Genetic diversity in Western Australian Sandalwood (*Santalum spicatum*)

Report to the Forest Products Commission

Margaret Byrne

Science Division

DEPARTMENT OF
Conservation
AND LAND MANAGEMENT

December 2001

Table of Contents

Introduction

This study investigated the level and structuring of genetic diversity within Western Australian Sandalwood, *S. spicatum* to provide information contributing to management stratgeies for natural populations, and development of an improvement program for planted stands. Nuclear RFLP markers were developed to allow investigation of genetic diversity using co-dominant genetic markers. The diversity in the chloroplast genome was also investigated to give a historical perspective to the genetic structure in the species. This section summarizes the main findings of the study and the implications that arise. The technical detail for the study is provided in two scientific papers, one discussing the nuclear diversity and the other the chloroplast diversity in sandalwood.

Outline of results

Samples were collected from twenty three populations across the geographic range of Sandalwood in both the wheatbelt and rangelands areas. The diversity in the nuclear genome was assessed with RFLP probes which are specific for *Santalum* and were developed for this study. These RFLPs assay random pieces of DNA throughout the nuclear genome of the species. The data generated were analysed for genetic parameters and genetic relationships between populations. The diversity in the chloroplast genome was assessed using specific chloroplast probes. These probes identify changes in the length of DNA fragments that are caused by mutations. The presence and absence of mutations were analysed to determine genetic relationships between individuals and populations.

Nuclear diversity

Sandalwood showed moderate levels of diversity for a woody perennial (see Table 1). Each individual has two alleles at each locus (bit of DNA), one inherited from the mother and one from the father. The two alleles can be the same or different. Heterozygosity is when an individual has two alleles that are different and therefore has greater variation than if the alleles were the same. The number of alleles and the heterozygosity values, were generally lower in the wheatbelt populations than in the rangelands populations. The similarity between the observed heterozygosity and the expected heterozygosity means that the populations are essentially outcrossing with random mating and no inbreeding. However, this reflects the state of the adult population and there may be inbreeding at the point of seed production but the inbred seeds are selected against during juvenile stages of development

Table 1. Genetic diversity parameters for populations of Sandalwood. P, percentage of loci that are polymorphic (0.99 criterion); A, mean number of alleles per locus; Ho, observed heterozygosity; He, Hardy-Weinberg expected panmictic heterozygosity; F_{IS}, mean fixation index over all loci; standard errors in parentheses.

Population	*P*	*A*	H_o	H_e	F_{IS}
Mean – Rangelands	68	2.1	0.216 (0.009)	0.220 (0.007)	-0.041 (0.014)
Mean – Wheatbelt	58	1.8	0.184 (0.013)	0.198 (0.012)	-0.008 (0.018)
Mean – Total	63.5	1.95	0.201 (0.009)	0.210 (0.008)	-0.026 (0.011)

Table 2. Genetic diversity parameters for Sandalwood and other Australian tree species. Parameters for individual species are from RFLP studies, mean of Australian trees is from isozyme studies. H_T, mean estimate of total gene diversity, H_S, mean gene diversity within populations, G_{ST}, relative degree of gene diversity between populations.

Species	H_T	H_S	G_{ST} (%)	Reference
S. spicatum	0.233	0.214	8.3	This study
E. nitens	0.445	0.373	16.2	Byrne et al. 1994
E. kochii	0.514	0.502	2.2	Byrne 1999
E. loxophleba	0.418	0.402	3.8	Hines and Byrne 2001
E. camaldulensis	0.530		7.8	Butcher et al. 2001
A. mangium	0.195	0.130	33.1	Butcher et al. 1998
Australian trees	0.254	0.203	23.0	Moran 1992

Comparison of genetic diversity in sandalwood and other Australian tree species (for which similar analysis has been carried out) is given in Table 2. H_T represents total genetic

diversity and Sandalwood showed lower diversity than the eucalypts but higher diversity than the tropical *Acacia mangium*. The eucalypts have higher levels of diversity because they maintain a higher level of rare alleles of low frequency (which means they have a lot of different variants of that bit of DNA). G_{ST} represents the proportion of diversity that exists between populations. Sandalwood maintained a relatively low level of diversity between populations, similar to the other widespread species *Eucalyptus camaldulensis*. The two eucalypt species with very low diversity between populations, *E. kochii* and *E. loxophleba*, have more localised distributions in Western Australia, and the two species with high levels of diversity between populations, *E. nitens* and *Acacia mangium*, have disjunct distributions which contribute to the high levels of diversity maintained between populations in these species. The level of differentiation between populations was higher in the wheatbelt than in the rangelands.

A test for association between genetic differentiation of populations and their geographic distance was not significant across the whole species ($p = 0.131$) or for the populations in the wheatbelt ($p = 0.379$). However, it was significant for populations in the rangelands ($p = 0.021$). This indicates that the rangelands populations have an equilibrium between drift and gene flow, whereas the wheatbelt populations did not show equilibrium. Drift (differences due to isolation) and gene flow (dispersal of pollen or seeds between populations) effect evolutionary processes which influence genetic structure. Drift and gene flow act as opposite forces, with drift increasing the differentiation of populations and gene flow decreasing the differentiation. Drift-gene flow equilibrium in the rangelands means that this region has experienced a stable environment for a long period of time. The pattern in the wheatbelt suggests more recent expansion in this area and/or fragmentation and isolation of populations in the region.

A diagram of the relationships between populations is shown in Figure 1. There is no strong evidence of genetic structuring within *S. spicatum*, although the rangelands populations showed a greater degree of association than the wheatbelt populations. The majority of rangelands populations were grouped into a cluster, although two wheatbelt populations were also clustered in this group. Some pairs of wheatbelt populations showed similarities but the majority of populations were divergent and showed little association.

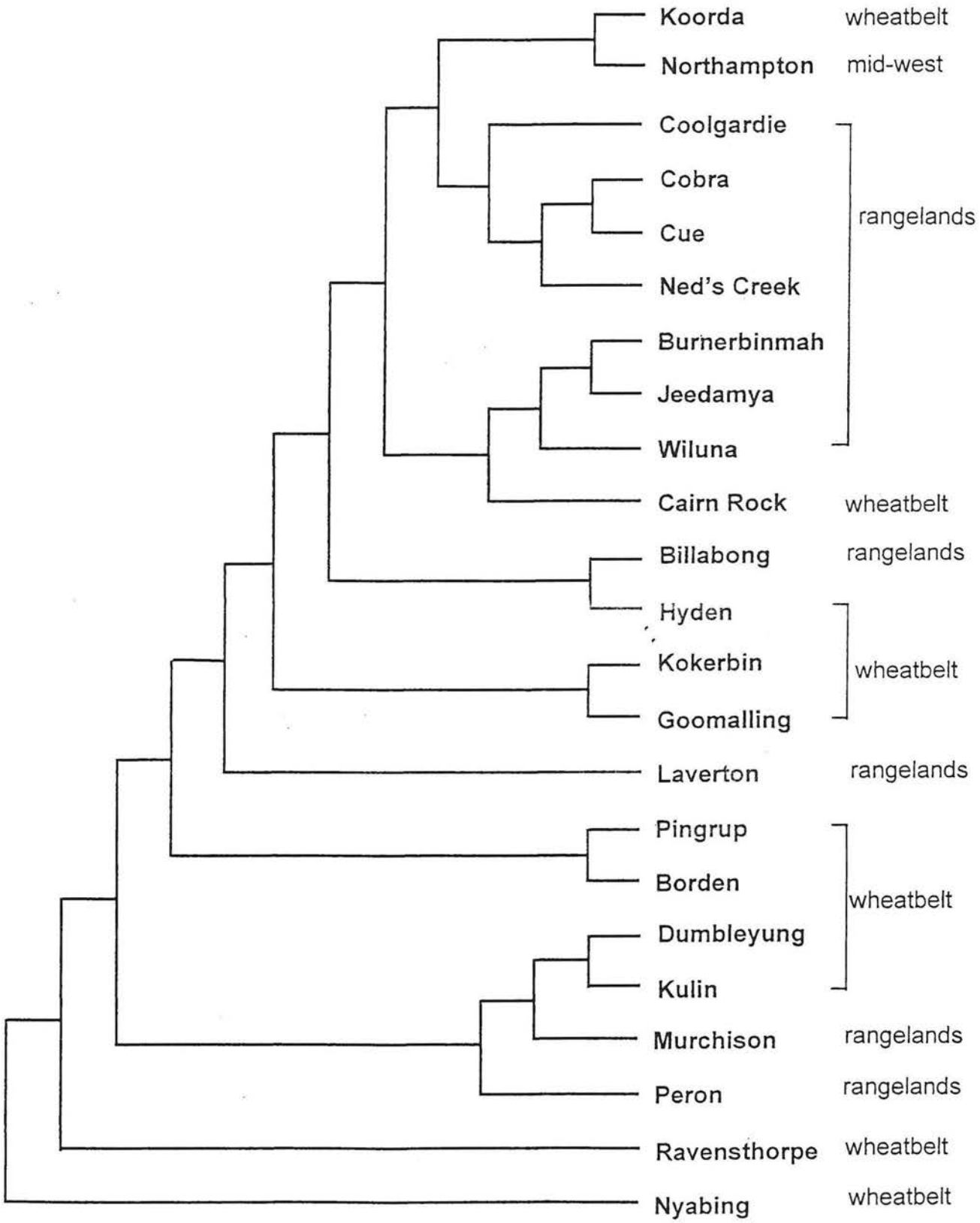


Fig. 1 Consensus tree for population phylogeny (continuous character maximum likelihood) analysis of populations of *Santalum spicatum*.

Chloroplast diversity

A moderate level of chloroplast diversity was detected in sandalwood. There were 14 mutations which were distributed over eleven haplotypes. A haplotype is a set of mutations found in the same individuals. There was one haplotype (B) that was more common than all others, occurring in 45% of the individuals sampled mainly from the wheatbelt populations. The level of diversity in the chloroplast genome of *S. spicatum* is high compared to other woody perennials but not as high as has been identified in other Australian tree species, including the main host, *Acacia acuminata* (Byrne et al. 2001).

The relationships between the haplotypes show two groups (Figure 2), one containing the wheatbelt populations plus the rangelands populations of Burnerbinmah and Coolgardie, and the other the remaining populations from the rangelands. The group of wheatbelt populations shows little divergence within the group, but the group of rangelands populations shows greater divergence between the haplotypes in the group. A nested clade analysis identified that the geographic structuring into two groups in the phylogeny was significant (p=0.000). Analysis of the pattern of relationships leads to an inference of past fragmentation as the main influence causing the geographic structuring of these groups.

