High genetic identities between three oil mallee taxa, *Eucalyptus kochii* ssp. *kochii*, ssp. *plenissima* and *E. horistes*, based on nuclear RFLP analysis

M. BYRNE

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

The eucalypt species *Eucalyptus kochii* ssp. *kochii*, *E. kochii* ssp. *plenissima* and *E. horistes*, are harvested for oil production. The genetic diversity in these taxa and their phylogenetic relationships were assessed using anonymous nuclear RFLP loci. Variation was determined for 20 individuals from 10 populations representing the three taxa. The level of diversity in the taxa was high compared to other species where RFLP surveys have been conducted, including another eucalypt species. The populations showed little differentiation, low genetic distance and high gene flow between them. Phylogenetic relationships reveal clustering of populations consistent with their geographical distribution. This analysis suggests that genetically these taxa represent a single widespread, variable species with some provenance delineation. Establishment of a breeding programme based on all populations would allow progress in genetic improvement for oil yield.

Keywords: essential oils, *Eucalyptus*, gene flow, genetic diversity, oil mallees, RFLP.

Introduction

Eucalypt oil has traditionally been used for pharmaceutical, industrial, and perfumery and flavouring purposes (Boland *et al*. 1991). Australia was formerly a major producer of eucalypt oil but world production is now dominated by China and total world production is estimated to be 3000 tonnes (ABARE, 1994). In Western Australia there is considerable interest in developing oil mallees (eucalypt species with multistemmed habit) to remedy extensive dryland salinity through planting trees that will provide a commercial return (Bartle *et al*., 1996). Salinity currently affects some 9% of the 18 million hectares of agricultural land in the south-west of the state and this is projected to expand to 32% within several decades (Ferdowsian *et al*., 1996). Planting of deep-rooted tree crops will improve water use and reduce the ground water accumulation that causes salinity. The scale of planting necessary to be significant in terms of salinity control requires a commercial return to the farmer. To be commercially attractive much larger markets

for eucalypt oil will be required than the current pharmaceutical markets. These are potentially available in industrial solvent markets where eucalypt oil can be used to replace the discontinued hydrocarbon, trichloroethane (Barton & Knight, 1997). However, the price of oil in the industrial solvent market is lower than the current pharmaceutical markets. Bartle *et al*. (1996) show that such prices can be met but this will require an increase in oil yield and a reduction in the cost of production. Genetic improvement has the potential to increase yield and reduce costs; therefore a major breeding programme has commenced.

The eucalypt oil industry in Western Australia is centred on two groups of species, the *Eucalyptus kochii*/*E. horistes* group and the *E. loxophleba*/*E. gratiae* group. These species produce an average of 2.5% oil yield of which around 90% is 1,8-cineole. An improvement programme is being developed with the aim of increasing oil yield through selection and breeding, whilst retaining any significant genetic distinctness of taxa or ecotypes, and maintaining appropriate levels of genetic variation in the production populations. The degree of genetic diversity and genetic relatedness within these groups of species

has not been investigated, and this information would be valuable in the design and structuring of the breeding programme.

In the *E. kochii/E. horistes* group there are three taxa, *E. kochii* ssp. *kochii* Brooker, ssp. *plenissima* Brooker and *E. horistes* Johnson & Hill, which are closely related taxonomically. They belong to the oleosa group of species which occurs in all mainland states of Australia (Brooker *et al.*, 1988). For ease of reading the three taxa will be referred to as kochii, plenissima and *E. horistes* from here on. The three taxa are morphologically very similar. Kochii and plenissima are distinguished by operculum shape; kochii has a conical operculum whereas plenissima has a short, obtusely conical operculum (Brooker *et al.*, 1988). *Eucalyptus horistes* has broader, glossier leaves than kochii, and a longer, more pointed operculum than plenissima (Chippendale, 1988). The three taxa have allopatric but adjacent distributions (Fig. 1). *Eucalyptus horistes* occurs on lateritic and sandy loam in mallee shrubland in the northern wheat belt and gold fields regions. Plenissima occurs on sandy loam in mallee shrubland in the northeastern wheat belt region and kochii occurs on sandy soil in open woodland and is restricted to the Wongan Hills–Dalwallinu area of the wheat belt.

