eastern Indonesia. *Tree Genetics and Genomes* **4,** 519-530

Dvorak WS, Hodge GR, PAYN KG (2008). The conservation and breeding of *Eucalyptus urophylla*: A case study to better protect important populations and improve productivity. *Southern Hemisphere Forestry Journal* (in press)

CHAPTER 1

Introduction

Importance of *Eucalyptus* in plantation forestry

There is an increasing demand for forest products from a diminishing forest resource. The world presently has marginally less than four billion hectares of forest, covering approximately 30 percent of the total land area (FAO 2007). This resource is reduced by deforestation at a rate of 13 million hectares a year. Concurrently, the planting of forests together with the natural expansion of some forests, largely through conservation efforts, have significantly reduced the net loss of forest area to approximately 7.3 million hectares per year or 20 000 hectares per day (FAO 2007).

Planted forests, comprising plantations and the planted component of semi-natural forests (enrichment planting of native species), are expanding and now contribute almost 50 percent of the total global wood production (FAO 2007). The genus *Eucalyptus* is the most widely grown commercial hardwood with about 18 million hectares planted worldwide (FAO 2000). The greatest industrial use for *Eucalyptus* wood is for the production of pulp and paper. However, the wood is also commonly used for a wide range of end products, including firewood, charcoal, furniture, mine props, plywood and chipboard (Campinhos 1999).

The genus *Eucalyptus* was first described in 1788 by the French botanist Charles Louis L'He´ritier de Brutelle, based on a specimen from Bruny Island, Tasmania, thought to be collected during James Cook's third expedition (Eldridge et al. 1993). The genus now consists of more than 600 species, primarily endemic to the Australian continent (Ladiges et al. 2003). *Eucalyptus* species naturally occur over a wide range of habitats, from semi-arid to rainforest. Their diverse adaptation enables different species to be successfully planted in different regions, experiencing tropical or temperate climates between latitudes 45°S and 40°N, in addition to a wide range in elevation (Campinhos 1999). Large areas of *Eucalyptus* plantations were first established in Brazil with the introduction of many species in the early 1900s (Eldridge et al. 1993). More recently it was estimated that Brazil had 3,000,000 hectares of planted areas, of which most comprised *E. grandis*. India has approximately 8 million hectares of mostly *E. tereticornis*. In more temperate areas such as Spain, Portugal and Chile, *E. globulus* is widely planted (Eldridge et al. 1993; Campinhos 1999; FAO 2000).

Eucalyptus hybrids form a significant component of *Eucalyptus* plantation forestry, particularly in tropical and sub-tropical regions, where cost-efficient, clonal propagation is key to their exploitation (Potts and Dungey 2004). They have been incorporated into forestry operations for several decades, with initial hybrid plantations being established from spontaneous hybridization occurring in exotic environments, including botanical gardens, species trials, plantations and seed orchards (Eldridge et al. 1993; Griffin et al. 1998). Such hybrids became eminent in tropical and sub-tropical regions through the widespread planting of *E. grandis* x *E. urophylla* clones in Brazil and the Congo (Republic of the Congo).

Manipulated hybridization through controlled pollination programs soon followed (Potts and Dungey 2004).

Most *Eucalyptus* hybrids deployed today are first generation (F_1) hybrids, with the majority from the sections *Maidenaria* (e.g. *E. globulus* and *E. nitens*), *Exsertaria* (e.g. *E. camaldulensis and E. tereticornis*) and *Transversaria* (renamed to *Latoangulatae*, Brooker 2000) (e.g. *E. pellita*, *E. grandis* and *E. urophylla*) of the subgenus *Symphyomyrtus* (Griffin et al. 1988; Eldridge et al. 1993). The most common inter-specific crosses (and reciprocals) consist of *E. grandis* x *E. urophylla*, *E. grandis* x *E. camaldulensis*, and hybrids including at least one of *E. saligna*, *E. pellita*, *E. exserta* and *E. tereticornis* (Potts and Dungey 2004). Such hybrids are planted widely in Brazil and the Congo, and large plantations have also been established in other countries including China, Colombia, Venezuela, Indonesia and South Africa.

Natural habitat and conservation status of *Eucalyptus urophylla*

The genus *Eucalyptus* is primarily endemic to the Australian continent with a number of the eucalypts also indigenous to countries north of the continent. For example, New Guinea hosts several species, including *E. alba*, *E. papuana*, *E. confertiflora*, *E. tereticornis* and *E. polycarpa* (Martin and Cossalter 1975). The only two *Eucalyptus* species that occur exclusively outside of Australia are *E. deglupta* and *E. urophylla*. *Eucalyptus deglupta* has a natural habitat that extends from New Guinea to the Philippines (Martin and Cossalter 1975), whereas the natural distribution of *E. urophylla* is limited to a series of disjunct populations

on the Lesser Sunda Islands in eastern Indonesia (Fig. 1). The largest natural stands are found on the islands of Timor and Wetar, which are the eastern-most occurrences recorded to date. More scattered stands occur to the west on the islands of Alor, Pantar, Lomblen (lembata), Adonara and Flores. Pryor et al. (1995) reported that *E. urophylla* once occurred on the island of Solor, but slash and burn agriculture has led to its disappearance on the island. The species has not been found on islands further west along the archipelago beyond Flores (Martin and Cossalter 1975).

The Lesser Sunda Islands form the western part of two island chains, the inner and outer Banda arcs, which lie between the Asian and Australian continental shelves (Fig. 2). The inner arc islands are principally volcanic and include, among others, the islands of Flores through Wetar. Timor is the largest island of the non-volcanic outer arc. The Banda arcs were formed by the collision and subduction of the Australian Plate beneath the Asian Plate during the Pliocene (Norvick 1979). Interestingly, this collision zone corresponds closely to the region of Wallacea, which is a biogeographic region of transition situated between areas with entirely Asiatic and Australian floras and faunas and is characterised by a high degrees of endemism (Myers et al. 2000). Several geological and environmental factors have influenced the contemporary distribution of Australasian and Asian biota on these islands. They include the continuing northward movement of the Australian continental margin into the region (Michaux 1991), the associated volcanic activity on many of the islands, and sea level fluctuations during the Pleistocene (Voris 2000) creating and breaking land bridges thus affecting migration among islands.

Eucalyptus urophylla S.T. Blake (Timor mountain gum) was formally described in 1977 (Blake 1977). Before then, it had been mistakenly labelled *E. decaisneana* (Blume 1849 in Eldrige et al. 1993). The species occurs from low elevation or at times almost sea level to high volcanic mountains with the best development in a broad band between approximately 500 m and 2200 m elevation (Pryor et al. 1995). Over most of its range it is a forest tree reaching 25-45 m in height and up to 2.5 m in diameter, with a straight bole for half to two thirds of the tree height. At an elevation of almost 3000 m, however, the species grows as a small, gnarled shrub less than one meter tall. In addition to the growth variation, there is variation in fruit size and bark color and texture, resulting in this species being described as one of the most variable of all the eucalypts (Eldridge et al. 1993). A remarkable feature of *E. urophylla* is that it is the only *Eucalyptus* species occupying such a wide altitudinal range and diverse microhabitats along the island archipelago; such a wide range in Australia would be occupied by a mosaic of several *Eucalyptus* species (Eldridge et al. 1993). However, on the lower slopes *E. urophylla* does form a mosaic distribution pattern with *E. alba* Reinw. ex Blume, which dominates these drier sites.

The importance of *E. urophylla* as a plantation tree species led to several large-scale seed collections from natural stands in an effort to capture the available genetic diversity and ensure a broad genetic base necessary for effective tree breeding programs (Martin and Cossalter 1975; Gunn and McDonald 1991). From 1996 to 2003, a new series of explorations and seed collections were conducted by research staff of PT Sumalindo Lestari Jaya, a

private Indonesian forestry company and Camcore, North Carolina State University, USA, an international tree conservation and domestication program. This series was the most comprehensive *E. urophylla* seed collection to date, comprising seeds from 1104 mother trees distributed across 62 provenances representing all seven islands on which *E. urophylla* is known to occur (Hodge et al. 2001; Pepe et al. 2004).

One of the exploration objectives was to assess and categorize the *in situ* conservation status of the species according to the rules developed by the International Union for Conservation of Nature and Natural Resources. Provenances classified as *Low risk* were located exclusively on the islands of Timor and Wetar (Pepe et al. 2004). Most Timor provenances were protected within the boundaries of Mt. Mutis Forest Park, and Wetar provenances were at low risk because the human population pressure on the island was minimal. However, the conservation situation on the remaining five islands was very different. On these islands, the provenance status ranged from *Vulnerable* to *Critically endangered*.

The depletion in the genetic resource was primarily the result of land conversion to agriculture, including the establishment of short-rotation crops like macadamia nut trees (Pepe et al. 2004). Camcore, through its industrial members, have established more than 100 conservation-banks/genetic-tests of *E. urophylla* distributed across Argentina, Brazil, Colombia, Mexico, South Africa and Venezuela (Dvorak et al. 2008). These e*x situ* plantings may be the last hope for the conservation and future restoration of many of the endangered *E. urophylla* provenances.

Population genetic structure of *Eucalyptus urophylla*

Fundamental to the effective implementation of successful conservation strategies, is a detailed understanding of the genetic diversity and population genetic structure of the species concerned. For example, the identification of genetically homogenous regions enables one to determine the number and location of primary gene-pool reserves (Bucci and Vendramin, 2000). Such gene-pools may be sufficiently differentiated to deserve management as separate units. Populations with low genetic diversity may be considered to be at higher risk because their ability to evolve by adapting to environmental change is compromised (Frankham et al 2004).

Information on the levels and distribution of genetic diversity is also central to long-term breeding programs, particularly when balancing the maintenance of genetic diversity and continuous genetic gains. Hybrid breeding programs may adopt recurrent selection (RS) or reciprocal recurrent selection (RRS) as the strategy of choice, primarily depending on whether additive or non-additive gene effects predominate (Eldridge et al 1993; Potts and Dungey 2004). For either strategy, it would be beneficial for the breeding population of each pure species to reflect inherent levels of genetic diversity and allelic richness present in their native populations, so as to enhance the ability to respond to different breeding objectives and changing environmental conditions (Burley and Knowski 2005).

Genetic variation of forest trees has traditionally been studied following the approach of progeny tests and provenance trials established in different environments with the focus being on quantitative phenotypic traits. Provenance trials of *E. urophylla* in tropical countries have demonstrated that low to mid-elevation seed sources generally have superior height growth compared to high elevation seed sources (Moura 1983; Wencelius 1983; Hodge et al. 2001). In a study of provenance variation in seedling morphology, seedlings from the island of Wetar were distinguished from the remaining islands on the basis of narrower leaves, more acute leaf base angles and square stems (Pinyopusarerk et al. 1993). Similar results were shown for leaf oil composition where provenances from the island of Wetar differentiated from other provenances of *E. urophylla* (Doran et al. 1995). A contemporary study of *in situ* floral and foliar morphological features resulted in several dry site provenances on Wetar and several high elevation provenances in East Timor being differentiated from other provenances of *E. urophylla* (Pryor et al. 1995). The morphological differentiation led to the proposed separation of two new species from *E. urophylla*, namely *E. wetarensis* and *E. orophila* on the island of Wetar and Timor, respectively (Pryor et al. 1995). This new taxonomic classification has not been unanimously accepted (e.g. Brooker 2000).

The utilization of molecular markers is a complimentary approach to the study of phenotypic traits and allows for direct genome sampling and precise estimation of genetic divergence, variability and discrimination among genotypes, all of which are useful for the implementation of effective and sustainable conservation and breeding strategies. Molecular

genetic analyses of *Eucalyptus* species have been used to investigate population genetic structure (Steane et al. 2006), the interaction among populations through gene flow (Jones et al. 2005), and the demographic histories of populations and their contribution to contemporary levels of diversity and population structure (Jones et al. 2006).

In order to study natural populations, one has to consider suitable strategies for sampling within and among both populations and genomes (nuclear and organellar). In addition, choices must be made with respect to which markers will be appropriate for each particular investigation (Lowe et al. 2004). In plants, genetic variation occurs in the nuclear as well as the chloroplast and mitochondrial genomes. These genomes differ in their evolutionary characteristics, which determine the types of genetic questions that they may be used to investigate. For example, the nuclear genome is typically biparentally inherited, undergoes recombination and is subject to a relatively high mutation rate (Wolfe et al. 1987), all of which result in potentially high levels of genetic polymorphism within populations. In contrast, plant chloroplast and mitochondrial genomes generally exhibit a more conserved pattern of genetic diversity due to their typical uniparental mode of inheritance, absence of meiotic recombination and lower mutation rates (Wolfe et al. 1987; Birky, 1995). Such characteristics of organellar genomes, particularly the chloroplast genome, typically result in significant phylogeographic structure (Demesure et al. 1996; King and Ferris 1998; Cavers et al. 2003).

Phylogeography is a field of study that analyses the geographical distribution of genealogical lineages, especially those at the intraspecific level (Avise 1998). The relationships of genealogies may be displayed geographically and analysed to deduce the evolutionary history of populations. This is of importance since contemporary levels of genetic variation and population genetic structure may be influenced by historical events such as range expansion, population fragmentation and population bottlenecks (Schaal et al. 1998). This is particularly true for species with ecological barriers to gene flow, as in the case of the tropical tree *Santalum austrocaledonicum* (sandalwood), endemic to the archipelago of New Caledonia (Bottin et al. 2005).

The chloroplast genome is maternally inherited in most angiosperms (Corriveau and Coleman, 1988), including *Eucalyptus* (Byrne et al. 1993; McKinnon et al. 2001), thus the geographical distribution of chloroplast DNA haplotypes across the native range of *E. urophylla* can be used to infer its historical seed migration patterns. On the basis of the proposed order of geological events, with the inner Banda arc arising before the outer Banda arc (Audley-Charles 2004), one might assume that *E. urophylla* may have colonised several of the inner arc islands, possibly starting with the older islands in the west (e.g. Flores and Adonara), before spreading east along the inner arc as the islands were formed. This would then have been followed by the colonisation of the younger outer arc island of Timor. Alternatively, *E. urophylla* may have initially colonised the islands of Timor and/or Wetar, since these islands are geographically closer to the natural distribution of related species such as *E. pellita* (Brooker 2000), indigenous to New South Wales, Queensland and New Guinea.

Then a westerly migration route along the archipelago toward Flores ensued. Support for either hypothesis would be gained by a gradient of decreasing chloroplast genetic diversity congruent with an increase in geographic distance from the source population (Demesure et al. 1996; Ferris et al. 1998; King and Ferris 1998).

Because *Eucalyptus* seeds are mainly dispersed by gravity (House 1997), volcanic peaks and seawater will likely limit inter-population seed dispersal. Therefore, a high degree of population structure at the chloroplast level is expected. Of interest is whether such gene flow barriers significantly affect vector-mediated pollen dispersal and thus affect the population structure of the biparentally inherited nuclear genome of *E. urophylla.*

Population genetic studies of tropical trees have long shown that many species investigated exhibit high levels of gene flow and carry much of the variation within rather than among populations (Hamrick and Loveless 1986; Loveless 1992). Island populations, however, may experience unique ecological factors that influence the spatial distribution of genetic variation. For example, islands are discrete geographic entities with defined oceanic boundaries. Despite their often-small size they contain a variety of habitats. In addition, islands are sometimes geologically dynamic with historical volcanic activity. It is the combination of these factors that is partly responsible for the typically high levels of endemism within oceanic island systems (Witter and Carr 1988; Emerson 2002).

The impact of island systems on intraspecific nuclear genetic diversity has not been widely reported for forest tree species (Sheely and Meagher 1996; Bottin et al. 2005; Karhu et al. 2006). Sheely and Meagher (1996) examined the distribution of genetic variation in four widely dispersed island populations of *Campnosperma brevipetiolata* (Anacardiaceae) in the Micronesian islands, north-east of Indonesia. Seeds of this species are dispersed by birds and possibly bats, and their pollinators are presumed to be insects such as flies, bees, or wasps (Sheely and Meagher 1996). The authors hypothesized that there would be a decrease in genetic variation with increasing distance from the presumed Indo-Malayan source of the species. Indeed, a trend of decreasing genetic variation from west to east was observed. Most of the genetic variation in the species was apportioned within populations, although 17% of the total variation was apportioned among islands. Jones et al. (2002) investigated the genetic diversity and population structure of *E. globulus*, which is indigenous to the south-eastern Australian mainland and islands further south. Biotic vectors such as insect and bird taxa disperse the pollen of *E. globulus*. However, unlike *C. brevipetiolata*, *E. globulus* seeds are likely dispersed by gravity (Hingston and Potts 1998). A low level of population structure was observed (F_{ST} = 0.08) and genetic diversity was evenly distributed across the species range (Jones et al. 2002). Since genetic variation in the *E. globulus* chloroplast genome is geographically structured (Freeman et al. 2001), it appears that gene flow via pollen may be predominantly contributing to the low level of population structure observed for the nuclear genome.

The perceived presence of gene flow barriers, restricting pollinator movements among the Lesser Sunda islands, are expected to increase the population subdivision of *E. urophylla* due to genetic drift and environmental stochasticity (volcanic activity and changes in sea level). Alternatively, if the island distribution does not prevent inter-population gene flow (e.g. pollen dispersed through wind blown insects) then limited population differentiation will be observed (Wright 1931).

Levels of nuclear genetic diversity and population differentiation of *E. urophylla* have been investigated using isozyme markers (House and Bell 1994). Limited population differentiation (G_{ST} = 0.12) was observed, suggesting that most of the genetic diversity in the species was found within rather than between populations. In addition, populations generally did not cluster according to geography, and there were large error bars associated with the higher order clustering represented in the UPGMA dendogram.

A potentially higher resolution class of molecular markers suitable for the assessment of genetic differentiation and gene flow of *E. urophylla* are nuclear microsatellites, or simple sequence repeats (SSR). They are typically highly polymorphic co-dominant markers that are distributed throughout eukaryotic genomes (Goldstein and Pollock, 1997). The primary disadvantage of microsatellites is their high cost of development. However, this class of molecular marker has been developed for several plantation forest tree genera including *Populus* (van der Schoot et al. 2000), *Pinus* (Smith and Devey 1994; Kostia et al. 1995), *Quercus* (Dow et al. 1995) and *Eucalyptus* (Brondani et al. 1998, 2002). In addition, the

transferability of these markers among closely related species within a subgenus, as is the case with *Symphyomyrtus* (Brondani et al. 1998; Marques et al. 2002), has been largely successful. Microsatellite markers have been used in a number of eucalypt population studies as a valuable tool for inferring relationships between individuals and population genetic structure (Byrne et al. 1996, McGowen et al. 2001; Jones et al. 2002; Holman et al. 2003; Steane et al. 2006; Tripiana et al. 2007).

Microsatellite analyses infer the genetic diversity and population structure at likely neutral loci. The advantage is that such markers enable population geneticists to investigate important neutral evolutionary forces such as migration and genetic drift (Black et al. 2001). The genetic diversity and population structure at genes that govern phenotypic variation in adaptive and/or economically important traits such as growth rate, cold hardiness and various wood properties, may be very different. Consequently, genetic markers that are closely linked with candidate genes might provide a better indicator of levels of functional variation within populations.

Direct sequencing of candidate gene regions from multiple individuals may result in the identification of single nucleotide polymorphisms (SNPs) (Shattuck-Eidens et al. 1990; Gaut and Clegg 1993). SNPs are typically biallelic and their expected heterozygosity is low relative to microsatellite markers. However, SNPs are most useful when several closely spaced SNPs completely define haplotypes in the gene region being examined. In the presence of linkage disequilibrium (LD), the nonrandom association of alleles at different

sites, some of the SNPs that form a haplotype are redundant and so only a subset of SNPs that completely defines the haplotypes of interest need to be assayed (Rafalski 2002). The occurrence of LD underlies association genetics, a population-based approach used to identify relationships between allelic forms of a gene and a phenotype. In association genetic studies, it is usually not feasible to genotype all the discovered SNPs in a candidate gene and so a subset of SNPs that will identify or ‘‘tag’’ common haplotypes is genotyped. The first association genetics study in plants was reported in 2001 when flowering time, a quantitative trait, could be associated with polymorphisms in the *dwarf8* gene of maize (Thornsberry et al. 2001). Since then, other marker-trait association studies in plants have been reported (Olsen et al. 2004; de Meaux et al. 2005; Thumma et al. 2005).

SNPs have recently been discovered in two *E. urophylla* wood formation genes (*sucrose synthase1*, *EuSuSy1* and *cinnamyl alcohol dehydrogenase2*, *EuCAD2*; Maleka et al. 2007) involved in cellulose and lignin biosynthesis. Understanding the genetic architecture of such candidate genes is of great importance to the pulp and paper industry since cellulose content is correlated with pulp yield, whereas lignin is an undesirable component causing paper to be weak, rigid and susceptible to discoloration. The discovered SNPs will be useful in the present study aimed at assaying haplotype diversity in *EuSuSy1* and *EuCAD2* across the natural distribution of *E. urophylla*, because conservationists and breeders are predominantly interested in conserving and managing functional genetic variation. This information may be useful for designing future association genetics studies that aim to associate alleles with phenotypic traits in *E. urophylla*.

Literature Cited

Audely-Charles MG (2004) Ocean trench blocked and obliterated by Banda forearc collision with Australian proximal continental slope. *Tectonophysics* **389**, 65–79.

Avise J (1998) The history and purview of phylogeography: a personal reflection. *Molecular Ecology* **7**, 371–379.

Birky CW (1995) Uniparental inheritance of mitochondrial and chloroplast genes: Mechanisms and evolution. *Proceedings of the National Academy of Sciences USA* **92**, 11331–11338.

Blake ST (1977) Four new species of *Eucalyptus*. *Austrobaileya* **1**, 7–9.

Bottin L, Verhaegen D, Tassin J, Olivieri I, Vaillant A, Bouvet JM (2005) Genetic diversity and population structure of an insular tree, *Santalum austrocaledonicum* in New Caledonian archipelago. *Molecular Ecology* **14**, 1979–1989.

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

Brondani RPV, Brondani C, Grattapaglia D (2002) Towards a genus wide reference linkage map for *Eucalyptus* based exclusively on highly informative microsatellite markers. *Molecular Genetics and Genomics* **267**, 338–347.

Brooker MIH (2000) A New Classification of the Genus *Eucalyptus* L'Hér. (Myrtaceae). *Australian Systematic Botany* **13**, 79–148.

Bucci G, Vendramin GG (2000) Delineation of genetic zones in European Norway spruce natural range: preliminary evidence. *Molecular Ecology* **9**, 923–934.

Burley J, Kanowski PJ (2005) Breeding strategies for temperate hardwoods. *Forestry* **78**, 199–208.

Byrne M, Moran GF, Tibbits WN (1993) Restriction map and maternal inheritance of chloroplast DNA in *Eucalyptus nitens*. *Journal of Heredity* **84**, 218–220.

Byrne M, Marques-Garcia MI, Uren T, Smith DS, Moran GF (1996) Conservation and genetic diversity of microsatellite loci in the genus *Eucalyptus*. *Australian Journal of Botany* **44,** 331–341.

Campinhos E (1999) Sustainable plantations of high-yield *Eucalyptus* trees for production of fiber: the Aracruz case. *New Forests* **17**, 129–143.

Cavers S, Navarro C, Lowe AJ (2003) Chloroplast DNA phylogeography reveals colonization history of Neotropical tree, *Cedrela odorata* L., in Mesoamerica. *Molecular Ecology* **12,** 1451–1460.

Corriveau JL, Coleman AW (1988) Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *American Journal of Botany* **75**, 1443–1458.

De Meaux J, Goebel U, Pop A, Mitchell-Olds T (2005) Allele-specific assay reveals functional variation in the chalcone synthase promoter of Arabidopsis thaliana that is compatible with neutral evolution. *Plant Cell* **17,** 676–690.

Demesure B, Comps B, Petit RJ (1996) Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica* L.) in Europe. *Evolution* **50,** 2515–2520.

Doran JC, Williams ER, Brophy JJ (1995) Patterns of variation in the seedling leaf oils of *Eucalyptus urophylla*, *E pellita* and *E scias*. *Australian Journal of Botany* **43,** 327–336.

Dow BD, Ashley MV, Howe HF (1995) Characterization of highly variable (Ga/CT) *n* microsatellites in the bur oak, *Quercus macrocarpa*. *Theoretical and Applied Genetics* **91**, 137–141.

Dvorak WS, Hodge GR, Payn KG (2008). The conservation and breeding of *Eucalyptus urophylla*: A case study to better protect important populations and improve productivity. *Southern Hemisphere Forestry Journal* (in press).

Eldridge K, Davidson J, Harwood C, Van Wyk G (1993) 'Eucalypt domestication and breeding.' (Oxford University Press: Oxford).

Emerson BC (2002) Evolution on oceanic islands: molecular phylogenetic approaches to understanding pattern and process. *Molecular Ecology* **11**, 951–966.

FAO (2000) Global forest resources assessment. Food and Agriculture Organisation, Rome.

FAO (2007) State of the world's forests. Food and Agriculture Organisation, Rome.

Frankham R, Ballou JD, Briscoe DA (2004) A primer of conservation genetics. Cambridge University Press, Cambridge, USA.

Freeman JS, Jackson HD, Steane DA, McKinnon GE, Dutkowski GW, Potts BM, Vaillancourt RE (2001) Chloroplast phylogeography of *Eucalyptus globulus*. *Australian Journal of Botany* **49,** 585–596.

Gaut BS, Clegg MT (1993) Nucleotide polymorphism in the *Adh1* locus of pearl millet (*Pennisetum glaucum*) (Poaceae). *Genetics* **135**, 1091–1097.

Goldstein DB, Pollock DD (1997) Launching microsatellites: A review of mutation processes and methods of phylogenetic inference. *Journal of Heredity* **88**, 33–342.

Gonzalez-Martinez SC, Ersoz E, Brown GR, Wheeler NC, Neale DB (2006) DNA sequence variation and selection of tag single-nucleotide polymorphisms at candidate genes for drought-stress response in Pinus taeda L. *Genetics* **172**, 1915–1926.

Griffin AR, Burgess IP, Wolf L (1988) Patterns of natural and manipulated hybridisation in the genus *Eucalyptus* L'Herit - a review. *Australian Journal of Botany* **36**, 41–66.

Gunn BV, McDonald MW (1991) *Eucalyptus urophylla* seed collections. Forest Genetic Resources 19, 34–37, FAO, Rome.

Hamrick JL, Loveless MD (1986) Isozyme variation in tropical trees: Procedures and preliminary results. *Biotropica* **18,** 201–207.

Hingston AB, Potts BM (1998) Floral visitors of *Eucalyptus globulus* subsp. *globulus* in eastern Tasmania. *Tasforests* **10,** 125–157.

Hodge GR, Pepe B, Wijoyo FS, Dvorak WS (2001) Early results of *Eucalyptus urophylla* provenance/progeny trials in Colombia and Venezuela. In: Developing the Eucalypts of the Future. Proc. IUFRO Working Party 2.08.03.

Holman JE, Hughes JM, Fensham RJ (2003) A morphological cline in *Eucalyptus*: a genetic perspective. *Molecular Ecology* **12**, 3013–3025.

House APN, Bell JC (1994) Isozyme variation and mating system in *Eucalyptus urophylla* ST Blake. *Silvae Genetica* **43,** 167–176.

House SM (1997) Reproductive biology of eucalypts. In 'Eucalypt Ecology: Individuals to Ecosystems' (Eds J Williams, J Woinarski) pp. 30–55. (Cambridge University Press: Cambridge).

Ingvarsson PK (2005) Nucleotide Polymorphism and Linkage Disequilibrium Within and Among Natural Populations of European Aspen (*Populus tremula* L., Salicaceae). *Genetics* **169**, 945–953.

Jones RC, Steane DA, Potts BM, Vaillancourt RE (2002) Microsatellite and morphological analysis of *Eucalyptus globulus* populations. *Canadian Journal of Forest Research* **32,** 59–66.

Jones RC, McKinnon GE, Potts BM, Vaillancourt RE (2005) Genetic diversity and mating system of an endangered tree *Eucalyptus morisbyi*. *Australian Journal of Botany* **53**, 367–377.

Jones ME, Shepherd M, Henry RJ, Delves A (2006) Chloroplast DNA variation and population structure in the widespread forest tree, *Eucalyptus grandis*. *Conservation Genetics* **7**, 691–703.

Karhu A, Vogl C, Moran GF, Bell JC, Savolainen O (2006) Analysis of microsatellite variation in *Pinus radiata* reveals effects of genetic drift but not recent bottlenecks. *Journal of Evolutionary Biology* **19**, 167–175.

King RA, Ferris C (1998) Chloroplast DNA phylogeography of *Alnus glutinosa* (L.) Gaertn. *Molecular Ecology* **7**, 1151–1161.

Kostia S, Varvio SL, Vakkari P, Pulkkinen P (1995) Microsatellite sequences in a conifer, *Pinus sylvestris*. *Genome* **38**, 1244–1248.

Krutovsky KV, Neale DB (2005) Nucleotide diversity and linkage disequilibrium in cold-hardiness-and wood quality-related candidate genes in Douglas fir. *Genetics* **171**, 2029–2041.

Ladiges PY, Udovicic F, Nelson G (2003) Australian biogeographical connections and the phylogeny of large genera in the plant family Myrtaceae. *Journal of Biogeography* **30**, 989–998.

Loveless MD (1992) Isozyme variation in tropical trees: patterns of genetic organization. *New Forests* **6**, 67–94.

Lowe A, Harris S, Ashton P (2004) Ecological genetics: Design, analysis, and application. Blackwell Publishing, Malden, MA USA.

Maleka MF (2007) Allelic diversity in cellulose and lignin biosynthetic genes of *Eucalyptus urophylla* S.T. Blake. M.Sc. thesis, University of Pretoria.

Marques CM, Brondani RPV, Grattapaglia D, Sederoff R (2002) Conservation and synteny of SSR loci and QTLs for vegetative propagation in four Eucalyptus species. *Theoretical and Applied Genetics* **105**, 474–478.

Martin B, Cossalter C (1975) Les eucalyptus des iles de la Sonde. *Bois et Foréts des Tropiques* **163**, 3–25.

McGowen MH, Wiltshire RJE, Potts BM, Vaillancourt RE (2001) The origin of *Eucalyptus vernicosa*, a unique shrub Eucalypt. *Biological Journal of the Linnean Society* **74,** 397–405.

Michaux B (1991) Distributional patterns and tectonic development in Indonesia. *Australian Systematic Botany* **4**, 25–36.

McKinnon GE, Vaillancourt RE, Tilyard PA, Potts BM (2001) Maternal inheritance of the chloroplast genome in *Eucalyptus globulus* and interspecific hybrids. *Genome* **44**, 831–835.

Moura VPG (1983) Resultados de pesquisa com várias procêdencias de *Eucalyptys urophylla* ST Blake, no centro-leste do Brasil. *Silvicultura, São Paulo* **31**, 474–480.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858.

Norvick MS (1979) The tectonic history of the Banda Arcs, eastern Indonesia: a review. *Journal of the Geological Society London* **136**, 519–527.

Olsen KM, Halldorsdottir SS, Purugganan MD, Stinchcombe JR, Schmitt J, Weinig C (2004) Linkage Disequilibrium mapping of Arabidopsis CRY2 flowering time alleles. *Genetics* **167**, 1361–1369.

Pepe B, Surata K, Suhartono F, Sipayung M, Purwanto A, Dvorak WS (2004) Conservation status of natural populations of *Eucalyptus urophylla* in Indonesia and international efforts to protect dwindling gene pools. *Forest Genetic Resources* **31**, 62–64, FAO Rome.

Pinyopusarerk K, Gunn BV, Williams ER, Pryor LD (1993) Comparative geographical variation in seedling morphology of three closely related red mahoganies, *Eucalyptus urophylla*, *E pellita* and *E scias. Australian Journal of Botany* **41**, 23–34.

Poke FS, Vaillancourt RE, Elliott RC, Reid JB (2003) Sequence variation in two lignin biosynthesis genes, cinnamoyl CoA reductase (*CCR*) and cinnamyl alcohol dehydrogenase 2 (*CAD2*). *Molecular Breeding* **12**, 107–118.

Potts BM, Dungey HS (2004) Interspecific hybridization of *Eucalyptus*: key issues for breeders and geneticists. *New Forests* **27**, 115–138.

Pryor LD, Williams ER, Gunn BV (1995) A morphometric analysis of *Eucalyptus urophylla* and some related taxa with descriptions of two new species. *Australian Systematic Botany* **8**, 57–70.

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. *Current Opinion in Plant Biology* **5**, 94–100.

Schaal BA, Hayworth DA, Olsen KM, Rauscher JT, Smith WA (1998) Phylogeographic studies in plants: problems and prospects. *Molecular Ecology* **7**, 465–474.

Shattuck-Eidens DM, Bell RN, Neuhausen SL, Helentjaris T (1990) DNA sequence variation within maize and melon: observations from polymerase chain reaction amplification and direct sequencing. *Genetics* **126**, 207–217.