Summary

In summary Sandalwood has moderate levels of genetic diversity with a significant level of differentiation between populations but no strong genetic structuring within the species. Whilst the wheatbelt and rangelands populations are not highly differentiated there are differences in the evolutionary history of these two regions and therefore differences in the genetic responses of the populations in the regions. The genetic analysis indicates that the rangelands populations have existed in a stable environment for a long period of time. The populations have higher levels of diversity than the wheatbelt populations, and higher levels of gene flow through pollen dispersal, although seed dispersal is limited to populations within broad areas. In contrast, the wheatbelt populations have had a more recent expansion of the species range, although they have been in place sufficiently long to experience fragmentation and isolation of populations. These populations have lower levels of diversity than the rangelands populations and greater divergence between populations due to more limited gene flow. The greater differentiation within the wheatbelt may also be related to the greater environmental gradiant across the wheatbelt compared to

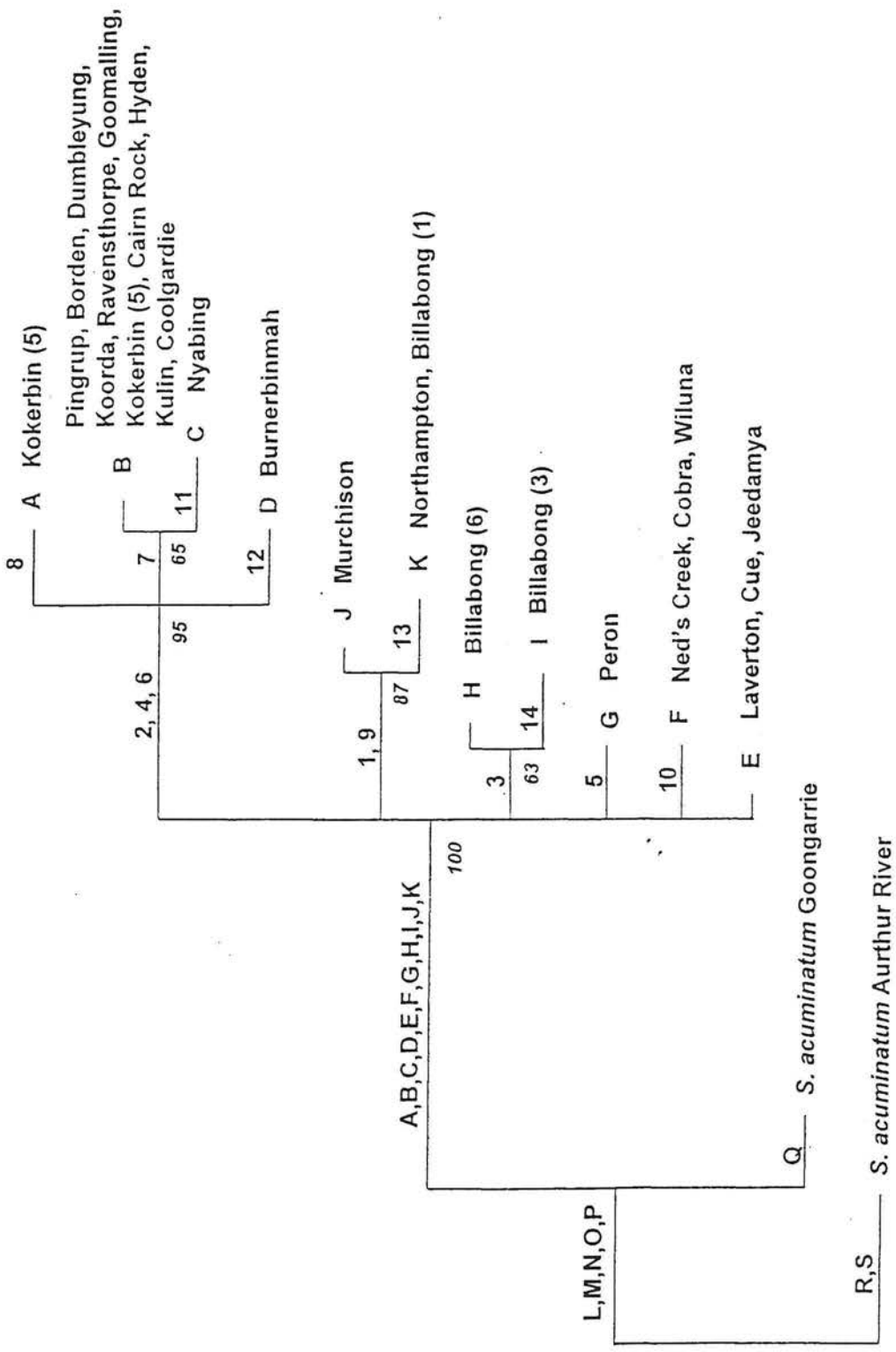


Fig. 2. Phylogenetic tree of haplotype relationships in *Santalum spicatum*. Numbers above lines refer to mutations. Numbers in italics below lines at nodes represent bootstrap confidence values (%). Numbers in brackets indicate number of individuals in each population with that haplotype.

the rangelands. The morphological differences between the rangelands and wheatbelt ecotypes most likely represent adaptation to environmental conditions, but may also reflect some effect of the different genetic processes associated with drift and gene flow that have occurred due to the divergence of the regions.

Implications

The results of this study have implications for the domestication of sandalwood, and for the management of natural populations of this important genetic resource. Whilst the level of differentiation between the rangelands and the wheatbelt forms of sandalwood are not large the different genetic history in the wheatbelt and the rangelands suggests that it may be prudent to establish planted stands in the wheatbelt region from germplasm collected from wheatbelt populations. This would also avoid any concerns about the possible genetic contamination of natural populations of sandalwood in the wheatbelt. There is sufficient diversity across the wheatbelt populations to ensure a broad genetic base for germplasm collections. The degree of differentiation of populations in the wheatbelt suggests that sampling strategies should encompass all of the current populations, particularly the more divergent southern most populations, to capture the maximum genetic diversity. The level of differentiation between populations also suggests that crossing between populations through establishment of seed orchards could lead to a heterosis effect which may be manifest in superior growth. Although there was no evidence of inbreeding in the adult populations that were assayed in this study, the low levels of diversity, particularly in the wheatbelt populations, suggest that inbreeding may be a significant factor in seed crops from small populations.

Regeneration in populations in the rangelands region is one of the issues for management strategies of these populations. The level of differentiation between populations in the rangelands suggests that seed used in regeneration of natural populations should be used from local areas so as not to disturb the balance between drift and gene flow that occurs in these populations. There is considerable debate about how to define "local area" and such definitions are often quite arbitrary. However, in order to give some guidance a recommendation would be that seed collections be made from within a 10km radius of the place of planting. This is based on the maximum distances that rare long distance gene flow may be expected to occur in Sandalwood.

References

Butcher PA, Moran GF, Perkins HD (1998) RFLP diversity in the nuclear genome of *Acacia mangium*. Heredity 81: 205-213.

Butcher PA, Otero A, McDonald MW, Moran GF (2001) Nuclear RFLP variation in *Eucalyptus camaldulensis* Dehnh. from northern Australia. Heredity in press.

Byrne M (1999) High genetic identities between three oil mallee taxa, *Eucalyptus kochii* ssp. *kochii* ssp. *plenissima* and *E. horistes*, based on nuclear RFLP analysis. Heredity 82: 205-211.

Byrne, M, Macdonald, B, Coates D (2001) Phylogeographic patterns within the *Acacia acuminata* (Leguminosae: Mimosoideae) complex in Western Australia. *J. Evol. Biol.* submitted

Byrne M, Parrish TL, Moran GF (1998) Nuclear RFLP diversity in *Eucalyptus nitens*. Heredity 81: 225-232

Hines B, Byrne M (2001) Genetic differentiation between mallee and tree forms in the *Eucalyptus loxophleba* complex. Heredity in press.

Moran GF (1992) Patterns of genetic diversity in Australian tree species. New For 6: 49-66.

Genetic diversity in Australian Sandalwood (*Santalum spicatum*) as revealed by nuclear RFLP analysis

M. Byrne, B. Macdonald, L. Broadhurst, J. Brand

Science Division, Department of Conservation and Land Management, Locked Bag 104, Bentley Delivery Centre, WA 6983, Australia.

Communicating author
M. Byrne. Science Division, Department of Conservation and Land Management, Locked Bag 104, Bentley Delivery Centre, WA 6983, Australia. Phone +618 93340503, Fax +618 93340515. Email margaretb@calm.wa.gov.au

Abstract

Western Australian sandalwood, *Santalum spicatum*, is harvested for sandalwood oil. The species is widespread in the semi-arid and arid regions of Western Australian and there is some morphological variation suggestive of two ecotypes. The level and structuring of genetic diversity within the species was investigated using anonymous nuclear RFLP loci. *Santalum spicatum* showed moderate levels of genetic diversity compared to other Australian tree species. The northern populations in the arid region showed greater levels of diversity and less population differentiation than the southern populations in the semi-arid region. Equilibrium between drift and gene flow in the northern populations indicated that these populations have been established for a long period of time with stable conditions conducive to gene flow. In contrast the southern populations showed a relationship between drift and gene flow indicative of a pattern of fragmentation and isolation where drift has greater effect than gene flow. The different pattern of diversity between the two regions suggests that the ecotypes have experienced differences in the relative influences of genetic processes during their evolutionary history.

Key Words: genetic diversity, RFLP, Sandalwood, equilibrium

Introduction

Sandalwood is highly valued for the sequiterpene essential oil contained in the heartwood (Adams et al. 1975) which is mainly used in the perfume industry, but is also used for incense and joss stick manufacture and for carving (Rai 1990). The majority of commercial sandalwood is obtained from the Indian Sandalwood (*Santalum album*) which is harvested from natural populations and plantations in India and Asia (Srinivasan et al. 1992). The Western Australian Sandalwood (*S. spicatum*) also produces aromatic wood similar to Indian sandalwood (Adams et al. 1975), and is exported from Western Australian to several Asian countries where it is used mainly in the manufacture of joss sticks (Jones 2001). There are several other species of *Santalum* that occur in Australia but they do not produce sufficient quantity of the sesquiterpenes that make up the Sandalwood fragrance to be commercially significant (Shea et al. 1998).

Western Australian sandalwood has a broad distribution throughout the semi-arid (approximately 300 – 600 mm rainfall) and arid (approximately 150 – 300 mm rainfall) areas of Western Australia but is absent from the high rainfall areas in the extreme south-west and the tropical areas in the north. It is a slow growing root hemi-parasite tree (Hewson and George 1984) which commonly occurs with nitrogen fixing plants such as *Acacia* or *Allocasuarina* that it uses as hosts (Loneragan 1990). Experimental data has shown *Acacia acuminata* to be a preferred host (Brand et al. 2000). Extensive harvesting of sandalwood, combined with poor recruitment and slow growth rate, have reduced the extent of naturally occurring Western Australian sandalwood and there is interest in commercial plantings of sandalwood in the agricultural areas of Western Australia (Shea et al. 1998).

Western Australian sandalwood occurs across a wide range of environmental conditions and shows morphological variation suggesting the presence of two ecotypes within the species (Fox and Brand 1993). In the higher rainfall areas in the south of the distribution the species exhibits larger leaves and nuts, higher chlorophyll content and more of a tree habit than in the arid areas in the north of the distribution (Fox and Brand 1993). Trees in the arid areas also show higher concentration of oil in the wood (Loneragan 1990). This morphological variation probably reflects adaptation to environmental conditions but may also have some genetic basis. Variation in growth rate between provenances has been observed in a field trial (Brand et al. 1999) indicating some genetic influence on growth

characteristics, and genetic differentiation of two lineages has been observed in cpDNA (Byrne unpubl. data).

Domestication and development of improvement programs are enhanced by a knowledge of the genetic structure of species in order to capture high levels of diversity in breeding populations and seed orchards. Appropriate management of natural populations and development of in situ conservation of genetic resources also requires information on the amount and distribution of genetic diversity. The domestication of Western Australian sandalwood for agroforestry will involve the development of an improvement program, and knowledge of genetic diversity and structure within the species will provide a basis for a selection program to identify superior germplasm. It will also provide information for the conservation of this important genetic resource, and the management of natural populations that are still harvested for sandalwood production. This study investigates the level and structuring of genetic diversity within *S. spicatum*. Isozyme analysis has proved unsuitable in this species (Brand 1993) therefore nuclear RFLP markers were developed to allow investigation of genetic diversity using co-dominant genetic markers.

Materials and Methods

Genomic library construction

A random genomic library was constructed from *S. spicatum* using standard methods (Sambrook et al. 1989). Total genomic DNA from a single sandalwood tree was digested with *Pst*I, treated with Gene Clean (Geneworks) then ligated into *Pst* I digested, dephosphorylated pUC19 plasmid. Plasmids were transformed into competent cells of the bacterial strain Dhα101 (Gibco) by heat shocking. Transformed colonies containing recombinant plasmids were selected using ITPG/X-gal screening and maintained as glycerol stocks at –80°C. Probes were prepared by PCR amplification of the insert using M13 universal primers and cleaned by passing through filter tips at 6000rpm in a microfuge, followed by ethanol precipitation. To screen for high copy DNA sequence, plasmid inserts were separated on 0.8% agarose gels, Southern blotted and hybridized with total genomic DNA. Inserts containing high copy sequence were discarded. Probe evaluation was carried out by screening 103 probes against DNA from six different individuals digested with each of three restriction enzymes, *Bgl* II, *Eco* RI, or *Eco* RV. Twenty-eight probe – enzyme combinations that showed scorable single locus patterns were selected to assess the diversity of populations.