An analysis of variation within and among the three taxa was undertaken using RFLP loci. This will enable determination of the genetic diversity in the taxa in order to establish a breeding programme that encompasses the genetic diversity of the species group. The phylogenetic relationships between populations of the taxa were determined, to assess whether genetic distinctness exists within or among the taxa, which would indicate the need for specific structuring within the breeding programme.

Materials and methods

Plant material

Leaf samples were collected from two populations of kochii, four populations of plenissima and four populations of *E. horistes*. The distribution of the species and the location of sampled populations is shown in Fig. 1. Only two populations of kochii were sampled because this taxon has a restricted distribution compared to plenissima and *E. horistes*. Collections were made of 20 individuals from each of the 10 populations, except that 18 individuals were sampled from the Three Springs population and 22 from the Mullewa population. DNA was extracted from the leaves of the 200 individuals as in Byrne *et al.* (1998), digested with either *Bgl*II or *Eco*RV, Southern blotted and hybridized with 38 RFLP probes (c021, c030, c087, c092, c113, c115, c116, c135, c136, c137, c170, c238, c288, c299, c333, c395, c411, c451, g059, g067, g086, g095, g099, g121, g142, g154, g174, g183, g195, g233, g243, g250, g256, g261, g409, g421, g425, g474). The probes are those used by Byrne *et al.* (1998) and hybridization was as in Byrne & Moran (1994). The probes were hybridized against DNA digested with *Bgl*II or *Eco*RV at random. Any fragment pattern that was unscorable with one enzyme–probe combination was reassayed with the alternative enzyme.

Data analysis

Alleles were scored at each locus and numbered 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). A hierarchical cluster analysis was carried
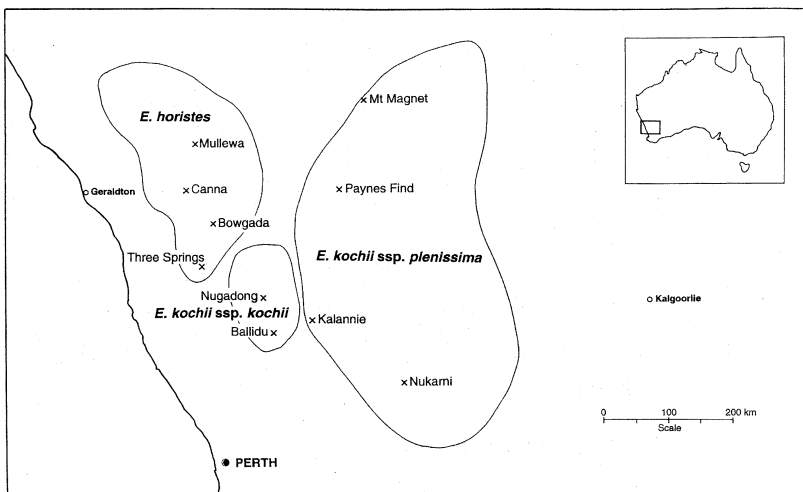

Fig. 1 Distribution and location of sampled populations of *Eucalyptus kochii* ssp. *kochii*, *E. kochii* ssp. *plenissima* and *E. horistes*.

out using UPGMA, based on unbiased genetic distance measures (Nei, 1978). A phylogenetic analysis was carried out using the CONTML program of PHYLIP (Felsenstein, 1993) and the consensus tree following 100 bootstraps of the data was produced. Estimates of gene flow as the mean number of migrants between populations per generation (Nm) were made using the F_{ST} method (Slatkin, 1987). Regression of F_{ST} and genetic distance against geographical distance was calculated and tested for significance using a Mantel randomization test (Mantel, 1967).