Sheely DL, Meagher TR (1996) Genetic diversity in Micronesian island populations of the tropical tree *Campnosperma brevipetiolata* (Anacardiaceae). *American Journal of Botany* **83**, 1571–1579.

Smith DN, Devey ME (1994) Occurrence and inheritance of microsatellites in *Pinus radiata*. *Genome* **37**, 977–983.

Steane DA, Conod N, Jones RC, Vaillancourt RE, Potts BM (2006) A comparative analysis of population structure of a forest tree, *Eucalyptus globulus* (Myrtaceae), using microsatellite markers and quantitative traits. *Tree Genetics and Genomes* **2**, 30–38.

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES (2001) *Dwarf8* polymorphisms associate with variation in flowering time. *Nature Genetics* **28**, 286–289.

Thumma BR, Nolan MF, Evans R, Moran GF (2005) Polymorphisms in *cinnamoyl CoA reductase* (*CCR*) are associated with variation in microfibril angle in *Eucalyptus* spp. *Genetics* **171**, 1257–1265.

Tripiana V, Bourgeois M, Verhaegen D, Vigneron P, Bouvet JM (2007) Combining microsatellites, growth, and adaptive traits for managing in situ genetic resources of *Eucalyptus urophylla*. *Canadian Journal of Forest Research* **37**, 773–785.

Van der Schoot J, Pospíżkova M, Vosman B, Smulders MJM (2000) Development and characterization of microsatellite markers in Black Poplar (*Populus nigra* L.). *Theoretical and Applied Genetics* **101**, 317–322.

Voris HK (2000) Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *Journal of Biogeography* **27**, 1153–1167.

Wencelius F (1983) *Eucalyptus urophylla* en Côte d'Ivoire. *Silvicultura, São Paulo* **31**, 515–518.

Witter MS, Carr GD (1988) Adaptive radiation and genetic differentiation in the Hawaiian silversword alliance (Compositae: Madiinae). *Evolution* **42**, 1278–1287.

Wolfe KH, Wen-Hsiung LI, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Science USA* **84**, 9054–9058.

Wright S (1931) Evolution in Mendelian populations. *Genetics* **28**, 114–138.

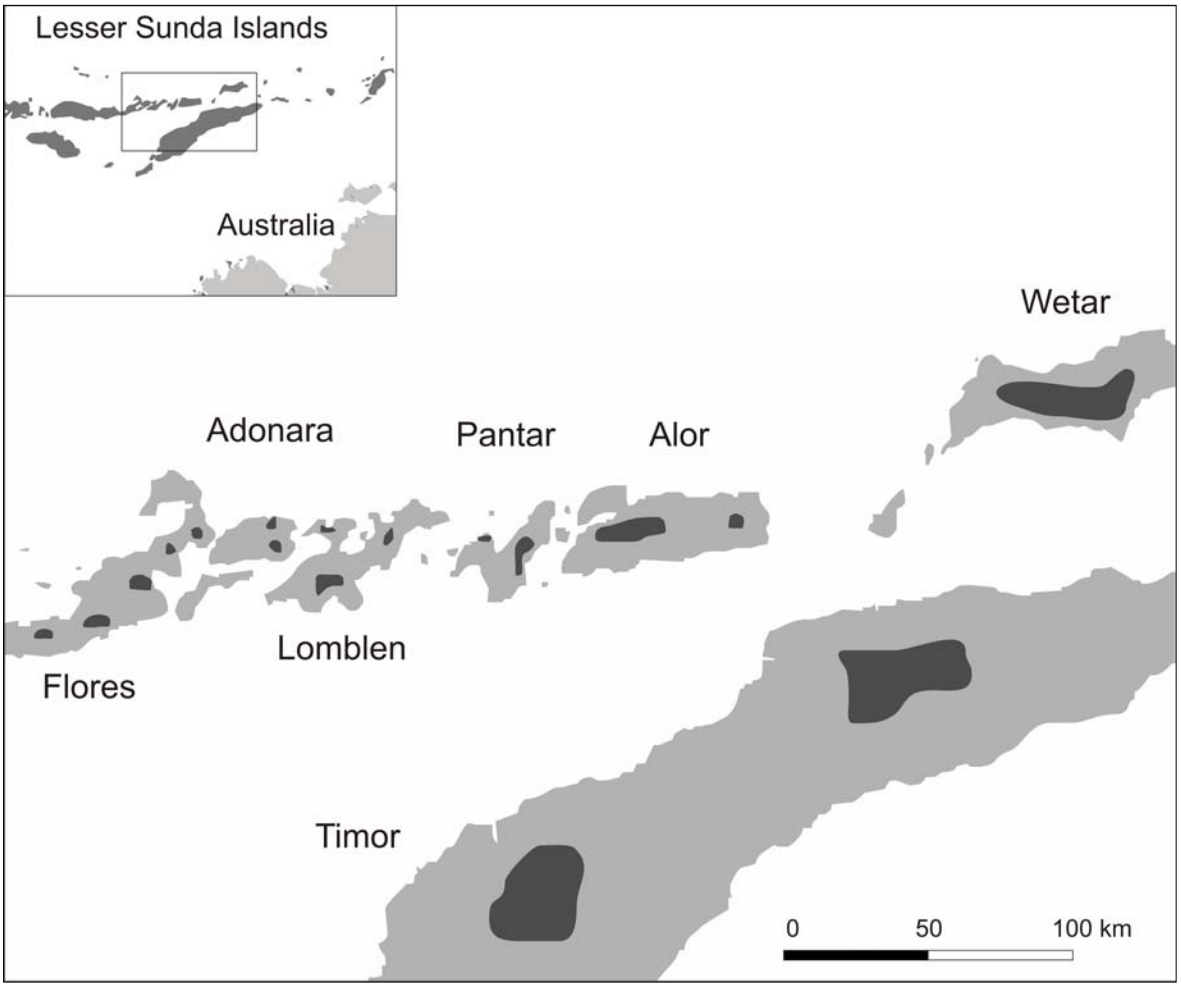


Figure 1. Natural distribution of *Eucalyptus urophylla* on the Lesser Sunda Islands (adapted from Dvorak et al. 2008).

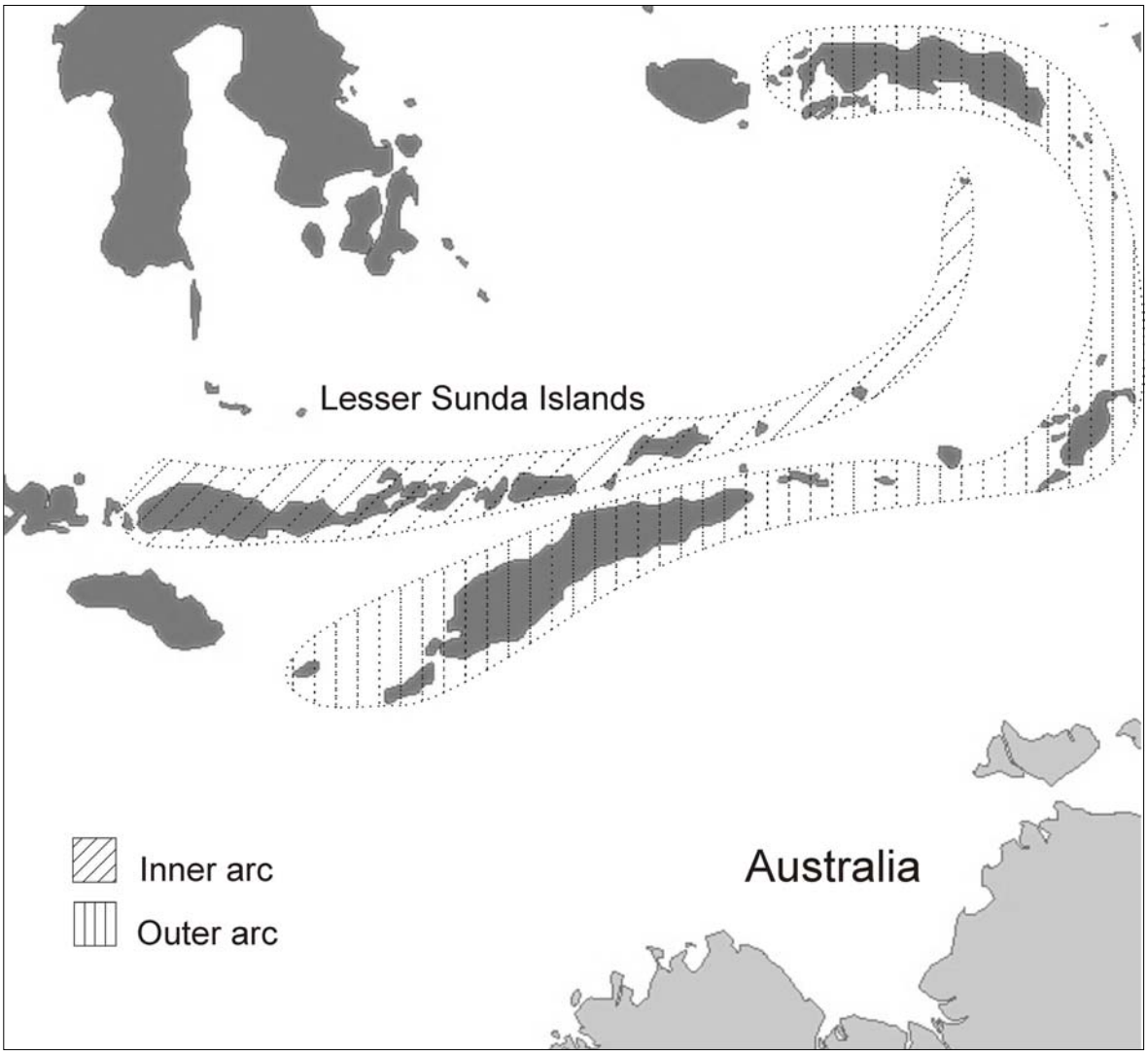


Figure 2. Map of the inner and outer Banda arcs.

CHAPTER 2

Chloroplast DNA phylogeography reveals the island colonisation route of *Eucalyptus urophylla* (Myrtaceae)

Kitt G. Payn[a, b], William S. Dvorak[b] and Alexander A. Myburg[a]

[a]*Forest Molecular Genetics Programme, Department of Genetics, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, 0002, South Africa.*

[b]*Camcore, North Carolina State University, PO Box 8008, Raleigh, NC 27695, USA.*

This chapter consists of a manuscript prepared for publication in the Australian Journal of Botany 55: 673-683, 2007, reproduced with permission from CSIRO Publishing, Melbourne Australia. I conducted all laboratory work, except for technical assistance acknowledged at the end of the paper, and wrote the manuscript. Alexander Myburg supervised the research project, and together with William Dvorak, provided extensive suggestions on the organization and content of the manuscript.

ABSTRACT

We present a study of the colonisation patterns of a tropical tree species among an island archipelago. *Eucalyptus urophylla* (S.T. Blake) is an economically important plantation species endemic to the volcanic slopes of seven islands in eastern Indonesia. In this study we investigated the geographical distribution of chloroplast DNA sequence variation in *E. urophylla* to gain insight into its historical seed migration routes. DNA sequence data were obtained from 198 plants from which 20 haplotypes were identified. A moderate to high level of chloroplast genetic differentiation (G_{ST} = 0.581, N_{ST} = 0.724) and significant phylogeographic structure ($N_{ST} > G_{ST}$; $P < 0.01$) were observed, suggesting low levels of recurrent seed-mediated gene flow among the islands. The highest levels of haplotype diversity were observed on the eastern islands of Wetar and Timor. The two most westerly islands, Flores and Lomblen, were fixed for what appeared to be the ancestral haplotype. Chloroplast haplotype diversity therefore exhibited a decreasing trend from east to west in the species' range, consistent with an east-to-west colonisation route across the seven islands. Environmental factors that may have contributed to the contemporary spatial distribution of chloroplast DNA haplotypes include island paleogeology, ocean currents, fluctuations in sea levels and possible hybridisation events.

INTRODUCTION

Situated at the interface of the Asian and Australian biotic realms are the islands of Indonesia. These islands contain some of the most diverse collections of flora and fauna on

earth and have long been a region of major biogeographical interest (Wallace 1860; Myers et al. 2000; Brown et al. 2004). The division between Asian and Australian biota in Indonesia, first described by Alfred Wallace in the nineteenth century, is now recognised as a biogeographic region of transition, named Wallacea (Fig. 1*A*). Wallacea, encompasses Sulawesi, the Lesser Sunda Islands, and the Moluccas. This area is not only recognized for its rich biodiversity but also exhibits a high level of species endemicity (Myers et al. 2000). Several geological and environmental factors have influenced the contemporary distribution of Indo-Malay and Australasian biota on the islands of Indonesia. They include the continuing northward drift of the Indo-Australian plate into the region (Michaux 1991), the associated volcanic activity on many of the islands, and sea level fluctuations during the Pleistocene (Voris 2000) creating land bridges that facilitate migration among islands.

The genus *Eucalyptus*, comprising more than 600 species, is primarily endemic to the Australian continent (Ladiges et al. 2003). *Eucalyptus urophylla* S.T. Blake is one of only two *Eucalyptus* species that occurs exclusively outside of Australia, the other being *E. deglupta* Blume. Its natural distribution is limited to a series of disjunct populations located on seven of the Lesser Sunda Islands in eastern Indonesia (Fig. 1*C*). It occurs from almost sea level to high volcanic mountain slopes (3000 m elevation) with the largest stands found on the islands of Timor and Wetar, whereas more scattered stands occur on the islands of Adonara, Alor, Flores, Lomblen (Lembata) and Pantar (Eldridge et al. 1993). On the lower slopes of these islands it forms a mosaic distribution pattern with *E. alba* Reinw. ex Blume, which, unlike *E. urophylla*, is also indigenous to northern Australia and Papua (Pryor et al.

1995). In tropical and subtropical regions of Africa, South America and Asia, *E. urophylla* is commercially planted to produce wood that is used for a diverse array of products including pulp, sawn timber and fuel wood. The species is often crossed with *E. grandis* to produce hybrid progeny displaying rapid growth and superior disease resistance compared to the *E. grandis* parent (Pepe et al. 2004).

The Lesser Sunda Islands form part of the Banda arc, which comprises the inner volcanic arc and the outer non-volcanic arc (Norvick 1979). The eastern region of the inner arc includes the islands of Flores through Wetar, whereas the island of Timor forms part of the outer arc. These arcs were formed by the collision and subduction of the Australian plate beneath the Asian plate during the Pliocene. The colonisation and historical migration patterns of *Eucalyptus urophylla* among the islands are unclear. An earlier isozyme study supports the hypothesis that *E. urophylla* existed on Timor before colonising the remaining islands because populations on Timor and nearby Alor contained the highest number alleles observed in the species (House and Bell 1994). The putative initial colonisation of Timor from an Australian source would have been aided by a period of low sea level during a glacial maximum, bringing emergent lands into close proximity (Ladiges et al. 2003). However, it is thought that there has never been a continuous land link between the Sunda Islands and Australia (Hall 2001), which suggests that historical seed migration would have required the successful crossing of ocean water.

Phylogeography is a field of study that analyses the geographical distribution of genealogical lineages, especially those at the intraspecific level (Avise 1998). The resolved phylogeographic structure of *E. urophylla* would provide insight into its evolutionary history and may be used to infer historical migration routes, and previous occurrences of genetic bottlenecks or population expansions. Plant phylogeographic studies predominantly make use of genetic variation in the chloroplast genome. The genome is inherited from a single parent, is effectively haploid and does not recombine (Birky 1995), thereby eliminating analytical complications involving interallelic recombination and heterozygosity. Notably, both the reduced ploidy and the uniparental mode of inheritance of organelle DNA decrease the effective population size, consequently increasing genetic drift and resulting in greater phylogeographic structure (McCauley 1995).

The chloroplast genome appears to be maternally inherited in most angiosperms (Corriveau and Coleman 1988), including *Eucalyptus* (Byrne et al. 1993; McKinnon et al. 2001*a*), thus the distribution of chloroplast DNA (cpDNA) haplotypes may be used to infer seed-mediated migration routes. For example, two studies based on cpDNA variation of *Quercus* spp. supported the existence of refugia located in southern Iberia, Italy and the Balkans and proposed several postglacial migration routes into northern Europe from each refugium (Dumolin-Lapegue et al. 1997; Ferris et al. 1998). Similar studies have been carried out on *Liriodendron tulipifera* in North America (Sewell et al. 1996) and *Cedrela odorata* in Measoamerica (Cavers et al. 2003). More recently, a large-scale project called CYTOFOR (http://www.pierroton.inra.fr/Cytofor), comprising nine research groups representing six

European countries, investigated the patterns of chloroplast genetic diversity and the postglacial recolonisation history of 22 widespread European trees and shrubs (Petit et al. 2003). Together these studies have shown that cpDNA markers are very useful in providing insight into the evolutionary history of a species.

Chloroplast DNA variation has also been extensively examined in *Eucalyptus* (Byrne and Moran 1994; Steane et al. 1998; Jackson et al. 1999; McKinnon et al. 1999). These studies were based on restriction fragment length polymorphism (RFLP) variation in cpDNA. More recently, Vaillancourt and Jackson (2000) found the J_{LA} region (an intergenic spacer on either side of the junction between the large single copy region and inverted repeat A of the chloroplast genome; Goulding et al. 1996) to be hypervariable in *Eucalyptus.* The DNA sequence data were shown to accurately identify haplotypes from divergent *E. globulus* cpDNA lineages previously identified by RFLP analysis (Jackson et al. 1999).

A subsequent study by Freeman *et al*. (2001) expanded the sampling of *E. globulus* to a finer geographic resolution and extended the J_{LA} region in the 3' direction to cover the complete *trnH* gene and the *trnH-psbA* intergenic spacer. Analysis of the extended hypervariable sequence, termed J_{LA}^{+} (Freeman et al. 2001), found the distribution of major haplotype clades to be broadly consistent with the former study in *E. globulus* (Jackson et al. 1999), but allowed for a greater resolution of the phylogenetic relationships between and within haplotype clades. A continental Australian origin of *E. globulus* was supported by the widespread distribution of the basal J_{Cg} haplotypes on continental Australia. There was also

evidence of glacial refugia in the coastal areas of eastern and south-eastern Tasmania with the most recent seed migration of *E. globulus* between Tasmania and continental Australia occurring along a western island migration route during a glacial maximum and accompanying reduced sea level (Freeman et al. 2001).

More recently, the J_{LA} region was used to investigate chloroplast variation and population structure in *E. grandis* (Jones et al. 2006), which predominantly occurs in subtropical eastern Australia, with smaller populations located in the tropical north. According to Jones *et al*. (2006), there was a low level of the chloroplast differentiation among populations (G_{ST} = 0.30) that was possibly due to a relatively recent geographical isolation of the northern populations of *E. grandis*. It was further suggested that the northern populations might have been colonised from the southern populations based on the greater number of haplotypes in the latter populations (Jones et al. 2006).

In contrast to the *Eucalyptus* species endemic to the Australian mainland, volcanic peaks and seawater geographically isolate the distributions of *E. urophylla* populations. Therefore, a high degree of population structure at the chloroplast level is expected. Furthermore, the hypothesis of an initial colonisation of Timor and/or Wetar, followed by a westerly migration route, would be supported by a gradient of decreasing chloroplast genetic diversity along the chain of islands toward Flores. With this view, the aim of this study was to investigate and describe the contemporary phylogeographic structure of *E. urophylla* among the seven islands of the Sunda archipelago. We report the first estimates of chloroplast sequence

variation in the species and infer historical seed migration routes among the islands based on the relationship between haplotypes, as defined by polymorphisms in the hypervariable J_{LA}^{+} region and their present geographical distribution. The results are interpreted in light of the known geological and oceanographic patterns of the region.

MATERIALS AND METHODS

Plant material and DNA isolation

Seed collections were conducted by the research staff of PT Sumalindo Lestari Jaya, a private Indonesian forestry company and Camcore, North Carolina State University, USA, an international tree conservation and domestication programme (Pepe et al. 2004). This series of collections comprised seed from 1104 mother trees distributed across 62 provenances (geographic locations) representing the natural distribution of *E. urophylla*, barring the region of East Timor, which was experiencing political unrest at the time of seed collection. A subset of 51 provenances was included in this study. Seeds were sown in a commercial nursery in South Africa (Mondi, South Africa). Leaf tissue was sampled from 198 plants of known family and provenance origin (Table 1).

Total genomic DNA was extracted from 50 mg of fresh leaf tissue using the DNeasy Plant Mini Kit (QIAGEN, Valencia, California, USA). Samples were homogenized for 30-60 sec in a FastPrep FP120 instrument (QBiogene, Carlsbad, California, USA) set at 4.0 m/s. In order to improve efficiency, cell lysis was performed at 65°C for 30 min. Thereafter, all steps

were performed as described in the DNeasy Plant Mini Kit manual. DNA quality and quantity were determined by agarose gel electrophoresis and spectrophotometry (Nanodrop Technologies, Wilmington, Delaware, USA).

Chloroplast DNA amplification and sequencing

The extended J_{LA}^{+} region used previously by Freeman *et al*. (2001) was further extended in the 5' direction in this study. A forward primer (*euro_rpl2*; GCGTCCTGTAGTAAGAGGAG) was designed to anneal to a conserved region 151 bp upstream of the forward primer *rpl2* previously developed by Goulding *et al*. (1996) and used in Freeman *et al*. (2001). We used this primer together with the reverse primer *eucpsbA* (*eucpsbA*; GGAGCAATAACCAACACTCTTG) developed by Freeman *et al*. (2001). The reverse primer anneals to a conserved region found in eucalypt species 45-66 bp downstream of the stop codon of the *psbA* gene (Freeman et al. 2001). PCR amplification reactions were performed in 20 μL volumes containing 5 ng of genomic DNA, 0.8 U of Exsel polymerase (Southern Cross), 1× PCR Exsel buffer, 0.2 mM dNTPs and 0.4 μM of each primer. PCR amplifications were performed with an iCycler (BIO-RAD Laboratories, Hercules, CA, USA) with the following cycling conditions: an initial denaturation step of 94°C for 1 min, followed by 25 cycles of 94°C for 20 s, 64°C for 30 s and 72°C for 40 s with a 1 s increase per cycle; and a final extension step of 68°C for 10 min. The total PCR product length was approximately 780 bp.

PCR products were cleaned with the QIAquick PCR Product purification kit (QIAGEN) and sequenced in both directions, using primers *euro_rpl2* and *eucpsbA*, with the Big Dye terminator kit (v3.1, Applied Biosystems, Foster City, California, USA) on an ABI 3100 Automated DNA sequencer (Applied Biosystems).

Genetic diversity analysis

Sequence data were assembled and aligned with the software package SeqScape (v2.1, Applied Biosystems). The sequence alignment length was reduced to 576 bp to ensure no missing data across 198 samples. Indel (insertion/deletion) mutations were further removed from the analysis as it was unknown whether the indels were produced by a single mutational event or several events. Consequently, the exclusion of indel mutations provided a more conservative estimate of sequence divergence.

Haplotype diversity (*h*) (Nei 1987) was calculated using DnaSP version 4.0 (Rozas et al. 2003). Two estimates of population differentiation, G_{ST} and $N_{ST,}$ were determined by the HAPSTEP program (version 2001, Pons and Petit 1996). The G_{ST} estimate depends only on the frequencies of the haplotypes, but both haplotype frequencies and the genetic distances between haplotypes influence N_{ST}. Provenances were treated as populations for the differentiation analysis. The HAPSTEP program requires populations with sample sizes smaller than three individuals to be excluded from the analysis. Therefore, differentiation parameter estimates (G_{ST} and N_{ST}) were based on 50 populations ($n < 3$ for Wasbila provenance from the island of Pantar, Table 1).

Hierarchical analysis of molecular variance (AMOVA, Excoffier et al. 1992) was implemented using ARLEQUIN software (version 2.000; Schneider et al. 2000) to apportion variance within provenances, between provenances within islands, and among islands. The significance of variance components was tested by a nonparametric permutation procedure with 1000 permutations.

Haplotype network and nested clade analysis

The program TCS (version 1.2.1, Clement et al. 2000) was used to construct a network of haplotypes using statistical parsimony (Templeton et al. 1992). Closed haplotype loops were removed using the procedures described by Crandall and Templeton (1993). The haplotype network was converted manually into a nested design according to the procedures defined by Templeton *et al*. (1987) and Templeton and Sing (1993).

A nested clade analysis (NCA) was performed using the program GEODIS (version 2.4; Posada et al. 2000) to assess geographical associations of haplotypes and infer historical patterns of colonisation and dispersal. Clades without geographical or genetic variation were not included in the following analyses. The geographical coordinates of each provenance were used to calculate two statistics, the clade distance (D_c), which measures the geographical spread of a clade, and the nested clade distance (D_n), which measures how a clade is geographically distributed relative to other clades in the same higher-level nesting category (Posada et al. 2000). In addition, for each nesting clade, the average differences in

D_c and D_n values between older (interior) and more recent (tip) clades were calculated, abbreviated (I-T)D_c and (I-T)D_n, respectively. To determine whether any of these distance parameters were significantly small or large, all clades within a nesting clade were permuted randomly across localities (1000 times) to generate a null distribution against which the observed values were tested. The output of significant parameters (D_c, D_n, (I-T)D_c, (I-T)D_n) were entered into a program called AUTOINFER (version 1.0; Zhang et al. 2006), which inferred biological events by implementing the algorithm of Templeton (2004).

RESULTS

DNA sequence data were obtained from 198 seedlings, representing seven islands and 51 provenances with a harmonic mean sample size of 26.3 and 3.8 respectively. Each seedling was derived from a unique maternal parent tree because in angiosperms the progeny from the same maternal parent typically share the same chloroplast genome (Corriveau and Coleman 1988).

Haplotype polymorphism and geographic distribution

In this study the forward primer was moved 151 bp in the 5' direction of the J_{LA}^{+} region in an effort to identify additional polymorphic sites. However, all observed nucleotide substitutions were found to occur within the J_{LA}^{+} region defined by Freeman *et al*. (2001). In total, 21 polymorphic sites comprising 18 parsimony informative sites and three singleton sites were detected. Twenty haplotypes (haplotypes I-XX) were identified from the polymorphic sites

(Table 2). Haplotype frequencies ranged from 0.005 to 0.535 (Table 2). Haplotype X was the most prevalent, with 106 observations. Five haplotypes were observed in a single individual each.

The geographic distribution of the haplotypes is listed in Table 1 and illustrated in Fig. 1*C*. Haplotype X was the most geographically widespread haplotype, occurring in provenances on all seven islands. Haplotype XX was observed in provenances on three adjacent islands, namely Timor, Wetar and Alor. Haplotypes V and VI were observed in provenances on Wetar and were also present on the islands of Adonara and Alor respectively. The remaining haplotypes were geographically restricted to single islands.

Haplotype diversity and population differentiation

Haplotype diversity (*h*) across the entire region was 0.689 (Table 3). At the island level, Wetar exhibited the greatest haplotype diversity of 0.878, whereas the populations on the western islands of Flores and Lomblen were fixed for haplotype X.

As defined in the Materials and Methods section, provenances were treated as populations for the population differentiation analysis. A moderate to high proportion of variation resulted from differences among populations, $G_{ST} = 0.581$. The parameter N_{ST} was used to investigate whether related haplotypes were clustered according to geographical location. The N_{ST} estimate of 0.724 was larger than the G_{ST} estimate and the difference was significant ($P < 0.01$). According to Pons and Petit (1996), a higher N_{ST} than G_{ST} usually indicates the

presence of phylogeographic structure with closely related haplotypes being found more often in the same area than less closely related haplotypes.

Hierarchical AMOVA revealed that cpDNA variation within provenances accounted for 25.1% of the total molecular variance (Table 4). A further 25.5% of the total variation was distributed among provenances within islands, whereas 49.4% of the total molecular variance occurred among islands.

Population history inferred from NCA

Chloroplast haplotypes were connected in a single most parsimonious network with 95% probability (Fig. 2). Two closed loops, each a consequence of more than one parsimonious connection of a haplotype to the rest of the network, were resolved following the criteria suggested by Crandal and Templeton (1993). Accordingly, a haplotype connection was maintained to high frequency haplotypes with an interior position in the network rather than to low frequency haplotypes located in tip clades. Secondly, connections between haplotypes occurring in the same geographical area were preferentially maintained.

Ancestral haplotypes are identifiable by their internal position in the network, by the number of lineages that arise from them, and by their commonness (Castelloe and Templeton 1994). Statistical parsimony implemented in the TCS programme identified haplotype X as the ancestral haplotype (Fig. 2). Haplotype X was connected to multiple lower frequency haplotypes, which is consistent with the expectation that older haplotypes have a higher

probability of producing mutational derivatives than do younger haplotypes, thereby becoming interior haplotypes (Crandall and Templeton 1993). Related haplotypes derived from haplotype X were mostly clustered according to geographical location (Fig. 2).

The nested clade analysis was performed manually on the resolved haplotype network according to the algorithm of Templeton *et al*. (1987) (Fig. 2). Haplotype XVII was the only observed haplotype symmetrically stranded and was grouped with the nesting category that had the smallest sample size, in accordance with Templeton and Sing (1993). Stranded haplotypes that were missing intermediates were left unnested (Templeton and Sing 1993). The nesting design resulted in a three-step hierarchy, with a total of 13 clades. Of these, 10 clades (Table 5) contained both geographical and genetic variation and could therefore be tested for geographical association.

The results of the nested clade analysis are presented in Fig. 3. Nested clades that showed significant spatial genetic structuring were used to infer biological events according to the algorithm of Templeton (2004). Restricted gene flow with isolation by distance was the inferred biological process leading to significant geographical-genetic associations for the haplotypes nested in clades 1-1 and 1-9, and for the one-step level clades nested in clade 2-1 (Table 5). The biological processes inferred for the remaining clades and for the total cladogram included: Contiguous range expansion (clade 1-4), and long distance colonisation and/or past fragmentation (clade 1-8 and total cladogram). In this study there was insufficient evidence to discriminate between long-distance colonisation and past fragmentation. It is

feasible that both may have played a role leading to the present day distribution of *E. urophylla* cpDNA haplotypes.

DISCUSSION

In this paper we report the first investigation of cpDNA variation in *E. urophylla,* a tropical forest species that occurs in a series of disjunct populations distributed on seven islands of the Sunda archipelago in eastern Indonesia. Our results allow us to infer possible seed dispersal routes that may explain the observed patterns in chloroplast and nuclear (House and Bell 1994) genetic diversity within *E. urophylla*. A moderate to high level of chloroplast genetic differentiation was found. The observed chloroplast genetic structure supports the hypothesis of an east to west historical seed migration route among the seven islands.

Limited population differentiation was observed for *E. urophylla* while using biparentally inherited nuclear DNA markers (G_{ST} = 0.12; House and Bell 1994), indicating that most of the nuclear genetic diversity in the species is contained within rather than between populations. However, the level of chloroplast differentiation among populations observed in this study was substantially higher (G_{ST} = 0.581, N_{ST} = 0.724) and was close to the average cytoplasmic differentiation for angiosperm species (G_{ST} = 0.64, Petit et al. 2005). There was also significant phylogeographic structure in the chloroplast variation ($N_{ST} > G_{ST}$; $P < 0.01$) with an estimated 49.37% of the total variance explained by differences among islands and 25.51% due to differences among provenances within islands (Table 4). Similar results were

reported for *Santalum austrocaledonicum*, an economically important forest tree species endemic to the New Caledonia and Vanuatu archipelagos, whereby populations were highly differentiated based on chloroplast microsatellite markers (F_{ST} = 0.66; Bottin *et al*. 2007) and the majority of the total variance was explained by differences among islands. The higher level of differentiation observed using chloroplast markers compared to nuclear markers in *E. urophylla* was expected since *Eucalyptus* seeds are mainly dispersed by gravity, where as pollen is typically dispersed by insect or even bird vectors (House 1997). *Eucalyptus urophylla* is predominantly insect pollinated whereas the seeds are mainly dispersed by gravity. The gravitational seed dispersal is relatively limited, particularly where high volcanic mountains and seawater form formidable dispersal barriers at the provenance and island level, respectively.

The islands of Wetar in the east and Timor in the south contained most of the chloroplast genetic diversity of *E. urophylla* (Table 3). Notably there was a high amount of morphological variation observed on these two islands that resulted in a proposed separation of two new species from *E. urophylla* sensu lato, namely *E. wetarensis* and *E. orophila* on the islands of Wetar and Timor, respectively (Pryor et al. 1995). However, a subsequent isozyme study did not fully support the proposal, although there was a large degree of allelic diversity on the two islands (House and Bell 1994). Numerous studies of forest trees have described the trend of comparatively high cpDNA diversity in glacial refugia and less diversity in regions colonised more recently following deglaciation (Demesure et al. 1996; Petit et al. 1997; Ferris et al. 1998; King and Ferris 1998; Marchelli et al. 1998). Our data are

therefore consistent with the original colonisation by *E. urophylla* of its present natural range occurring on the eastern and/or southern islands, followed by a more recent colonisation of the western region.