Plant material and DNA procedures

Leaf samples were collected from ten individuals, from each of 23 populations across the geographic range of the species in both the southern (semi-arid) and northern (arid) areas (Figure 1). Samples were also collected from ten individuals from each of two populations (one southern, one north-eastern) of the related species *S. acuminatum* (Quandong). DNA was extracted from the leaves of the 250 individuals following the methods outlined in Byrne et al. (1994) with the addition of 0.1M sodium sulphite to the extraction buffers (Byrne et al. 2001). DNA samples were digested with either *Bgl* II, *Eco* RI, or *Eco* RV, Southern blotted and hybridized with 28 RFLP probes. Hybridization conditions were as in Byrne and Moran (1994).

Data analysis

Data analysis was carried out at the species level for all populations and also at a regional level (southern/semi arid region versus northern/arid region) where appropriate. The population at Northampton has a semi-arid, coastal environment but is in the northern area and was included in the northern group of populations for the regional analysis. Alleles were numbered at each locus according to the size of the fragment, with the largest fragment designated allele 1. Allelic diversity parameters were calculated using Popgene (Yeh et al. 1997). Gene diversity parameters, mean estimate of total genetic diversity (H_T), mean gene diversity within populations (H_S), mean gene diversity between populations (D_{ST}) and the relative degree of gene diversity between populations (G_{ST}), were calculated according to Nei (1987) using FSTAT. Population differentiation (θ) was calculated using FSTAT (Goudet 2000) for all populations and for the northern and southern populations separately. Regression of pairwise population differentiation against geographic distance was calculated and tested for significance using a Mantel randomisation test with 1000 iterations (Mantel 1967) for all populations, and for the northern and southern populations separately. Where a significant result was obtained the residuals of the regression were also tested against geographic distance. Population relationships were determined using a continuous character maximum likelihood analysis in Phylip (Felsenstein 1993). The southern population of Nyabing was specified as the outgroup as it was one of the three populations that had the greatest average genetic distance from the rest of the populations. The significance of the population relationships was assessed by bootstrapping with 100

replications. Banding patterns in the two populations of *S. acuminatum* were unable to be scored and were not included in the data analysis.

Results.

Information derived from twenty-seven probes was used in the analysis, since one probe produced a banding pattern that could not be reliably scored. Polymorphism in *S. spicatum* was high, with only one locus monomorphic in all populations. There were no loci that were polymorphic in all populations. The number of alleles detected at a locus ranged from one to 10 with the maximum number of alleles in any one population being six. The distribution of alleles in frequency classes showed that the majority of alleles were common with frequency >0.5 (mean of all populations = 51%) and the proportion of rare alleles of frequency <0.1 was much lower (mean of all populations = 17%).

Measures of allelic diversity among populations of *S. spicatum* showed moderate levels of diversity for a woody perennial. The number of alleles per locus (A), observed heterozygosity (Ho), and heterozygosity expected under Hardy-Weinberg equilibrium (He), were generally lower in the southern populations than in the northern populations, and were significantly lower than the majority of populations in two southern populations, Goomalling and Borden (Table 1). Borden is the southern most population at the edge of the species range, and is a very small population. The inbreeding coefficient, F_{IS}, was small and negative for most populations suggesting a possible heterozygote advantage.

The mean estimates of total genetic diversity (H_T) showed moderate levels of diversity with the majority of this diversity within populations (H_S), although some diversity was maintained between populations (G_{ST} = 8.3%) (Table 2). Although diversity between populations was not high, the level of population differentiation was significantly different from zero (θ = 0.087, 95% confidence interval 0.067 – 0.108). The level of population differentiation was higher in the southern populations (θ = 0.108, 95% confidence interval 0.079 – 0.140) than in the northern populations (θ = 0.055, 95% confidence interval 0.038 – 0.074). There was no significant association between population differentiation (F_{ST}) and geographic distance across the whole species as assessed by a Mantel test (g = 1.213, p = 0.131). There was also no significant association for the populations in the south (g = 0.276, p = 0.379), however there was a significant association for populations in the north

($g = 2.167$, $p = 0.021$). For the northern populations the residuals from the regression also showed a highly significant association with geographic distance ($g = 7.546$, $p = 0.001$).

Unbiased estimates of genetic distance between all pairwise comparisons of *S. spicatum* populations were low, with the highest distance (0.073) occurring between the southern populations at Nyabing and Pingrup. The lowest genetic distance (0.006) occurred between the northern populations of Ned's Creek and Cue. The average genetic distance across the whole species was low (0.057).

A population phylogeny based on maximum likelihood analysis showed no strong evidence of genetic structuring within *S. spicatum* (Figure 2), although the northern populations showed a greater degree of association than the southern populations. The majority of northern populations were grouped into a clade, although two southern populations were also clustered in this clade. Some pairs of wheatbelt populations showed similarities but the majority of populations were divergent and showed little association. None of the nodes were significant with all bootstrap values less than 40% (not shown).

The related species *S. acuminatum*, showed multiple bands that were not consistent with a standard diploid allelic pattern. The Arthur River population (south) showed up to six bands at some loci which were all identical between individuals. The Goongarrie population (north-east) showed up to four bands in each individual and these were variable between individuals.

Discussion

The level of genetic diversity in *S. spicatum* is lower than that reported for most other Australian tree species that have been assessed using RFLP loci (Table 2). *Santalum spicatum* showed lower levels of diversity than the eucalypt species studied but higher diversity levels than *Acacia mangium*. The lower level of gene diversity in *S. spicatum* compared to eucalypts is probably related to the distribution of allele frequency. *Santalum spicatum* had a high level of common alleles and low levels of rare alleles compared to eucalypt species where the frequency of alleles is skewed towards rare alleles (*E. kochii* allele frequency $<0.5 = 11\%$ allele frequency $<0.1 = 66\%$ (Byrne 1999); *E. loxophleba* allele frequency $<0.5 = 22\%$ allele frequency $<0.1 = 33\%$ (Hines and Byrne 2001); *E. nitens* allele frequency $<0.5 = 19\%$ allele frequency $<0.1 = 28\%$ (Byrne et al. 1998)).

Santalum spicatum maintained a relatively low but significant proportion of diversity between populations, similar to the other widespread species *E. camaldulensis*. The two eucalypt species with very low diversity between populations have more localised distributions in Western Australia, and *E. nitens* and *A. mangium* have disjunct distributions which contribute to the high levels of diversity maintained between populations in these species. The level of diversity maintained between populations of *S. spicatum* was lower than the average for all other Australian tree species studied with isozymes (Table 2).

Within *S. spicatum* there is evidence for some genetic differences between the northern and southern regions. Populations in the north showed higher levels of genetic diversity and less differentiation between populations compared with those in the south. The northern populations also showed isolation by distance which is indicative of an equilibrium between drift and gene flow, whereas the southern populations did not show equilibrium. Drift and gene flow, as well as population size, affect evolutionary processes which influence genetic structure. Drift and gene flow act as opposite forces, with drift increasing the differentiation of populations and gene flow decreasing the differentiation. When populations are at equilibrium between gene flow and drift, genetic differentiation is inversely related to gene flow. However, populations will only reach an equilibrium between gene flow and drift if suitable conditions exist for a long enough period of time (Hutchinson and Templeton 1999). Species where population structure has been influenced by recent historic events, such as range expansion since glaciation, will not be expected to show drift-gene flow equilibrium (Schaal and Olsen 2000). However, species occupying ancient landscapes that have had a stable evolutionary history, such as that in south-west Western Australia, may be expected to show drift-gene flow equilibrium and several species in south-west Western Australia do show drift-gene flow equilibrium (Coates 2000).

Detection of drift-gene flow equilibrium in the northern region implies that the population structure in this region has been stable for a long period of time. In contrast, lack of equilibrium in the southern region suggests that either the region has not been occupied for sufficient time to allow equilibrium to be reached or that stable conditions have not been present throughout the history of the region. Either of these scenarios are feasible. The first would suggest that the southern populations represent a recent expansion of the species

from the longer established populations in the northern region. The second scenario is consistent with proposals of climatic instability from cyclic expansion and contraction of the mesic and arid zones during the Plieostocene, leading to fragmentation and isolation in the intermediate areas (Hopper et al 1996). The southern region is between the mesic and arid zones of south-west western Australia in the region where the influences of climatic instability have been predicted and therefore would have experienced reduced gene flow due to historic fragmentation and isolation. The variance in estimates of divergence for the southern region show a wide degree of scatter independent of geographic distance which also indicates that drift has been more influential than gene flow, and this would arise if populations in the region had been fragmented into small isolated populations (Hutchinson and Templeton 1999). The population phylogeny is consistent with higher levels of differentiation between southern populations than between the northern populations. In contrast, the northern region would not have experienced such climatic instability and would have maintained suitable conditions for gene flow. This difference in genetic patterns between the regions implies that there have been differences in the relative influences of genetic processes in the regions, and supports the identification of southern and northern ecotypes based on morphological variation. The pattern of differentiation in the southern region suggests that there may be variation in quantitative traits for which genetic gain could be achieved by a program of selection and breeding for sandalwood in this region.

The genetic patterns observed in the related Quandong, *S. acuminatum*, are not consistent with a diploid genome and suggest that this species may be a polyploid. The patterns observed are not due to degradation or poor quality DNA since the same filters used in this study have been hybridized with chloroplast probes and there is no inconsistency between the cpDNA banding patterns observed in *S. acuminatum* and those in *S. spicatum*. There were also differences between the two populations of *S. acuminatum* in their genetic patterns. The patterns in the Goongarrie (north-east) population were consistent with the individuals being tetraploid, while the patterns in the Arthur River (southern) population had a greater number of RFLP bands suggestive of higher order polyploidy such as hexaploids. Chromosome counts for *S. spicatum* and *S. acuminatum* have not been made but *S. album* has been reported as having a diploid chromosome number of 2n=20 (Kulkarni et al. 1998). There was also no variation between the individuals in the Arthur River population which suggests that there may be fixed heterozygosity at duplicate loci in

this population, or alternatively that all individuals are representative of a single clone. The Arthur River population is small and the ten individuals sampled were collected over a geographic range of 30 m and could be stems of the same clone. Clonality has also been observed in remnant isolated populations of *S. lancelolatum* at the edge of the species range in northern Victoria (Warburton et al. 2000). The different genetic patterns observed between the two populations of *S. acuminatum* suggests that there is significant genetic differentiation within this species that would warrant further investigation.

Acknowledgments

We thank the Forest Products Commission of Western Australia for funding this study.

References

Adams DR, Bhatnagar SP, Cookson RC (1975) Sequiterpenes of *Santalum album* and *Santalum spicatum*. Phytochemical Reports 14:1459-1460

Brand JE (1993) Preliminary observations on ecotypic variation in *Santalum spicatum* 2. Genotypic variation. Mulga Research Journal 11: 13-19.

Brand JE, Crombie DS, Mitchell, M.D (2000) Establishment and growth of sandalwood (*Santalum spicatum*) in south-western Australia: the influence of host species. Aust For 63: 60-65

Brand JE, Ryan PC, Williams MR (1999) Establishment and growth of sandalwood (*Santalum spicatum*) in south-western Australia: the Northampton pilot trial. Aust For 62: 33-37

Butcher PA, Moran GF, Perkins HD (1998) RFLP diversity in the nuclear genome of *Acacia mangium*. Heredity 81: 205-213.