Results

Thirty of the probes produced interpretable fragment patterns that were scored according to an alleles/locus model. The remaining eight probes produced multiple banded patterns, resulting from hybridization to two or more loci, and were not scored. One locus was monomorphic across all taxa. Within single populations up to three loci were monomorphic. The number of alleles detected at a locus ranged from 1 to 26, with the maximum number of alleles in any one population being 19. The distributions of alleles in frequency classes across loci for each population are shown in Table 1. The observed distribution is U-shaped but strongly skewed towards alleles with low frequency (<0.1). The proportion of alleles in each frequency class was similar in all populations.

The allelic diversity measures for each population, and means over all populations and for each taxon, are presented in Table 2. The values for mean number of alleles per locus (A), mean number of polymorphic loci (P), observed heterozygosity (H_o) and the Hardy–Weinberg expected panmictic heterozygosity (H_e) were high for each population. The level of diversity is similar in all populations. The allelic diversity measures over all populations and for each taxon were high (Table 3) and were similar for each taxon and overall. The number of alleles was higher overall than for each taxon, indicating the presence of some alleles specific to populations or taxa. A large number of alleles were specific to each population: 22% of the total number of alleles, with 4–12 specific alleles in each population. All population-specific alleles were rare alleles. There was only one allele that was specific to a taxon and present in all populations of the taxon, *E. horistes*. No alleles were specific to kochii, 4.5% of alleles were specific to some but not all populations of plenissima, and 2.4% of alleles were specific to some but not all populations of *E. horistes*.

The fixation index was small and positive for all populations indicating a small degree of inbreeding. The differentiation between populations (F_{ST}) was low overall, and low for each taxon (Table 4). Plenissima had more diversity between populations than the other two taxa. The estimates of the mean number of migrants (Nm) are shown in Table 4. Gene flow was high across all populations and for each taxon, with the highest values in kochii which has the closest populations geographically. The gene flow in plenissima, although high, was less than in kochii and *E. horistes*.

Unbiased estimates of genetic distance between all pair-wise comparisons of populations were low, with the highest distance (0.0853) between the plenissima populations at Nukarni and Mt. Magnet, and the lowest distance (0.0196) between the two kochii populations (Ballidu and Nugadong).

Table 1 Allele frequency distribution for all RFLP loci for each population of *Eucalyptus kochii* ssp. *kochii*, *E. kochii* ssp. *plenissima* and *E. horistes*. Percentages in parentheses

Population	<0.1	0.1–0.2	0.2–0.3	0.3–0.4	0.4–0.5	>0.5
Nukarni	111 (65)	24 (14)	6 (4)	6 (4)	4 (2)	19 (11)
Kalannie	125 (68)	22 (12)	5 (3)	6 (3)	5 (3)	20 (11)
Mt Magnet	112 (64)	23 (13)	6 (4)	9 (5)	7 (4)	17 (10)
Paynes Find	135 (69)	22 (11)	12 (6)	3 (2)	6 (3)	17 (9)
Mullewa	161 (73)	24 (11)	9 (4)	4 (2)	5 (2)	17 (8)
Three Springs	93 (57)	29 (18)	13 (8)	5 (3)	5 (3)	18 (11)
Bowgada	114 (66)	21 (12)	10 (6)	6 (4)	4 (2)	18 (10)
Canna	110 (63)	25 (14)	9 (5)	8 (5)	4 (2)	19 (11)
Ballidue	115 (66)	25 (14)	7 (4)	4 (2)	3 (2)	21 (12)
Nugadong	104 (65)	21 (13)	8 (5)	3 (2)	3 (2)	20 (13)
Mean	118	23.6	8.5	5.4	4.6	18.6