The hypothesis of a historical east to west migration pattern is further supported by the haplotype network (Fig. 2). Both Wetar and Timor had clades of haplotypes that were private to each island and exhibited considerable divergence from the ancestral haplotype X. This suggests that *E. urophylla* was present on both islands for a relatively long time before colonizing the other islands to the west. The island of Alor appeared most similar to Wetar and Timor in terms of shared and related haplotypes but the haplotypes were not highly diverged, suggesting a shorter period of occupation on Alor. The observed haplotype distribution on Pantar was quite different from its neighbouring islands in that it had a high frequency of private haplotypes I and IX. Notably, haplotype I was closely related to haplotye II, which was observed on the neighbouring island of Alor (Fig. 2). One suggestion is that Pantar was colonised by individuals with haplotype II from Alor, which subsequently gave rise to haplotype I. Other colonisation events may have included individuals with haplotype X from which haplotype IX was likely derived. Flores, Lomblen and Adonara appeared to be the most recently colonized islands. A likely source would have been from the island of Alor, which had a high frequency of haplotype X. The absence of observed cpDNA variation for the islands of Flores and Lomblen (Table 3) is the signature of a recent founder event. Several environmental factors may have contributed to the historical seed migration patterns that lead to the contemporary distribution of cpDNA variation. These include,

amongst others, island paleogeology and subsequent proximity to the mainland, ocean currents, fluctuations in sea levels and possible hybridisation events.

The Lesser Sunda Islands form part of the Banda arc that represents the convergence zone between the still northward drifting Australian continental margin and the inner arc. The inner arc, which started to appear approximately 12 million years ago (Audley-Charles 2004), was built up before the collision with the age of inception decreasing eastward (van der Werff 1995). The outer arc arose in the front part of the convergence zone where low-density sedimentary rocks were uplifted by their buoyancy. The emergence of Timor island occurred after the arc-continent collision at approximately 3.5 to 2 million years ago (Audely-Charles 2004). On the basis of the proposed order of geological events, one might assume that *E. urophylla* may have colonised several of the inner arc islands, possibly starting with the older islands in the west, before the emergence and subsequent colonisation of the outer arc island of Timor. However, Ladiges *et al*. (2003) proposed that *E. urophylla* diverged from Australian taxa in the subgenus *Symphyomyrtus* relatively recently, approximately 5 to 2 million years ago during the compression of Timor between the inner Banda arc and the north-west region of the Australian continental crust. The putative initial colonisation of Timor from an Australian or New Guinea source would have been assisted by a period of lower sea levels occurring during the Quaternary, bringing emergent lands closer together. The greater number of nuclear DNA alleles on Timor and nearby Alor (House and Bell 1994), together with the greater chloroplast haplotype diversity on the islands of Wetar,

Timor and Alor (Table 3), further support the hypothesis of an initial southern or eastern colonisation followed by a more recent east to west colonisation process.

Sea-surface currents are another environmental factor that may be important for predicting ecological and genetic connections among island populations. The currents proximal to the Lesser Sunda Islands primarily comprise North Pacific water flowing from the Makassar Strait into the Flores and Banda Seas before curling southwards into the Timor Sea and Indian Ocean (Fig. 1*B*; Gordon and Fine 1996). In addition, there are currents that are channelled into the Ombai Strait between Alor and Timor islands and are generally directed toward the Savu (Sawu) Sea, but there is an occasional reversal of flow in a northeast direction entering the Savu Sea from the Indian Ocean (Molcard et al. 2001). The highly structured distribution of cpDNA at the island level suggests that seed migration across bodies of water is an uncommon occurrence (Fig. 1*C*). However, if plant material drifted among the islands at the mercy of prevailing currents it could be assumed that long distance colonisation (Table 5) would more likely occur in a westerly direction.

During the Pleistocene, glaciation and deglaciation led to fluctuating sea levels that greatly affected landmass configurations in Southeast Asia (Voris 2000). At the peak of the glacial maxima, sea levels were at a minimum and many of the present islands, currently separated by shallow seas, merged to form composite islands. Of the seven islands upon which *E. urophylla* naturally occurs, Flores, Adonara and Lomblen were connected when sea levels were 60-120 m below the present level, while the other four islands remained separated

(Heaney 1991; Voris 2000). According to How *et al*. (1996), a land bridge between Flores and Lomblen during glacial maxima was considered to be a major factor explaining why populations of several species of snake on Flores and Lomblen were more similar to one another than they were to conspecific populations on adjacent islands to the west (Lombok, Sumba) and east (Alor). Furthermore, the most pronounced morphological differentiation occurred among snake populations existing on different islands that remained separate throughout the Pleistocene. Notably, all of the *E. urophylla* samples obtained from the islands of Flores and Lomblen, and the majority of samples from Adonara, were fixed for haplotype X (Fig. 1*C*). These data suggest a recent colonisation of the western region followed by a founder effect. The likelihood of haplotype X being fixed in samples from both Flores and Lomblen would have been increased if long distance seed colonisation occurred during a relatively recent era when these western islands were joined.

Chloroplast DNA variation in *Eucalyptus* generally appears to be geographically structured, but does not always conform to species boundaries as a result of hybridisation (Steane et al. 1998; Jackson et al. 1999; McKinnon et al. 1999, 2001*b*). For example, intraspecific cpDNA polymorphism in 14 of 17 species sampled in Tasmania was coupled with extensive sharing of identical haplotypes across populations of different species in the same geographic area (McKinnon et al. 2001*b*). They concluded that sharing of cpDNA haplotypes among Tasmanian species of *Eucalyptus* subgenus *Symphyomyrtus* is the rule rather than the exception. *Eucalyptus urophylla*, which occupies a wide altitudinal range on volcanic slopes (180–3000 m, Pepe et al. 2004), forms a mosaic distribution pattern with *E. alba* on low

elevation sites. Here, natural *E. urophylla* x *E. alba* hybrids do exist but they are considered rare as mature trees (Martin and Cossalter 1976).

Putative *E. urophylla* x *E. alba* hybrids have been observed in both first generation *E. urophylla* and *E. alba* provenance trials established in South Africa, suggesting that hybridisation is bi-directional (Payn KG, unpubl.data). We obtained three *E. alba* samples from each of the islands of Flores, Wetar and Timor, and a single sample from New Guinea. All the samples from Flores, Wetar and Timor, possessed haplotype X (Payn KG, unpubl.data), the putative ancestral haplotype of *E. urophylla*. The sample from New Guinea had a highly related haplotype, with only two additional substitutions. These findings suggest that haplotype sharing does occur between *E. urophylla* and *E. alba*. Hence it raises the question whether natural hybridisation events at lower elevation have influenced the distribution of chloroplast haplotypes observed in *E. urophylla*, particularly haplotype X (Fig. 1*C*). It is however important to note that we presently do not have enough information on the cpDNA haplotype diversity within *E. alba* to determine whether haplotype X is ancestral to both species, which also could explain the high prevalence of this haplotype in the small number of *E. alba* samples that we have analysed.

Conclusions

This study demonstrates the capacity of cpDNA variation to reveal the phylogeographic history of island dispersed plant species such as *E. urophylla* and to draw inferences regarding past migratory routes and possible interactions with other species. The

geographical distribution of chloroplast haplotype diversity suggests an east to west colonisation pattern. Timor was likely the first island to be colonised based on its high haplotype diversity and proximity to Australia or New Guinea. The haplotype diversity observed on the islands of Wetar and Alor suggests that they too could be islands of early colonisation, whereas the lack of chloroplast haplotype diversity on the islands of Flores and Lomblen suggest a more recent colonisation event. Restricted gene flow with isolation by distance and long distance colonisation events, possibly assisted by sea currents, are considered largely responsible for the spatial distribution of cpDNA haplotypes within extant populations of the species.

Pollen flow among provenances and even among islands is likely to be largely responsible for the low estimate of population differentiation using nuclear markers (House and Bell 1994). However, a gradient of decreasing nuclear genetic diversity from east to west was also observed with the exception being the populations on the island of Flores. On the basis of our chloroplast data, we propose that the high nuclear genetic diversity reported for Flores may be a result of hybridisation with *E. alba.* This hypothesis is supported by the observation that provenances from Flores appear to have a higher frequency of putative hybrids in first generation *E. urophylla* provenance trials established in South Africa (Payn KG, unpubl.data).

Proficient management of this valuable genetic resource, with respect to conservation and breeding strategies, will benefit from the knowledge of the nature and distribution of the

chloroplast and nuclear genetic variation across the native range of *E. urophylla*. In addition, an understanding of the spatial distribution of cpDNA variability in *E. urophylla* may be used for practical applications such as seed-source certification and the determination of geographic origin of unknown samples.

ACKNOWLEDGEMENTS

The authors thank G.E. McKinnon for valuable comments on the manuscript and D. Posada for the review of the nested clade design. We are grateful to PT Sumalindo Lestari Jaya for the seed collections, Camcore for locality and reference information, Mondi for maintenance of the plant materials, Sappi Forestry South Africa for providing the *E. alba* plant material, and F. Maleka for technical assistance. Financial support for this work was provided by Mondi through the Wood and Fibre Molecular Genetics Programme and by Camcore, Raleigh, NC, USA. Additional funding was provided by the Technology and Human Resources for Industry Programme (THRIP) in South Africa.

LITERATURE CITED

Audely-Charles MG (2004) Ocean trench blocked and obliterated by Banda forearc collision with Australian proximal continental slope. *Tectonophysics* **389**, 65–79.

Avise J (1998) The history and purview of phylogeography: a personal reflection. *Molecular Ecology* **7**, 371–379.

Birky CW (1995) Uniparental inheritance of mitochondrial and chloroplast genes: Mechanisms and evolution. *Proceedings of the National Academy of Sciences USA* **92**, 11331–11338.

Bottin L, Tassin J, Nasi R, Bouvet JM (2007) Molecular, quantitative and abiotic variables for the delineation of evolutionary significant units: case of sandalwood (*Santalum austrocaledonicum* Vieillard) in New Caledonia. *Conservation Genetics* **8**, 99–109.

Brown P, Sutikna T, Morwood MJ, Soejono RP, Jatmiko, Saptomo EW, Due RA (2004) A new small-bodied hominin from the late Pleistocene of Flores, Indonesia. *Nature* **431**, 1055–1061.

Byrne M, Moran GF, Tibbits WN (1993) Restriction map and maternal inheritance of chloroplast DNA in *Eucalyptus nitens*. *Journal of Heredity* **84**, 218–220.

Byrne M, Moran GF (1994) Population divergence in the chloroplast genome of *Eucalyptus nitens*. *Heredity* **73**, 18–28.

Castelloe J, Templeton AR (1994) Root probabilities for intraspecific gene trees under neutral coalescent theory. *Molecular Phylogenetics and Evolution* **3**, 102–113.

Cavers S, Navarro C, Lowe AJ (2003) Chloroplast DNA phylogeography reveals colonization history of Neotropical tree, *Cedrela odorata* L., in Mesoamerica. *Molecular Ecology* **12**, 1451–1460.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9**, 1657–1659.

Corriveau JL, Coleman AW (1988) Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *American Journal of Botany* **75**, 1443–1458.

Crandall KA, Templeton AR (1993) Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics* **134**, 959–969.

Demesure B, Comps B, Petit RJ (1996) Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica* L.) in Europe. *Evolution* **50**, 2515–2520.

Donnelly P, Tavaré S (1986) The ages of alleles and a coalescent. *Advances in Applied Probability* **18**, 1–19.

Dumolin-Lapegue S, Demesure B, Fineschi S, Le Corre V, Petit RJ (1997) Phylogeographic structure of White Oaks throughout the European continent. *Genetics* **146**, 1475–1487.

Eldridge K, Davidson J, Harwood C, Van Wyk G (1993) 'Eucalypt domestication and breeding.' (Oxford University Press: Oxford)

Excoffier L, Smouse P, 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.

Ferris C, King RA, Vainola R, Hewitt GM (1998) Chloroplast DNA recognizes three refugial sources of European oaks and suggests independent eastern and western immigrations to Finland. *Heredity* **80**, 584–594 .

Freeman JS, Jackson HD, Steane DA, McKinnon GE, Dutkowski GW, Potts BM, Vaillancourt RE (2001) Chloroplast phylogeography of *Eucalyptus globulus*. Australian *Journal of Botany* **49**, 585–596.

Gordon AL, Fine RA (1996) Pathways of water between the Pacific and Indian oceans in the Indonesian seas. *Nature*, **379**, 146–149.

Goulding SE, Olmstead RG, Morden CW, Wolfe KH (1996) Ebb and flow of the chloroplast inverted repeat. *Molecular and General Genetics* **252**, 195–206.

Hall R (2001) Cenozoic reconstruction of SE Asia and the SW Pacific: changing patterns of land and sea. In 'Faunal and floral migrations and evolution in SE Asia-Australasia'. (Eds I Metcalfe, JMB Smith, M M orwood, ID Davidson) pp. 35–56. (A.A. Balkema: Lisse)

Hare MP (2001) Prospects for nuclear gene phylogeography. *Trends in Ecology and Evolution* **16**, 700–706.

Heaney LR (1991) A synopsis of climatic and vegetational change in Southeast Asia. *Climate Change* **19**, 53–61.

House APN, Bell JC (1994) Isozyme variation and mating system in *Eucalyptus urophylla* S.T. Blake. *Silvae Genetica* **43**, 167–176.

House SM (1997) Reproductive biology of eucalypts. In 'Eucalypt Ecology: Individuals to Ecosystems' (Eds J Williams, J Woinarski) pp. 30–55. (Cambridge University Press: Cambridge)

How RA, Schmitt LH, Suyanto A (1996) Geographical variation in the morphology of four snake species from the Lesser Sunda Islands, eastern Indonesia. *Biological Journal of the Linnean Society* **59**, 439–456.

Jackson HD, Steane DA, Potts BM, Vaillancourt RE (1999) Chloroplast DNA evidence for reticulate evolution in *Eucalyptus* (Myrtaceae). *Molecular Ecology* **8**, 739–751.

Jones ME, Shepherd M, Henry RJ, Delves A (2006) Chloroplast DNA variation and population structure in the widespread forest tree, *Eucalyptus grandis*. *Conservation Genetics* **7**, 691–703.

King RA, Ferris C (1998) Chloroplast DNA phylogeography of *Alnus glutinosa* (L.) Gaertn. *Molecular Ecology* **7**, 1151–1161.

Ladiges PY, Udovicic F, Nelson G (2003) Australian biogeographical connections and the phylogeny of large genera in the plant family Myrtaceae. *Journal of Biogeography* **30**, 989–998.

Marchelli P, Gallo L, Scholz F, Ziegenhagen B (1998) Chloroplast DNA markers reveal a geographical divide across Argentinean southern beech *Nothofagus nervosa* (Phil.) Dim. et Mil. distribution area. *Theoretical and Applied Genetics* **97**, 642–646.

Martin B, Cossalter C (1976) Les eucalyptus des iles de la Sonde. *Bois et Foréts des Tropiques* **164**, 3–14.

McCauley DE (1995) The use of chloroplast DNA polymorphism in studies of gene flow in plants. *Trends in Ecology and Evolution* **10**, 198–202.

McKinnon GE, Steane DA, Potts BM, Vaillancourt RE (1999) Incongruence between chloroplast and species phylogenies in *Eucalyptus* subgenus *Monocalyptus* (Myrtaceae). *American Journal of Botany* **86**, 1038–1046.

McKinnon GE, Vaillancourt RE, Tilyard PA, Potts BM (2001*a*) Maternal inheritance of the chloroplast genome in *Eucalyptus globulus* and interspecific hybrids. *Genome* **44**, 831–835.

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

Michaux B (1991) Distributional patterns and tectonic development in Indonesia. *Australian Systematic Botany* **4**, 25–36.

Molcard R, Fieux M, Syamsudin F (2001) The throughflow within Ombai Strait. *Deep-Sea Research I* **48**, 1237–1253.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858.

Nei M (1987) 'Molecular evolutionary genetics.' (Columbia University Press: New York)

Norvick MS (1979) The tectonic history of the Banda Arcs, eastern Indonesia: a review. *Journal of the Geological Society London* **136**, 519–527.

Pepe B, Surata K, Suhartono F, Sipayung M, Purwanto A, Dvorak WS (2004) Conservation status of natural populations of *Eucalyptus urophylla* in Indonesia and international efforts to protect dwindling gene pools. *Forest Genetic Resources* No. 31, FAO Rome, 62–64.

Petit RJ, Pineau E, Demesure B, Bacilierir R, Ducousso A, Kremer A (1997) Chloroplast DNA footprints of postglacial recolonization by oaks. *Proceedings of the National Academy of Sciences USA* **94**, 9996–10001.

Petit RJ, Aguinagalde I, de Beaulieu JL, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M, Mohanty A, Műller-Starck G, Demesure-Musch B, Palmé A, Martin JP, Rendell S, Vendramin GG (2003) Glacial refugia: Hotspots but not melting pots of genetic diversity. *Science* **300**, 1563–1565.

Petit RJ, Duminil J, Fineschi S, Hampe A, Salvini D, Vendramin GG (2005) Comparative organisation of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology* **14**, 689–701.

Pons O, Petit RJ (1996) Measuring and testing genetic differentiation with ordered vs. unodered alleles. *Genetics* **144**, 1237–1245.

Posada D, Crandall KA, Templeton AR (2000) GEODIS: a program for cladistic nested analysis of the geographical distribution of genetic haplotypes. *Molecular Ecology* **9**, 487–488.

Pryor LD, Williams ER, Gunn BV (1995) A morphometric analysis of *Eucalyptus urophylla* and some related taxa with descriptions of two new species. *Australian Systematic Botany* **8**, 57–70.

Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496–2497.

Schneider S, Roessli D, Excoffier L (2000) Arlequin: A software for population genetic data analysis Ver 2.000. Genetics and Biometry Lab, Department of Anthropology, University of Geneva.

Sewell MM, Parks CR, Chase MW (1996) Intraspecific chloroplast DNA variation and biogeography of North American *Liriodendron* L. (Magnoliaceae). *Evolution* **50**, 1147–1154.

Steane DA, Byrne M, Vaillancourt RE, Potts BM (1998) Chloroplast DNA polymorphism signals complex interspecific interactions in *Eucalyptus* (Myrtaceae). *Australian Systematic Botany* **11**, 25–40.

Templeton AR (2004) Statistical phylogeography: methods of evaluating and minimizing inference errors. *Molecular Ecology* **13**, 789–809.

Templeton AR, Sing CF (1993) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analysis with cladogram uncertainty and recombination. *Genetics* **134**, 659–669.

Templeton AR, Boerwinkle E, Sing CF (1987) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in *Drosophila*. *Genetics* **117**, 343–352.

Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* **132**, 619–633.

Vaillancout RE, Jackson HD (2000) A chloroplast DNA hypervariable region in eucalypts. *Theory of Applied Genetics* **101**, 473–477.

van der Werff W (1995) Cenozoic evolution of the Savu Basin, Indonesia: forearc basin response to arc-continent collision. *Marine and Petroleum Geology* **12**, 247–262.

Voris HK (2000) Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *Journal of Biogeography* **27**, 1153–1167.

Wallace AR (1860) On the zoological geography of the Malay archipelago. *Journal of the Linnean Society of London* **4**, 172–184.

Zhang AB, Tan S, Sota T (2006) AUTOINFER 1.0: a computer program to infer biogeographical events automatically. *Molecular Ecology Notes*, **6**, 597-599.

Table 1. Island, provenance, sample size, location and altitude of sampled individuals, and observed chloroplast DNA haplotypes for *Eucalyptus urophylla.* Sequence polymorphisms of haplotypes I-XX are listed in Table 2.

Island	Provenance	Sample size	Location	Altitude (m)	Haplotypes (no. of individuals)
Flores		32			
	Hokeng	4	08°31'S, 122°47'E	575	X (4)
	Ile Meak	4	08°37'S, 122°15'E	680	X (4)
	Ile Nggele	4	08°39'S, 122°27'E	685	X (4)
	Kilawair	4	08°41'S, 122°29'E	378	X (4)
	Kolibuluk	4	08°28'S, 122°42'E	648	X (4)
	Lere-Baukrenget	4	08°39'S, 122°23'E	725	X (4)
	Natakoli	4	08°37'S, 122°24'E	900	X (4)
	Palueh	4	08°40'S, 122°35'E	570	X (4)
Adonara		26			
	Doken	4	08°21'S, 123°18'E	800	VIII (1), X (3)
	Gonehama	4	08°20'S, 123°16'E	687	X (4)
	Kawela	4	08°21'S, 123°03'E	600	X (4)
	Lamahela	3	08°21'S, 123°15'E	856	VII (1), X (2)
	Lamalota	4	08°16'S, 123°18'E	735	X (4)
	Muda	4	08°21'S, 123°16'E	750	X (4)
	Watololong	3	08°19'S, 123°15'E	630	V (1), X (2)
Lomblen		28			
	Bunga Muda	4	08°16'S, 123°32'E	650	X (4)
	Ile Ape	4	08°29'S, 123°30'E	860	X (4)
	Ile Kerbau	4	08°29'S, 123°29'E	740	X (4)
	Jontona	4	08°16'S, 123°25'E	788	X (4)
	Labalekan	4	08°32'S, 123°30'E	770	X (4)
	Padekluwa	4	08°30'S, 123°26'E	800	X (4)
	Puor	4	08°34'S, 123°24'E	940	X (4)
Pantar		17			
	Beangonong	3	08°20'S, 124°12'E	565	IX (2), X (1)
	Delaki	4	08°28'S, 124°11'E	810	I (4)
	Lalapang	4	08°20'S, 124°12'E	575	I (3), IX (1)
	Mauta	4	08°26'S, 124°10'E	620	I (4)
	Wasbila	2	08°20'S, 124°03'E	380	IX (2)
Alor		23			
	Apui	4	08°16'S, 124°44'E	1200	II (1), X (3)
	Mainang	4	08°14'S, 124°39'E	1175	X (3), XX (1)
	Manabai	3	08°14'S, 124°45'E	400	VI (3)
	Molpui	4	08°15'S, 124°44'E	400	II (1), IV (3)
	Pintu Mas	4	08°17'S, 124°33'E	385	X (4)
	Watakika	4	08°18'S, 124°30'E	475	X (4)
Wetar		28			
	Alasannaru	4	07°51'S, 126°23'E	596	VI (1), XI (2), XII (1)
	Elun Kripas	4	07°51'S, 126°16'E	733	XII (3), XIV (1)
	Nakana Ulam	4	07°51'S, 126°21'E	715	XIII (3), XIV (1)
	Nesunhuhun	4	07°52'S, 126°15'E	621	XIII (3), XIV (1)
	Puaanan	4	07°51'S, 126°26'E	485	III (1), X (2), XX (1)
	Remamea	4	07°52'S, 126°26'E	476	V (1), VI (2), X (1)
	Talianan	4	07°52'S, 126°28'E	521	VI (3), X (1)

Table 1. Continued

Island	Provenance	Sample size	Location	Altitude (m)	Haplotypes (no. of individuals)
Timor		44			
	A. Esrael	4	09°36'S, 124°14'E	1655	X (1), XX (3)
	Bonleu	4	09°33'S, 124°04'E	1700	XV (1), XVI (1), XX (2)
	Fatumnase	4	09°34'S, 124°13'E	1850	XVII (2), XIX (1), XX (1)
	Lelobatan	4	09°43'S, 124°10'E	1525	XX (4)
	Lelobatang	4	09°41'S, 124°14'E	1300	XVII (1), XX (3)
	Leloboko	4	09°37'S, 124°10'E	1500	X (2), XX (2)
	Mollo	4	09°41'S, 124°11'E	1400	XX (4)
	Naususu	4	09°38'S, 124°13'E	1325	XX (4)
	Nuafin	4	09°31'S, 124°11'E	1900	XVIII (3), XX (1)
	Tune	4	09°33'S, 124°19'E	1250	XVII (2), XIX (1), XX (1)
	Tutem	4	09°35'S, 124°17'E	1300	X (1), XVII (2), XIX (1)
Total		198			

Table 2. Summary of informative polymorphic sites in the J_{LA}^{+} region of the cpDNA of *Eucalyptus urophylla*. All sequences are compared to the reference sequence (haplotype I). Positions of the polymorphic sites are relative to the start of the aligned DNA sequences (GenBank accession EF507880-EF507899), before indels were removed.

Hap ID	Count	Position:	70	101	111	119	133	154	181	261	262	366	373	423	446	454	467	472	530	537	558	568	575
Hap I	11		G	A	A	A	T	C	G	C	C	T	T	G	T	C	T	A	A	A	G	C	C
Hap II	2		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
Hap III	1		A	C	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	A	.	.
Hap IV	3		.	C	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	A	.	.
Hap V	2		.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
Hap VI	9		.	.	.	T	.	.	.	.	.	.	.	.	.	G	.	.	.	C	.	.	.
Hap VII	1		.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	T	.
Hap VIII	1		.	T	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
Hap IX	5		.	.	.	T	.	.	.	G	G	.	.	.	.	.	.	.	.	C	.	.	.
Hap X	106		.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
Hap XI	2		.	.	T	T	.	.	A	.	.	G	.	.	G	.	.	T	.	C	.	.	.
Hap XII	4		.	.	T	T	.	.	A	.	.	.	.	.	G	.	.	T	.	C	.	.	.
Hap XIII	6		.	.	T	T	.	.	.	G	G	.	.	.	G	.	.	T	.	C	.	.	.
Hap XIV	3		.	.	T	T	.	.	.	.	.	.	.	.	G	.	.	T	.	C	.	.	.
Hap XV	1		.	.	.	T	G	.	.	.	.	.	G	A	.	.	.	.	C	C	.	.	T
Hap XVI	1		.	.	.	T	G	.	.	.	.	.	.	A	.	.	.	.	C	C	.	.	T
Hap XVII	7		.	.	.	T	G	.	.	.	.	.	.	.	.	.	.	.	C	C	.	.	T
Hap XVIII	3		.	.	.	T	G	.	.	.	.	.	.	.	.	.	A	.	C	C	.	.	.
Hap XIX	3		.	.	.	T	G	.	.	.	.	G	.	.	.	.	.	.	C	C	.	.	.
Hap XX	27		.	.	.	T	G	.	.	.	.	.	.	.	.	.	.	.	C	C	.	.	.

Table 3. The number of individuals investigated ($N_{ind.}$), number of haplotypes detected ($N_{hap.}$) and estimates of haplotype diversity (h) for *Eucalyptus urophylla* on each of the seven islands.

Region	$N_{ind.}$	$N_{hap.}$	h (SD)
Flores	32	1	0.000 (0.000)
Adonara	26	4	0.222 (0.106)
Lomblen	28	1	0.000 (0.000)
Pantar	17	3	0.522 (0.101)
Alor	23	5	0.613 (0.104)
Wetar	28	9	0.878 (0.029)
Timor	44	7	0.648 (0.071)
Total	198	20	0.689 (0.034)

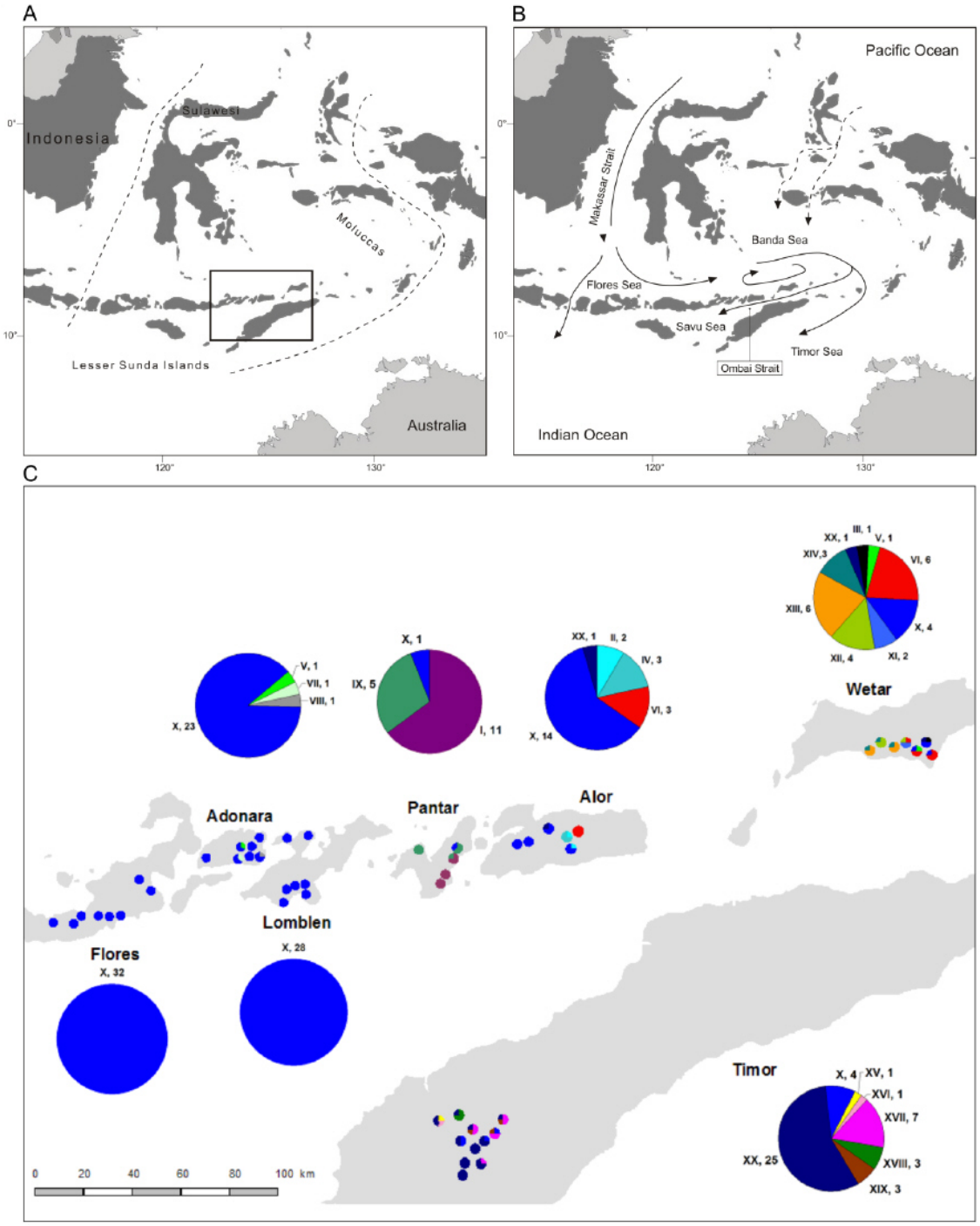
Table 4. Analysis of molecular variance (AMOVA) for provenances of *Eucalyptus urophylla* based on cpDNA sequences. All variance components were significant at the 0.01 level.

Source of variation	d.f.	Sum of Squares	Variance components	Percentage of variation
Among islands	6	98.096	0.53698	49.37
Among provenances within islands	44	59.409	0.27748	25.51
Within provenances	147	40.167	0.27324	25.12

Table 5. Interpretation of the results presented in Fig. 3 according to the key of biological inference of Templeton (2004).

Clade	Chain of inference	Demographic event inferred
1-1	1-2-3-4-NO	Restricted gene flow with isolation by distance.
1-2	No significant clade distances	
1-4	1-2-11-12-NO	Contiguous range expansion
1-6	No significant clade distances	
1-8	1-2-3-5-6-13-14-NO	Long distance colonisation and/or past fragmentation.
1-9	1-2-3-4-NO	Restricted gene flow with isolation by distance.
2-1	1-2-3-4-NO	Restricted gene flow with isolation by distance.
2-2	No significant clade distances	
2-3	No significant clade distances	
Total cladogram	1-2-3-5-6-13-14-NO	Long distance colonisation and/or past fragmentation.

Figure 1. Maps showing Indonesian islands, ocean current pathways, *Eucalyptus urophylla* collection sites and geographical distribution of cpDNA haplotypes. (*A*) The natural range of *E. urophylla* is indicated by a rectangle. Wallacea is the region between the dotted lines. (*B*) Schematic representation of ocean current pathways between the Pacific and Indian oceans in the Indonesian seas (adapted from Gordon and Rana 1996, and Molcard et al. 2001). Solid lines highlight dominant currents, dashed lines highlight minor currents. (*C*) Location of provenances sampled across the natural range of *Eucalyptus urophylla* and distribution of cpDNA haplotypes at the island and provenance level. Haplotype frequencies at the island level are indicated by large pie charts. Haplotypes are colour coded and labelled in Roman Numerals followed by the number of observations in each island. Haplotype frequencies at the provenance level and provenance locations are indicated by small pie charts.