Butcher PA, Otero A, McDonald MW, Moran GF (2001) Nuclear RFLP variation in *Eucalyptus camaldulensis* Dehnh. from northern Australia. Heredity in press.

Byrne M (1999) High genetic identities between three oil mallee taxa, *Eucalyptus kochii* ssp. *kochii* ssp. *plenissima* and *E. horistes*, based on nuclear RFLP analysis. Heredity 82: 205-211.

Byrne M, Macdonald B, Francki M. (2001) Incorporation of sodium sulphite into extraction protocol minimizes degradation of *Acacia* DNA. BioTechniques. 30: 742-744

Byrne M, Moran GF (1994) Population divergence in the chloroplast genome of

Eucalyptus nitens. Heredity 73: 18-28

Byrne M, Parrish TL, Moran GF (1998) Nuclear RFLP diversity in *Eucalyptus nitens*. Heredity 81: 225-232

Coates DJ (2000) Defining conservation units in a rich and fragmented flora: implications for management of genetic resources and evolutionary processes in south-west Australian plants. Aust J Bot 48: 329-339.

Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package) version 3.5c. Department of Genetics, The University of Washington, Seattle.

Fox JED, Brand JE (1993) Preliminary observations on ecotypic variation in *Santalum spicatum* 1. Phenotypic variation. Mulga Research Journal 11: 1-12

Hewson HJ and Geogre AS 1984 Flora of Australia Vol. 22. Australian Government Publishing Service, Canberra, pp 61 – 66.

Hines B, Byrne M (2001) Genetic differentiation between mallee and tree forms in the *Eucalyptus loxophleba* complex. Heredity in press.

Hopper SD, Harvey MS, Chappill JA, Main AR, Main BY (1996) The Western Australian biota as Gondwanan heritage - a review. In: S.D. Hopper, J.A. Chappill, M.S. Harvey and A.S. George (eds) Gondwanan Heritage. Past present and Future of the Western Australian Biota. Surrey Beatty & Sons, Chipping Norton, pp. 1-46..

Hutchinson DW, Templeton AR (1999) Correlation of pairwise genetic and geographic distance measures inferring the relative influences of gene flow and drift on the distribution of genetic variability. Evol 53: 1898-1914.

Jones P (2001) Sandalwood re-visited in Western Australia. Sandalwood Research Newsletter 12: 3-4.

Kulkarni RS, Fakrudin B, Shashidhar KS (1998) Tree improvement efforts in sandal: the need to employ novel strategies. In: Radomiljac AM, Ananthapadmanabho HS, Welbourn RM, Satyanarayana Rao K (eds) Sandal and it's products ACIAR, Canberra, pp 151-153.

Loneragan OW 1990 Historical review of sandalwood (*Santalum spicatum*) research in Western Australia. Research Bulletin No. 4 Department of Conservation and Land Management, Perth.

Mantel N. (1967) The detection of disease clustering and a generalised regression approach. Cancer Res 27: 209-220.

Moran GF (1992) Patterns of genetic diversity in Australian tree species. New For 6: 49-66.

Nei M (1987) Molecular Evolutionary Genetics. Columbia University Press, New York.

Rai SN (1990) Status and cultivation of sandalwood in India. In: Hamilton L, Conrad CE, (eds). Sandalwood in the Pacific. USDA Forest Service Technical Report PSW-122, pp 66-71.

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Schaal BA, Olsen KM (2000) Gene genealogies and population variation in plants. Proc Nat Acad Sci USA 97: 7024-7029.

Shea SR, Radomiljac, Brand J, Jones P. (1998) An overview of sandalwood and the development of sandal in farm forestry in Western Australia. In: Radomiljac AM, Ananthapadmanabho HS, Welbourn RM, Satyanarayana Rao K (eds) Sandal and it's products ACIAR, Canberra, pp 9-15.

Srinivasan VV, Sivaramakrishnan VR, Rangaswamy CR, Ananthapadmanabha HS, Shankaranarayana KH (1992) Sandal (*Santalum album* L.) Indian Council of Forestry Research and Education, Dehra Dun.

Warburton CL, James EA, Fripp YJ, Trueman SJ, Wallace HM 2000. Clonality and sexual reproductive failure in remnant populations of *Santalum lanceolatum* (Santalaceae). Biol Cons 96: 45-54.

Yeh FC, Yang R-C, Boyle BJ, Ye Z-H, Mao JX (1997) POPGENE, the user-friendly shareware for population genetic analysis. Molecular Biology and Biotechnology Centre, University of Alberta, Canada.

List of Figures

Figure 1. Location of sampled populations of *Santalum spicatum* (●) and *Santalum acuminatum* (▲) Dotted line indicates separation of rangelands (north) and wheatbelt (south) regions.

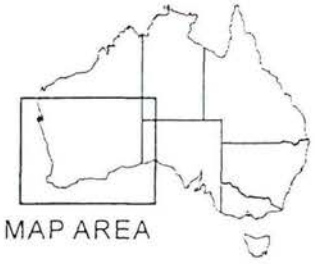
Figure 2. Consensus tree for population phylogeny (continuous character maximum likelihood) analysis of populations of *Santalum spicatum*.

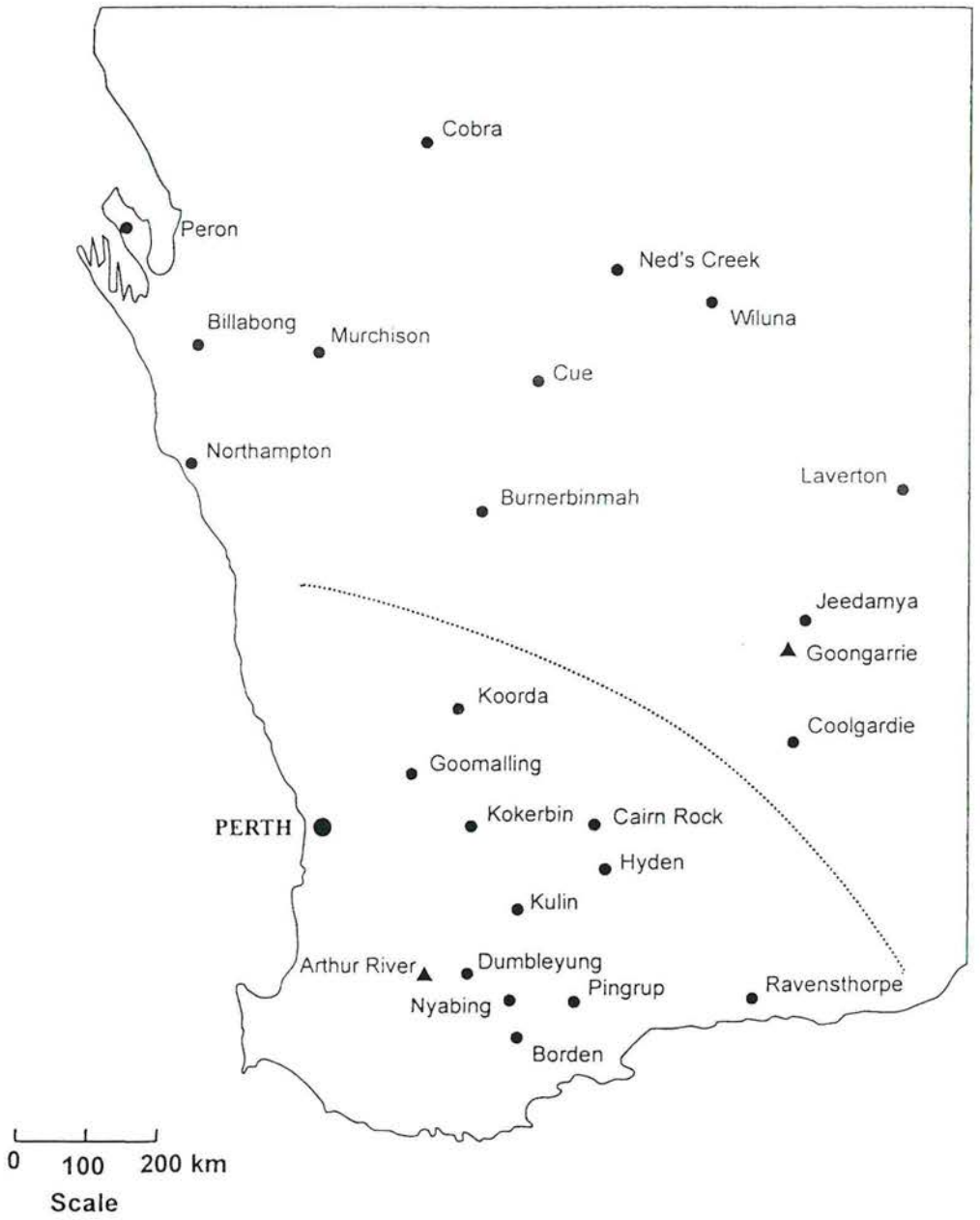
Table 1. Allelic diversity parameters for populations of *Santalum spicatum*. P, percentage of loci that are polymorphic (0.99 criterion); A, mean number of alleles per locus; Ho, observed heterozygosity; He, Hardy-Weinberg expected panmictic heterozygosity; F_{IS}, mean fixation index over all loci; standard errors in parentheses.