Table 2 Genetic diversity parameters for populations and taxa of *Eucalyptus kochii* ssp. *kochii*, *E. kochii* ssp. *plenissima* and *E. horistes*. *A*, mean number of alleles per locus; *P*, mean number of polymorphic loci (0.99 criterion); H_o, observed heterozygosity; H_e, Hardy–Weinberg expected panmictic heterozygosity; F_{IS}, fixation index. Standard errors in parentheses

Population	A	P	H_o	H_e	F_{IS}
Nukarni	5.6 (0.6)	90	0.475 (0.062)	0.480 (0.064)	0.005 (0.001)
Kalannie	6.1 (0.7)	93	0.508 (0.067)	0.511 (0.062)	0.019 (0.004)
Mt Magnet	5.8 (0.6)	93	0.502 (0.066)	0.516 (0.062)	0.026 (0.006)
Paynes Find	6.5 (0.7)	93	0.488 (0.063)	0.513 (0.063)	0.039 (0.009)
Mullewa	6.7 (0.8)	93	0.492 (0.070)	0.513 (0.067)	0.057 (0.013)
Three Springs	5.4 (0.6)	90	0.431 (0.067)	0.479 (0.068)	0.101 (0.022)
Bowgada	5.8 (0.6)	93	0.471 (0.067)	0.474 (0.065)	0.028 (0.006)
Canna	5.8 (0.7)	93	0.453 (0.063)	0.479 (0.066)	0.032 (0.007)
Ballidu	5.9 (0.7)	93	0.410 (0.061)	0.465 (0.064)	0.101 (0.023)
Nugadong	5.6 (0.6)	90	0.430 (0.064)	0.455 (0.067)	0.034 (0.007)
Means:					
Overall	5.9 (0.1)	92	0.466 (0.011)	0.489 (0.007)	0.044 (0.010)
kochii	5.8 (0.2)	92	0.415 (0.010)	0.460 (0.005)	0.067 (0.033)
plenissima	6.0 (0.2)	92	0.493 (0.007)	0.505 (0.008)	0.022 (0.007)
horistes	5.9 (0.2)	92	0.462 (0.013)	0.486 (0.009)	0.054 (0.016)

A UPGMA analysis of population relationships is presented in Fig. 2. The populations of *E. horistes* cluster together and the populations of kochii cluster together. The populations of plenissima are spread out and encompass the *E. horistes* and kochii populations. A continuous character maximum likelihood analysis of population phylogeny is presented in Fig. 3. The populations separate into two clusters, one containing the two kochii populations, and the other containing the *E. horistes* and plenissima populations. Within the horistes/plenissima cluster the two northern populations of plenissima form a significant group.

There was a highly significant association between F_{ST} and geographical distance, with a Mantel test probability = 0.0021 for 10 000 iterations, indicating isolation by distance for the populations. This result was checked by testing genetic distance against geographical distance, and the association was also highly significant ($P = 0.0008$ for 10 000 iterations). However, removing the most northern population, Mt Magnet, from the analysis resulted in a nonsignificant association ($P = 0.268$), indicating that the Mt Magnet population is isolated, whereas the remaining populations are not.

Discussion

There is a very high level of variation in the nuclear genome of the oil mallee taxa *E. kochii* ssp. *kochii*, ssp. *plenissima* and *E. horistes*. The diversity in these species is greater than the diversity detected in

Table 3 Genetic diversity parameters for taxa of *Eucalyptus kochii* ssp. *kochii*, *E. kochii* ssp. *plenissima* and *E. horistes*. *A*, mean number of alleles per locus; *P*, mean number of polymorphic loci (0.99 criterion); H_o, observed heterozygosity; H_e, Hardy–Weinberg expected panmictic heterozygosity; F_{IS}, fixation index. Standard errors in parentheses