A
Indonesia
Sulawesi
Moluccas
Lesser Sunda Islands
Australia
0°
10°
120°
130°
B
Pacific Ocean
Makassar Strait
Banda Sea
Flores Sea
Savu Sea
Timor Sea
Ombai Strait
Indian Ocean
C
Wetar
Adonara
Pantar
Alor
Lomblen
Flores
Timor
0 20 40 60 80 100 km

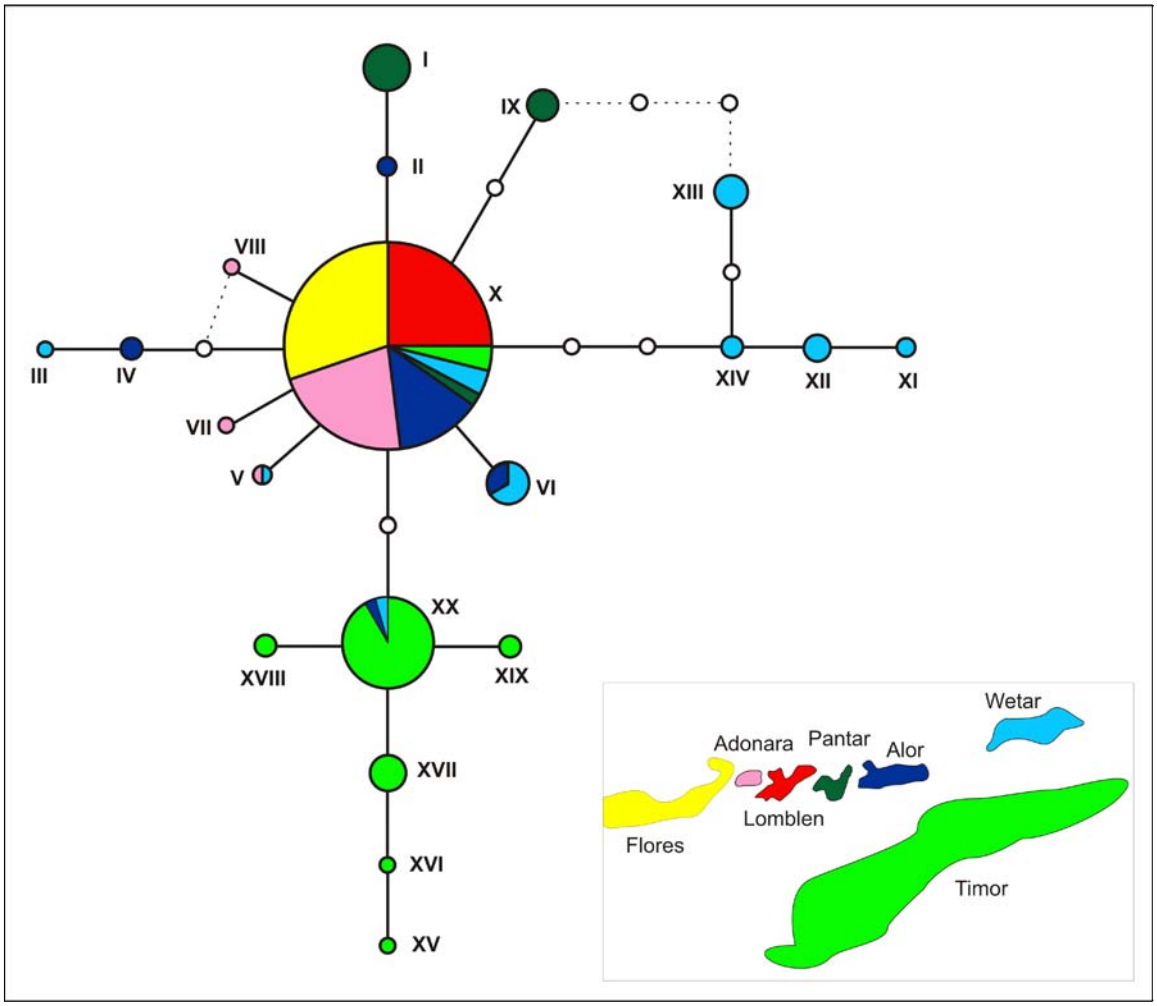


Figure 2. The resolved cpDNA haplotype network and resulting set of nested clades for *Eucalyptus urophylla*. Observed haplotypes are identified by Roman Numerals (I – XX). The size of each circle is approximately proportional to the haplotype frequency, and is colour coded according to sample locality at the island level (Table 1). Small empty circles indicate intermediate haplotypes that were not observed in the data set. Each solid line represents a single mutational change that interconnects two haplotypes. Thick-lined boxes and polygons enclose one-step clades, which are designed by "1-x" where x is a number assigned to identify the clade; and thin-lined boxes enclose two-step clades ("2-x"). Two initially closed loops indicated by dotted lines were resolved using the frequency and geography criteria (Crandall and Templeton 1993).

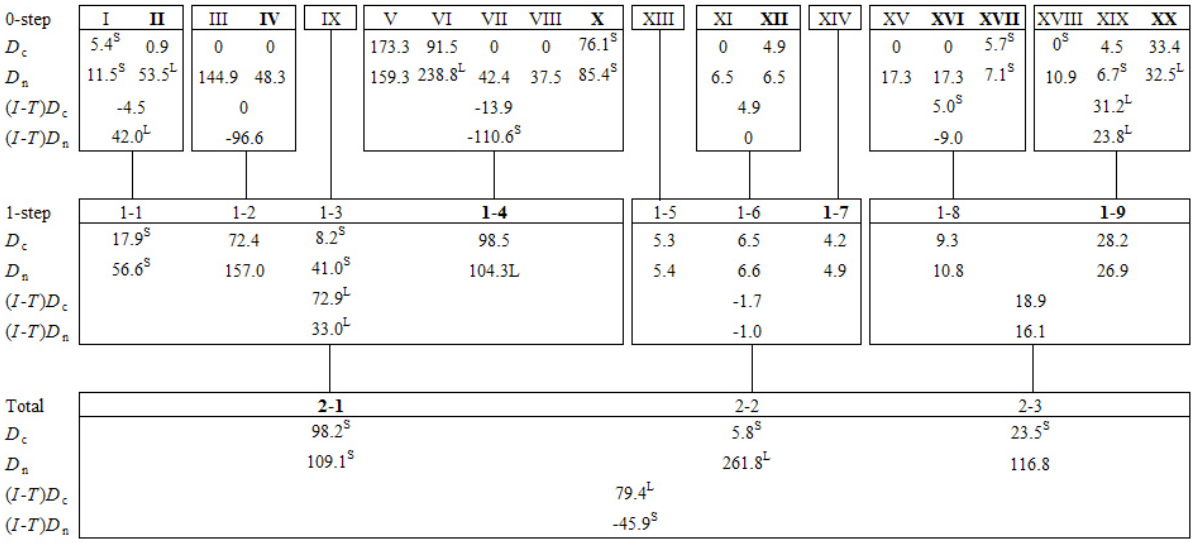


Figure 3. Results of the nested clade analysis of geographical distances for cpDNA haplotypes I – XX. Haplotypes and clades are boxed according to the nesting design given in Fig. 2. Straight lines are drawn between nesting levels when only one clade is included in the next higher level clade. Interior haplotypes/clades are in bold. Significantly smaller or larger values for D_c, D_n, (I-T)D_c and (I-T)D_n than expected at the 5% level based on 1000 permutations are indicated by an 'S' or 'L' superscript.

CHAPTER 3

Microsatellite diversity and genetic structure of the commercially important tropical tree species *Eucalyptus urophylla*, endemic to seven islands in eastern Indonesia

Kitt G. Payn[a,b,c], William S. Dvorak[a], Bernard J.H. Janse[b] and Alexander A. Myburg[c]

[a]*Camcore, North Carolina State University, P.O. Box 8008, Raleigh, NC 27695, USA*

[b]*Mondi South Africa, P.O. Box 31024, Merebank, 4059, South Africa*

[c]*Forest Molecular Genetics Programme, Department of Genetics, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, 0002, South Africa*

This chapter consists of a manuscript prepared for publication in the journal of Tree Genetics and Genomes 4: 519-530, 2008, reproduced with permission from Springer-Verlag. I conducted the laboratory work and data analyses, and wrote the manuscript. Alexander Myburg supervised the research project, and together with William Dvorak, provided extensive suggestions on the organization and content of the manuscript. All co-authors provided critical review of the manuscript. Other people, institutions and private companies that provided assistance, materials, and funding are listed in the Acknowledgements section at the end of the manuscript.

ABSTRACT

Eucalyptus urophylla (Timor mountain gum) is an economically important plantation species that occurs naturally in a series of disjunct populations on the volcanic slopes of seven islands in eastern Indonesia. Twelve microsatellite markers were used to investigate the distribution of nuclear genetic diversity among 19 geographically defined *E. urophylla* populations. High levels of gene diversity were observed throughout the geographic range (H_E = 0.703 to 0.776). The level of genetic differentiation among populations was low (F_{ST} = 0.031), but the amount of differentiation increased with geographic distance. A phenogram produced by a neighbor-joining analysis illustrated that populations clustered according to islands. However, a Bayesian clustering approach revealed a more cryptic population structure comprising two genetically homogeneous groups. Gene flow among the populations is likely responsible for the apparent weak influence of geographic insularity on the genetic diversity and structure of the island species. These findings provide direction for conservation and breeding strategies in *E. urophylla*.

INTRODUCTION

Eucalyptus urophylla (S.T. Blake) is an important plantation species in tropical and subtropical regions of South America, Africa and Asia (Eldridge et al. 1993). Its natural distribution is limited to a series of disjunct populations located on the volcanic slopes of seven of the Lesser Sunda islands in eastern Indonesia. The largest stands are found on the islands of Timor and Wetar, whereas more scattered stands occur on the islands of Adonara,

Alor, Flores, Lomblen (Lembata) and Pantar. Of concern is the apparent depletion in the species' genetic resource, primarily on the latter five islands, as a consequence of land conversion to agriculture (Pepe et al. 2004). Several large-scale seed collections from natural stands of *E. urophylla* have been carried out in an effort to capture the available genetic diversity and ensure a broad genetic base necessary for effective *ex situ* conservation and breeding (Martin and Cossalter 1975; Gunn and McDonald 1991; Hodge et al. 2001; Pepe et al. 2004). Proficient management of the genetic resource requires a detailed understanding of the level and structure of genetic variation across the species' native range.

In its natural habitat, *E. urophylla* is frequently the dominant species of advanced secondary mountain forest (Eldridge et al. 1993). It has the greatest altitudinal range among all *Eucalyptus* species, occurring from almost sea level on Wetar to nearly 3000 m elevation on Timor (Gunn and McDonald 1991). Over most of its range it is a large tree reaching 25-45 m in height, but at the highest elevation the species grows as a small shrub less than two meters tall (Turnbull and Brooker 1978). Added to its diverse growth form is the commensurate variation in fruit size and bark morphology (Martin and Cossalter 1975). This has led to *E. urophylla* being described as one of the most variable of all *Eucalyptus* species (Eldridge et al. 1993).

Traditionally, the genetic variation of forest trees is studied using provenance trials established in different environments and the focus is on quantitative traits. *Ex situ E. urophylla* provenance trials have demonstrated that low elevation seed sources have superior

height growth compared to high elevation seed sources (Moura 1983; Wencelius 1983; Hodge et al. 2001). Provenance variation in seedling morphology (Pinyopusarerk et al. 1993) and essential leaf oils (Doran et al. 1995) has also been described. In both of these studies the sampled provenances from the island of Wetar differentiated from other provenances of *E. urophylla*. A contemporary study of adult morphology *in situ* resulted in several dry site provenances on Wetar and several high elevation provenances in East Timor being differentiated from other provenances of *E. urophylla* (Pryor et al. 1995). The morphological differentiation led to the proposed separation of two new species from *E. urophylla*, namely *E. wetarensis* and *E. orophila* on the island of Wetar and Timor, respectively (Pryor et al. 1995). However, this new taxonomic classification has not been unanimously accepted (e.g. Brooker 2000).

The study of morphological variability as a method of assessing the nature and distribution of genetic diversity is complicated by environmental and/or maturation effects. Molecular techniques offer an alternate approach that is not subject to environmental variation. Notably, plants accumulate genetic differences in the nuclear, as well as the two organellar genomes i.e. the mitochondrial and chloroplast genome. The chloroplast genome is maternally inherited in most angiosperms (Corriveau and Coleman 1988), including *Eucalyptus* (Byrne et al. 1993; McKinnon et al. 2001). The geographical distribution of chloroplast DNA sequence variation in *E. urophylla* was analyzed to gain insight into the seed-mediated island colonisation routes of the species (Payn et al. 2007). Provenances from the islands of Wetar in the east and Timor in the south contained most of the chloroplast genetic diversity and

there was evidence to suggest that *E. urophylla* was present on both islands for a period of time before colonizing the remaining islands of its extant distribution.

The level and geographical distribution of nuclear genetic diversity in *E. urophylla* is less clear. A species-wide study of the genetic diversity of this species has been performed using isozyme markers (House and Bell 1994). The investigation revealed limited population differentiation (G_{ST} = 0.12) suggesting that most of the nuclear genetic diversity in the species was found within rather than between populations. A cluster analysis revealed no striking patterns that could be related to geography, with the exception of the populations from the island of Wetar, which clustered together on the basis of isozyme genotypes (House and Bell 1994). A recent microsatellite study of the genetic structure of *E. urophylla* (Tripiana et al. 2007) confirmed that there is low genetic differentiation among populations (F_{ST} = 0.04). However, this study was not able to resolve the clustering of populations within islands due to limited sampling, particularly on the island of Wetar (Tripiana et al. 2007).

It is important for *ex situ* conservation and breeding of *E. urophylla* to obtain more detailed insight into the nature and distribution of genetic diversity at the population and island levels. It is not clear whether there is significant pairwise differentiation among some populations and whether the amount of differentiation increases with geographic distance. Our recent analysis of chloroplast genetic diversity in *E. urophylla* revealed that a greater percentage of the total variance could be explained by differences among islands compared to differences among populations within islands. A similar analysis of molecular variance is required to

determine how nuclear genetic diversity is apportioned between and within islands. Finally, if genetic differentiation is low at the population and island levels, is it possible to identify genetically homogenous groups of individuals at a higher hierarchical level? We present the results of a study that aimed to (1) assess the nuclear genetic diversity of 19 geographically defined populations distributed across the natural range of *E. urophylla*, (2) determine the level of genetic differentiation at the island and population levels and assess its relationship to geographical distribution, and (3) provide recommendations for conservation and breeding strategies for the species.

MATERIALS AND METHODS

Plant material

Seed collections were conducted by Camcore (North Carolina State University, Raleigh, NC, USA), an international tree conservation and domestication program, in collaboration with PT Sumalindo Lestari Jaya, a private Indonesian forestry company (Pepe et al. 2004). This series of collections comprised seed from 1104 mother trees distributed across 62 sites representing the natural distribution of *E. urophylla*, barring East Timor, which was experiencing political unrest at the time of seed collection. A subset of 45 collection sites was included in this molecular study. Seed collection sites were grouped according to geographic location to form 19 populations with a sample size $N \geq 12$ (Table 1; Fig. 1). A test for differentiation among pairwise combinations of seed collection sites within each geographically defined population revealed no significant differentiation (results not shown).

Seeds were sown in a commercial nursery in South Africa (Mondi, South Africa). Leaf tissue was sampled from 357 plants, each originating from a different mother tree.

None of the populations included in this study comprised high elevation sites in East Timor that would be classified by Pryor et al. (1995) as *E. orophila.* There is the possibility that some populations sampled from the island of Wetar would be classified as *E. wetarensis*. However, this study follows the taxonomic classification of *E. urophylla* according to Brooker (2000), whereby all samples are considered to be *E. urophylla*.

DNA isolation and microsatellite analysis

Total genomic DNA was extracted from 50 mg of fresh leaf tissue using the DNeasy Plant Mini Kit (QIAGEN, Valencia, California, USA). Samples were homogenized for 30-60 sec in a FastPrep FP120 instrument (QBiogene, Carlsbad, California, USA) set at 4.0 meters/sec. In order to improve efficiency, cell lysis was performed at 65°C for 30 minutes. Thereafter, all steps were performed as described in the DNeasy Plant Mini Kit manual. DNA quality and quantity were determined by agarose gel electrophoresis and spectrophotometry (Nanodrop Technologies, Wilmington, Delaware, USA).

A total of 12 microsatellite loci with different levels of polymorphism were selected for this study: nine previously described loci (EMBRA 3, Brondani et al. 1998; EMBRA 28, EMBRA 37, EMBRA 48, EMBRA 69, Brondani et al. 2002; EMBRA 125, EMBRA 219, EMBRA 227, Brondani et al. 2006; FMRSA 4, van der Nest et al. 2000) and three

unpublished loci (Supplementary material Table S1). Polymerase chain reactions (PCRs) were performed in a total volume of 13 μl with the following components: 6 ng of genomic DNA, 0.2 μM of each primer, 0.2 mM of each dNTP, 0.15 U of Exsel High Fidelity DNA Polymerase (JMR Holdings, London, UK), and 1× PCR Exsel buffer. PCR amplifications were performed in an iCycler (BIO-RAD Laboratories, Hercules, CA, USA) with the following cycling conditions: an initial denaturation step of 94°C for 1 min, followed by 30 cycles of 94°C for 1 min, 56°C (54°C for locus EMBRA 48, Brondani et al. 2002) for 1 min and 72°C for 1 min, followed by an elongation step at 72°C for 5 min, and a final step at 60°C for 45 min to ensure the complete non-template addition of adenine (Magnuson et al. 1996). One primer in each of the sets was labeled with a phosphoramidite fluorescent label (FAM, HEX, TET or VIC). PCR products were analyzed on an ABI PRISM® 3100 Genetic Analyzer (Applied Biosystems, Foster City, USA) using Genescan™-500 LIZ™ (Applied Biosystems) as an internal standard. Electropherograms were analyzed with the ABI PRISM® Genemapper® 3.0 software (Applied Biosystems).

Statistical analysis

The number of alleles (A) and the unbiased estimates of observed (H_O) and expected (H_E) heterozygosity were calculated for each locus and averaged for each population over all loci using GDA 1.1 (Lewis and Zaykin 2001). Alleles were deemed private if they showed a frequency of more than 5% in one population and did not occur in any other population. Allelic richness (A_R) as a standardized measure of the number of alleles per locus corrected

by the sample size was calculated using FSTAT v2.9.3.2 (Goudet 1995). Mean A_R and H_E were compared among groups of populations (populations grouped by island) using a two-sided permutation test implemented in FSTAT v2.9.3.2. The Pearson product-moment correlation between estimates of genetic diversity (H_E and A_R) and geographical parameters (longitude and elevation) was calculated using JMP v4.04 (SAS Institute, Cary, USA). This analysis was performed at the seed collection site level and only included sites with a sample size $N \geq 7$ (Table 1).

Populations were tested for recent reduction of their effective population size based on the approach of Cornuet and Luikart (1996) with the program BOTTLENECK (Piry et al. 1999). The Wilcoxon sign-rank test was used to test the significance of heterozygote excess under two different models, the infinite allele model (IAM) and the stepwise mutation model (SMM). All pairs of loci were tested for genotypic linkage disequilibrium based on a Markov chain method using GENEPOP v3.4 (Raymond and Rousset 1995). All tests were adjusted for multiple comparisons using a sequential Bonferroni correction (Rice 1989).

Population genetic structure was analyzed through hierarchical F coefficients (Weir and Cockerham, 1984) using FSTAT v2.9.3.2. F_{ST} values (θ-statistics of Weir and Cockerham 1984) between pairs of populations were estimated and the significances of pairwise population differentiation were tested using the log-likelihood statistic G (Goudet et al. 1996) implemented in FSTAT v2.9.3.2. Reported significance values are after sequential Bonferroni correction (Rice 1989). The Mantel test (Mantel 1967) was used to test the

correlation between $F_{ST}/(1 - F_{ST})$ and the natural logarithm of geographic distance between sample sites using GENEPOP v3.4. Significance of the correlation was assessed by 1000 random permutations.

An Analysis of Molecular Variance (AMOVA, Excoffier et al. 1992) was carried out using the program ARLEQUIN v2.0 (Schneider et al. 2000) to apportion variance within populations, among populations within islands, and among islands, applying the estimator of Weir and Cockerham (1984). Genetic distances between populations were estimated using *Nei* D_A genetic distance (Nei et al. 1983) as implemented in Powermarker v3.25 (Liu and Muse 2005). The resulting distance matrix was used to construct a neighbor-joining (NJ) phenogram that was bootstrapped 10 000 times. The phenogram topology was examined using MEGA v2 (Kumar et al 1993).

The above tests for genetic structure (*F* statistics and AMOVA) require predefined groups of individuals. An alternate approach, implemented in the program STRUCTURE v2.0 (Pritchard et al. 2000), can estimate the number of genetically homogeneous populations (*K*) using a Bayesian model-based clustering method that does not require prior information of the number of locations and from which location each individual was sampled. Admixture among populations and correlated allele frequencies were assumed for the algorithm. A burn-in period of 100,000 iterations was followed by 100,000 iterations of the Markov chain. The model was run for a range of *K*-values from 1 to 20, with five replications each. The best *K*-value supported by the data was assessed according to the recommendations of Evanno et al.

(2005), whereby the statistic ΔK was calculated based on the rate of change in the log probability of data between successive *K*-values.

Multilocus estimates of the effective number of migrants (*Nm*) between pairs of populations based on private alleles (Slatkin 1985) were computed using GENEPOP v3.4. The mean *Nm* between paired populations within islands was compared to the mean *Nm* between paired populations among islands by means of an ANOVA, implemented in JMP v4.04.

RESULTS

Microsatellite polymorphism and population genetic diversity

In a total of 357 *E. urophylla* seedlings surveyed, 302 alleles were identified at 12 polymorphic microsatellite loci (Supplementary material Table S2). The number of alleles per locus ranged from six alleles at locus FMG-EUCSSR3 to a maximum of 56 alleles at EMBRA48, with an average of 25 alleles per locus. Expected heterozygosity (H_E) values ranged from 0.223 at locus FMG-EUCSSR3 to 0.947 at locus EMBRA48, while the observed heterozygosity (H_O) values ranged from 0.192 to 0.919 at the same loci.

For the 19 populations surveyed, the mean number of alleles per locus per population (*A*) ranged from 7.7 in Lomblen-1 to 12 in Alor-2 (Table 2). The overall mean allelic richness (A_R) was 8.2, and varied from 7.5 in Lomblen-1 to 9.2 in Flores-3. Private alleles were identified in only three populations, namely Adonara-2, Pantar-1 and Pantar-2. The mean

expected heterozygosity across all loci over all populations was 0.739, and ranged from 0.703 in Wetar-1 to 0.776 in Timor-2. The observed heterozygosity across all loci was highest in Pantar-2 with a value of 0.737 and lowest in Adonara-1 with a value of 0.610 (Table 2).

Overall, there was no clear pattern with respect to genetic diversity and geography. More specifically, the tests for differences in mean A_R and H_E among groups of populations (populations grouped by island) were not significant ($\alpha = 0.05$). The correlations between the genetic diversity parameters (H_E and A_R) and select geographical parameters (longitude and elevation) were not significant (results not shown). The Wilcoxon sign-rank test ($\alpha = 0.05$) revealed evidence for recent bottlenecks in two populations (Lomblen-1 and Timor-3) according to the IAM. However, after Bonferroni correction, none of the tests remained significant at the experimental level. None of the populations revealed heterozygote excess when assuming a SMM.

The test for genotypic linkage disequilibrium for pair-wise combinations of the 12 microsatellite loci over all populations were not significant ($P < 0.05$). Forty out of 1254 tests of genotypic linkage disequilibrium within populations were nominally significant ($P < 0.05$; 63 significant values are expected at the 5% level), however, after Bonferroni correction, none of the combinations remained significant at the experimental level (results not shown). These findings suggest that the marker loci used in this study segregated independently.

Population genetic structure

The microsatellite markers applied in this study revealed a low degree of genetic differentiation among the 19 *E. urophylla* populations with an overall F_{ST} value of 0.031 (Supplementary material Table S2) and a 95% confidence interval of 0.027 - 0.035. Pairwise F_{ST} values (Table 3) ranged from -0.003 (Lomblen-1 vs. Lomblen-3) to 0.066 (Pantar-1 vs. Wetar-2). According to Weir (1996), estimates of θ (F_{ST}) may be negative if the true value of θ is positive but small or if the parameter is in fact negative, which corresponds to a negative intraclass correlation i.e. alleles are more related between than within populations. Significant genetic differentiation was found in 145 out of 171 population pairs after sequential Bonferroni correction. Genetic differentiation observed between populations occurring on the same island within Adonara, Lomblen, Pantar and Wetar were not significant (NS). The relationship between genetic differentiation and geographical distance among pairwise comparisons of populations was investigated using the Mantel test. There was a significant correlation between $F_{ST}/(1 - F_{ST})$ and the natural logarithm of distance (r = 0.43, $P < 0.01$).

An AMOVA of hierarchical gene diversity revealed that genetic variation among individuals within populations accounted for 96.6% of the total molecular variance (Table 4). A further 1.5% of the total variance was distributed among populations within islands, whereas 1.9% of the variance occurred among islands. The relationships between the 19 *E. urophylla* populations based on Nei's (1983) genetic distances were summarized in an unrooted

neighbor-joining phenogram (Fig. 2). Overall, the topology of the phenogram coincided well with geographic distribution, and the populations generally clustered at the island level.

The Bayesian cluster analysis performed using the STRUCTURE software showed that the log-likelihoods of the number of population clusters reached a plateau at $K = 7$. The methodology of Evanno et al. (2005) strongly supported $K = 2$ as the correct number of groups. For this scenario the proposed membership of population cluster-1 mostly comprised samples located on the western islands of Flores, Adonara, Lomblen and Pantar (Fig. 1). The proposed membership of population cluster-2 predominantly comprised samples located on the eastern island of Wetar and southern island of Timor. The island of Alor appears to represent an east to west population cluster transition zone.

Gene flow among populations

Estimates of *Nm* between pairs of populations (Table 3) ranged from 1.23 (between Lomblen-1 and Timor-3) to 4.69 (between Alor-2 and Timor-2). The difference in the mean *Nm* between paired populations within islands (*Nm* = 3.20, SE = 0.15) and the mean *Nm* between paired populations among islands (*Nm* = 2.57, SE = 0.05) was significant ($P < 0.01$).

DISCUSSION

The mean level of gene diversity ($H_E = 0.74$) for *E. urophylla* obtained in this study was comparable to gene diversity estimates reported in other *Eucalyptus* species based on

microsatellite markers (Byrne et al. 1996; McGowen et al. 2001; Jones et al. 2002; Butcher et al. 2005). Although the range of diversity across the 19 *E. urophylla* populations was small (H_E = 0.703 - 0.776), the highest estimates were recorded for populations from the island of Timor. The higher level of gene diversity for these populations may be due to the relatively large populations present on the island (Pepe et al. 2004). In addition, Timor was likely one of the first of the seven islands colonized by *E. urophylla* (Payn et al. 2007) and genetic diversity is expected to be higher in an older source population.

An isozyme study by House and Bell (1994) also reported high levels of gene diversity for *E. urophylla* populations from Timor despite the seed samples being collected from a disproportionately low number of mother trees relative to the other islands. Populations from the island of Wetar were also reported to have among the highest levels of diversity (House and Bell 1994). In the present study Wetar populations exhibited genetic diversity at the lower end of the observed range. An explanation for these dissimilar results may be due to the sampling of different Wetar populations and different genetic loci in the two studies.

The observed level of genetic differentiation among the 19 populations (F_{ST} = 0.031) was similar to that recently reported for *E. urophylla* based on microsatellite markers (F_{ST} = 0.04; Tripiana et al. 2007), but was lower than the level of genetic differentiation observed in other *Eucalyptus* species (*E. globulus*, F_{ST} = 0.08, Jones et al. 2002; *E. benthamii*, F_{ST} = 0.105, Butcher et al. 2005; *E. morrisbyi*, F_{ST} = 0.19, Jones et al. 2005). These species were sampled as mature trees *in situ*, whereas nursery-grown seedlings were sampled in the present study.

It is possible that differentiation is more pronounced in mature trees, because of selection against non-adapted migrants after the seedling stage (Kalisz et al. 2001).

In this study, population differentiation increased with increasing geographic distance and low but significant population pairwise F_{ST} values were found for 145 out of 171 population pairs ($F_{ST} \leq 0.066$; Table 3). The significant pairwise values were mostly for populations from different islands. This was illustrated in the neighbor-joining phenogram whereby most populations grouped according to island (Fig. 2). Our recent analysis of chloroplast DNA variation in *E. urophylla* (Payn et al. 2007) suggested that bottlenecks, possibly associated with founder effects, have contributed to the high level of genetic differentiation at the chloroplast level. Microsatellite data generated in this study revealed no evidence for bottlenecks at the nuclear genome level. The lower level of population structure inferred from the microsatellite data may be explained by the nuclear genome having a larger effective population size relative to the chloroplast genome thus making it less susceptible to genetic drift (Birky et al. 1989). In addition, the nuclear genome migrates through pollen and seed, whereas dispersal of the chloroplast genome is limited to the latter, which is typically restricted to a distance of twice the parent tree height (Cremer 1977).

The estimated mean number of migrants between populations on different islands was significantly ($P < 0.01$) less than the mean number of migrants between populations within islands. This was not surprising because in most cases the distance between islands is greater than the distance that insect vectors might normally carry pollen. There is a possibility that

wind blown insects could carry pollen between islands, or that birds may potentially be pollen vectors. Tropical cyclones occasionally pass through the Lesser Sunda islands. Cyclonic events overlapping with the time of pollen dispersal would promote long-distance gene flow within and among islands (Dvorak et al. 2007). According to Wright (1931), migration rates of greater than one migrant per generation may be sufficient to prevent population differentiation due to drift. The migration rates estimated between the pairs of island populations (Table 3) were above that level and appeared to be sufficiently high to counteract the effect of island distribution on population structure (Table 4).

The low level of population differentiation may also be due to historically recent population fragmentation. Volcanic lava flow on the Lesser Sunda islands for example, may have resulted in the fragmentation of some *E. urophylla* populations. Such fragmented populations could diverge through genetic drift if inter-population gene flow was sufficiently low. Population fragmentation may also have occurred at the island level. During the Pleistocene, glaciation and deglaciation led to fluctuating sea levels that greatly affected landmass configurations in Southeast Asia (Voris 2000). There is evidence to suggest that the islands of Flores, Adonara and Lomblen were connected when sea levels were 60-120 m below the present level, however, the other four islands on which *E. urophylla* naturally occurs remained separated (Heaney 1991; Voris 2000). The phylogeographic structure of *E. urophylla* observed at the chloroplast level (Payn et al. 2007) provides some evidence that the species was present on the islands of Timor and Wetar for a period of time before colonizing the remaining five islands. The low level of nuclear differentiation between the

distantly located Timor and Wetar populations is likely the result of gene flow through pollen migration, and suggests that gene flow would also occur between fragmented populations within islands. Therefore the low level of population differentiation observed in the present study may primarily be the result of extant pollen flow rather than recent population fragmentation.

A Bayesian statistical method (Pritchard et al. 2000) that provides a framework for clustering individuals into genetically homogeneous groups predicted two population clusters. The individuals from the islands in the west were primarily members of cluster-1, whereas those from Wetar and Timor were primarily members of cluster-2. Population age and geographic distance may partly explain the clustering of Wetar and Timor samples. Wetar populations, in particular, which had the highest proportion of membership of cluster-2, have previously been differentiated at the morphological level (Pinyopusarerk et al. 1993; Doran et al. 1995; Pryor et al. 1995). The natural range of *E. urophylla* experiences two distinct climatic types defined by average monthly precipitation and temperature at provenance locations (Dvorak et al. 2007). Climatic type-1 comprised all provenances except those on the island of Wetar and was characterized by a dry season of 5 to 6 months. Climatic type-2 comprised the provenances of Wetar and was distinguished by a lower annual precipitation and a more evenly distributed rainfall pattern throughout the year (Dvorak et al. 2007). Interestingly, the frequency of membership in cluster-1 vs. cluster-2 was strongly correlated with provenance (seed collection site) volume growth assessed in Camcore trials planted in Mexico ($r = 0.75$, $P = 0.0011$) and Venezuela ($r = 0.74$, $P = 0.0012$) (Dvorak et al. 2007).