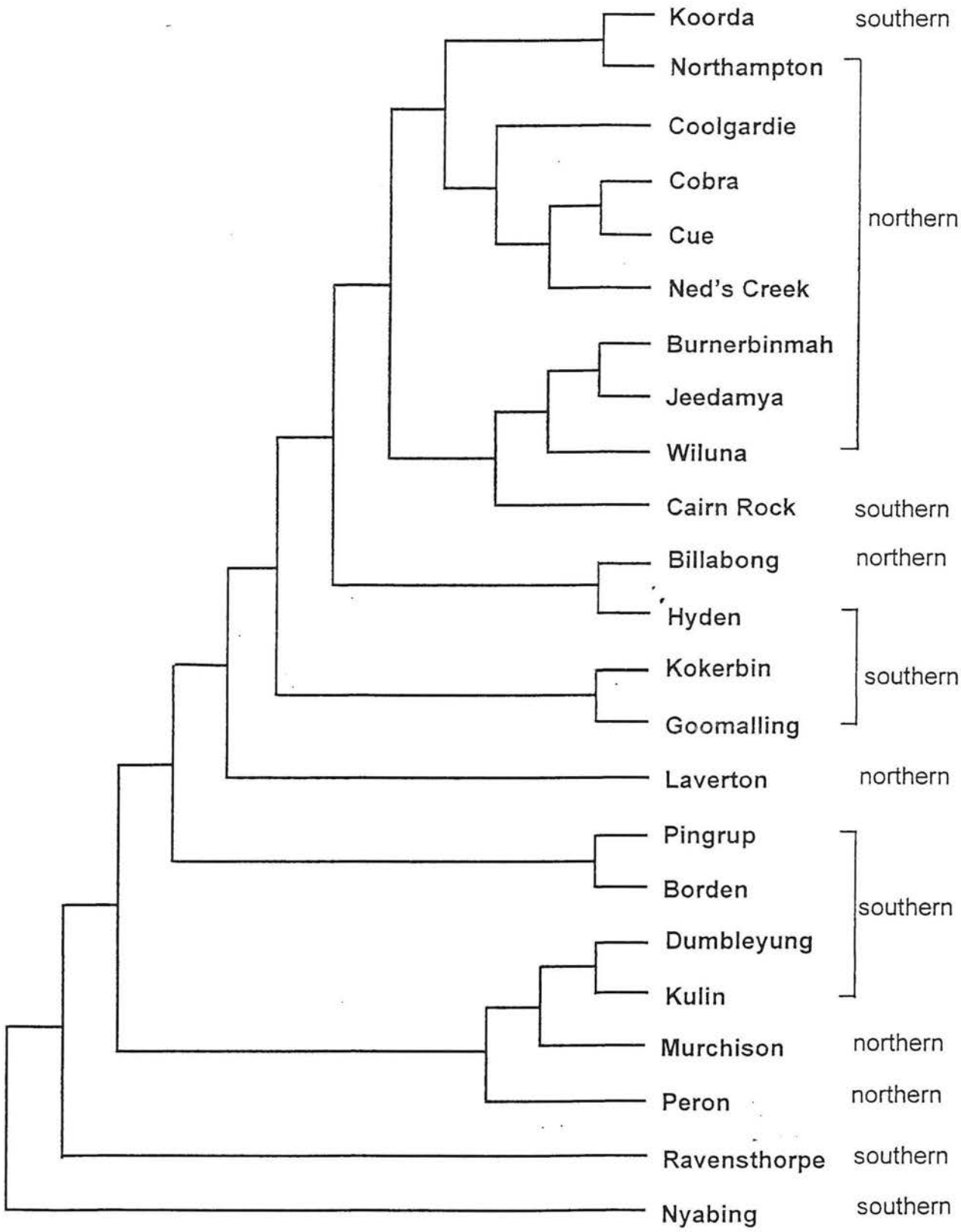
Population	*P*	*A*	H_o	H_e	F_{IS}
Cobra	70	2.2	0.236 (0.018)	0.247 (0.017)	-0.041 (0.027)
Peron	67	2.1	0.230 (0.017)	0.234 (0.017)	-0.087 (0.029)
Ned's Creek	70	2.6	0.226 (0.017)	0.230 (0.017)	-0.050 (0.028)
Wiluna	75	2.0	0.180 (0.015)	0.206 (0.017)	0.005 (0.027)
Murchison River	70	2.0	0.197 (0.015)	0.212 (0.016)	-0.031 (0.027)
Billabong R	56	1.9	0.185 (0.018)	0.171 (0.017)	-0.114 (0.029)
Cue	67	1.9	0.221 (0.017)	0.205 (0.017)	-0.106 (0.024)
Laverton	67	2.1	0.278 (0.020)	0.2696 (0.018)	0.021 (0.042)
Northampton	67	2.0	0.222 (0.019)	0.2226 (0.017)	-0.010 (0.036)
Burnabinmah	70	2.0	0.160 (0.014)	0.1875 (0.016)	0.007 (0.027)
Jeedamya	70	2.3	0.237 (0.017)	0.2316 (0.017)	-0.086 (0.025)
Coolgardie	70	1.9	0.224 (0.017)	0.2280 (0.017)	-0.003 (0.030)
Mean – Rangelands	68	2.1	0.216 (0.009)	0.220 (0.007)	-0.041 (0.014)
Koorda	70	2.1	0.237 (0.017)	0.243 (0.017)	-0.039 (0.025)
Goomalling	48	1.6	0.126 (0.016)	0.152 (0.016)	0.084 (0.050)
Cairn Rock	63	1.9	0.167 (0.016)	0.188 (0.016)	0.008 (0.038)
Kokerbin	67	1.8	0.174 (0.016)	0.200 (0.016)	0.061 (0.030)
Hyden	60	1.7	0.217 (0.018)	0.209 (0.0172)	-0.102 (0.023)
Kulin	63	2.0	0.195 (0.016)	0.216 (0.0165)	0.017 (0.036)
Dumbleyung	63	1.8	0.157 (0.015)	0.193 (0.0160)	0.071 (0.035)
Pingrup	52	1.7	0.191 (0.018)	0.198 (0.0173)	-0.052 (0.045)
Nyabing	63	1.9	0.225 (0.019)	0.228 (0.0181)	-0.076 (0.032)
Ravensthorpe	63	2.0	0.235 (0.02)	0.243 (0.0191)	-0.021 (0.033)
Borden	30	1.4	0.104 (0.017)	0.114 (0.0165)	-0.042 (0.084)
Mean – Wheatbelt	58	1.8	0.184 (0.013)	0.198 (0.012)	-0.008 (0.018)
Mean - Total	63.5	1.95	0.201 (0.009)	0.210 (0.008)	-0.026 (0.011)

Table 2. Gene diversity parameters for *S. spicatum* and other Australian tree species. Parameters for individual species are from RFLP studies, mean of Australian trees is from isozyme studies. H_T, mean estimate of total genetic diversity, H_S, mean gene diversity within populations, G_{ST}, relative degree of gene diversity between populations.

Species	H_T	H_S	G_{ST} (%)	Reference
S. spicatum	0.233	0.214	8.3	This study
E. nitens	0.445	0.373	16.2	Byrne et al. 1998
E. kochii	0.514	0.502	2.2	Byrne 1999
E. loxophleba	0.418	0.402	3.8	Hines and Byrne 2001
E. camaldulensis	0.530		7.8	Butcher et al. 2001
A. mangium	0.195	0.130	33.1	Butcher et al. 1998
Australian trees	0.254	0.203	23.0	Moran 1992


MAP AREA


Cobra
Peron
Ned's Creek
Wiluna
Billabong
Murchison
Cue
Northampton
Laverton
Burnerbinmah
Jeedamya
Goongarrie
Koorda
Coolgardie
Goomalling
PERTH
Kokerbin
Cairn Rock
Hyden
Kulin
Arthur River
Dumbleyung
Pingrup
Ravensthorpe
Nyabing
Borden
0 100 200 km
Scale


Koorda
Northampton
Coolgardie
Cobra
Cue
Ned's Creek
Burnerbinmah
Jeedamya
Wiluna
Cairn Rock
Billabong
Hyden
Kokerbin
Goomalling
Laverton
Pingrup
Borden
Dumbleyung
Kulin
Murchison
Peron
Ravensthorpe
Nyabing
southern
northern
northern
southern
northern
southern
northern
northern
southern
southern

Diversity in the chloroplast genome of Western Australian Sandalwood (*Santalum spicatum*)

M. Byrne [A], B. Macdonald and J. Brand

Science Division, Department of Conservation and Land Management, Locked Bag 104, Bentley Delivery Centre, WA 6983, Australia.

[A] Corresponding author: email margaretb@calm.wa.gov.au, Phone +61893340503, Fax +61893340515

Running Heading: Phylogeography in Sandalwood

Abstract

Western Australian sandalwood (*Santalum spicatum*) is widespread throughout Western Australia across the semi-arid and arid regions. The diversity and phylogeographic patterns within the chloroplast genome of *S. spicatum* were investigated using RFLP analysis of 23 populations. The chloroplast diversity was structured into two main clades that were geographically separated, one centered in the southern (semi-arid region) and the other in the northern (arid) region. Fragmentation due to climatic instability was identified as the most likely influence on the differentiation of the lineages. The lineage in the southern region showed a greater level of differentiation than that in the arid region suggesting a higher level of gene flow or a more recent range expansion of sandalwood in the southern region. The phylogeographic pattern in the cp genome is congruent with that detected in the nuclear genome which identified different genetic influences between the regions and also suggested a more recent expansion of sandalwood in the southern region.

Key words: Sandalwood, cpDNA, phylogeny, phylogeography

Introduction

There are five species of the genus *Santalum* endemic to Australia. Of these *Santalum spicatum* (Western Australian Sandalwood) is found in Western Australia and South Australia, whilst *S. lanceolatum* (Plumbush), *S. acuminatum* (Quandong) and *S. murrayanum* (Bitter Quandong) are distributed throughout Australia, and *S. obtusifolium* is confined to the eastern coast (Hewson and George 1984). The Indian Sandalwood, *S. album* also occurs in Australia in the most northern parts of the Northern Territory. Among the Australian endemic species only Western Australian Sandalwood has commercial significance as it produces a sesquiterpene oil similar to Indian Sandalwood (Shea et al. 1998). Western Australian sandalwood has been harvested commercially since 1845 and was a primary industry in the early years of european settlement in Western Australia (Loneragan 1990).

Western Australian Sandalwood occurs across a wide range of environmental conditions as it has a broad distribution throughout the semi-arid (300 – 600 mm rainfall) and arid (150 – 300 mm rainfall) areas of Western Australia. Morphological variation has been observed across the range suggesting the presence of two ecotypes within the species (Fox and Brand 1993). In the higher rainfall areas in the south-west of the distribution the species shows larger leaves and nuts, has higher chlorophyll content and forms more of a tree habit than in the arid areas in the north and east of the distribution (Fox and Brand 1993). Trees in the arid areas also show higher concentration of oil in the wood (Loneragan 1990). A genetic study using nuclear RFLP markers showed some evidence of differentiation of ecotypes but it was not significant (Byrne et al. 2002). However, the study did show show significant differences in the genetic influences within the two regions, with the northern region showing stable history with equilibrium between drift and gene flow, and the southern region showing no drift-gene flow equilibrium and variances consistent with fragmented, isolated populations under greater influence of drift than gene flow.

This pattern of historic differentiation reflects an ancient landscape that has experienced geological stability with no major glaciation or extinction events, where fragmentation of populations has occurred through climatic instability during the Pleistocene due to cyclic expansion and contraction of the mesic and arid zones (Hopper 1979, Hopper et al. 1996). This has led to a naturally fragmented system with a high level of endemism and both relictual and derived species.

Analysis of genetic variation in the chloroplast genome allows the elucidation of historical factors influencing genetic variation that are not confounded by the effects of recent gene flow (Schaal et al 1998). The reduced effective population size of haploid genomes combined with the restricted nature of gene flow through seed dispersal makes maternally inherited organelle markers more likely to record the effects of population history in present day genetic patterns than nuclear markers (Ennos et al. 1999). Analysis of cpDNA in widespread species, and restricted species with disjunct distributions, in Western Australia has revealed the presence of historical lineages that have been differentiated through isolation occurring during the Pleistocene (Byrne et al. 1999, Byrne et al. 2001b, Byrne et al. 2001c). The presence of two ecotypes in Western Australian Sandalwood influenced by different genetic processes may represent similar patterns of historic separation in this species.

This study investigates the evolutionary and phylogeographic patterns within *S. spicatum* using RFLP analysis of the chloroplast genome. The mode of inheritance of chloroplasts has not been determined in *Santalum* but is maternal in most angiosperms (Harris and Ingram 1991; Birky 1995).

Materials and Methods

Plant Collections

Leaf samples were collected from twenty three populations of *S. spicatum.* The distribution of the species and location of sampled populations is shown in Fig. 1. Collections were made of 10 individuals from each of the populations. Leaf samples were also collected from 10 individuals in each of two populations of *S. acuminatum* (Quandong) as outgroups (Figure 1). DNA was extracted from the leaves of the 250 individuals as in Byrne and Moran (1994) with the addition of 0.1M sodium sulphite to the extraction buffers (Byrne et al. 2001a). DNA quality was good but the yield was low probably due to high levels of sesquiterpene oils. Initially 3 ug DNA from five individuals per population was digested with six restriction enzymes (*Bcl*I, *Bgl*II, *Eco*RI, *Eco*RV, *Hin*dIII, *Xba*I) and hybridized with heterologous probes covering the majority of the chloroplast genome. Six petunia cpDNA probes were used, P1, P3, P4, P6, P8, P10, (details given in Sytsma and Gottlieb 1986), plus one tobacco cpDNA probes, pTBa1 (Suguira *et al.* 1986, Shinozaki *et al.* 1986). Restriction digestion and hybridisation were as described in Byrne and Moran

(1994), and probe plasmids were linearized then labelled with ^{32}P using the random priming method. After analysis of the data for the first 125 individuals, DNA of the remaining five individuals per population was digested with two of the enzymes (*EcoR*I and *EcoR*V) and probe-enzyme combinations for mutations detected in the first 125 individuals were assayed for the second 125 individuals.