Taxon	A	P	H_o	H_e	F_{IS}
Overall	12.6 (1.1)	97	0.466 (0.019)	0.512 (0.021)	0.080
kochii	7.2 (0.8)	97	0.420 (0.043)	0.468 (0.047)	0.083
plenissima	9.8 (1.0)	97	0.493 (0.030)	0.527 (0.031)	0.058
horistes	9.6 (0.9)	97	0.462 (0.032)	0.501 (0.033)	0.080

another eucalypt species, *E. nitens* (Byrne *et al.*, 1998). Both studies used similar sampling strategies and the same RFLP probes, although the current study analysed 30 loci whereas the *E. nitens* study used 40 loci. The diversity in *E. nitens* was high in comparison to other species that have been surveyed with RFLP probes and analysed using a locus/allele model, including the long-lived woody species walnut, poplar and cocoa (Aldrich & Doebley, 1992; McGrath & Quiros, 1992; Liu & Furnier, 1993; Brubaker & Wendel, 1994; Fjellstrom & Parfitt, 1994; Lerceteau *et al.*, 1997). The diversity detected in each of the three oil mallee taxa here is higher again than these species and *E. nitens*. It was suggested that the high levels of diversity in *E. nitens* may result from the large range and regional differentiation of *E. nitens* (Byrne *et al.*, 1998). The oil mallee species have a smaller range and little geographical differentiation in comparison to *E. nitens*, yet maintain greater genetic diversity. The high level of allelic diversity is caused by a large number of alleles of low frequency, which appear to be generated within populations but maintained at low frequencies. The high numbers of rare alleles in all populations may reflect a high mutation rate in equilibrium with drift eliminating alleles and preventing their elevation to higher frequency.

The level of diversity in each of the oil mallee taxa is high yet the differentiation between populations within each taxon is low, probably because of the high levels of gene flow. Plenissima, which has the largest distribution of the three taxa, showed slightly more differentiation and lower gene flow than kochii and *E. horistes*. Over all populations the level of differentiation is low and the level of gene flow is high, indicating significant interaction between the populations. The high level of alleles present in only one population is suggestive of low levels of gene flow. The alleles present in only one population, private alleles, are all of low frequency and estimates of gene flow based on private alleles were similar to those based on F_{ST}. The differentiation that does exist between the populations is probably because of the presence of the rare alleles rather than a change in allele frequency of common alleles.

The phenetic analysis shows close relationships between most of the populations and some differentiation of the two northern populations of plenissima. This grouping reflects the geographical distribution of the populations. The arc of populations from Mullewa to Nukarni essentially follows the rainfall isohyet, with the north-eastern populations in the more arid zone. The isolation by distance detected between the most north-eastern population and the other populations is also consistent with the geographical distribution. Although the

Table 4 Differentiation and gene flow for populations of *Eucalyptus kochii* ssp. *kochii*, *E. kochii* ssp. *plenissima* and *E. horistes*, and for all populations

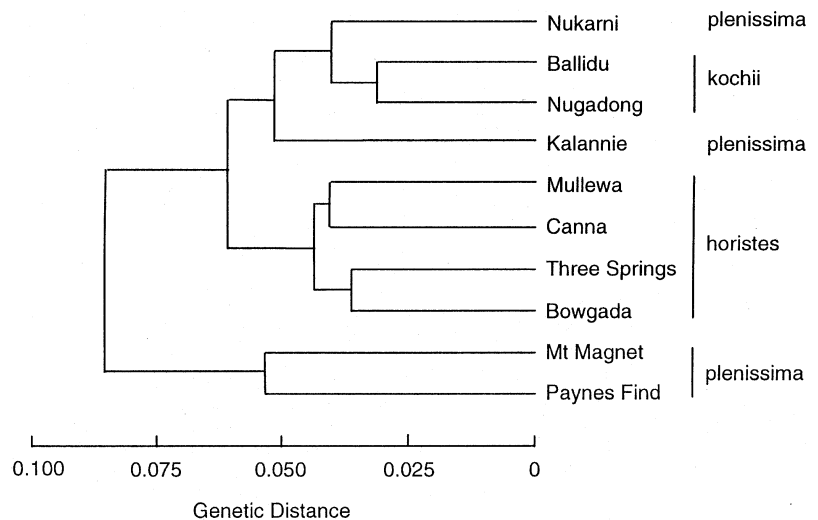| Taxon	F_{ST}	Nm
Overall	0.045	5.2
kochii	0.017	14.2
plenissima	0.042	5.6
horistes	0.029	8.4