The proposed separation of two new species from *E. urophylla*, namely *E. wetarensis* and *E. orophila* (Pryor et al. 1995) is yet to be strongly supported at the molecular level. Although populations from Wetar clustered together on the basis of their isozyme genotypes (House and Bell 1994), it was concluded that there was no strong evidence to support the establishment of a new species. Due to the political unrest in East Timor, the present study does not comprise samples likely to be classified by Pryor et al. (1995) as *E. orophila.* However, there is the possibility that populations sampled from the island of Wetar would be classified as *E. wetarensis*. Wetar populations were shown to cluster together based on microsatellite data, but the same applied for most of the other populations sampled from the same islands (Fig. 2). The results of the Bayesian analysis further suggested that Wetar populations generally cluster into a genetically homogenous group (cluster 2, Fig. 1), however members of cluster 2 were also prominent on the islands of Timor and Alor. Therefore, there is no strong evidence on the basis of our microsatellite data to support the establishment of new species. New molecular techniques such as DArT (Diversity Arrays Technology; Jaccoud et al. 2001) that sample the genome more intensively could possibly be used to further investigate the proposed classification of two new species from *E. urophylla.*

Management implications and recommendations for future studies

The majority of genetic variation observed in this study resides within rather than between populations suggesting that a subset of populations need only be included in an *ex situ* conservation or breeding program in order to capture most of the genetic variation within the

species. In light of the two genetically homogeneous population clusters, it would be prudent to keep genotypes representing each cluster separate. Within each cluster, it appears appropriate to group populations that are not significantly different. For example, the grouping of populations could be implemented at the island level for Lomblen, Pantar and Wetar populations (Table 3).

Microsatellite analyses infer the genetic diversity and population structure at likely neutral loci. The advantage is that such markers enable population geneticists to investigate important neutral evolutionary forces such as migration and genetic drift (Black et al. 2001). The genetic diversity and population structure at genes conferring adaptive potential may be very different. Therefore neutral molecular marker data should be used in combination with phenotypic information provided by *ex situ* field trials to characterize the genetic resource. In this regard, populations exhibiting superior growth and adaptability in exotic environments have great economic importance and should be included in the genetic resource. For example, several *E. urophylla* populations from the western islands of Flores and Lomblen have been characterized by superior growth and adaptability (survival) in Mexico and Venezuela (Dvorak et al. 2007). However, as these authors point out, the best performing populations may differ among countries and will result in the prioritization of different populations.

Finally, we recommend that a future molecular study specifically investigates the effect elevation has on the population genetic structure of *E. urophylla*, since *ex situ* field trials

have demonstrated that low elevation seed sources have superior growth compared to high elevation seed sources (Moura 1983; Wencelius 1983; Hodge et al. 2001). *Eucalyptus alba* also occurs on the Lesser Sunda islands on low elevation sites. Morphological data indicate that hybrids do exist *in situ* but are considered rare (Martin and Cossalter 1976). It would be of value for an elevation study to include *E. alba* samples and estimate the level of introgression.

ACKNOWLEDGEMENTS

We are grateful to PT Sumalindo Lestari Jaya for the seed collections, Camcore for locality and reference information, and Mondi for maintenance of the plant materials. Financial support for this work was provided by Mondi through the Wood and Fiber Molecular Genetics Programme and by Camcore, Raleigh, NC, USA. Additional funding was provided by the Technology and Human Resources for Industry Programme (THRIP) in South Africa.

LITERATURE CITED

Birky Jr CW, Fuerst P, Maruyama T (1989) Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics* **103**, 513–527.

Black WC, Baer CF, Antolin MF, DuTeau NM (2001) Population genomics: Genome-wide sampling of insect populations. *Annual Review of Entomology* **46**, 441–469.

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

Brondani RPV, Brondani C, Grattapaglia D (2002) Towards a genus wide reference linkage map for *Eucalyptus* based exclusively on highly informative microsatellite markers. *Molecular Genetics and Genomics* **267**, 338–347.

Brondani RPV, Williams ER, Brondani1 C, Grattapaglia D (2006) A microsatellite-based consensus linkage map for species of *Eucalyptus* and a novel set of 230 microsatellite markers for the genus. *BMC Plant Biology* **6,** 1–16.

Brooker MIH (2000) A New Classification of the Genus *Eucalyptus* L'Hér. (Myrtaceae). *Australian Systematic Botany* **13**, 79–148.

Butcher PA, Skinner AK, Gardiner CA (2005) Increased inbreeding and inter-species gene flow in remnant populations of the rare *Eucalyptus benthamii. Conservation Genetics* **6**, 213–226.

Byrne M, Moran GF, Tibbits WN (1993) Restriction map and maternal inheritance of chloroplast DNA in *Eucalyptus nitens. Journal of Heredity* **84**, 218–220.

Byrne M, Marques-Garcia MI, Uren T, Smith DS, Moran GF (1996) Conservation and genetic diversity of microsatellite loci in the genus *Eucalyptus. Australian Journal of Botany* **44**, 331–341.

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**, 2001–2014.

Corriveau JL, Coleman AW (1988) Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *American Journal of Botany* **75**, 1443–1458.

Cremer KW (1977) Distance of seed dispersal in eucalypts estimated from seed weights. *Australian Forest Research* **7**, 225–228.

Doran JC, Williams ER, Brophy JJ (1995) Patterns of variation in the seedling leaf oils of *Eucalyptus urophylla*, *E pellita* and *E scias. Australian Journal of Botany* **43**, 327–336.

Dvorak WS, Hodge GR, Payn KG (2007) The conservation and breeding of *Eucalyptus urophylla*: A case study to better protect important populations and improve productivity. In: Eucalypts & Diversity: Balancing Productivity and Sustainability. Proc. IUFRO Working Party 2.08.03.

Eldridge K, Davidson J, Harwood C, Van Wyk G (1993) Eucalypt domestication and breeding. Oxford University Press, Oxford.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **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.

Goudet J (1995) FSTAT: a computer program to calculate F-statistics. *Journal of Heredity* **86,** 485–486.

Goudet J, Raymond M, Demeeus T, Rousset F (1996) Testing differentiation in diploid populations. *Genetics* **144**, 1933–1940.

Gunn BV, McDonald MW (1991) *Eucalyptus urophylla* seed collections. *Forest Genetic Resources* **19**, 34–37, FAO, Rome.

Heaney LR (1991) A synopsis of climatic and vegetational change in Southeast Asia. *Climate Change* **19**, 53–61.

Hodge GR, Pepe B, Wijoyo FS, Dvorak WS (2001) Early results of *Eucalyptus urophylla* provenance/progeny trials in Colombia and Venezuela. In: Developing the Eucalypts of the Future. Proc. IUFRO Working Party 2.08.03.

House APN, Bell JC (1994) Isozyme variation and mating system in *Eucalyptus urophylla* ST Blake. *Silvae Genetica* **43**, 167–176.

Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity Arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Research* **29**, 25–31.

Jones RC, Steane DA, Potts BM, Vaillancourt RE (2002) Microsatellite and morphological analysis of *Eucalyptus globulus* populations. *Canadian Journal of Forest Research* **32**, 59–66.

Jones RC, McKinnon GE, Potts MB, Vaillancourt RE (2005) Genetic diversity and mating system of an endangered tree *Eucalyptus morrisbyi. Australian Journal of Botany* **53**, 367–377.

Kalisz S, Nason JD, Hanzawa FM, Tonsor SJ (2001) Spatial population genetic structure in *Trillium grandiflorum*: The roles of dispersal, mating, history, and selection. *Evolution* **55**, 150–1568.

Kumar S, Tamura K, Nei M (1993) MEGA version 10: Molecular Evolutionary Genetic Analysis. Pennsylvania State University, Pennsylvania.

Lewis PO, Zaykin D (2001) Genetic Data Analysis: Computer program for the analysis of allelic data Version 11. Free program distributed by the authors over the Internet from http://hydrodictyoneebuconnedu/people/plewis/softwarephp.

Liu K, Muse S (2005) PowerMarker: new genetic data analysis software Version 323. Free program distributed by the author over the internet from http://wwwpowermarkernet.

Magnuson VL, Ally DS, Nylund SJ, Karanjawala ZE, Rayman JB, Knapp JI, Lowe AL, Ghosh S, Collins FS (1996) Substrate nucleotide-determined non-templated addition of adenine by Taq DNA polymerase: Implications for PCR-based genotyping and cloning. *Biotechniques* **21**, 700–709.

Mantel N (1967) The detection of disease clustering and generalized regression approach. *Cancer Research* **27**, 209–220.

Martin B, Cossalter C (1975) Les eucalyptus des iles de la Sonde. *Bois et Foréts des Tropiques* **163**, 3–25.

Martin B, Cossalter C (1976) Les eucalyptus des iles de la Sonde. *Bois et Foréts des Tropiques* **164**, 3–14.

McGowen MH, Wiltshire RJE, Potts BM, Vaillancourt RE (2001) The origin of *Eucalyptus vernicosa*, a unique shrub Eucalypt. *Biological Journal of the Linnean Society* **74**, 397–405.

McKinnon GE, Vaillancourt RE, Tilyard PA, Potts BM (2001) Maternal inheritance of the chloroplast genome in *Eucalyptus globulus* and interspecific hybrids. *Genome* **44**, 831–835.

Moura VPG (1983) Resultados de pesquisa com várias procêdencias de *Eucalyptys urophylla* ST Blake, no centro-leste do Brasil. *Silvicultura, São Paulo* **31**, 474–480.

Nei M, Tajima F, Tateno Y (1983) Accuracy of genetic distances and phylogenetic trees from molecular data. *Journal of Molecular Evolution* **19**, 153–170.

Payn KG, Dvorak WS, Myburg AA (2007) Chloroplast DNA phylogeography reveals the island colonisation route of *Eucalyptus urophylla* (Myrtaceae). *Australian Journal of Botany* **55**, 673–683.

Pepe B, Surata K, Suhartono F, Sipayung M, Purwanto A, Dvorak WS (2004) Conservation status of natural populations of *Eucalyptus urophylla* in Indonesia and international efforts to protect dwindling gene pools. *Forest Genetic Resources* **31**, 62–64, FAO Rome.

Pinyopusarerk K, Gunn BV, Williams ER, Pryor LD (1993) Comparative geographical variation in seedling morphology of three closely related red mahoganies, *Eucalyptus urophylla*, *E pellita* and *E scias. Australian Journal of Botany* **41**, 23–34.

Piry S, Luikart G, Cornuet JM (1999) BOTTLENECK: a computer programme for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity* **90**, 502–503.

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

Pryor LD, Williams ER, Gunn BV (1995) A morphometric analysis of *Eucalyptus urophylla* and some related taxa with descriptions of two new species. *Australian Systematic Botany* **8**, 57–70.

Raymond M, Rousset F (1995) GENEPOP (version 12): population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**, 248–249.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution* **43**, 223–225.

Schneider S, Roessli D, Excoffier L (2000) Arlequin ver 2000: A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Slatkin M (1985) Rare alleles as indicators of gene flow. *Evolution* **39**, 53–65.

Tripiana V, Bourgeois M, Verhaegen D, Vigneron P, Bouvet JM (2007) Combining microsatellites, growth, and adaptive traits for managing in situ genetic resources of *Eucalyptus urophylla*. *Canadian Journal of Forest Research* **37**, 773–785.

Turnbull JW, Brooker MIH (1978) Timor mountain gum, Eucalyptus urophylla ST Blake. *Forest Tree Leaflet* 214 CSIRO, Melbourne.

Van Der Nest MA, Steenkamp ET, Wingfield BD, Wingfield MJ (2000) Development of simple sequence repeat (SSR) markers in *Eucalyptus* from amplified inter-simple sequence repeats (ISSR). *Plant Breeding* **119,** 433–436.

Voris HK (2000) Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *Journal of Biogeography* **27**, 1153–1167.

Wencelius F (1983) *Eucalyptus urophylla* en Côte d'Ivoire. *Silvicultura, São Paulo* **31,** 515–518.

Weir BS, Cockerham C (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370.

Weir BS (1996) Genetic data analysis II. Sinauer Associates Inc., Sunderland, USA.

Wright S (1931) Evolution in Mendelian populations. *Genetics* **28**, 114–138.

Table 1. Details of the 19 *Eucalyptus urophylla* populations sampled. The population name, seed collection site name, location, elevation and sample size per population.

Population	Collection site (no. of individuals)	Location	Elevation (m)	Sample size
Flores-1				23
	Ile Meak (6)	08°37'S, 122°15'E	680	
	Lere-Baukrenget (10)	08°39'S, 122°23'E	725	
	Natakoli (7)	08°37'S, 122°24'E	900	
Flores-2				25
	Ile Nggele (10)	08°39'S, 122°27'E	685	
	Kilawair (11)	08°41'S, 122°29'E	378	
	Palueh (4)	08°40'S, 122°35'E	570	
Flores-3				19
	Kolibuluk (7)	08°28'S, 122°42'E	648	
	Hokeng (12)	08°31'S, 122°47'E	575	
Adonara-1				20
	Watololong (4)	08°19'S, 123°15'E	630	
	Gonehama (7)	08°20'S, 123°16'E	687	
	Doken (9)	08°21'S, 123°18'E	800	
Adonara-2				19
	Lamahela (10)	08°21'S, 123°15'E	856	
	Muda (9)	08°21'S, 123°16'E	750	
Lomblen-1				12
	Jontona (7)	08°16'S, 123°25'E	788	
	Bunga Muda (5)	08°16'S, 123°32'E	650	
Lomblen-2				20
	Ile Ape (6)	08°29'S, 123°30'E	860	
	Padekluwa (7)	08°30'S, 123°26'E	800	
	Ile Kerbau (7)	08°29'S, 123°29'E	740	
Lomblen-3				15
	Puor (7)	08°34'S, 123°24'E	940	
	Labalekan (8)	08°32'S, 123°30'E	770	
Pantar-1				12
	Beangonong (4)	08°20'S, 124°12'E	565	
	Lalapang (8)	08°20'S, 124°12'E	575	
Pantar-2				17
	Mauta (9)	08°26'S, 124°10'E	620	
	Delaki (8)	08°28'S, 124°11'E	810	
Alor-1				20
	Watakika (10)	08°18'S, 124°30'E	475	
	Pintu Mas (10)	08°17'S, 124°33'E	385	

Table 1. Continued

Population	Collection site (no. of individuals)	Location	Elevation (m)	Sample size
Alor-2				24
	Mainang (7)	08°14'S, 124°39'E	1175	
	Molpui (4)	08°15'S, 124°44'E	400	
	Apui (10)	08°16'S, 124°44'E	1200	
	Manabai (3)	08°14'S, 124°45'E	400	
Wetar-1				21
	Nesunhuhun (10)	07°52'S, 126°15'E	621	
	Elun Kripas (11)	07°51'S, 126°16'E	733	
Wetar-2				14
	Nakana Ulam (7)	07°51'S, 126°21'E	715	
	Alasannaru (7)	07°51'S, 126°23'E	596	
Wetar-3				21
	Puaanan (7)	07°51'S, 126°26'E	485	
	Remamea (8)	07°52'S, 126°26'E	476	
	Talianan (6)	07°52'S, 126°28'E	521	
Timor-1				17
	Bonleu (7)	09°33'S, 124°04'E	1700	
	Nuafin (10)	09°31'S, 124°11'E	1900	
Timor-2				16
	Leloboko (6)	09°37'S, 124°10'E	1500	
	Naususu (10)	09°38'S, 124°13'E	1325	
Timor-3				20
	Tutem (10)	09°35'S, 124°17'E	1300	
	Tune (10)	09°33'S, 124°19'E	1250	
Timor-4				22
	Mollo (11)	09°41'S, 124°11'E	1400	
	Lelobatang (11)	09°41'S, 124°14'E	1300	
Total				357

Table 2. Microsatellite diversity indices for 19 *Eucalyptus urophylla* populations. Sample size (N), average number of alleles per locus (A), mean allelic richness per population (A_R), number of private alleles (A_P), expected (H_E) and observed (H_O) heterozygosity, and Wright's fixation index (F).

Population	N	A	A_R	A_p	H_E	H_O	F
Adonara-1	20	10.3	8.0	-	0.715	0.610	0.150
Adonara-2	19	9.9	8.0	1	0.727	0.677	0.070
Alor-1	20	9.5	7.6	-	0.721	0.687	0.049
Alor-2	24	12.0	8.7	-	0.736	0.666	0.096
Flores-1	23	11.2	8.2	-	0.717	0.699	0.026
Flores-2	25	11.2	8.2	-	0.758	0.709	0.066
Flores-3	19	11.8	9.2	-	0.768	0.698	0.093
Lomblen-1	12	7.7	7.5	-	0.750	0.720	0.042
Lomblen-2	20	10.0	7.8	-	0.735	0.661	0.103
Lomblen-3	15	8.6	7.6	-	0.743	0.683	0.082
Pantar-1	12	7.9	7.6	2	0.750	0.708	0.058
Pantar-2	17	10.3	8.6	1	0.749	0.737	0.017
Timor-1	17	10.3	8.6	-	0.743	0.723	0.029
Timor-2	16	10.8	9.1	-	0.776	0.713	0.085
Timor-3	20	9.3	7.8	-	0.769	0.713	0.075
Timor-4	22	11.3	8.5	-	0.743	0.639	0.144
Wetar-1	21	9.9	7.8	-	0.703	0.661	0.061
Wetar-2	14	8.7	7.9	-	0.719	0.684	0.051
Wetar-3	21	11.3	8.6	-	0.718	0.642	0.108
Mean	19	10.1	8.2		0.739	0.686	0.074

Table 3. Pairwise F_{ST} values between 19 *Eucalyptus urophylla* populations and the significance level of population differentiation after sequential Bonferroni correction corresponding to the nominal *P* value of 0.05* and 0.01** (above diagonal). Pairwise multilocus estimates of the effective number of migrants (*Nm,* Slatkin 1985, below diagonal).

	Adonara		Alor		Flores			Lomblen			Pantar		Timor				Wetar		
Population	Ado 1	Ado 2	Alor 1	Alor 2	Flor 1	Flor 2	Flor 3	Lom 1	Lom 2	Lom 3	Pan 1	Pan 2	Tim 1	Tim 2	Tim 3	Tim 4	Wet 1	Wet 2	Wet 3
Ado 1		0.014 NS	0.032 **	0.022 **	0.028 **	0.022 **	0.019 *	0.020 NS	0.024 **	0.031 *	0.039 **	0.028 **	0.024 **	0.027 **	0.048 **	0.039 **	0.044 **	0.051 **	0.035 **
Ado 2	3.18		0.034 **	0.035 **	0.035 **	0.029 **	0.033 **	0.040 **	0.043 **	0.051 **	0.042 **	0.032 **	0.030 **	0.040 **	0.064 **	0.044 **	0.062 **	0.061 **	0.051 **
Alor 1	2.92	2.47		0.013 **	0.031 **	0.028 **	0.019 **	0.030 **	0.019 **	0.036 **	0.032 **	0.024 **	0.020 **	0.028 **	0.048 **	0.032 **	0.054 **	0.054 **	0.039 **
Alor 2	2.92	2.43	3.73		0.019 **	0.019 **	0.008 NS	0.011 NS	0.018 **	0.027 *	0.035 *	0.026 **	0.006 NS	0.018 NS	0.031 **	0.026 *	0.020 **	0.015 NS	0.010 NS
Flor 1	2.66	2.51	2.68	3.54		0.014 *	0.012 NS	0.021 **	0.023 **	0.038 **	0.045 **	0.028 **	0.024 **	0.023 **	0.048 **	0.043 **	0.038 **	0.039 **	0.028 **
Flor 2	2.93	2.38	2.64	3.15	4.03		0.014 **	0.020 **	0.022 **	0.028 **	0.042 **	0.013 **	0.020 **	0.024 **	0.042 **	0.040 **	0.048 **	0.049 **	0.036 **
Flor 3	3.53	2.34	3.40	4.28	3.37	3.46		0.007 NS	0.017 *	0.026 **	0.038 **	0.020 **	0.015 NS	0.016 *	0.028 **	0.029 **	0.023 **	0.026 **	0.014 *
Lom 1	2.47	2.04	2.21	3.47	2.34	2.18	3.22		0.005 NS	-0.003 NS	0.028 *	0.023 **	0.013 NS	0.022 **	0.036 **	0.048 **	0.044 **	0.037 **	0.026 **
Lom 2	3.06	2.21	3.21	3.27	2.71	2.83	4.01	2.90		0.006 NS	0.028 **	0.029 **	0.015 **	0.025 **	0.052 **	0.041 **	0.039 **	0.038 **	0.036 **
Lom 3	2.60	1.92	2.70	3.62	2.20	2.61	3.28	2.68	2.67		0.042 **	0.039 **	0.014 NS	0.025 *	0.041 **	0.048 **	0.048 **	0.044 **	0.037 **
Pan 1	2.35	1.63	2.39	3.01	2.06	2.04	2.45	1.82	3.04	1.87		0.024 NS	0.034 **	0.039 **	0.066 **	0.057 **	0.063 **	0.066 **	0.052 **
Pan 2	2.69	2.24	2.61	2.44	2.73	3.25	3.26	2.43	2.29	2.61	2.81		0.034 **	0.038 **	0.055 **	0.052 **	0.050 **	0.055 **	0.044 **
Tim 1	2.47	2.68	2.72	4.06	2.88	2.53	2.93	2.75	3.42	3.34	2.47	2.12		0.004 NS	0.024 *	0.025 NS	0.020 **	0.023 **	0.014 NS
Tim 2	2.97	2.41	3.02	4.69	2.86	2.82	3.02	2.36	3.12	2.30	2.52	2.64	3.43		0.009 NS	0.008 NS	0.032 **	0.028 **	0.021 NS
Tim 3	1.95	1.53	1.90	2.80	2.07	1.94	2.58	1.23	1.75	1.74	1.30	1.79	2.57	2.99		0.031 **	0.046 **	0.039 **	0.039 **
Tim 4	2.14	2.26	2.51	4.28	2.27	2.01	2.39	1.73	2.65	2.48	1.63	2.06	3.02	3.27	2.31		0.045 **	0.046 **	0.039 **
Wet 1	2.54	1.82	2.33	3.71	2.15	2.34	3.09	1.77	2.78	2.25	1.69	1.95	3.68	2.69	1.87	2.96		0.010 NS	0.008 NS
Wet 2	1.55	1.94	2.12	3.59	2.41	2.19	2.80	1.62	2.30	1.67	1.72	1.57	2.79	2.65	1.60	2.43	3.04		0.005 NS
Wet 3	2.33	2.06	2.52	3.59	2.89	2.62	3.21	2.35	2.82	2.62	2.20	2.16	3.55	3.61	2.74	3.60	4.57	3.52	

Table 4. Analysis of molecular variance (AMOVA) based on 12 microsatellite loci and 19 populations of *Eucalyptus urophylla*. Significance levels were based on 1000 permutations (*** $P < 0.001$).

Source of variation	d.f.	Sum of Squares	Variance components	Percentage of variation
Among islands	6	96.152	0.088 ***	1.9
Among populations within islands	12	85.239	0.071 ***	1.5
Within populations	695	3097.536	4.457 ***	96.6
Total	713	3278.927	4.616	100

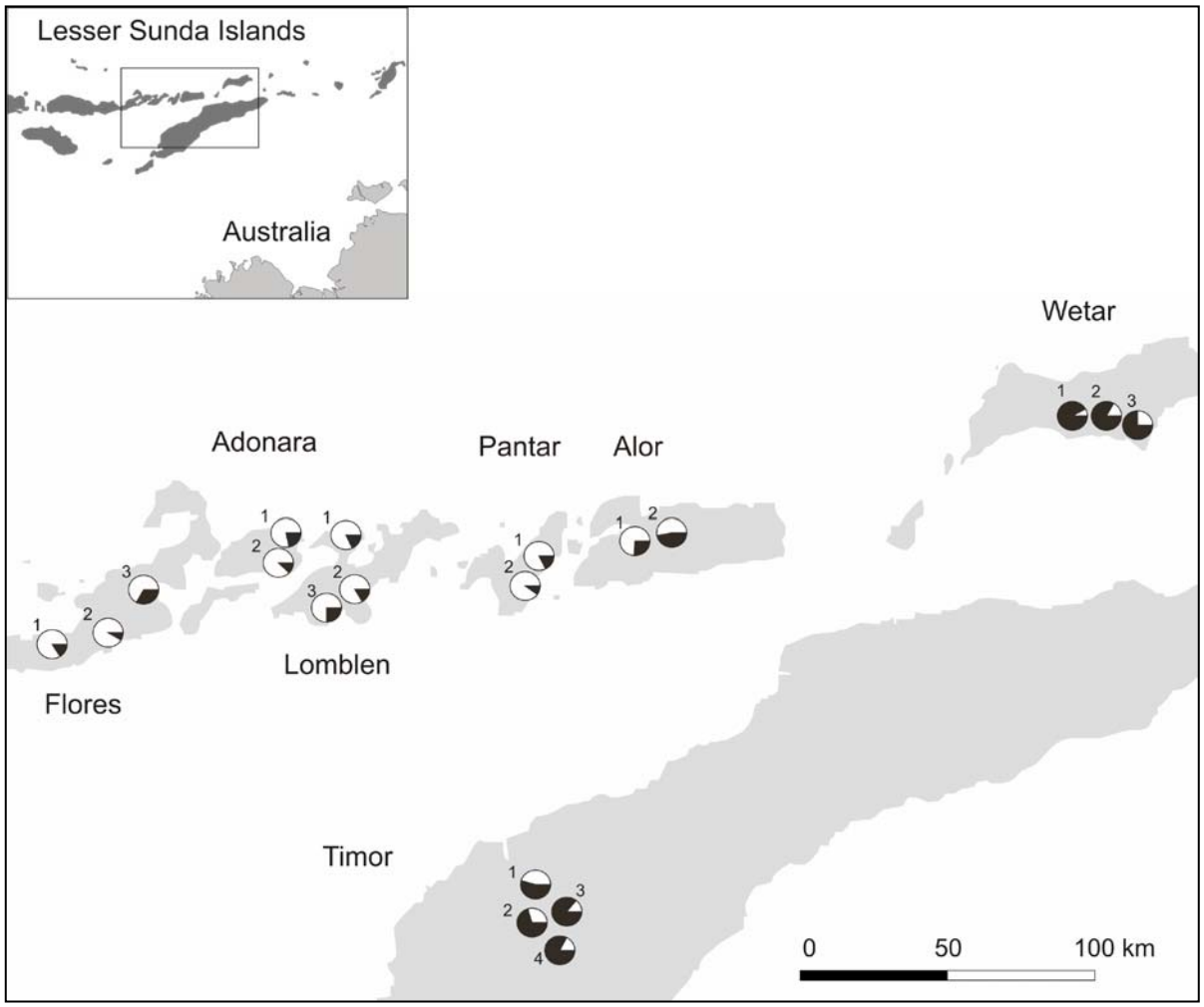


Figure 1. Geographic distribution of 19 *Eucalyptus urophylla* populations (Table 1) sampled in the Lesser Sunda islands, eastern Indonesia. Pie charts represent estimated proportion of membership in each of two population clusters according to a Bayesian cluster analysis (Pritchard et al. 2000). Population clusters are identified as cluster-1 (white) and cluster-2 (black).

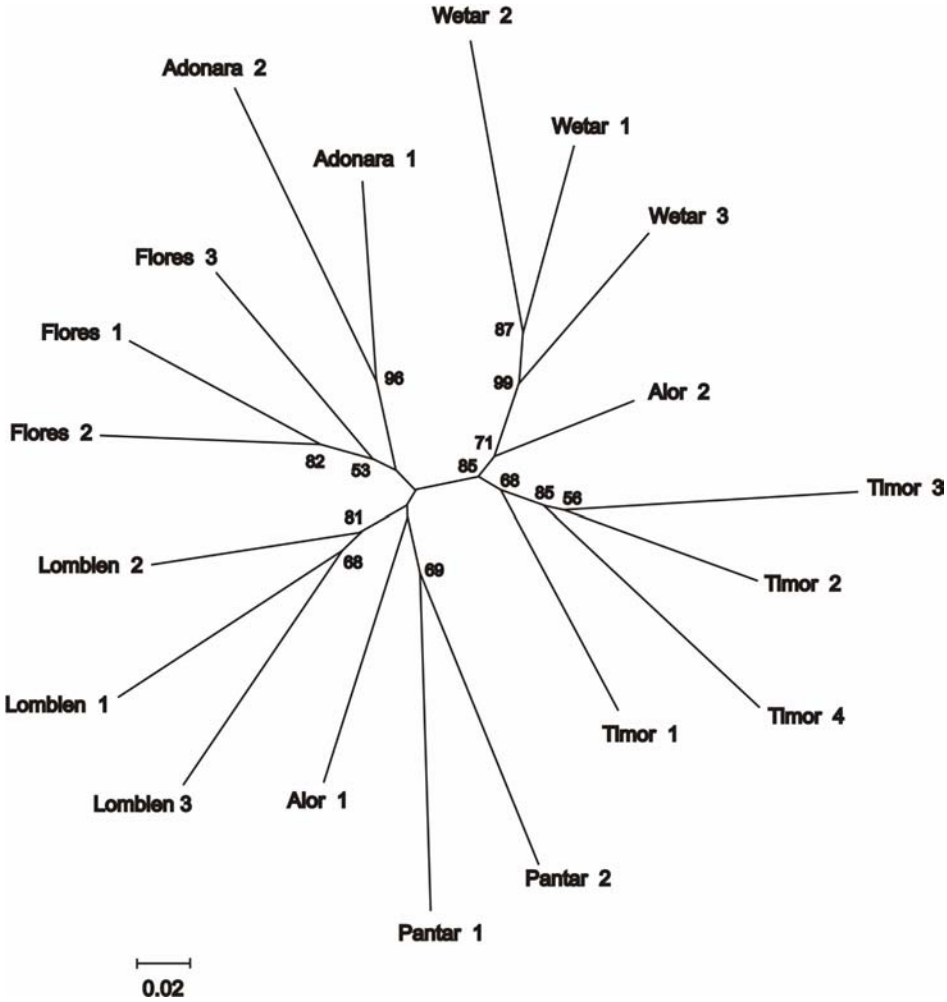


Figure 2. Unrooted neighbor-joining (NJ) phenogram based on Nei D_A (Nei et al. 1983) genetic distance for 19 *Eucalyptus urophylla* populations. Numbers indicate nodes with bootstrap support of more than 50% for 10 000 replications.

SUPPLEMENTARY MATERIAL

Table S1 Flanking primers, repeat motif and allele size range of three unpublished microsatellites.

Locus	Forward and reverse primers	Repeat motif	Allele size range (bp)
FMG-EUCSSR1	CCAGCCCAACAGAACCGTTT GCGGTGGATCAGGGATTCAT	$(CT)_{10}$	188-236
FMG-EUCSSR3	ATGTGCGTGCGAGTGCGTTG AGCTCGTTGCGGCTGTGAGA	$(CCT)_6$	220-235
FMG-EUCSSR5	ATCCTCCATTCTCCTCTC GTCTGGACCGTGATCTACAGAA	$(GA)_{19}$	334-369

Table S2 Genetic parameters for 12 polymorphic microsatellite loci in *Eucalyptus urophylla*. Observed number of alleles per locus (A), observed heterozygosity (H_O), expected heterozygosity (H_E) and inbreeding coefficients (F_{IS}, F_{IT} and F_{ST}).

Locus	N	A	H_E	H_O	F_{IT}	F_{IS}	F_{ST}
FMG-EUCSSR1	355	22	0.566	0.510	0.105	0.057	0.051
FMG-EUCSSR3	355	6	0.223	0.192	0.148	0.105	0.048
FMRSA4	353	11	0.331	0.297	0.104	0.077	0.030
FMG-EUCSSR5	350	18	0.858	0.694	0.192	0.165	0.032
EMBRA3	348	41	0.907	0.862	0.052	0.015	0.038
EMBRA28	356	32	0.925	0.820	0.114	0.093	0.023
EMBRA37	356	28	0.926	0.854	0.079	0.054	0.027
EMBRA48	344	56	0.947	0.919	0.032	0.000	0.032
EMBRA69	349	23	0.907	0.811	0.107	0.076	0.034
EMBRA125	346	21	0.924	0.749	0.191	0.176	0.017
EMBRA219	354	27	0.908	0.811	0.109	0.079	0.032
EMBRA227	353	17	0.719	0.683	0.053	0.026	0.027
Mean	351.6	25.2	0.762	0.683	0.104	0.076	0.031

CHAPTER 4

Diversity and population genetic structure of SNP haplotypes in two wood formation genes of *Eucalyptus urophylla*

Kitt G. Payn[a,c,d], Mathabatha F. Maleka[b], William S. Dvorak[c], Bernard J.H. Janse[a] and Alexander A. Myburg[d]

[a]*Mondi South Africa, P.O. Box 31024, Merebank, 4059, South Africa*

[b]*Department of Plant Sciences, University of the Free State, Bloemfontein, 9300, South Africa*

[c]*Camcore, North Carolina State University, P.O. Box 8008, Raleigh, NC 27695, USA*

[d]*Forest Molecular Genetics Programme, Department of Genetics, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, 0002, South Africa*

This chapter consists of a SNP genotyping study that will be combined with a SNP discovery study (Maleka et al. 2007) prior to submission to a peer-reviewed journal. I conducted the laboratory work, data analyses, and wrote the chapter. Alexander Myburg supervised the research project and provided suggestions on the organization and content of the chapter. Mathabatha Maleka performed the SNP discovery work that formed the basis of this study. Members of my PhD. advisory committee provided critical comments on the chapter.