Data Analysis

Banding patterns obtained were interpreted in terms of restriction site or length mutations, and assessed as presence or absence of mutations (not presence or absence of bands). Fragment patterns for consecutive cp probes were compared to ensure that each mutation was correctly interpreted and counted only once. Where a length mutation was detected by more than one restriction enzyme it was counted as only one mutation. Nucleotide diversity was calculated for restriction site mutations using HAPLO (Lynch and Crease 1990), and partitioned within and between populations. Haplotype diversity was calculated by considering haplotypes as alleles at one locus using Nei's gene diversity measures (Nei 1977). A parsimony analysis of haplotype relationships characterised by presence or absence of each mutation was carried out using PAUP (Swofford 1991). Bootstrap analysis used 1000 replications and heuristic search with TBR branch swapping and MULPARS on.

Results

Variation in cp DNA

This study analysed restriction sites in the cp genome and revealed polymorphism with all enzymes used. Within *S. spicatum* 14 mutations were detected, of these five were restriction site mutations and nine were length mutations (Table 1). Half of the mutations occurred in the small single copy region and half occurred in the large single copy region. The first assay of five individuals per population detected the majority of the mutations. The assay of an additional five individuals per population detected one additional mutation (number 14) which was restricted to three individuals from the Billabong population.

The 14 mutations were distributed over eleven haplotypes. There was one haplotype (B) that was more common than all others occurring in 45% of the individuals sampled. The remaining haplotypes occurred with frequencies ranging from 2.6 to 13%. The most common haplotype (B) was present in all southern populations except Nyabing and the

north-eastern population of Coolgardie. The remaining haplotypes occurred in one to three populations. Intra-population variation was present in two populations, Kokerbin in the south had two haplotypes present and Billabong in the north had three haplotypes within the population.

There were 11 mutations that differentiated *S. spicatum* from the two populations of the related species *S. acuminatum*, three site mutations and eight length mutations. There was also eight mutations detected between the two populations of *S. acuminatum*, two restriction site mutations and six length mutations. Of these eight mutations five were shared between the Goongarrie (north-eastern) population of *S. acuminatum* and all individuals of *S. spicatum*. The other three mutations were specific to either the Goongarrie population or the Arthur River (southern) population of *S. acuminatum*.

Nucleotide and Haplotype Diversity

Nucleotide diversity, the average number of nucleotide differences per site between two sequences (Nei 1978), can be determined for restriction site but not length mutations. Nucleotide diversity, averaged over all pairs of individuals in *S. spicatum* was 0.00097%. The mean diversity within populations was 0.0001% and the mean diversity between populations was 0.00096%. The proportion of nucleotide diversity between populations, N_{ST}, was 99%. The mean diversity between lineages (see below) was 0.116%.

A measure of haplotype diversity can be determined by treating the whole cp genome as a single locus with each haplotype as an allele. Total haplotype diversity (H_T) was 0.749 and haplotype diversity within populations (H_S) was 0.7014. The proportion of haplotype diversity between populations, G_{ST}, was 94%.

Haplotype and Population Relationships

A phylogenetic parsimony analysis of haplotypes gave one tree of length 32, with a consistency index of 1.0. The cladogram showed all individuals of *S. spicatum* clustered into a clade clearly differentiated from the outgroup *S. acuminatum* (Figure 2). Within this *S. spicatum* clade the tree shows a star pattern with an unresolved polytomy of six clades. One clade (I) shows structuring of four related haplotypes. This clade is differentiated from the other clades and characterised by three mutations, with an additional three mutations that differentiate the haplotypes within the clade. This clade contains the common

haplotype and occurs in all of the southern populations plus the northern populations of Burnerbinmah and Coolgardie. Two more clades in the polytomy (II and IV) show some structure, being characterised by one to two mutations, and having two haplotypes differentiated by one mutation. These clades occur in the north west of the distribution. The remaining three clades (III – VI) show little differentiation, characterised by up to one mutation and containing one haplotype each. These clades contain the remaining northern populations. Doubling the number of individuals assayed did not result in a change to the distribution of the mutations among populations and did not change the structure of the phylogenetic tree except for a tip clade in the Billabong population resulting from the additional detection of mutation 14. Therefore the sampling in this study has been sufficient to detect the phylogenetic patterns with this species.

The phylogeny also shows that *S. spicatum* is not monophyletic in relation to the related species *S. acuminatum* since it shares mutations with the Goongarrie population of *S. acuminatum* (Figure 2) although there were a large number of mutations (12) that separated *S. spicatum* from *S. acuminatum* at Goongarrie. There was a high level of variation between the two *S. acuminatum* populations with 8 mutations differentiating them. This is higher than the average number of mutations separating populations of *S. spicatum* (3.6) although there was one pair of *S. spicatum* populations which were differentiated by 8 mutations (Nyabing and Northampton).

A nested clade analysis identified significant geographic structuring in the phylogeny of *S. spicatum* (p=0.000). The cladogram showing the nested clades is presented in Figure 3 and the *Dc* and *Dn* values for the clades are shown in Table 3. All clades showed significant structuring. Analysis could not be carried out for Clade 1-5 as there was no populations variation in the clade. All other one-step and two-step clades identified restricted gene flow through isolation by distance as the main influence on the geographic structuring within these clades, except for Clade 1-4 where an inference of long distance colonisation was made. At the highest nesting level past fragmentation was identified as the main influence on the geographic structuring between the two two-step clades. These two clades are geographically separated with Clade 2-1 contain haplotypes characterising the southern populations and Clade 2-2 containing haplotypes characterising the northern populations except for the Burnerbinmah population which contains haplotypes nested in Clade 2-1. The two clades are also separated by a higher than average number of mutational steps.

Clade 2-1 is equivalent to Clade I in the parsimony analysis and Clade 2-2 is equivalent to Clades II – VI in the parsimony analysis (Figure 2).

Discussion

The level of diversity in the chloroplast genome of *S. spicatum* is high compared to other woody perennials but not as high as has been identified in other Australian tree species, including the main host of *S. spicatum*, *Acacia acuminata* (Byrne et al. 2001b). Significant structuring of the diversity occurs with two main clades that are geographically separated, one centred on the southern region and one centred in the north of the distribution. This differentiation of two lineages is consistent with the identification of differences between the regions in the influence of genetic processes on the nuclear genome (Byrne et al. 2002). It is also consistent with the identification of two ecotypes in these regions based on morphological variation. The level of differentiation between the lineages suggests that they have been isolated during the Pleistocene. Geographical structuring due to historical isolation has also been observed in a common host of Sandalwood, *Acacia acuminata* (Byrne et al 2001b). It has been proposed that cyclic contraction and expansion of the arid region in the north-east, and the mesic region in the south-west, during the Pleistocene led to fragmentation and isolation in the intermediate area between the arid and mesic zones (Hopper 1979, Hopper et al. 1996). This is a plausible explanation for the isolation and differentiation of lineages observed within the species that have been investigated. The large degree of differentiation between the southern and northern populations of *S. acuminatum* also suggests that there may be lineages within Quandong which have been differentiated due to historical isolation. The greater degree of differentiation suggests that the effects of isolation may be more pronounced in Quandong than in Sandalwood.

There were differences in the level of differentiation within the two lineages in Sandalwood. The southern lineage had high similarity within the lineage with the majority of the populations having one of the four haplotypes present in the lineage. In comparison the northern lineage had greater differentiation within the lineage with the six haplotypes having lower frequencies and distributed into three main areas of geographical diversity, although populations within these broad areas show little differentiation. The higher similarity between populations in the southern lineage compared to the northern lineage could be due to a higher level of gene flow or recent establishment of populations in the region. The seeds of *S. spicatum* are large nuts which would result in low dispersal,

although emus are known to eat the seeds and would be a means of seed dispersal, and Havel (1993) hypothesised that the woylie, a small mammal, may have cached seeds which would increase dispersal. Emus and woylies were likely to have been more abundant in the southern region due to higher abundance of food source and may have lead to greater seed dispersal in the south than in the north. The populations sampled in the south are also geographically closer than those sampled in the north, an may have had a higher abundance of *S. spicatum* compared to those in the north. There are no direct comparisons of abundance of *S. spicatum* in the southern and northern regions at the time of settlement, but the tonnage of Sandalwood harvested from the southern region in the early years of settlement indicates that the species was abundant and greater than the current abundance in the northern region. Higher abundance and more continuous distribution would lead to greater gene flow and hence less differentiation between populations in the south. The greater similarity of the southern populations could also indicate a more recent range expansion of *S. spicatum* in this region compared to the north. The northern clade is the interior clade and the southern clade is a tip clade which indicates that the haplotypes in the northern clade represent the ancestral haplotypes in the species, and suggests that sandalwood may have evolved in the northern region and spread into the southern region more recently in evolutionary history.

The genetic patterns in the chloroplast data are both concordant and discordant with the patterns identified in the nuclear genome. Genetic analysis of the nuclear genome identified an equilibrium between drift and gene flow in the northern region indicating that the region had existed under stable conditions of dispersal for a long period of time (Byrne et al. 2002). In contrast, the southern region showed a genetic pattern indicating that the populations have been more recently established, and that they have been fragmented and influenced to a greater degree by genetic drift than by gene flow. The southern region also showed greater differentiation between populations than those in the northern region. The genetic patterns in the nuclear and chloroplast genomes both suggest that the southern populations have been established more recently than the northern populations. However the nuclear genome showed greater differentiation in southern populations whereas the chloroplast genome identified greater differentiation in the northern populations. This suggests that the lower differentiation in the chloroplast genome is not likely to be a result of higher seed dispersal or greater abundance, but may be due to closer population

sampling and less time for fragmentation to influence divergence in the chloroplast genome.

The phylogenetic analysis shows that *S. spicatum* and the northern form of *S. acuminatum* have a shared ancestry that distinguishes them from the southern populations of *S. acuminatum*. The shared ancestry of *S. spicatum* and the northern *S. acuminatum* is consistent with a northern evolution of *S. spicatum* and a more recent expansion into the southern region. The level of differentiation between the southern and northern forms of Quandong, and the shared ancestry of northern Quandong and sandalwood suggests that the two forms in Quandong may represent different species. This is also consistent with the morphological differences that have been observed between the two forms of Quandong and the potentially different ploidy levels identified from the nuclear RFLP banding patterns (Byrne et al. 2002).

Acknowledgments

We thank the Forest Products Commission of Western Australia for funding this study.

References

BIRKY, C. W. 1995. Uniparental inheritance of mitochomdrial and chloroplast genes, mechanisms and evolution. *Proceedings of the National Academy of Sciences USA* **92**, 11331-11338.

BYRNE, M. AND MORAN, G. F. 1994. Population divergence in the chloroplast genome of *Eucalyptus nitens*. *Heredity* **73**, 18-28.

BYRNE, M., MACDONALD, B. AND COATES, D. 1999. Divergence in the chloroplast genome and nuclear rDNA of the rare Western Australian plant *Lambertia orbifolia* Gardner (Proteaceae). *Molecular Ecology* 8: 1789-1796

BYRNE, M., MACDONALD, B., BROADHUSRT, L. AND BRAND, J. 2002 Genetic diversity in Australian Sandalwood (*Santalum spicatum*) as revealed by nuclear RFLP analysis. Theor. Appl. Genet. submitted

BYRNE M, MACDONALD B, FRANCKI M. 2001a Incorporation of sodium sulphite into extraction protocol minimizes degradation of *Acacia* DNA. *BioTechniques*. **30**. 742-744.

BYRNE, M, MACDONALD, B, AND COATES D. 2001b Phylogeographic patterns within the *Acacia acuminata* (Leguminosae: Mimosoideae) complex in Western Australia. *J. Evol. Biol.* submitted

BYRNE, M., TISCHLER, G., MACDONALD. B., COATES. D. J. AND McCOMB, J. 2001c Phylogentic relationships between two rare acacias and their common widespread relatives in south-western Australia. *Conservation Genetics* **2**, 157-166.

ENNOS, R. A., SINCLAIR, W. T., HU, X-S., AND LANGDON, A. 1999. Using organelle markers to elucidate the history, ecology and evolution of plant populations. In: Hollingsworth PM, Bateman RM, Gornall RJ (eds) *Molecular Systematics and Plant Evolution*, pp. 1-19. Taylor & Francis, London.