ABSTRACT

Eucalyptus urophylla is a commercially important plantation species that is native to seven islands in eastern Indonesia. In order to gain insight into the diversity and population structure of genes associated with economically important traits, we investigated the geographical distribution of SNP haplotypes in two wood formation genes (*sucrose synthase1*, *EuSuSy1* and *cinnamyl alcohol dehydrogenase2*, *EuCAD2*). Three haplotype-tag SNPs at the 5' and 3' ends of each gene were genotyped using a polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) approach. High levels of SNP haplotype diversity were observed throughout 19 geographically defined *E. urophylla* natural populations (H_d = 0.58 to 0.76). Low levels of population genetic structure were observed for the SNP haplotypes of each gene ($F_{ST} \leq 0.08$). At the individual SNP level, the degree of population differentiation estimated for the A/T SNP (pos 415) of the *EuSuSy1* 5' region was more than twice the level of population differentiation estimated for the remaining SNPs. These findings, together with previous data on chloroplast haplotype diversity and nuclear microsatellite studies from the same *E. urophylla* samples provide improved direction for conservation and breeding strategies for this species.

INTRODUCTION

Eucalyptus urophylla is a commercially important plantation tree species that is often crossed with *E. grandis* to produce hybrid progeny displaying rapid growth and superior disease resistance. It is native to seven islands in eastern Indonesia with the largest stands located on

the islands of Timor and Wetar, and smaller stands occurring on the islands of Adonara, Alor, Flores, Lomblen (Lembata) and Pantar (Fig. 1). Seed collections on these islands were carried out by several international organizations over the last 30 years to provide a broad genetic resource necessary for effective breeding and conservation programs (Martin and Cossalter 1975; Gunn and McDonald 1991; Hodge et al. 2001; Pepe et al. 2004). Management of the genetic resource will benefit from a greater understanding of the nature and distribution of genetic diversity within the species. This may be achieved using molecular marker techniques, which provide tools for the characterization of genetic diversity and population genetic structure.

Genetic diversity in plants resides in the nuclear, mitochondrial and chloroplast genomes. A recent molecular study based on chloroplast genetic diversity investigated the island colonisation route of *E. urophylla* (Payn et al. 2007). Historical events such as range expansion and genetic bottlenecks can leave an imprint on contemporary levels of genetic diversity. Sequence variation of the maternally inherited chloroplast genome was used to infer historical seed migration patterns along the island archipelago. A high level of chloroplast genetic differentiation (G_{ST} = 0.58) and significant phylogeographic structure were detected. The majority of chloroplast genetic diversity was observed on the islands of Wetar and Timor, with a decreasing trend moving further west, consistent with an east-to-west island colonisation pattern. Biparentally inherited microsatellite markers were further used to infer contemporary genome-wide levels of nuclear genetic diversity and population structure of *E. urophylla* (Tripiana et al. 2007; Payn et al. 2008). These studies revealed a

high level of nuclear genetic diversity across the species range and a low level of population structure (F_{ST} = 0.04, Tripiana et al. 2007; F_{ST} = 0.03, Payn et al. 2008) compared to the chloroplast data. This was largely explained by the nuclear genome migrating through pollen as well as seed, with sufficient pollen flow among populations to limit population divergence through genetic drift (Payn et al. 2008).

Of further interest is the level of genetic diversity and population structure of genes putatively controlling economically important and/or adaptive traits. The comparative sequencing of candidate gene regions across multiple individuals may result in the discovery of single nucleotide polymorphisms (SNPs). These are naturally occurring alleles found at a single nucleotide site. SNP alleles may occur in different combinations along a gene region to form SNP haplotypes. The continued existence of different SNP haplotypes largely depends on the presence of linkage disequilibrium (LD), which is the non-random association of alleles at different sites. The rate of decay of LD, and hence the reshuffling of SNP haplotypes, is determined by the interplay of several factors, including mutation and recombination rates, selection, population structure and demographic history (Flint-Garcia et al. 2003; Gupta et al. 2005).

Estimates of nucleotide diversity for genes in forest trees are becoming more widely reported. There are typically large differences in estimates of nucleotide diversity between different genes and gene regions in different species (Poke et al. 2003; Ingvarsson 2005; Krutovsky and Neale 2005; Gonzalez-Martinez et al. 2006). This highlights the need to

specifically investigate the different genes of interest in the species being studied, and suggests that broader whole genome studies may be required to understand the basis of adaptation or performance. In addition, the diversity at a gene region may be geographically structured. Ingvarsson (2005) detected significant population genetic differentiation for five genes surveyed in *Populus tremula*, with a substantial range of differentiation across the different genes (F_{ST} range: 0.04 - 0.16). There was an even greater range of population genetic differentiation for six wood formation genes in *Pinus pinaster* (F_{ST} range: -0.055 to 0.453, Pot et al. 2005), of which two were significant (*KORRIGAN* and *Pp1*). The markedly higher level of differentiation for these two genes, relative to the other genes and previously reported neutral markers (Mariette et al. 2001), was ascribed to diversifying selection (Pot et al. 2005).

The level of LD along a gene sequence varies greatly among plant species, ranging from 200 to1500 bp in maize and up to 50 to100 kb in *Arabidopsis* (Rafalski and Morgante 2004). LD estimates in out-crossing forest tree species are generally at the lower end of the observed range and decay within the length of a gene (Brown et al. 2004; Ingvarsson 2005; Krutovsky and Neale 2005; Thumma et al. 2005; Gonzalez-Martinez et al. 2006). The presence of LD underlies association genetics, a population-based approach used to identify relationships between allelic forms of a gene and a phenotype. If LD extends across a large region, for example over 50 kb, as it does in humans (Reich et al. 2001), then genome-wide association genetic studies are feasible. However, if LD declines within several kilobases, the number of markers required for genome-wide analyses becomes very large and then a candidate gene-

based association genetics approach is recommended (Neale and Savolainen 2004). The selection of candidate genes is typically based on prior knowledge of biochemical pathways, further supported by the interrogation of EST and microarray databases.

The level of nucleotide diversity and LD was surveyed in two *E. urophylla* wood formation genes (*sucrose synthase1*, *EuSuSy1* and *cinnamyl alcohol dehydrogenase2*, *EuCAD2*) (Maleka et al. 2007). *Eucalyptus* wood is widely used for pulp and paper production. High cellulose content is desirable in pulpwood because there is a positive correlation between cellulose content and pulp yield. Conversely, lignin is undesirable in pulpwood because it causes the discoloration of paper and has to be removed during the pulping process at great expense. Sucrose synthase is an enzyme that catalyzes the reversible conversion of sucrose into fructose and UDP-glucose. The latter is an important substrate in the cellulose biosynthesis process. Cinnamyl alcohol dehydrogenase is an enzyme in the lignin biosynthesis pathway, directly responsible for the reduction of the monolignol aldehydes to their alcohol counterparts. Diversity within these gene regions may potentially affect the biosynthetic pathways and ultimately the phenotype (e.g. *cad-nl* mutant allele, MacKay et al. 1997). Therefore, understanding the diversity and population structure of wood formation genes such as *EuSuSy1* and *EuCAD2* may be important for the future genetic improvement of trees planted for the production of pulp and paper.

Maleka et al. (2007) sequenced the 5' and 3' ends (approx. 1 kb from each end) of *EuSuSy1* and *EuCAD2* in an *E. urophylla* SNP discovery panel comprising 25 individuals broadly

representing the geographical distribution of the species (Supplementary Table S1.). A high level of nucleotide diversity per site for *EuSuSy1* ($\pi = 0.0175$) and *EuCAD2* ($\pi = 0.0134$) was detected. LD declined to negligible levels ($r^2 < 0.20$) within approximately 600 bp in *EuSuSy1*, but remained present ($r^2 > 0.20$) along the length of *EuCAD2* (~3200 bp). We extended the SNP discovery study of Maleka et al. (2007) and investigated the species-wide geographic distribution of SNP haplotype diversity and the level of population structure of these SNP haplotypes. The SNP discovery panel was increased to form a SNP genotyping sample of 343 individuals, representing 19 geographically defined populations distributed across much of the native range of *E. urophylla.* Subsets of SNPs discovered by Maleka et al. (2007) were selected and genotyped at the 5' and 3' ends of *EuSuSy1* and *EuCAD2*, and their SNP haplotypes were statistically inferred.

This study was undertaken with the following objectives: (1) quantify the haplotype diversity across the natural range of *E. urophylla*, (2) determine whether the haplotypes are geographically structured and (3) determine whether the distribution of haplotype diversity and population structure of the gene regions mirror patterns of genome-wide genetic diversity previously reported for the same populations using microsatellite markers (Payn et al. 2008). This information may be useful for designing future studies that aim to associate alleles with phenotypic traits in *E. urophylla*.

MATERIALS AND METHODS

Sampling and DNA extraction

Eucalyptus urophylla seed collections were conducted by Camcore (North Carolina State University, Raleigh, NC, USA), an international tree conservation and domestication program, in collaboration with PT Sumalindo Lestari Jaya, a private Indonesian forestry company (Pepe et al. 2004). Seeds were obtained from 1104 mother trees distributed across 62 sites representing the natural distribution of *E. urophylla*, barring East Timor, due to political unrest at that time. A subset of 45 collection sites was included in this study. Seed collection sites were grouped according to geographic location to form 19 populations with a sample size $N \geq 12$ (Fig. 1; Supplementary Table S2). Seeds were sown in a commercial nursery in South Africa (Mondi, South Africa). Leaf tissue was sampled from seedlings, each seedling originating from a different mother tree. Total genomic DNA from leaf tissue was extracted as described in Payn et al. (2008).

SNP Genotype Determination

Nested PCR-RFLP assays were used to genotype three SNPs located in the 5' and 3' regions of *EuSuSy1* and *EuCAD2* (Table 1, Fig. 2). The first step was the amplification of the 5' and 3' end fragments of each gene from genomic DNA using the gene-specific primers reported in Maleka et al. (2007) (Supplementary Table S3). PCR amplification reactions were performed in 20 μL volumes containing 5 ng of genomic DNA, 0.8 U of Exsel polymerase (Southern Cross), 1× PCR Exsel buffer, 0.2 mM dNTPs and 0.4 μM of each primer. PCR

amplifications were performed with an iCycler (BIO-RAD Laboratories, Hercules, CA, USA) with the following cycling conditions: an initial denaturation step of 94°C for 1 min, followed by 30 cycles of 94°C for 30 s, 60°C for 30 s and 72°C for 60 s, followed by an elongation step at 72°C for 10 min. The PCR product was diluted 1:100 and used as a template for the subsequent nested PCR reactions.

The nested PCR amplification reactions were performed in 20 μL volumes containing 5 μL of the diluted PCR product, 0.8 U of Exsel polymerase, 1× PCR Exsel buffer, 0.2 mM dNTPs and 0.4 μM of each primer (Supplementary Table S4). PCR amplifications were performed with the following cycling conditions: an initial denaturation step of 94°C for 1 min, followed by 25 cycles of 94°C for 30 s, 56°C for 30 s and 72°C for 30 s, and a final elongation step of 72°C for 10 min.

The PCR products were digested with restriction endonucleases listed in Table 1. Enzymes were selected based on the enzyme recognition site overlapping with a SNP, preferably a haplotype-tag SNP (htSNP) identified using the program HaploBlockFinder v0.7 (Zhang and Jin 2003) at default settings and a threshold of $r^2 = 0.2$ to define LD blocks. Some of the genotyped SNPs were located in the non-coding regions (5' UTRs, introns), while others were located in exons (Fig. 2). SNPs genotyped in the coding regions were all classified as synonymous substitutions. Restriction digests were performed separately for each enzyme. An aliquot of 3 μL of PCR product was digested overnight in a total reaction volume of 20

μL with 2 units of enzyme and the prescribed restriction enzyme buffer. The digestion products were analysed on a 2% agarose gel to infer each SNP genotype (Supplementary Fig. S1 and Fig. S2). Fragment sizes were verified against the sequences reported by Maleka et al. (2007).

Data Analyses

Haplotypes at the 5' and 3' ends of the *EuSuSy1* and *EuCAD2* genes were inferred from the individual SNP genotypes using the statistical software package PHASE v2.1 (Stephens et al. 2001; Stephens and Scheet 2005). This program was selected because its accuracy was found to be higher than other methods used for inferring and reconstructing haplotypes (Adkins 2004). Haplotype diversity, defined as the probability that two randomly chosen haplotypes are different in a population sample (Nei, 1987), was estimated with the ARLEQUIN v2.0 software (Schneider et al. 2000). Mean haplotype diversity was compared among groups of populations (populations grouped by island) by an analysis of variance implemented in JMP v4.04 (SAS Institute, Cary, USA).

Coefficients of population differentiation, G_{ST} and $N_{ST,}$ were estimated using the HAPSTEP program (version 2001, Pons and Petit 1996). The G_{ST} coefficient depends on the frequencies of the haplotypes, whereas N_{ST} is influenced by haplotype frequencies and the genetic distance between haplotypes. A permutation test with 1000 permutations was performed in HAPSTEP to determine whether N_{ST} differed significantly from G_{ST}. A higher N_{ST} than G_{ST} indicates the presence of phylogeographic structure (Pons and Petit 1996), with closely

related haplotypes being found more often in the same geographic area than less closely related haplotypes. F_{ST} values for each gene region (based on genetic distance between haplotypes) and each individual SNP were estimated by an analysis of molecular variance (AMOVA, Excoffier et al. 1992), as implemented in ARLEQUIN v2.0. An AMOVA was also used to apportion genetic variation within populations, among populations within islands, and among islands.

RESULTS

SNP frequencies

The 24 samples comprising the SNP discovery panel (Maleka et al. 2007) were included in the present SNP genotyping study. The SNP genotyping data of the discovery panel based on restriction enzyme analyses accurately matched the sequence data determined by Maleka et al. (2007). The minor allele frequency (MAF) across all SNPs ranged from 0.14 (position 213 of *EuSuSy1* 3' region) to 0.49 (position 593 of *EuCAD2* 3' region) in the 343 individuals (Supplementary Table S5).

Haplotype diversity

The number of inferred haplotypes and haplotype diversity for each population are listed in Table 2. The 5' and 3' ends of *EuSuSy1* had a total of five and six haplotypes, respectively. The 5' and 3' ends of *EuCAD2* had a total of eight and seven haplotypes, respectively. The 5' end of *EuCAD2* had the highest total haplotype diversity of 0.76 and the 5' end of *EuSuSy1*

had the lowest total haplotype diversity of 0.58. There was a relatively wide range in haplotype diversity at the population level for the 5' end of *EuSuSy1* compared to the other gene regions, with the highest diversity of 0.74 observed for the Timor-4 population and the lowest diversity of 0.30 observed for the Wetar-1 population. A test for differences in mean haplotype diversity among groups of populations (populations grouped by island) was significant ($\alpha = 0.05$) for the 5' end of *EuSuSy1*. No statistically significant results were observed for the other gene regions.

Population structure

A low degree of genetic differentiation among the 19 *E. urophylla* populations was observed for all gene regions (Table 3). The parameter N_{ST}, which takes into consideration the genetic distances between haplotypes, was not significantly greater than G_{ST} ($\alpha = 0.05$), thus indicating a lack of phylogeographic structure (Pons and Petit, 1996). Levels of population differentiation were also estimated for each SNP to determine the locus effect on the observed population haplotype structure. The highest estimate of population differentiation was $F_{ST} = 0.148$ for the A/T SNP (pos 415) of the *EuSuSy1* 5' region. The remaining SNPs had substantially lower levels of population differentiation ($F_{ST} \leq 0.069$, Supplementary Table S6).

For each gene region, an analysis of molecular variance revealed that genetic variation among individuals within populations accounted for the majority of the total molecular

variance (92.11 - 98.08%, Table 4). The remainder of the total molecular variance was apportioned among populations within islands (0.41 - 2.25%), and among islands (0.18 - 7.48%).

DISCUSSION

The aim of this research was to determine the level of diversity and population structure of SNP haplotypes in two wood formation genes of *E. urophylla*, and to compare these data to previous *E. urophylla* molecular marker studies based on the same samples. The protocol used to discover SNPs for such a population study may potentially introduce ascertainment bias. This bias may arise from the selection of SNP loci from an unrepresentative sample of individuals, or if only highly heterozygous (variable) loci are selected for further analyses (Brumfield et al. 2003). In an effort to reduce ascertainment bias, the SNPs used in this study were identified in a panel of individuals broadly representative of the geographic distribution of *E. urophylla* (Maleka et al. 2007). In addition, an effort was made to select SNPs with a range in heterozygosity with respect to each gene region. This is because highly heterozygous SNPs are expected to be older than SNPs with low heterozygosity. Older polymorphisms are likely to have had more time to be distributed among the population by gene flow and may lead to an underestimation of the magnitude of population genetic structure (Morin et al. 2004).

Haplotype diversity estimates for each gene region (Table 2) were evenly distributed across the species range with no clear pattern with respect to diversity and geographic location. This was in marked contrast to the distribution of chloroplast haplotype diversity, where high diversity in the east and limited diversity in the west was observed (Payn et al. 2007). Rather, the geographic distribution of haplotype diversity for each gene region was similar to the distribution of genome-wide diversity estimates based on microsatellite markers (Payn et al. 2008).

No haplotype diversity estimates based on SNP genotyping along a nuclear gene region in forest trees are available in the literature. However, there are estimates of haplotype diversity based on gene sequence data and some are similar to the diversity estimates reported in this study (mean $H_d = 0.83$ in *Pseudotsuga menziesii*, Krutovsky and Neale 2005; mean $H_d = 0.68$ in *Pinus taeda* Gonzalez-Martinez et al. 2006). The comparison of gene haplotype diversity among different studies is difficult due to different genes, different size fragments, and different genotyping techniques used in these studies. In particular, the number of variable sites generally influences haplotype diversity and so sequence data containing many SNP loci may yield higher haplotype diversity estimates than a subset of SNPs used to genotype the same region. This was evident by the high haplotype diversity estimates of *EuSuSy1* (H_d = 1.0) and *EuCAD2* (H_d = 1.0) based on the DNA sequence data of the SNP discovery panel (Maleka et al. 2007).

The degree of population differentiation estimated for the four gene regions was low (Table 3), with >90% of the genetic diversity apportioned within populations rather than among populations or among islands (Table 4). In addition, the test for phylogeographic structure was not significant. These findings are in contrast to the population genetic structure of the chloroplast genome of *E. urophylla* (G_{ST} = 0.5, Payn et al. 2007). The maternally inherited chloroplast genome is haploid and so the lower effective population size results in an increase in genetic differentiation due to genetic drift. The population differentiation estimates observed in this present study were more similar to the estimates obtained using nuclear microsatellite markers assayed across the same populations (F_{ST} = 0.03, Payn et al. 2008). Low levels of population differentiation observed in the nuclear genome are likely the result of pollen flow among populations by means of insect and/or bird vectors ($Nm \geq 1.2$, Payn et al. 2008).

At the individual SNP level, the degree of population differentiation estimated for the A/T SNP (pos 415) of the *EuSuSy1* 5' region was more than twice the level of population differentiation estimated for the remaining SNPs and for the neutral microsatellite markers (Payn et al. 2008). Of particular interest was the high allele frequency of A (pos 415) in the populations from the island of Timor (freq = 0.5, averaged across Timor populations), in comparison to the other populations (freq = 0.1, averaged across other populations; Supplementary Table S5). The geographical structure of this SNP may be the result of random gene flow and/or genetic drift. However, the deviation of a population differentiation

estimate for one locus relative to other loci may also be a signature of selection, as proposed by Pot et al. (2005) for *KORRIGAN* and *Pp1* in *P. pinaster*.

The application of SNP genotyping across a natural population extends beyond the investigation of haplotype diversity and phylogeographic structure for genes of interest. SNPs are the molecular marker of choice in association genetics studies (Rafalski 2002). The presence of LD along the length of a gene or a gene region makes it possible to detect association between genetic and phenotypic variation. In the first association genetics study in plants, Thornsberry et al. (2001) found an association between *Dwarf8* (*d8*) polymorphisms and variation in flowering time in maize. Other marker-trait association studies in plants have since been reported (e.g. Olsen et al. 2004; de Meaux et al. 2005; Szalma et al. 2005; Thumma et al. 2005). The identification of htSNPs, SNPs that capture most of the haplotypic variation, is a key step in association genetics. By selectively genotyping only htSNPs, one will save costs by genotyping a subset of SNPs capable of distinguishing the most common haplotypes (Gonzalez-Martinez et al. 2006).

Of the 81 SNPs reported by Maleka (2007), 51 were identified as htSNPs using the program HaploBlockFinder (Fig. 3). In this study, 10 of the 12 SNPs genotyped were htSNPs. The T/C (pos 527) and A/C (pos 689) SNPs of the *EuCAD2* 3' region were not identified as htSNPs. These two SNPs were redundant with respect to the most common haplotypes (common haplotypes are those representing more than 80% frequency of chromosomes

sampled in the discovery panel) of their haplotype block. By increasing the criteria to 100% frequency, the two SNPs were no longer considered redundant (data not shown).

In the event of a future association genetics study in *E. urophylla* based on candidate genes, including *EuSuSy1* and *EuCAD2*, it would be prudent to genotype a much larger percentage of htSNPs with respect to each gene region. This would certainly be required for genes that comprise many short LD blocks, within which common variants are in strong LD but among which recombination has left little LD. Indeed, genotyping strategies based on htSNPs in outcrossing forest trees would likely produce only moderate reductions in genotyping effort (e.g. ~35% reduction in genotyping effort in *Pinus taeda*, Gonzalez-Martinez et al. 2006). Consequently, larger SNP genotyping studies now use high-throughput SNP genotyping technology (eg. the GoldenGate genotyping assay, Fan et al. 2003; Shen et al. 2005).

The use of high-throughput SNP genotyping technology could provide a more in-depth analysis of haplotype diversity and structure in *E. urophylla* candidate genes. The results could then be compared to the results of this present study, which suggest that haplotype diversity in *EuSuSy1* and *EuCAD2* is evenly distributed across the species range and limited population genetic structure is present. In this context, only a subset of provenances would need to be included in a conservation or base breeding population without compromising the breeder's ability to select for improved wood properties. These findings agree with those based on microsatellite markers genotyped across the same populations (Payn et al. 2008). However, it is possible that some loci in these and in other gene regions are subject to

adaptive selection. This would potentially result in the genetic variation at such loci being geographically structured across the *E. urophylla* natural range. Evidence of this would warrant a more structured genetic resource population for conservation and breeding purposes.

LITERATURE CITED

Adkins RM (2004) Comparison of the accuracy of methods of computational haplotype inference using a large empirical dataset. *BMC Genetics* **5**, 22–29.

Brown GR, Gill GP, Kuntz RJ, Langley CH, Neale DB (2004) Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proceedings of the National Academy of Science USA* **101**, 15255–15260.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology and Evolution* **18**, 249–256.

de Meaux J, Goebel U, Pop A, Mitchell-Olds T (2005) Allele-specific assay reveals functional variation in the *chalcone synthase* promoter of *Arabidopsis thaliana* that is compatible with neutral evolution. *Plant Cell* **17**, 676–690.

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.

Fan JB, Oliphant A, Shen R, Kermani BG, Garcia F, Gunderson KL, Hansen M, Steemers F, Butler SL, Deloukas P, Galver L, Hunt S, McBride C, Bibikova M, Rubano T, Chen J, Wickham E, Doucet D, Chang W, Campbell D, Zhang B, Kruglyak S, Bentley D, Haas J, Rigault P, Zhou L, Stuelpnagel J, Chee MS (2003). Highly parallel SNP genotyping. *Cold Spring Harbor Symposium on Quantitative Biology* **68**, 69–78.

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. *Annual Review of Plant Biology* **54,** 357–374.

Gonzalez-Martinez SC, Ersoz E, Brown GR, Wheeler NC, Neale DB (2006) DNA sequence variation and selection of tag single-nucleotide polymorphisms at candidate genes for drought-stress response in *Pinus taeda* L. *Genetics* **172**, 1915–1926.

Gunn BV, McDonald MW (1991) *Eucalyptus urophylla* seed collections. *Forest Genetic Resources* **19**, 34–37, FAO, Rome.

Gupta PK, Rustgi S, Kulwal PL (2005) Linkage disequilibrium and association studies in higher plants: Present status and future prospects. *Plant Molecular Biology* **57**, 461–485.

Hodge GR, Pepe B, Wijoyo FS, Dvorak WS (2001) Early results of *Eucalyptus urophylla* provenance/progeny trials in Colombia and Venezuela. In: Developing the Eucalypts of the Future. Proc. IUFRO Working Party 2.08.03.

Ingvarsson PK (2005) Nucleotide Polymorphism and Linkage Disequilibrium Within and Among Natural Populations of European Aspen (*Populus tremula* L., Salicaceae). *Genetics* **169**, 945–953.

Krutovsky KV, Neale DB (2005) Nucleotide diversity and linkage disequilibrium in cold-hardiness-and wood quality-related candidate genes in Douglas fir. *Genetics* **171,** 2029–2041.

Kumar S, Tamura K, Nei M (1993) MEGA version 10: Molecular Evolutionary Genetic Analysis. Pennsylvania State University, Pennsylvania.

MacKay JJ, O'Malley DM, Presnell T, Booker FL, Campbell MM, Whetten RW, Sederoff RR (1997) Inheritance, gene expression, and lignin characterization in a mutant pine deficient in cinnamyl alcohol dehydrogenase. *Proceedings of the National Academy of Science USA* **94**, 8255–8260.

Maleka MF (2007) Allelic diversity in cellulose and lignin biosynthetic genes of *Eucalyptus urophylla* S.T. Blake. M.Sc. thesis, University of Pretoria.

Mariette S, Chagné D, Lézier C, Pastuszka P, Raffin A, Plomion C, Kremer A (2001) Genetic diversity within and among *Pinus pinaster* populations: comparison between AFLP and microsatellite markers. *Heredity* **86**, 469–479.

Martin B, Cossalter C (1975) Les eucalyptus des iles de la Sonde. *Bois et Foréts des Tropiques* **163**, 3–25.

Morin PA, Luikart G, Wayne RK, Allendorf FW, Aquadro CF, Axelsson T, Beaumont M, Chambers K, Durstewitz G, Mitchell-Olds T, Palsboll PJ, Pionar H, Przeworski M, Taylor B,

Nei M, Tajima F, Tateno Y (1983) Accuracy of genetic distances and phylogenetic trees from molecular data. *Journal of Molecular Evolution* **19**, 153–170.

Neale DB, Savolainen O (2004) Association genetics of complex traits in conifers. *Trends in Plant Science* **9**, 325–330.

Nei M (1987) 'Molecular evolutionary genetics.' (Columbia University Press: New York).
Olsen KM, Halldorsdottir SS, Stinchcombe JR, Weinig C, Schmitt J, Purugganan MD (2004) Linkage disequilibrium mapping of Arabidopsis *CRY2* flowering time alleles. *Genetics* **167**, 1361–1369.

Payn KG, Dvorak WS, Myburg AA (2007) Chloroplast DNA phylogeography reveals the island colonisation route of *Eucalyptus urophylla* (Myrtaceae). *Australian Journal of Botany* **55**, 673–683.

Payn KG, Dvorak WS, Janse BJH, Myburg AA (2008) Microsatellite diversity and genetic structure of the commercially important tropical tree species *Eucalyptus urophylla*, endemic to seven islands in Indonesia. *Tree Genetics and Genomes* **4**, 519– 530.

Pepe B, Surata K, Suhartono F, Sipayung M, Purwanto A, Dvorak WS (2004) Conservation status of natural populations of *Eucalyptus urophylla* in Indonesia and international efforts to protect dwindling gene pools. *Forest Genetic Resources* **31**, 62–64, FAO Rome.

Poke FS, Vaillancourt RE, Elliott RC, Reid JB (2003) Sequence variation in two lignin biosynthesis genes, cinnamoyl CoA reductase (*CCR*) and cinnamyl alcohol dehydrogenase 2 (*CAD2*). *Molecular Breeding* **12**, 107–118.

Pons O, Petit RJ (1996) Measuring and testing genetic differentiation with ordered vs. unodered alleles. *Genetics* **144**, 1237–1245.

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. *Current Opinion in Plant Biology* **5**, 94–100.

Rafalski A, Morgante M (2004) Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. *Trends in Genetics* **20**, 103–111.

Schneider S, Roessli D, Excoffier L (2000) Arlequin ver 2000: A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Shen R, Fan JB, Campbell D, Chang W, Chen J, Doucet D, Yeakley J, Bibikova M, Wickham Garcia E, McBride C, Steemers F, Garcia F, Kermani B G, Gunderson K, Oliphant

A (2005) High-throughput SNP genotyping on universal bead arrays. *Mutation Research* **573**, 70–82.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* **68**, 978–989.

Stephens M, Scheet P (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing data imputation. *American Journal of Human Genetics* **76**, 449–462.

Szalma SJ, Buckler ES, Snook ME, McMullen MD (2005) Association analysis of candidate genes for maysin and chlorogenic acid accumulation in maize silk. *Theoretical and Applied Genetics* **110**, 1324–1333.

Thumma BR, Nolan MF, Evans R, Moran GF (2005) Polymorphisms in *cinnamoyl CoA reductase* (*CCR*) are associated with variation in microfibril angle in *Eucalyptus* spp. *Genetics* **171**, 1257–1265.

Tripiana V, Bourgeois M, Verhaegen D, Vigneron P, Bouvet JM (2007) Combining microsatellites, growth, and adaptive traits for managing in situ genetic resources of *Eucalyptus urophylla*. *Canadian Journal of Forest Research* **37**, 773–785.

Wakeley J (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution* **19**, 208–216.

Zhang K, Jin L (2003) HaploBlockFinder: haplotype block analyses. *Bioinformatics* **19**, 1300–1301.

Table 1. Design of PCR-RFLP assays for SNP genotyping.

Gene fragment	SNP[a] (Pos)	Region	Enzyme[b]	Diagnostic frag. profile (bp)	
				Un-cut	Cut
SuSy1 5'	C/T (168)	Upstream	MwoI (GCN$_7$G**C**)	478	115, 363
	A/T (415)	Intron 1	AluI (**A**GCT)	478	367, 111
	T/C (609)	Intron 2	AclI (AA**C**GTT)	772	559, 213
SuSy1 3'	A/G (156)	Exon 12	PvuII (C**A**GCTG)	422	124, 298
	C/T (213)	Exon 12	MvaI (**C**CWGG)	270	181, 89
	C/T (496)	Exon 13	Fnu4HI (G**C**NGC)	415	218, 197
CAD2 5'	C/T (196)	Promoter	MspI (C**C**GG)	286	177, 109
	T/C (599)	Intron 1	MboII (N$_8$**T**CTTC)	295	99, 196
	C/T (931)	Intron 2	Eam1104I (**C**TCTTCN)	299	171, 128
CAD2 3'	T/C (527)	Exon 5	AatII (GACGT**C**)	412	150, 262
	T/C (593)	3' UTR	Alw26I (GT**C**TCN)	412	221, 191
	A/C (689)	3' UTR	Bsu15I (**A**TCGAT)	412	315, 97

[a]First nucleotide listed for each SNP has the major allele frequency. SNP positions are given relative to sequenced gene fragments (Appendix I - IV).
[b]SNP position within enzyme recognition site is indicated in bold.

Table 2. Haplotype count and haplotype diversity for the 5' and 3' regions of *EuSuSy1* and *EuCAD2*.