FOX, J. E. D. AND BRAND, J. E. 1993. Preliminary observations on ecotypic variation in *Santalum spicatum* 1. Phenotypic variation. *Mulga Research Journal* **11**, 1-12

HARRIS, S. A. AND INGRAM, R. 1991 Chloroplast DNA and biosystematics: the effects of intraspecific diversity and plastid transmission. *Taxon* **40**, 393-412.

HAVEL, J.J. 1993 Review of research needs of the sandalwood industry in Western Australia. Department of Conservation and Land management, Western Australia.

HEWSON, H. J. AND GEORGE, A. S. 1984. Flora of Australia Vol. 22. Australian Government Puclishing Service, Canberra. Pp 61 – 66.

HOPPER, S. D. 1979. Biogeographical aspects of speciation in the southwest Australian flora. *Annual Review of Ecology and Systematics* **19**, 399-422.

HOPPER, S. D., HARVEY, M. S., CHAPPILL, J. A., MAIN, A. R. AND MAIN B. Y. 1996. The Western Australian biota as Gondwanan heritage - a review. In: S. D. Hopper, J. A. Chappill, M. S. Harvey and A. S. George (eds) *Gondwanan Heritage. Past present and Future of the Western Australian Biota*, pp. 1-46. Surrey Beatty & Sons: Chipping Norton.

LONERAGAN O. W. 1990. Historical review of Sandalwood (*Santalum spicatum*) Research in Western Australia. *Research Bulletin No. 4* Department of Conservation and Land Management, Perth.

LYNCH, M., AND CREASE, T. J. 1990. The analysis of population survey data on DNA sequence variation. *Molecular Biology and Evolution* **7**, 377-394.

NEI, M. 1977. F-statistics and analysis of gene diversity in sub-divided populations. *Annals of Human Genetics* **41**, 225-233.

SCHAAL, B. A., HAYWORTH, D. A., OLSEN, K. M., RAUSCHER, J. T. AND SMITH, W. A. 1998. Phylogeographic studies in plants: problems and prospects. *Molecular Ecology* **7**, 465-474.

SHINOZAKI, K., OHME, M., TANAKA, M., WAKASUGI, T., HAYASHIUDA, N., MATSUBAYASHI, T., ZAITA, N., CHUNWONGSE, J., OBOKATA, J., YAMAGUCHI-SINOSAKI, K., OHTO, C., TORAZAWA, K., MENG, B. Y., SUGITA, M., DENO, H., KAMAGASHIRA, T., YAMADA, K., KUSUDA, J., TAKAIWA, F., KATO, A., TOHDOH, N., SHIMADA, H. AND SUGIURA, M. 1986. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *European Molecular Biology Organisation Journal* **5**, 2043-2049.

SHEA S.R., RADOMILJAC, A., BRAND J. AND JONES P. 1998. An overview of sandalwood and the development of sandal in farm forestry in Western Australia. In: Radomiljac AM, Ananthapadmanabho HS, Welbourn RM, Satyanarayana Rao K (eds) *Sandal and it's products*, pp 9-15. ACIAR, Canberra.

SUGUIRA, M., SHINOZAKI, K., ZAITA, N., KUSUDA, M. AND KUMANO, M. 1986. Clone bank of the tobacco (*Nicotiana tabacum*) chloroplast genome as a set of overlapping restriction endonuclease fragments: mapping of eleven ribosomal protein genes. *Plant Science* **44**, 211-216.

SWOFFORD, D. L. 1991. PAUP: Phylogenetic Analysis Using Parsimony, Version3.1.1. Computer program distributed by the Illinois Natural History Survey, Champaign, Illinois.

List of Figures

Fig. 1. Location of sampled populations of *Santalum spicatum* (●) and *Santalum acuminatum* (▲) Dotted line indicates separation of rangelands (north) and wheatbelt (south) regions.

Fig. 2. Phylogenetic parsimony tree of haplotype relationships in *Santalum spicatum*. Consensus tree based on 1000 boot strap replications. Numbers above lines refer to mutations in Table 1. Numbers in italics below lines at nodes represent bootstrap confidence values (%). Numbers in brackets indicate number of individuals in each population with that haplotype.

Fig. 3. Nested cladogram for haplotypes in *S. spicatum*.

Table 1. Restriction fragment length polymorphisms detected in the chloroplast genome of *Santalum spicatum*. Populations detailed possess the fragment sizes listed second. kb, kilobases.

Probe/enzyme	Mutation type	Fragment sizes (kb)	Population and individual numbers
1. pTBa1/*Hind* III	site	5.65 = 3.35 + 2.3	Northampton, Murchison, Billabong (3)
2. P8/*Eco* RV	site	5.2 + 2.8 = 8.0	Burnerbinmah, Goomalling, Hyden, Ravensthorpe, Pingrup, Cairn, Koorda, Nyabing, Borden, Kulin, Coolgardie Dumbleyung, Kokerbin
3. P8/*Eco* RV	site	8.0 + 2.3 = 10.3	Billabong (1,2,4,5)
4. P6/P10/*Bcl* I	site	4.5 + 2.8 = 7.3	Burnerbinmah, Goomalling, Hyden, Ravensthorpe, Pingrup, Cairn, Koorda, Nyabing, Borden, Kulin, Coolgardie Dumbleyung, Kokerbin
5. P1/P4/*Eco* RV	site	9.2 + 2.2 = 11.4	Peron
6. pTBa1/*Eco* RI	length	2.05 v 2.5	Burnerbinmah, Goomalling, Hyden, Ravensthorpe, Pingrup, Cairn, Koorda, Nyabing, Borden, Kulin, Coolgardie Dumbleyung, Kokerbin
7. P6/*Eco* RI	length	1.9 v 1.5	Burnerbinmah, Goomalling, Hyden, Ravensthorpe, Pingrup, Cairn, Koorda, Nyabing, Borden, Kulin, Coolgardie, Dumbleyung, Kokerbin (1,2,8,9,10)
8. P6/*Eco* R I	length	2.55 v 2.65	Kokerbin (3,4,5,6,7)
9. P1/P4/*Bcl* I	length	5.1 v 7.8	Northampton, Murchison, Billabong (3)
10. P1/P4/*Bcl* I	length	9.1 v 8.8	Ned's Creek, Cobra, Wiluna
11. P1/P4/*Bcl* I	length	9.1 v 8.9	Nyabing
12. P1/P4/*Bgl* II	length	2.15 v 1.95	Burnerbinmah
13. P1/P4/*Hind* III	length	1.4 v 0.9	Northampton, Billabong (3)
14. P1/P4*Eco*R I	length	4.2 v 3.9	Billabong (7,8,9)

Table 2. Restriction fragment length variation detected between the chloroplast genomes of *Santalum spicatum* and *S acuminatum*. Taxa detailed possess the fragment sizes listed second. kb, kilobases.

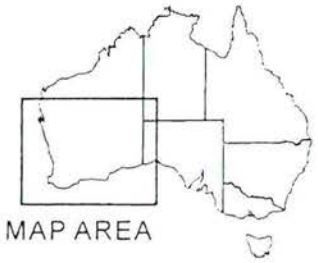
Probe - enzyme	Mutation	Fragment sizes (kb)	Taxon and population
A. P8/*Bgl* II	site	3.2 = 1.7 + 1.5	*S. spicatum*
B. P8/Eco RV	site	15.9 = 8.0 + 7.9	*S. spicatum*
C. P6/*Bgl* II	site	7.6 + 2.5 = 11.1	*S. spicatum*
D. pTBa1/*Bcl* I	length	3.75 v 2.65	*S. spicatum*
E. pTBa1/*Bgl* II	length	6.5 v 6.0	*S. spicatum*
F. pTBa1/*Eco* RI	length	1.5 v 1.6	*S. spicatum*
G. P3/*Bcl* I	length	7.8 v 5.2	*S. spicatum*
H. P6/*Eco* RI	length	3.3 v 2.45	*S. spicatum*
I. P8/*Eco* RI	length	2.9 v 3.1	*S. spicatum*
J. P8/*Eco* RV	length	1.8 v 2.3	*S. spicatum*
K. P10/*Eco* RV	length	17.0 v 8.3	*S. spicatum*
L. P3/*Bcl* I Goongarrie	length	3.8 v 3.5	*S. spicatum*, *S. acuminatum*
M.P3/*Bgl* II Goongarrie	length	2.9 v 3.9	*S. spicatum*, *S. acuminatum*
N. P3/*Hind* III Goongarrie	length	5.1 v 6.7	*S. spicatum*, *S. acuminatum*
O. P8/*Bcl* I Goongarrie	length	9.2 v 6.5	*S. spicatum*, *S. acuminatum*
P. P8/*Xba* I Goongarrie	length	5.3 v 4.3	*S. spicatum*, *S. acuminatum*
Q. pTBa1/*Bcl* I	length	3.75 v 2.8	*S. acuminatum* Goongarrie
R. P8/*Xba* I	site	5.25 = 4.0 + 1.25	*S. acuminatum* Arthur River
S. P10/*Bcl* I	site	6.8 + 1.6 = 8.4	*S. acuminatum* Arthur River

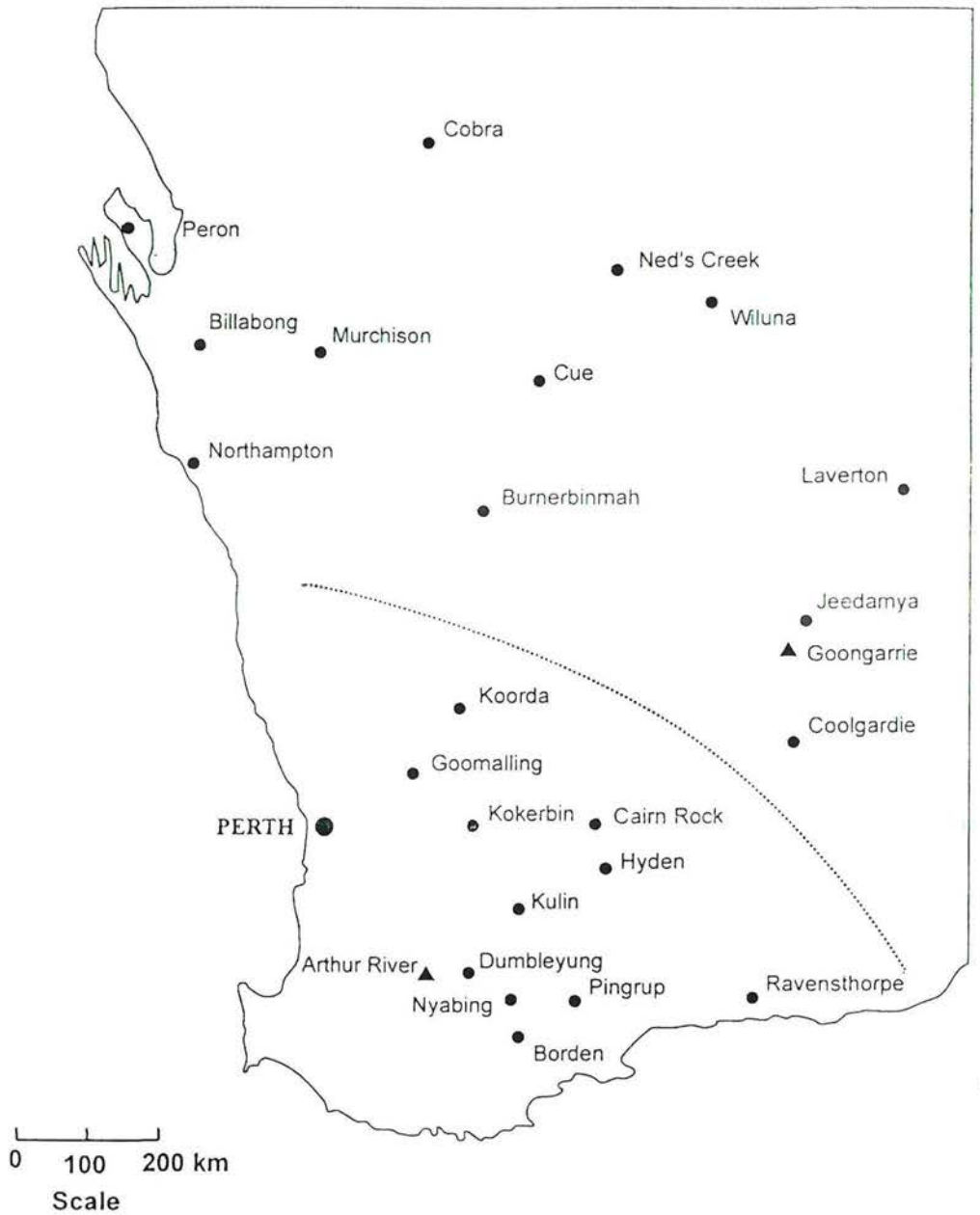
Table 3. Levels of clade, nested clade and interior-tip distance in clades with significant geographic association in *Santalum spicatum*. *Dc* clade distance (dispersion), *Dn* nested clade distance (displacement). L, probability of larger than expected value, S, probability of smaller than expected value. Significant probabilities in bold.