Haplotype[a]	Flores			Adonara		Lomblen			Pantar		Alor		Timor				Wetar			Overall
	1	2	3	1	2	1	2	3	1	2	1	2	1	2	3	4	1	2	3	
EuSuSy1 5'																				
H1 (TTT)	2	6	2	3	10	5	6	3	1	1	4	9	8	7	12	12	1		8	92
H2 (TAT)																		1		1
H3 (CTT)	1		1	1			2			5			3	5	7	6				31
H4 (CAT)	28	31	25	22	18	13	22	19	17	16	31	32	13	15	19	6	35	21	27	383
H5 (CAC)	15	13	10	10	10	6	10	8	6	12	5	7	4	1	2	6	6	6	7	137
Diversity	0.53	0.54	0.51	0.56	0.65	0.63	0.63	0.54	0.45	0.65	0.38	0.51	0.70	0.64	0.67	0.74	0.30	0.40	0.54	0.58
EuSuSy1 3'																				
H1 (GTT)	7		4	3	2	1	3	5	1	2	1	2	3	1	1	3			1	39
H2 (GCT)	1						1										1			3
H3 (GCC)	11	13	13	15	4	7	8	6	6	8	5	11	8	10	15	7	11	8	14	166
H4 (ATC)	4	5	1	3	9	2	1	1		3	10	2		4	6	1	1		1	53
H5 (ACT)	7	15	11	6	10	8	13	12	5	8	5	15	7	3	4	6	14	8	12	157
H6 (ACC)	16	17	9	9	13	6	14	6	12	13	19	18	10	10	14	13	15	12	14	226
Diversity	0.78	0.73	0.75	0.74	0.76	0.76	0.74	0.76	0.67	0.75	0.70	0.72	0.74	0.74	0.72	0.73	0.71	0.68	0.71	0.74

Table 2. Continued

Haplotype[a]	Flores			Adonara		Lomblen			Pantar		Alor		Timor				Wetar			Overall
	1	2	3	1	2	1	2	3	1	2	1	2	1	2	3	4	1	2	3	
EuCAD2 5'																				
H1 (TCT)	1										2	2					2		2	7
H2 (TCC)	3		3				4		1			1					3	1	1	16
H3 (TTT)	8	16	17	13	17	7	11	6	11	12	19	13	7	8	13	7	8	8	12	201
H4 (TTC)	12	16	8	12	8	5	9	14	2	1	1	6	4	5	11	6	5	6	10	131
H5 (CCT)					1								1					1		3
H6 (CCC)	9	5	7	2	9	8	7	8	8	10	8	21	15	12	9	10	17	8	6	173
H7 (CTT)							1												1	1
H8 (CTC)	13	13	3	9	3	4	8	2	2	11	10	5	1	3	7	7	7	4	10	112
Diversity	0.80	0.73	0.73	0.71	0.71	0.76	0.81	0.69	0.69	0.70	0.69	0.72	0.65	0.72	0.76	0.77	0.77	0.80	0.80	0.76
EuCAD2 3'																				
H1 (TTC)			1		1		3										1	1		7
H2 (TTA)	26	27	13	21	15	13	17	11	11	15	15	30	17	15	19	21	24	13	18	323
H3 (TCC)	6	6	9	4	8	2	1	5	3	7	6	6	1	1	5	4	8	6	11	88
H4 (TCA)	5	6	3	7	9	3	4		8	7	9	6	3	7	2	1	4	4	4	88
H5 (CTC)						1														1
H6 (CCC)	9	11	12	4	5	5	14	14	2	5	10	5	6	4	14	4	5	4	9	133
H7 (CCA)							1					1	1	1						4
Diversity	0.63	0.64	0.74	0.61	0.75	0.67	0.70	0.64	0.68	0.72	0.74	0.58	0.59	0.65	0.65	0.49	0.63	0.72	0.71	0.67

[a] Haplotype defined by three ht-SNPs listed in parenthesis

Table 3. Coefficients of population differentiation for the 5' and 3' regions of *EuSuSy1* and *EuCAD2*.

	EuSuSy1 5'	*EuSuSy1* 3'	*EuCAD2* 5'	*EuCAD2* 3'
G_{ST}	0.05	0.01	0.03	0.02
N_{ST}	0.08	0.02	0.03	0.03
F_{ST}	0.08	0.02	0.03	0.02

Table 4. Analysis of molecular variance for the 5' and 3' regions of *EuSuSy1* and *EuCAD2*.

Source of variation	d.f.	Percentage of variation			
		EuSuSy1 5'	*EuSuSy1* 3'	*EuCAD2* 5'	*EuCAD2* 3'
Among islands	6	7.48	0.91	0.67	0.18
Among populations within islands	12	0.41	1.00	2.19	2.25
Within populations	667	92.11	98.08	97.13	97.57

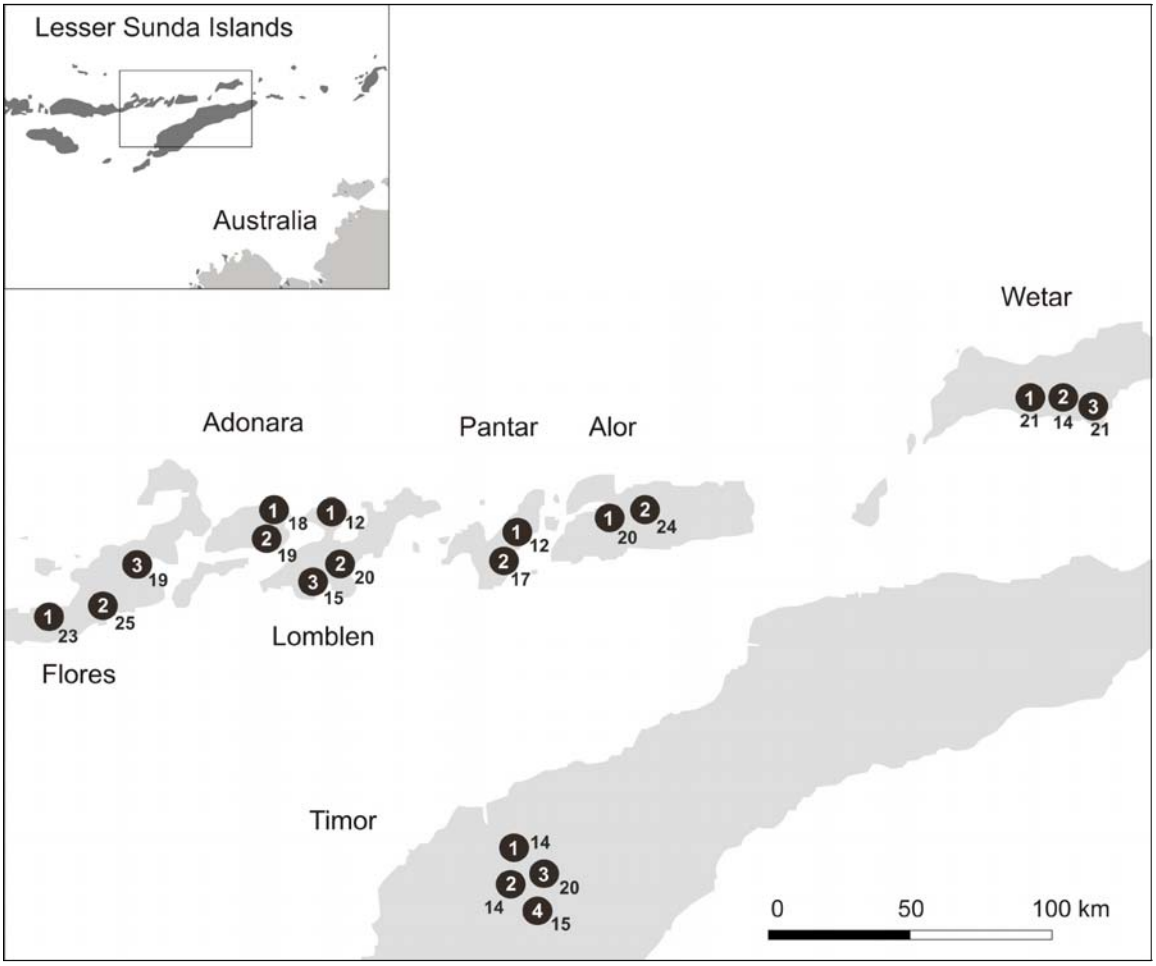


Figure 1. Geographical distribution of sampled *Eucalyptus urophylla* populations across the Lesser Sunda Island archipelago. Population identification number with respect to island is written in white font, and population sample size in black font.

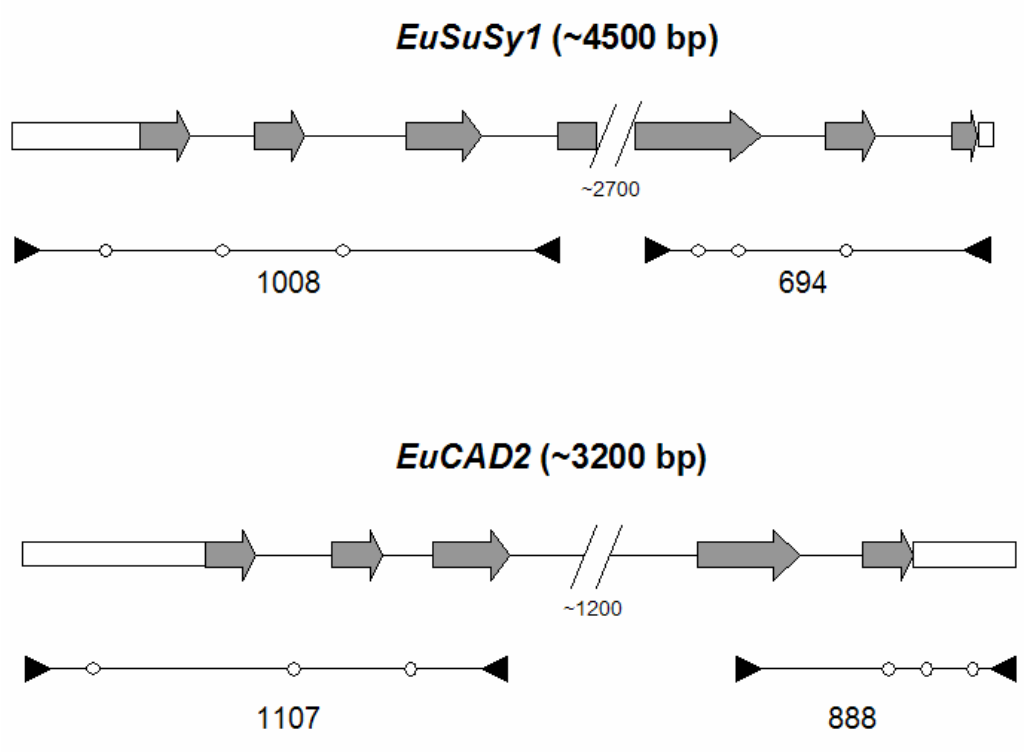


Figure 2. Gene maps of *EuSuSy1* and *EuCAD2* and location of ht-SNPs. The sizes of each region (in bp) as well as the genomic sizes of the full-length genes are indicated. Grey block arrows represent exons and lines connecting them are introns. Untranslated upstream and downstream regions are indicated with open boxes. Lines with inverted arrowheads represent the positions of the primers and the lengths of the sequenced 5' and 3' amplicons within each gene. Circles represent the position of each SNP.

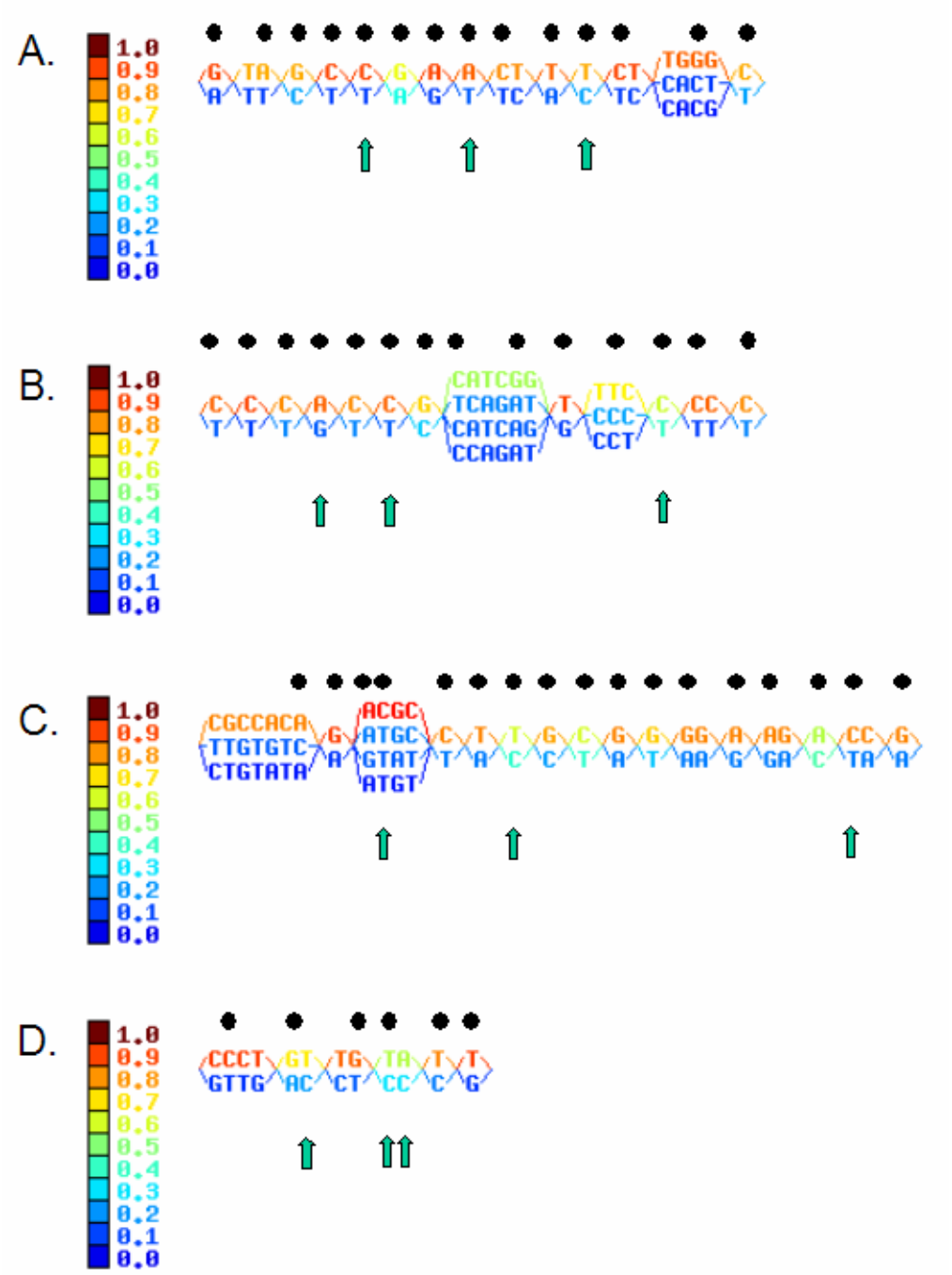


Figure 3. SNP haplotype block structure identified using HaploBlockFinder (Zhang and Jin 2003). (A) *EuSuSy1* 5' region, (B) *EuSuSy1* 3' region, (C) *EuCAD2* 5' region, (D) *EuCAD2* 3' region. Circles indicate haplotype-tag SNPs and arrows indicate SNPs genotyped in this study.

SUPPLEMENTARY MATERIAL

Table S1. Geographical information of *Eucalyptus urophylla* (*n* = 25) individuals that formed the SNP discovery panel.

Sample[a]	Island	Provenance	Latitude	Longitude
1	Timor	Naususu	09°38'S	124°13'E
2	Timor	Tutem	09°35'S	124°17'E
3	Timor	Tune	09°33'S	124°19'E
4	Timor	Lelobatan	09°43'S	124°10'E
5	Alor	Mainang	08°14'S	124°39'E
6	Alor	Apui	08°16'S	124°44'E
7	Alor	Pintumas	08°17'S	124°33'E
8	Alor	Watakika	08°18'S	124°30'E
9	Flores	Ille Ngelle	08°39'S	122°27'E
10	Flores	Lere Baukrenget	08°39'S	122°23'E
11	Flores	Kilawair	08°41'S	122°29'E
12	Flores	Kilawair	08°41'S	122°29'E
13	Flores	Hokeng	08°31'S	122°47'E
14	Adonara	Doken	08°21'S	123°18'E
15	Adonara	Lamalota	08°16'S	123°18'E
16	Adonara	Lamahela	08°21'S	123°15'E
17	Wetar	Nesunhuhun	07°52'S	126°15'E
18	Wetar	Elun Kripas	07°51'S	126°16'E
19	Wetar	Remamea	07°52'S	126°26'E
20	Lomblen	Bunga Muda	08°16'S	123°32'E
21	Lomblen	Labalekan	08°32'S	123°30'E
22	Lomblen	Puor	08°34'S	123°24'E
23	Pantar	Delaki	08°28'S	124°11'E
24	Pantar	Lalapang	08°20'S	124°12'E
25	Pantar	Mauta	08°26'S	124°10'E

[a]The samples were selected to ensure maximal representation of islands and provenances. In the single case where more than one sample was selected from a provenance (Kilawair), different families were sampled.

Table S2. Island, provenance (sample size), location, altitude and population sample size

Population	Provenance	Location	Alt. range (m)	Elevation (m)	*N*
Flores 1					23
	Ile Meak (6)	08°37'S, 122°15'E	680 – 680	680	
	Lere-Baukrenget (10)	08°39'S, 122°23'E	700 – 750	725	
	Natakoli (7)	08°37'S, 122°24'E	820 – 980	900	
Flores 2					25
	Ile Nggele (10)	08°39'S, 122°27'E	570 – 800	685	
	Kilawair (11)	08°41'S, 122°29'E	225 – 530	378	
	Palueh (4)	08°40'S, 122°35'E	540 – 600	570	
Flores 3					19
	Kolibuluk (7)	08°28'S, 122°42'E	648 – 648	648	
	Hokeng (12)	08°31'S, 122°47'E	350 – 800	575	
Adonara 1					18
	Watololong (4)	08°19'S, 123°15'E	630-630	630	
	Gonehama (5)	08°20'S, 123°16'E	687 – 687	687	
	Doken (9)	08°21'S, 123°18'E	600 - 1000	800	
Adonara 2					19
	Lamahela (10)	08°21'S, 123°15'E	856 – 856	856	
	Muda (9)	08°21'S, 123°16'E	600 – 900	750	
Lomblen 1					12
	Jontona (7)	08°16'S, 123°25'E	675 – 900	788	
	Bunga Muda (5)	08°16'S, 123°32'E	600 – 700	650	
Lomblen 2					20
	Ile Ape (6)	08°29'S, 123°30'E	840 – 880	860	
	Padekluwa (7)	08°30'S, 123°26'E	700 – 900	800	
	Ile Kerbau (7)	08°29'S, 123°29'E	730 – 750	740	
Lomblen 3					15
	Puor (7)	08°34'S, 123°24'E	900 – 980	940	
	Labalekan (8)	08°32'S, 123°30'E	620 – 920	770	
Pantar 1					12
	Beangonong (4)	08°20'S, 124°12'E	450 – 680	565	
	Lalapang (8)	08°20'S, 124°12'E	500 – 650	575	
Pantar 2					17
	Mauta (9)	08°26'S, 124°10'E	600 – 640	620	
	Delaki (8)	08°28'S, 124°11'E	780 – 840	810	
Alor 1					20
	Watakika (10)	08°18'S, 124°30'E	350 – 600	475	
	Pintu Mas (10)	08°17'S, 124°33'E	320 - 450	385	
Alor 2					24
	Mainang (7)	08°14'S, 124°39'E	1100 - 1250	1175	
	Molpui (4)	08°15'S, 124°44'E	400 - 400	400	
	Apui (10)	08°16'S, 124°44'E	1100 - 1300	1200	
	Manabai (3)	08°14'S, 124°45'E	400 - 400	400	
Wetar 1					21
	Nesunhuhun (10)	07°52'S, 126°15'E	600 - 642	621	
	Elun Kripas (11)	07°51'S, 126°16'E	715 - 750	733	
Wetar 2					14
	Nakana Ulam (7)	07°51'S, 126°21'E	680 - 750	715	
	Alasannaru (7)	07°51'S, 126°23'E	580 - 612	596	

Table S2. Continued

Population	Provenance	Location	Alt. range (m)	Elevation (m)	*N*
Wetar 3					21
	Puaanan (7)	07°51'S, 126°26'E	453 - 516	485	
	Remamea (8)	07°52'S, 126°26'E	409 - 542	476	
	Talianan (6)	07°52'S, 126°28'E	483 - 559	521	
Timor 1					14
	Bonleu (4)	09°33'S, 124°04'E	1600 - 1800	1700	
	Nuafin (10)	09°31'S, 124°11'E	1800 - 2000	1900	
Timor 2					14
	Leloboko (4)	09°37'S, 124°10'E	1400 - 1600	1500	
	Naususu (10)	09°38'S, 124°13'E	1200 - 1450	1325	
Timor 3					20
	Tutem (10)	09°35'S, 124°17'E	1200 - 1400	1300	
	Tune (10)	09°33'S, 124°19'E	1100 - 1400	1250	
Timor 4					15
	Mollo (11)	09°41'S, 124°11'E	1200 - 1600	1400	
	Lelobatang (4)	09°41'S, 124°14'E	1200 - 1400	1300	
Total					343

Table S3. Primers used to amplify the 5' and 3' fragments of *EuSuSy1* and *EuCAD2*. The melting temperature (T_m) of each primer is indicated

Gene amplified	Primer Name[a]	Primer sequence	T_m (ºC)
EuSuSy1	EuSuSy1-5'F	5'CCAACCGAGATCATCACCTA'3	60
	EuSuSy1-5'R	5'GTCATGGAAGAGCTTAGCGGAGAG'3	66
	EuSuSy1-3'F	5'GTGCCGACATGAGCATCTACTT'3	64
	EuSuSy1-3'R	5'CCAATGCTTCCGTCTTCTGGTA'3	64
EuCAD2	EuCAD2-5'F	5'GAACTCACGATGGTTCCAGAAAGG'3	65
	EuCAD2-5'R	5'TCGCCAACCACTATCTCACCAG'3	66
	EuCAD2-3'F	5'CACTGATTCGCTCGACTACG'3	62
	EuCAD2-3'R	5'AGAGTCGTATCCACCAAGAA'3	59

[a]5' and 3' refer to a gene fragment that was amplified from either the 5'- or 3'-end of each gene

Table S4. Primers used to amplify short fragments of *EuSuSy1* and *EuCAD2* prior to restriction enzyme digestion

Gene fragment	SNP (Pos)	Primer name	Primer sequence
SuSy1 5'	C/T (168)	EuSuSy1-5'1F	5'AATCCGGCCGCTCTA3'
		EuSuSy1-5'1R	5'CCTCAGAGATGGCCTCAA3'
	A/T (415)	EuSuSy1-5'1F	5'AATCCGGCCGCTCTA3'
		EuSuSy1-5'1R	5'CCTCAGAGATGGCCTCAA3'
	T/C (609)	EuSuSy1-5'1F	5'AATCCGGCCGCTCTA3'
		EuSuSy1-5'2R	5'TGGACGTTCACACGGATG3'
SuSy1 3'	A/G (156)	EuSuSy1-3'1F	5'CCGCTACATCTGTGACACGAAG3'
		EuSuSy1-3'1R	5'GTCAGGTTCAACAGCCTCTC3'
	C/T (213)	EuSuSy1-3'2F	5'CCGCTACATCTGTGAC3'
		EuSuSy1-3'2R	5'TCTCTGCATGGCACC3'
	C/T (496)	EuSuSy1-3'3F	5'GGTGCCATGCAGAGAAT3'
		EuSuSy1-3'3R	5'CTGCAAGGAAGAGCAAGAAC3'
CAD2 5'	C/T (196)	EuCAD2-5'1F	5'ATGGAGCTTCTCCATCACTT3'
		EuCAD2-5'1R	5'GAAACGAGACGGACCAATAG3'
	T/C (599)	EuCAD2-5'2F	5'TGGGCAGTCTTGAGAAGGAG3'
		EuCAD2-5'2R	5'CTGGTCCCGTGTTTCTGCAT3'
	C/T (931)	EuCAD2-5'3F	5'CAGAAACACGGGACCAGA3'
		EuCAD2-5'3R	5'GCAGCACCCAACCACTATAC3'
CAD2 3'	T/C (527)	EuCAD2-3'1F	5'TAGGGAGCATGAAGGAAACA3'
		EuCAD2-3'1R	5'CACAGCACAACCCAATTCAA3'
	T/C (593)	EuCAD2-3'1F	5'TAGGGAGCATGAAGGAAACA3'
		EuCAD2-3'1R	5'CACAGCACAACCCAATTCAA3'
	A/C (689)	EuCAD2-3'1F	5'TAGGGAGCATGAAGGAAACA3'
		EuCAD2-3'1R	5'CACAGCACAACCCAATTCAA3'

Table S5. SNP count with respect to each *Eucalyptus urophylla* population

SNP (pos.)	Flores			Adonara		Lomblen			Pantar		Alor		Timor				Wetar			Overall	Overall
	1	2	3	1	2	1	2	3	1	2	1	2	1	2	3	4	1	2	3	count	freq.
SuSy1 5'																					
C (168)	44	44	36	33	28	19	34	27	23	33	36	39	20	21	28	18	41	27	34	585	0.85
T	2	6	2	3	10	5	6	3	1	1	4	9	8	7	12	12	1	1	8	101	0.15
A (415)	43	44	35	32	28	19	32	27	23	28	36	39	17	16	21	12	41	28	34	555	0.81
T	3	6	3	4	10	5	8	3	1	6	4	9	11	12	19	18	1		8	131	0.19
T (609)	31	37	28	26	28	18	30	22	18	22	35	41	24	27	38	24	36	22	35	542	0.79
C	15	13	10	10	10	6	10	8	6	12	5	7	4	1	2	6	6	6	7	144	0.21
SuSy1 3'																					
A (156)	27	37	21	18	32	16	28	19	17	24	34	35	17	17	24	20	30	20	27	463	0.67
G	19	13	17	18	6	8	12	11	7	10	6	13	11	11	16	10	12	8	15	223	0.33
C (213)	35	45	33	30	27	21	36	24	23	29	29	44	25	23	33	26	41	28	40	592	0.86
T	11	5	5	6	11	3	4	6	1	5	11	4	3	5	7	4	1		2	94	0.14
C (496)	31	35	23	27	26	15	23	13	18	24	34	31	18	24	35	21	27	20	29	474	0.69
T	15	15	15	9	12	9	17	17	6	10	6	17	10	4	5	9	15	8	13	212	0.31
CAD2 5'																					
T (196)	24	32	28	25	25	12	24	20	14	13	22	22	11	13	24	13	18	15	25	380	0.55
C	22	18	10	11	13	12	16	10	10	21	18	26	17	15	16	17	24	13	17	306	0.45
T (599)	33	45	28	34	28	16	29	22	15	24	30	24	12	16	31	20	20	18	33	478	0.70
C	13	5	10	2	10	8	11	8	9	10	10	24	16	12	9	10	22	10	9	208	0.30
C (931)	37	34	21	23	20	17	28	24	13	22	19	33	20	20	27	23	32	19	27	459	0.67
T	9	16	17	13	18	7	12	6	11	12	21	15	8	8	13	7	10	9	15	227	0.33
CAD2 3'																					
T (527)	37	39	26	32	33	18	25	16	22	29	30	42	21	23	26	26	37	24	33	539	0.79
C	9	11	12	4	5	6	15	14	2	5	10	6	7	5	14	4	5	4	9	147	0.21
T (593)	26	27	14	21	16	14	20	11	11	15	15	30	17	15	19	21	25	14	18	349	0.51
C	20	23	24	15	22	10	20	19	13	19	25	18	11	13	21	9	17	14	24	337	0.49
A (689)	31	33	16	28	24	16	22	11	19	22	24	37	21	23	21	22	28	17	22	437	0.64
C	15	17	22	8	14	8	18	19	5	12	16	11	7	5	19	8	14	11	20	249	0.363

Table S6. F_{ST} estimates for each SNP locus

Gene region	SNP	F_{ST}
EuSuSy1 5'	C/T (pos. 168)	0.069
	A/T (pos. 415)	0.148
	T/C (pos. 609)	0.021
EuSuSy1 3'	A/G (pos. 156)	0.011
	C/T (pos. 213)	0.031
	C/T (pos. 496)	0.022
EuCAD2 5'	C/T (pos. 196)	0.018
	T/C (pos. 599)	0.060
	C/T (pos. 931)	0.011
EuCAD2 3'	T/C (pos. 527)	0.036
	T/C (pos. 593)	0.007
	A/C (pos. 689)	0.035

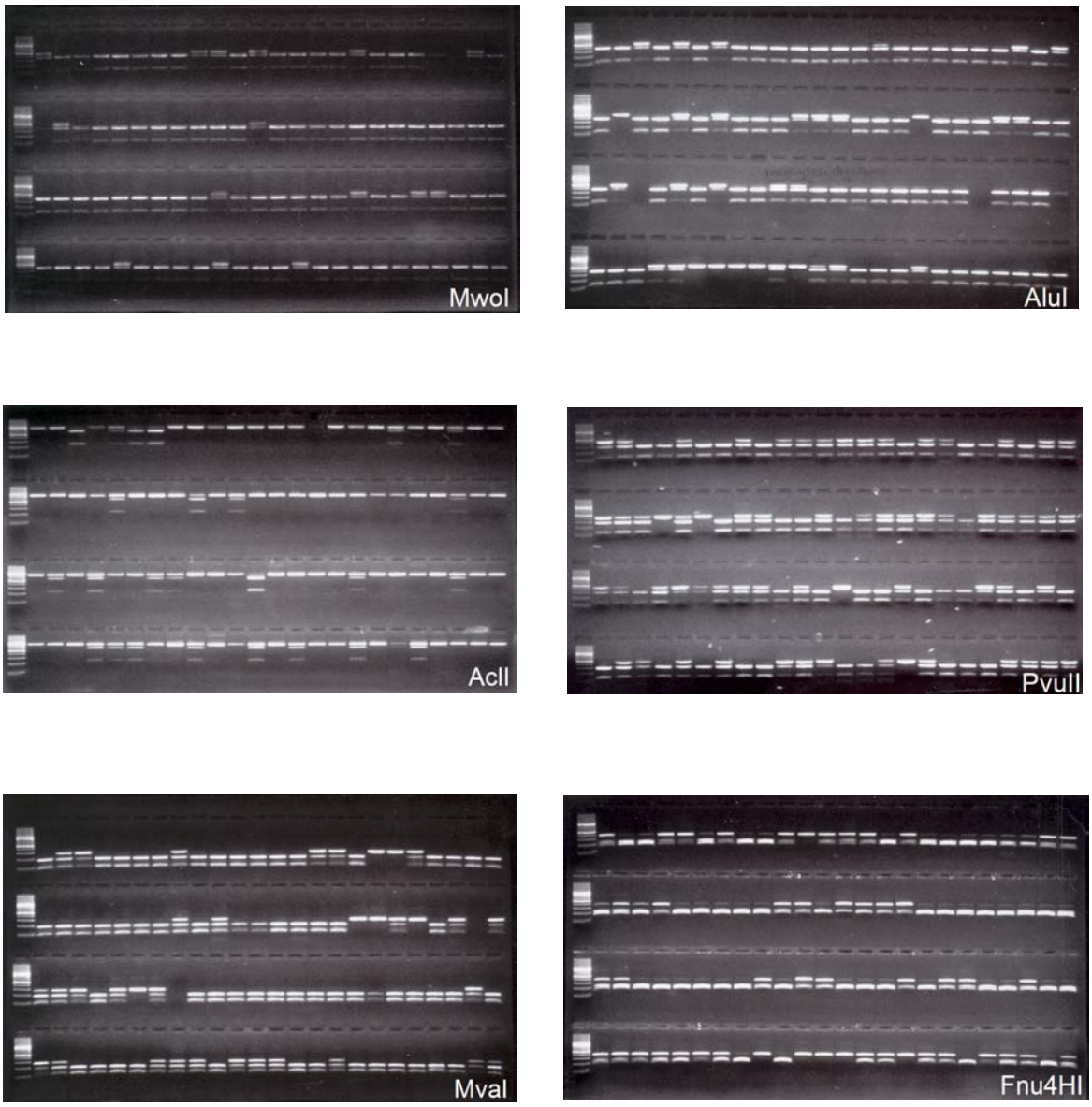


Figure S1. Illustration of 96 well agarose gels used to score SNP genotypes for *EuSuSy1*. Each gel is annotated with the name of the enzyme used to generate the profile.