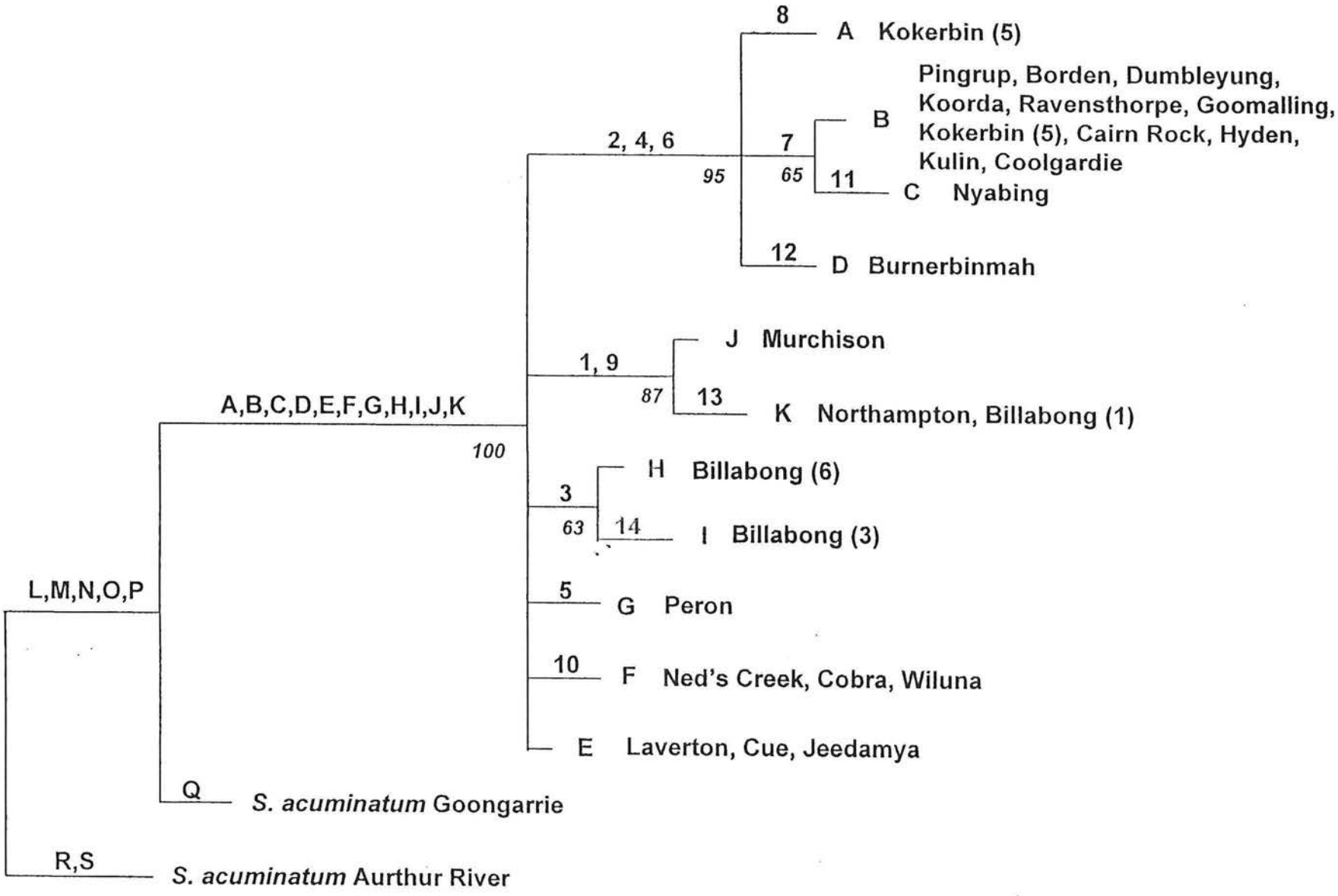
Nested clade	Clade/Haplotype	*Dc*	Probability	*Dn*	Probability
clade 1-1	hap B (interior)	151.7931	**L0.0450**	114.9499	0.0750
	hap C (tip)	0.0000	**S0.0000**	150.9620	0.0750
	interior vs. tip	151.7931	**L0.0000**	36.0122	0.0750
clade 1-2	hap A (tip)	0.0000	0.0790	227.2059	**L0.0010**
	hap D (tip)	0.0000	**S0.0010**	113.6161	**S0.0000**
clade 1-3	hap J (interior)	0.0000	**S0.0000**	99.9913	**S0.0000**
	hap K (tip)	27.2736	**S0.0000**	102.3298	**L0.0000**
	interior vs. tip	-27.2736	**S0.0000**	-2.3385	**S0.0000**
clade 1-4	hap E (interior)	207.2355	**S0.0000**	296.2069	0.4330
	hap G (tip)	0.0000	**S0.0000**	512.2734	**L0.0000**
	hap F (tip)	191.0068	**S0.0000**	232.6730	**S0.0010**
	interior vs. tip	63.9803	0.0510	-6.3662	0.4330
clade 2-1	clade 1-1 (tip)	147.8305	**S0.0000**	155.0542	**S0.0000**
	clade 1-2 (interior)	151.4793	0.3080	284.5015	**L0.0000**
	interior vs. tip	3.6488	0.3550	129.4473	**L0.0000**
clade 2-2	clade 1-3 (tip)	101.2162	**S0.0000**	249.0229	**S0.0150**
	clade 1-4 (interior)	299.8447	0.3430	323.0370	**L0.0160**
	clade 1-5 (tip)	0.0000	**S0.0000**	292.4753	0.3790
	interior vs. tip	228.9934	**L0.0000**	60.9784	**L0.0160**
clade 3	clade 2-1 (tip)	169.9904	**S0.0000**	298.3975	**S0.0000**
	clade 2-2 (interior)	304.7435	**S0.0020**	449.0316	**L0.0000**
	interior vs. tip	134.7531	**L0.0000**	150.6340	**L0.0000**

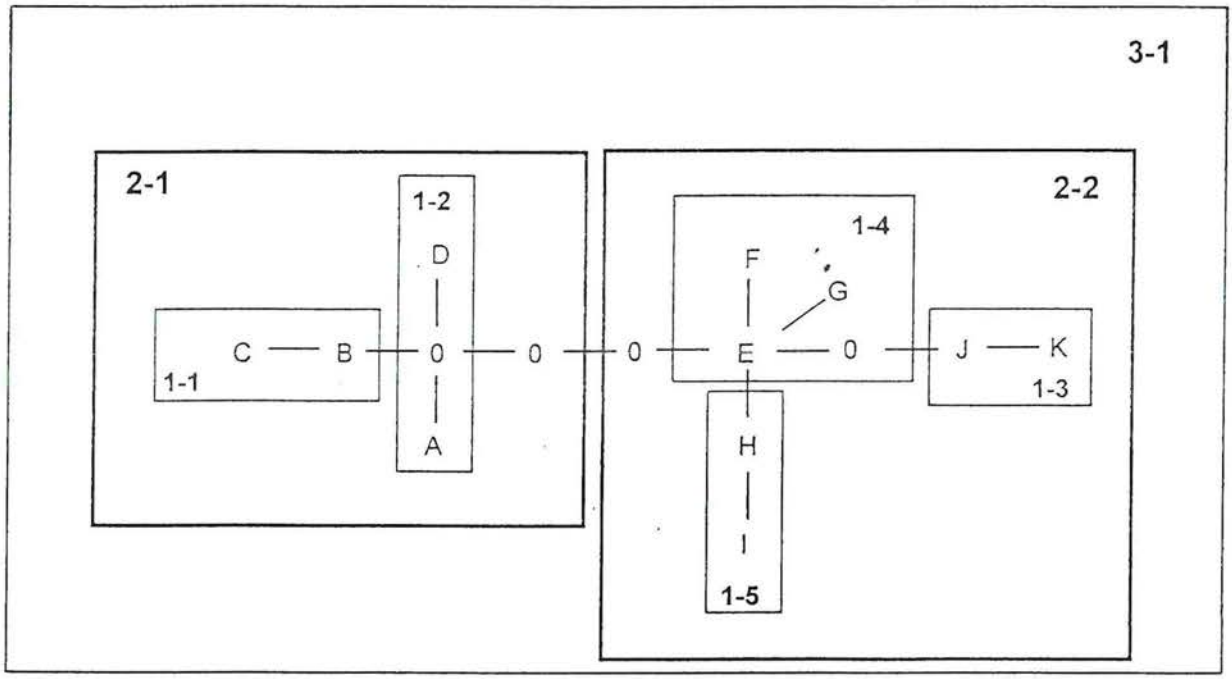
Table 4. Phylogeographic inferences inferred from nested clade analysis of *Santalum spicatum*. Numbers in Key refer to options in inference key in Templeton et al. (1995).

Nested Clade	Key	Inference
Clade 1-1	1, 2, 3, 4 no	restricted gene flow
Clade 1-2	1, 2, 3, 4, no	restricted gene flow
Clade 1-3	1, 2, 3, 4 no	restricted gene flow
Clade 1-4	1, 2, 5, 6, 13 yes	long distance colonization
Clade 2-1	1, 2, 3, 4 no	restricted gene flow
Clade 2-2	1, 2, 3, 4 no	restricted gene flow
Clade 3	1, 2, 3, 5, 15 no	past fragmentation


MAP AREA


Cobra
Peron
Ned's Creek
Wiluna
Billabong
Murchison
Cue
Northampton
Laverton
Burnerbinmah
Jeedamya
Goongarrie
Koorda
Coolgardie
Goomalling
PERTH
Kokerbin
Cairn Rock
Hyden
Kulin
Arthur River
Dumbleyung
Pingrup
Ravensthorpe
Nyabing
Borden
0 100 200 km
Scale


8
A Kokerbin (5)
Pingrup, Borden, Dumbleyung,
Koorda, Ravensthorpe, Goomalling,
Kokerbin (5), Cairn Rock, Hyden,
Kulin, Coolgardie
B
2, 4, 6
7
95
65
11
C Nyabing
12
D Burnerbinmah
J Murchison
1, 9
87
13
K Northampton, Billabong (1)
A,B,C,D,E,F,G,H,I,J,K
100
H Billabong (6)
3
63
14
I Billabong (3)
L,M,N,O,P
5
G Peron
10
F Ned's Creek, Cobra, Wiluna
E Laverton, Cue, Jeedamya
Q
S. acuminatum Goongarrie
R,S
S. acuminatum Aurthur River


3-1
2-1
1-2
D
C
B
0
0
A
1-1
0
2-2
1-4
F
G
E
0
J
K
1-3
H
I
1-5