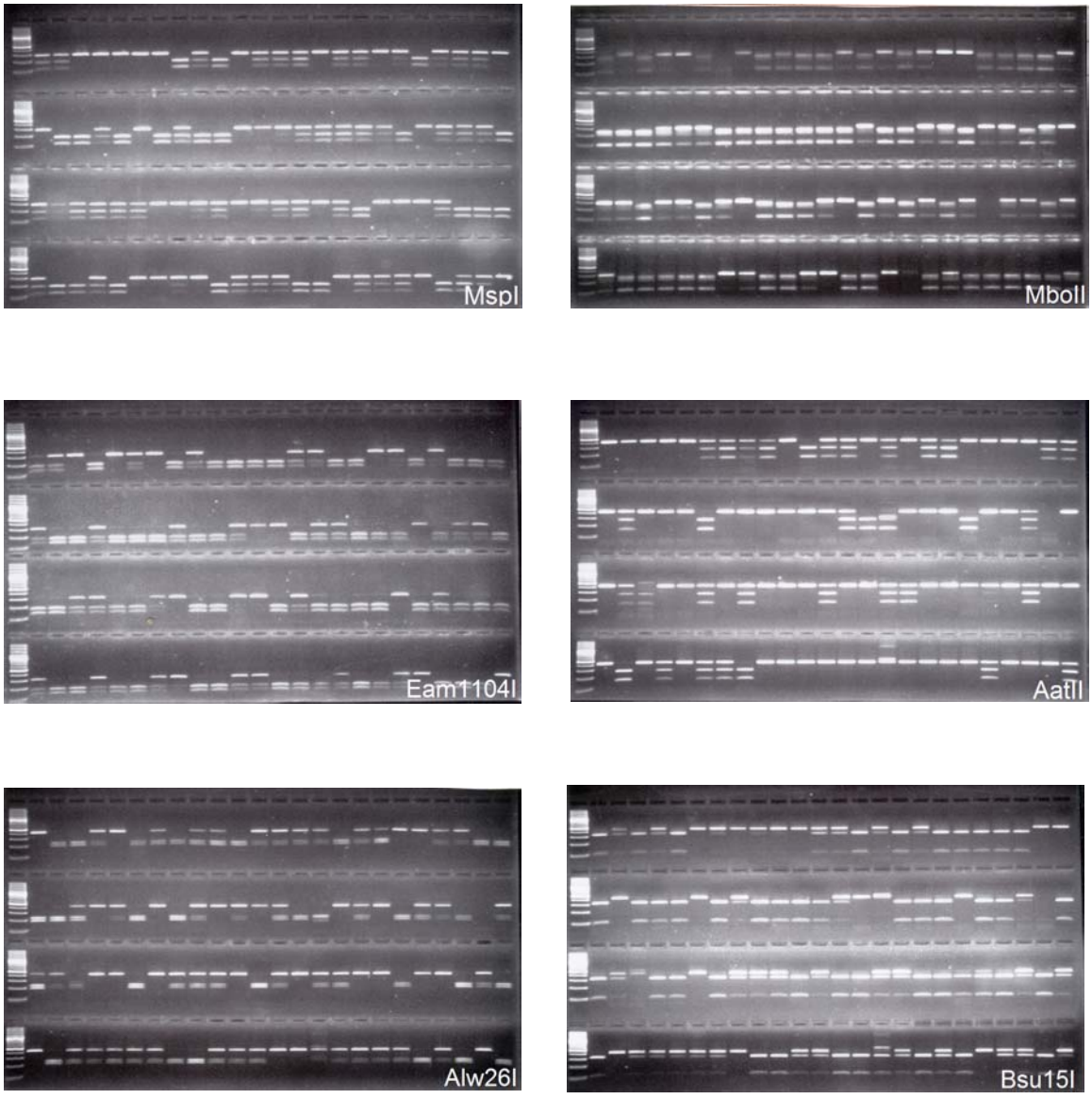


Figure S2. Illustration of 96 well agarose gels used to score SNP genotypes for *EuCAD2*. Each gel is annotated with the name of the enzyme used to generate the profile.

CHAPTER 5

Molecular genetic diversity of *E. urophylla*: Implications for breeding and conservation

The results of this study have important implications for the genetic conservation and breeding of *E. urophylla,* an important tropical plantation tree species. Many native *E. urophylla* populations are under threat from agricultural expansion in Indonesia (Pepe et al. 2004). Ideally, conservation efforts should concentrate on maintaining genetic diversity *in situ*. This approach has the advantage that it allows for the continuation of genetic processes *in situ*, such as selection and gene flow. In this way, there is the possibility of conserving a dynamic gene pool capable of responding to the changing natural environment. Many *E. urophylla* populations on the island of Timor are currently protected within the boundaries of Mt. Mutis Forest Park. Wetar populations are also at low risk because the human population pressure on the island is minimal but there is concern about increased mining activities. The conservation status of numerous populations on the remaining five islands range from *Vulnerable* to *Critically endangered* (Pepe et al. 2004). Here, low elevation populations are being subjected to annual fires in an effort to grow grass for livestock, and trees are being removed to establish crops like macademia nuts with higher economic returns. *In situ* conservation therefore requires the co-operation of local communities and government, which in turn requires appropriate incentives, ideally provided by those benefiting from the conservation of this species, such as international forestry companies.

A complementary or alternative approach to the *in situ* conservation is the implementation of *ex situ* conservation programs. *E. urophylla* seed collections were carried out by several international organizations over the last 30 years to provide a broad genetic resource necessary for effective conservation and breeding programs (Martin and Cossalter 1975; Gunn and McDonald 1991; Hodge et al. 2001; Pepe et al. 2004). The collection series conducted by Camcore, in collaboration with PT Sumalindo Lestari Jaya, comprised seeds from 1104 mother trees distributed across 62 sites representing the natural range of *E. urophylla* (Table 1). The seed lots were distributed among the Camcore members in Brazil, Colombia, Mexico, Venezuela, and South Africa. In total, 114 provenances trials were established in a compact family design where half-sib families were nested within provenances (see Hodge et al. 2001 for details). These trials currently serve as *ex situ* conservation banks, as well as base populations in tree breeding programs, which are now entering their second generation. The planting of two additional *E. urophylla* conservation parks in South Africa has recently been proposed (Dvorak et al. 2008). The challenge when designing successive generation conservation banks and breeding populations is to maximise the genetic representation of the species gene pool within the limits of a finite population size. Indeed, it is the presence and maintenance of genetic diversity that enables a species to adapt to future environmental conditions. Similarly a broad genetic base enhances the potential for domestication and improves the efficiency of breeding programs.

At the population or species level, immediate response to selection is mostly due to alleles at intermediate frequencies (e.g. 0.1 – 0.9), while long-term response may arise as favourable

alleles that are present at low frequencies (e.g. below 0.1) progressively increase in frequency (Namkoong et al. 2000). These low frequency alleles may have had little or no significance for adaptation in the past, but they may increase in importance when environments change, when trees are planted outside of their natural range, or when introduced diseases threaten. Such alleles provide the necessary variation for future adaptive responses and/or for the improvement of economically important traits, and should thus be conserved. Camcore aims to capture alleles that occur at a frequency of 0.05 or greater in natural populations (Dvorak et al. 2008). Therefore, an in depth understanding of the geographic distribution of genetic diversity of *E. urophylla* will assist in prioritising populations for conservation and domestication.

In the present study, chloroplast haplotype diversity was analysed to determine the phylogeographic history of *E. urophylla* and to draw inferences regarding past seed migratory routes (Chapter 2). The geographical distribution of chloroplast haplotype diversity suggested an east to west island colonisation pattern. Timor was likely the first island colonised based on its high haplotype diversity and proximity to Australia or New Guinea. The haplotype diversity observed on the islands of Wetar and Alor suggested that they were also colonized early on, whereas the lack of chloroplast haplotype diversity in the western islands of Flores and Lomblen suggested more recent colonisation events. Restricted gene flow with isolation by distance and long distance colonisation events, possibly assisted by westerly sea currents, are considered largely responsible for the spatial distribution of chloroplast haplotypes within extant populations of the species.

Since seed dispersal rates were shown to be low enough to allow population differentiation (Chapter 2), pollen flow among populations is likely responsible for the low estimates of population differentiation observed based on nuclear microsatellite (Chapter 3) and SNP haplotype markers (Chapter 4). The high level of gene diversity (H_E = 0.74) for *E. urophylla* obtained in this study was comparable to gene diversity estimates reported in other *Eucalyptus* species based on microsatellite markers (Byrne et al. 1996; McGowen et al. 2001; Jones et al. 2002; Butcher et al. 2005). Timor populations generally had the highest levels of genetic diversity, but contrary to the chloroplast genome, no east to west gradient of nuclear genetic diversity was observed. This result may possibly have been different if *E. urophylla* samples were also obtained from East Timor, assuming these populations had equally high levels of diversity.

The present study showed limited population genetic structure in the nuclear genome, with >90% of the microsatellite marker and SNP haplotype variation apportioned within populations rather than among populations or among islands. In addition, the tests for phylogeographic structure of the nuclear genetic variation were not significant. This suggests that only a subset of populations need to be included in an *ex situ* conservation or breeding program in order to capture most of the nuclear genetic variation within the species. In light of the two genetically homogeneous population clusters inferred using STRUCTURE v2.0 (Pritchard et al. 2000), it would be prudent to keep genotypes representing each cluster separate. Within each cluster, it appears appropriate to choose between or group populations

from similar environments that are not significantly differentiated based on molecular and genetic field trial results.

Genetic resource management practices should not be based entirely on genetic diversity assessments based on likely neutral molecular markers. Such markers enable population geneticists to investigate important neutral evolutionary forces such as migration and genetic drift. The genetic diversity and population structure of genes conferring adaptive potential may be very different, especially for a species with a disjunct distribution across different environments. There may be *E. urophylla* populations that have favourable alleles in relatively high frequencies (frequency > 0.20) that are rare on a species-wide basis (overall frequency < 0.05). Under these circumstances, a geographically structured sample would be recommended in order to capture alleles that are common in one area (one or a few adjacent populations) but rare across the species as a whole. In this present study, six of the twelve microsatellite loci had marker alleles (Table 2) that were common in one population but rare across the species distribution. These included populations from the islands of Pantar, Timor and Wetar. The latter two regions are environmentally diverse, with Timor populations occupying high elevation sites, and Wetar populations generally occupying dry sites (Martin and Cossalter, 1976). Geographically different environments likely host adapted genotypes at genetic loci that will exhibit geographic structure if sampled directly or through tightly linked marker loci. Indeed, the population structure of genes conferring adaptive potential may be far greater than that observed using neutral markers. High rates of recombination (low linkage disequilibrium) may isolate these structured loci in otherwise unstructured nuclear

genomes, which suggests that high-resolution, genome-wide marker analysis would be required to detect such adaptive loci. The composition and management of an *E. urophylla* genetic resource for conservation or breeding should account for the putative population genetic structure of adaptive loci by considering the specific environmental conditions (e.g. elevation and rainfall) of the seed source.

Since the forestry industry currently funds the management of *E. urophylla* genetic resources, priority is given to better performing provenances. However, the best performing provenances in one country are not necessarily the best in other countries (Dvorak et al. 2008). For example, when comparing provenance performance for growth of 27 *E. urophylla* provenances tested in Mexico and Colombia, the top seven performing provenances in Mexico included only one of the top seven performing provenances in Colombia (Dvorak et al. 2008). Priority for the inclusion of provenances in a genetic resource should also be given to provenances classified as critically endangered. Notably, all of the highly productive provenances on Flores, Adonara, Lomblen, Pantar and Alor are also those in urgent need of conservation (Dvorak et al. 2008).

The observation of poor performing provenances may not necessarily be indicative of a lack of desirable alleles in the population, but rather a response to inbreeding if the native stand is genetically isolated and small. Molecular markers can be used to test for recent bottlenecks and inbreeding (Cornuet and Luikart, 1996). The chloroplast haplotype data presented in this study suggest that *E. urophylla* was present on the islands of Timor and Wetar for a period of

time before colonizing the islands to the west. The lack of chloroplast haplotype diversity on the islands of Flores and Lomblen is likely the signature of a recent founder event. In contrast, the nuclear genetic diversity of *E. urophylla* is evenly distributed across the species natural range and there was no evidence of recent bottlenecks. The implication is that populations on the boundary of the species distribution need not be managed differently from other more central populations since they do not appear to be at greater risk of inbreeding in subsequent generations.

An additional benefit of genotyping the base population (Camcore collection) is that it provides a baseline for tracking genetic consequences of forest domestication (Williams et al. 2004). The present study will provide a baseline against which later generation breeding populations can be tested for changes in genetic diversity. For example, the 222 first generation selections made from a Camcore *E. urophylla* provenance trial series established in South Africa by Mondi (this particular trial series comprised provenances from the islands of Timor, Alor and Flores) could be genotyped and their genetic diversity compared to the diversity estimates presented in this study. In a similar study by Jones et al. (2006), the genetic diversity within the Australian national *E. globulus* breeding program was compared to the diversity estimated in native local populations. More specifically, the gene diversity (expected heterozygosity) and allelic richness of 140 first-generation selections from trials established using open-pollinated seed collected from native *E. globulus* trees were compared to a sample of 340 mature individuals collected from throughout the native distribution of the species. While gene diversity was high in the breeding population sample, and similar to that

displayed in the native sample, allelic richness was lower, suggesting a loss of rare alleles during selection (Jones et al. 2006). Indeed, allelic richness is strongly influenced by the presence or absence of rare alleles, and is regarded as a better parameter than gene diversity for measuring loss of genetic variation due to genetic bottlenecks (Luikart et al. 1998).

Another practical application arising from this study is for seed-source certification and the determination of geographic origin of unknown samples. This requires the identification of diagnostic alleles common in one population but not in another. The strong phylogeographic structure of *E. urophylla* chloroplast haplotypes may be useful in this respect.

The large difference in population genetic structure of the maternally inherited chloroplast genome compared to the biparentally inherited nuclear genome suggests that most of the gene flow among populations is a result of pollen flow. For *in situ* conservation, one has to ensure that the pollinators are also conserved (Boshier 2000). Many pollinators require alternate host species to provide food and so one may have to manage the ecosystem rather than just the species. Fortunately, *E. urophylla* flowers appear unspecialised and have generalist insect pollinators (Dvorak et al. 2008). Therefore, problems regarding the lack of generalist pollinators may only occur if there is a general loss of fauna in the area.

Pollination of an *ex situ E. urophylla* genetic resource may include pollen originating from surrounding *Eucalyptus* species. For example, tree breeders in South Africa working for Mondi found a high percentage of *E. grandis* x *E. urophylla* hybrids in a second generation

E. urophylla progeny trial (original *in situ* seed source provided by CSIRO) that was consequently discontinued and selections were rather made for the hybrid clonal program. Precautions can be taken to reduce pollen contamination of an *E. urophylla* genetic resource. Spatial isolation may be implemented whereby the genetic resource is established in areas far removed from external pollen sources. An alternative approach to the genetic isolation of a genetic resource can be accomplished by surrounding it with a buffer zone that provides a barrier to foreign pollen. However, such barriers are not always effective (Adams and Burczyk, 2000). Pollen contamination from surrounding stands can also be reduced by harvesting the surrounding stands several years before regenerating the genetic resource. Finally, it is also possible to control mating by applying controlled pollination procedures. This option is expensive and the financial costs may be prohibitive for large-scale use.

In conclusion, populations of *E. urophylla* will only be conserved if private industry takes a leadership role in doing so. The gene pool of *E. urophylla* is rich, offering a rewarding base for genetic improvement. There was no evidence of depleted nuclear genetic diversity in remote populations as a consequence of genetic drift, associated with population fragmentation or due to a founding event following long-distance colonisation. A subset of populations from the island of Timor should be included in a genetic resource because they appear to be the source populations, having the highest nuclear genetic diversity, even though they generally perform poorly in provenance assessments (Dvorak et al. 2008). A subset of populations from the island of Wetar should be included because they represent cluster-2 and have been shown to differentiate from the remaining *E. urophylla* populations based on

morphological characteristics, possibly in response to dryer environmental conditions. Such adaptations may be particularly beneficial if global climatic fluctuations become important for species survival. Several populations from the western islands of Flores, Adonara, Lomblen and Pantar should be included because they represent cluster-1, possibly as a result of sharing a more recent evolutionary history. These western islands, together with Alor, also demonstrate superior productivity *ex situ*, and are most at risk of extinction. Pollen contamination is possibly the biggest risk to the integrity of *ex situ E. urophylla* populations for conservation and domestication. However, there are methods of reducing this risk by applying controlled pollination procedures.

LITERATURE CITED

Adams WT, Burczyk J (2000) Magnitude and implications of gene flow in gene conservation reserves. In 'Forest conservation genetics: Principles and practice'. (Eds AG Young, DH Boshier, TJ Boyle) pp. 101-111. (CSIRO Publishing).

Boshier DH (2000) Mating Systems. In 'Forest conservation genetics: Principles and practice'. (Eds AG Young, DH Boshier, TJ Boyle) pp. . (CSIRO Publishing).

Butcher PA, Skinner AK, Gardiner CA (2005) Increased inbreeding and inter-species gene flow in remnant populations of the rare *Eucalyptus benthamii. Conservation Genetics* **6**, 213–226.

Byrne M, Marques-Garcia MI, Uren T, Smith DS, Moran GF (1996) Conservation and genetic diversity of microsatellite loci in the genus *Eucalyptus. Australian Journal of Botany* **44**, 331–341.

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**, 2001–2014.

Dvorak WS, Hodge GR, Payn KG (2008). The conservation and breeding of *Eucalyptus urophylla*: A case study to better protect important populations and improve productivity. *Southern Hemisphere Forestry Journal* (in press).

Gunn BV, McDonald MW (1991) *Eucalyptus urophylla* seed collections. *Forest Genetic Resources* **19**, 34–37, FAO, Rome.

Hodge GR, Pepe B, Wijoyo FS, Dvorak WS (2001) Early results of *Eucalyptus urophylla* provenance/progeny trials in Colombia and Venezuela. In: Developing the Eucalypts of the Future. Proc. IUFRO Working Party 2.08.03.

Jones RC, Steane DA, Potts BM, Vaillancourt RE (2002) Microsatellite and morphological analysis of *Eucalyptus globulus* populations. *Canadian Journal of Forest Research* **32**, 59–66.

Jones TH, Steane DA, Jones RC, Pilbeam D, Vaillancourt RE, Potts BM (2006) Effects of domestication on genetic diversity in Eucalyptus globulus. *Forest Ecology and Management* **234**: 78–84.

Luikart G, Sherwin WB, Steele BM, Allendorf FW (1998) Usefulness of molecular markers for detecting population bottlenecks via monitoring genetic change. *Moleclar Ecology* **7**, 963–974.

Martin B, Cossalter C (1975) Les eucalyptus des iles de la Sonde. *Bois et Foréts des Tropiques* **163**, 3–25.

Martin B, Cossalter C (1976) Les eucalyptus des iles de la Sonde. *Bois et Foréts des Tropiques* **165**, 3–14.

McGowen MH, Wiltshire RJE, Potts BM, Vaillancourt RE (2001) The origin of *Eucalyptus vernicosa*, a unique shrub Eucalypt. *Biological Journal of the Linnean Society* **74**, 397–405.

Namkoong G, Koshy MP, Aitken S (2000) Selection. In 'Forest conservation genetics: Principles and practice'. (Eds AG Young, DH Boshier, TJ Boyle) pp. 101-111. (CSIRO Publishing).

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

Williams CG, Bridgwater FE, Nelson CD (2004) A founder project: Marking the domestication baseline for forest trees. *Forest Genetic Resources* No. 31, FAO Rome, 20–24.

Table 1. *Eucalyptus urophylla* provenances sampled by Camcore/Sumalindo from 1996 to 2003.

Island	No. of Provenance Sampled	No. of Trees Selected	Altitudinal Range (m)
Timor	16	299	1100-2000
Wetar	8	118	409-750
Alor	6	100	320-1300
Pantar	5	97	380-840
Lomblen	8	137	540-980
Adonara	9	142	494-1000
Flores	10	211	225-800

Table 2. Microsatellite allele frequency for localised common alleles (frequency > 0.2) that are rare across the species distribution (frequency < 0.05).

Locus	Population	Allele	Allele frequency	
			Population	Species-wide
FMG-EUCSSR1	Timor-4	224	0.205	0.020
FMG-EUCSSR3	Pantar-1	226	0.250	0.037
EMBRA3	Timor-4	136	0.250	0.039
EMBRA37	Timor-3	119	0.275	0.048
EMBRA69	Wetar-1	089	0.238	0.026
EMBRA69	Wetar-2	105	0.286	0.043
EMBRA69	Timor-3	111	0.225	0.032
EMBRA227	Timor-3	311	0.225	0.028

APPENDIX

I: Partial (5'-end) genomic sequence of the *Eucalyptus urophylla SuSy1* gene (Adapted from Maleka et al. 2007). SNP position within enzyme recognition site (underlined) is indicated in bold italic.

```
LOCUS       EuSuSy1        1008 bp    DNA linear       30-OCT-2006
DEFINITION  Eucalyptus urophylla sucrose synthase 1 (SuSy1), partial gene.
ACCESSION   EuSuSy1
SOURCE      Eucalyptus urophylla
  ORGANISM  Eucalyptus urophylla
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta;
Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons;          core
eudicotyledons; rosids; Myrtales; Myrtaceae; Eucalyptus.
REFERENCE   1  (bases 1 to 1008)
  AUTHORS   Maleka,F.M., Bloomer,P. and Myburg,A.A.
  TITLE     Genetic diversity and linkage disequilibrium in wood
             biosynthetic genes of Eucalyptus urophylla S.T. Blake
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1008)
  AUTHORS   Maleka,F.M., Bloomer,P. and Myburg,A.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-OCT-2006) Department of Genetics, University of
            Pretoria, Faculty of Natural and Agricultural Sciences,
             Forestry and Agricultural Biotechnology Institute,
 Lunnon Street, Hillcrest, Pretoria, Gauteng 0001,
 Republic of South Africa
FEATURES             Location/Qualifiers
     source          1..1008
                     /organism="Eucalyptus urophylla"
                     /mol_type="genomic DNA"

     upstream        1...226

                     /gene="EuSuSy1"

     exons           join (227..325,425..583,739..891)

                     /gene="EuSuSy1"

     introns         join (326..424,584..738,892...>)

                     /gene="EuSuSy1"

BASE COUNT      218 a    184 c    249 g    357 t
ORIGIN
      1 ttcgcaattt taataccttc gtacatgctt agttggtaag gtttgaaaaa tccggccgct
     61 ctagaaaaga tcgatttttc caacgatttg actttttttgt tgctctgttt tgtgagatta
    121 ttcaaaaccc ctcctttatt agtggaagtt gggttttgct tctaatgcac ggtgtgtttc
    181 actttggtgg ttgttgcagt tctttttctg agagaagaat ttagacatgg ctgatcgcat
    241 gttgactcga agccacagcc ttcgcgagcg tttggacgag accctctctg ctcaccgcaa
    301 cgatattgtg gccttccttt caaggtaaaa agcaaggacg gaaggggata tattcaagaa
    361 atcttcaaag agagcatctt gatgagtggt tttaacataa agttggtgaa agggagctta
```

```
    421 aaaagtgttt tgatcccttt tgttgtcatg ttgaagggtt gaagccaagg gcaaaggcat
    481 cttgcagcgc caccagattt ttgctgagtt tgaggccatc tctgaggaga gcagagcaaa
    541 gcttcttgat ggggcctttg gtgaagtcct caaatccact caggtattat gaactccctt
    601 catgtcaaacg ttttttcggtt ctttacgctc ttgaaatcta ctcttctata gtgataatgg
    661 gttgattttt gcttttcttt gacctttttt gatttaaatt ctcaaggaat ttcttttgct
    721 ctaaattttg gggtttagga agcgattgtg tcgcctccat gggttgctct tgctgttcgt
    781 ccaaggccgg gcgtgtggga gcacatccgt gtgaacgtcc atgcgcttgt tcttgagcaa
    841 ttggaggttg ctgagtatct gcacttcaaa gaagagcttg ctgatggaag gtcagaatct
    901 ttatttttcc ttggtgatct cagatctctg ggtcatgttc ttttttgctg ttcttgtttt
    961 tggtcgtttt gggggtgtta atgagagtta ttcgtcgtgt ggttcagc
//
```

II: Partial (3'-end) genomic sequence of the *Eucalyptus urophylla SuSy1* gene (Adapted from Maleka et al. 2007). SNP position within enzyme recognition site (underlined) is indicated in bold italic.

```
LOCUS       EuSuSy1            694 bp    DNA linear       30-OCT-2006
DEFINITION  Eucalyptus urophylla sucrose synthase 1 (SuSy1), partial gene.
ACCESSION   EuSuSy1
VERSION
KEYWORDS    .
SOURCE      Eucalyptus urophylla
  ORGANISM  Eucalyptus urophylla
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta;
Tracheophyta;  Spermatophyta;  Magnoliophyta;  eudicotyledons;         core
eudicotyledons; rosids; Myrtales; Myrtaceae; Eucalyptus.
REFERENCE   1  (bases 1 to 694)
  AUTHORS   Maleka,F.M., Bloomer,P. and Myburg,A.A.
  TITLE     Genetic diversity and linkage disequilibrium in wood
             biosynthetic genes of Eucalyptus urophylla S.T. Blake
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 694)
  AUTHORS   Maleka,F.M., Bloomer,P. and Myburg,A.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-OCT-2006) Department of Genetics, University of
            Pretoria, Faculty of Natural and Agricultural Sciences,
             Forestry and Agricultural Biotechnology Institute,
 Lunnon Street, Hillcrest, Pretoria, Gauteng 0001,
 Republic of South Africa
FEATURES             Location/Qualifiers
     source          1..694
                     /organism="Eucalyptus urophylla"
                     /mol_type="genomic DNA"

     exons           join (<...315,404..541,665..694)

                     /gene="EuSuSy1"

     introns         join (316..403,542..664)

                     /gene="EuSuSy1"

BASE COUNT      163 a    156 c    163 g    212 t
ORIGIN
      1 cagatgaacc gggtgaggaa tggagagctc taccgctaca tctgtgacac gaagggagtc
     61 ttcgttcaac cggctatcta tgaagctttc gggttgactg tggttgaggc catgacttgt
    121 ggattgccaa cctttgccac ttgcaatggt ggacca gctg agatcattgt gcatggtaaa
    181 tcgggctacc acattgatcc ttaccatggt gacc aggcgg ccgagcttct tgtagatttc
    241 ttcaacaagt gcaagcttga ccagagccac tgggacaaga tctcaaaggg tgccatgcag
    301 agaattgaag agaagtaagc gttttcagat taaaatgatg ttcacttttt ttgaaataca
    361 atttttctaa tttaatttac tctttttttt gcttttgata aggtatacat ggaaaatata
    421 ctctgagagg ctgttgaacc tgactgccgt gtatggcttc tggaagcatg tgactaacct
    481 tgatcggcgc gagagccgcc ggtaccttga aatgttctat gccctcaagt atcgcccact
```

```
    541 ggtaagttcc tgcttgaacc ttatccgatc ctacattctt cattcaaatt ggtgcctggt
    601 tccttggcat acgtagtatg tttctcgcaa tccactcatg ctttgttctt gctcttcctt
    661 gcaggcacag tctgctcctc cggctgtcga gtaa
//
```

III: Partial (5'-end) genomic sequence of the *Eucalyptus urophylla CAD2* gene (Adapted from Maleka et al. 2007). SNP position within enzyme recognition site (underlined) is indicated in bold italic.

```
LOCUS       EuCAD2        1107 bp    DNA linear       30-OCT-2006
DEFINITION  Eucalyptus urophylla cinnamyl alcohol dehydrogenase 2 (CAD2),
partial gene.
ACCESSION   EuCAD2
SOURCE      Eucalyptus urophylla
  ORGANISM  Eucalyptus urophylla
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta;
Tracheophyta;  Spermatophyta;  Magnoliophyta;  eudicotyledons;       core
eudicotyledons; rosids; Myrtales; Myrtaceae; Eucalyptus.
REFERENCE   1  (bases 1 to 1107)
  AUTHORS   Maleka,F.M., Bloomer,P. and Myburg,A.A.
  TITLE     Genetic diversity and linkage disequilibrium in wood
             biosynthetic genes of Eucalyptus urophylla S.T. Blake
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1107)
  AUTHORS   Maleka,F.M., Bloomer,P. and Myburg,A.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-OCT-2006) Department of Genetics, University of
            Pretoria, Faculty of Natural and Agricultural Sciences,
             Forestry and Agricultural Biotechnology Institute,
 Lunnon Street, Hillcrest, Pretoria, Gauteng 0001,
 Republic of South Africa
FEATURES             Location/Qualifiers
     source          1..1107
                     /organism="Eucalyptus urophylla"
                     /mol_type="gene"

     Upstream        1..369

                     /gene="EuCAD2"

     5'UTR           370..485

                     /gene="EuCAD2"

     exons           join (486..572,767..880,967...>)

                     /gene="EuCAD2"

     introns         join (573..766,881..966)

                     /gene="EuCAD2"

BASE COUNT      270 a    253 c    253 g    331 t
ORIGIN
      1 gacagatgga gcgttggatg gagcttctcc atcacttaat ttgtcccttc aagatgaaaa
     61 aagtaagagg tccactgtac caaaacattc ttccacccag aagaaaacca tagtcgctgg
    121 agggagtcaa gcatgtcaga agcacagaaa ctgggaatgg ctaaaaagca agtcttgacc
```

```
   181 cttaacccac cccactggtt cacctaccgc acctggggtt aggtattgct tgctgaggtg
   241 tctgtcactt ttcgccaaag tcatgtctct cttttggatt cttcctattg gtccgtctcg
   301 tttcctcgtt gcaggttgct ggtagcgttt ttgtccatat atatatgcag tccatatgtt
   361 tccccgtcac tcctcatcta tgctcctacc cggcaacttc ccactacgat aagcagcaag
   421 tcttcggctc tgtcgaatct ctctccgagc accactttga aaaaagcttg gatctttgag
   481 caaaaatggg cagtcttgag aaggagagga ccaccacggg ttgggctgca agggacccgt
   541 ctggcgttct ctctccttac acttatagcc tcaggtagat tcaagaactt gccttcttca
   601 ggattgataa agatagctaa gaatctaagt tttcgttgtg cttgtgatgt cgttctttaa
   661 ttcttgtttt tgcttgttcg atcaattacg tattaatcaa tattcgattg attaatctga
   721 ggttatcgac aaaaaagatt tgtctaagtc acttcccaac aaatgcagaa acacgggacc
   781 agaagatctt tacatcaagg tgttgagctg cgggatttgc cacagtgaca ttcaccagat
   841 caagaatgat cttggcatgt cccactaccc tatggttcct gggtaggtct tttcttgcat
   901 taatcatgac taattcttcc tcgtctgtgt ttcttcatat tctaattatt ctttcccctc
   961 tttttcaggc atgaagtggt gggtgaggtt ctggaggtgg gatcagaggt gacaaagtac
  1021 agagttggtg accgagtggg gaccggtata gtggttgggt gctgcagaag ctgtggccct
  1081 tgcaattcgg accaggagca ataccgc
//
```

IV: Partial (3'-end) genomic sequence of the *Eucalyptus urophylla CAD2* gene (Adapted from Maleka et al. 2007). SNP position within enzyme recognition site (underlined) is indicated in bold italic.

```
LOCUS       EuCAD2           888 bp    DNA linear       30-OCT-2006
DEFINITION  Eucalyptus urophylla cinnamyl alcohol dehydrogenase 2 (CAD2),
partial gene
ACCESSION   EuCAD2
SOURCE      Eucalyptus urophylla
  ORGANISM  Eucalyptus urophylla
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta;
 Tracheophyta;    Spermatophyta;    Magnoliophyta;    eudicotyledons;    core
eudicotyledons; rosids; Myrtales; Myrtaceae; Eucalyptus.
REFERENCE   1  (bases 1 to 888)
  AUTHORS   Maleka,F.M., Bloomer,P. and Myburg,A.A.
  TITLE     Genetic diversity and linkage disequilibrium in wood
             biosynthetic genes of Eucalyptus urophylla S.T. Blake
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 888)
  AUTHORS   Maleka,F.M., Bloomer,P. and Myburg,A.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-OCT-2006) Department of Genetics, University of
            Pretoria, Faculty of Natural and Agricultural Sciences,
             Forestry and Agricultural Biotechnology Institute,
 Lunnon Street, Hillcrest, Pretoria, Gauteng 0001,
 Republic of South Africa
FEATURES             Location/Qualifiers
     source          1..888
                     /organism="Eucalyptus urophylla"
                     /mol_type="genomic DNA"

     exons           join (<...51,354..548)

                     /gene="EuCAD2"

     intron          52..353

                     /gene="EuCAD2"

     3'UTR           549..888

                     /gene="EuCAD2"

BASE COUNT      235 a    160 c    201 g    292 t
ORIGIN
      1 actggtgtca tcaatgctcc tcttcaattt atctctccca tggttatgct tggtaaattc
     61 tctatactcc ctttctcttg agcgctgttt ttgaatggat tagtccatgc atcaatgaag
    121 gcataggcag ccacactgca caaggaaatt tatacagcct gtgtaccata tgaaaatcca
    181 ttgtgaagcc tgtcataatt tactctaaaa tggctattac atcattttgt gatcacggtc
    241 cgatgttttt ttgctggcat tttgcgaaca aatgcaaaat cttctcttgg attgacggtc
    301 tttcaaagaa attgtatgtc acctcatttg tgtggttata acatgcaggg aggaagtcaa
```

361 tcactgggag tttcataggg agcatgaagg aaacagagga gatgcttgag ttctgcaaag
421 aaaagggatt gacttcccag atcgaagtga tcaagatgga ttatgtcaac acggccctag
481 agaggctcga gaagaatgat gtcaggtaca ggttcgtcgt ggacgt***t***gcg ggaagcaagc
541 ttgattagtt ttttcctttc cccataatta aacaagaaat cgacgtgctt gt***c***tctcaat
601 tcgagttcct catgccctct gttgtatcat tgtttgttat accgagagta ttattttctt
661 ctgtcttcgt attgaaacca tagacctt***c***t cgattgtgta ttcaatgatg aaggtgttaa
721 tgattttatc acttaagaaa tttgactatt tggattctgg aagcattttg aattgggttg
781 tgctgtgttt ccaagagggg tgtgttttca agaggggtgg gtgagggttc tctttcttga
841 cagtgaccca acaacaaact cggatgaata aaagtgacac gatgtgtt//