Fig. 2 UPGMA analysis of genetic distance between populations of *Eucalyptus kochii* ssp. *kochii*, *E. kochii* ssp. *plenissima* and *E. horistes*.
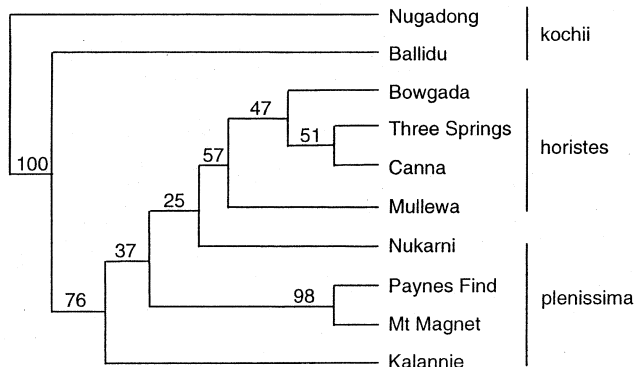
Fig. 3 Consensus tree for phylogenetic (continuous character maximum likelihood) analysis of populations of *Eucalyptus kochii* ssp. *kochii*, *E. kochii* ssp. *plenissima* and *E. horistes*. Numbers on nodes represent bootstrap confidence values for 100 replications.

Mt Magnet population appears to be isolated, it does not show strong genetic differentiation from the other populations. The taxa are insect-pollinated and pollen dispersal would be expected to be lepto-kurtic. Although long-distance gene flow may be infrequent it appears to be sufficient to prevent genetic differentiation of the populations, even where isolation may occur. The isolation by distance of the Mt Magnet population also suggests that the presence of the species in the arid north-eastern area is not the result of recent colonization, but rather indicates a more ancient presence in the area.

The phylogenetic analysis shows a close relationship between the two kochii populations. The remaining populations of plenissima and *E. horistes* form a cluster within which the grouping of populations is quite variable, except for the close relationship of the two north-eastern populations. The historic association of the kochii populations may be related to their close geographical distance; however, there are other populations (e.g. Ballidu–Kalannie, Three Springs–Bowgada) that have similar geographical distances but do not show a strong historic association. The close relationship between the kochii populations may be because they are recently derived from the main population group, or they may represent the ancestral populations from which the rest of the group has spread. Analysis of the chloroplast genome in these species may indicate the direction of evolution within the group.

The phenetic and phylogenetic relationships between the populations suggest three groupings: (i) the four *E. horistes* populations and the two southern populations of plenissima; (ii) the two kochii populations; and (iii) the two northern populations of plenissima. This grouping of populations essentially reflects their geographical positions where the *E. horistes* and southern plenissima populations form a central band, with the Mt Magnet and Paynes Find populations to the north-east and the kochii populations to the south-west.

The high diversity, high gene flow and low differentiation between the populations, combined with a correlation between populations and their geographical positioning, suggests that they represent a single, large, genetically and morphologically variable species. Within this taxon the grouping of the populations could be considered as provenances or subspecies.

There are several implications of considering these taxa as a single species for the development of a breeding programme. The high degree of genetic relatedness between the populations implies that a single breeding programme can be established without genetic contamination of the current taxa. This represents a substantial benefit, in terms of both economy and effort, for the oil mallee industry. The high levels of genetic diversity and low differentiation between populations means that a large proportion of the diversity of the taxa can be incorporated into the breeding programme through standard collection practices without extra efforts to make special collections. It would be advisable to include all taxa (provenances) in the collections. Although the average oil yield of kochii is higher than plenissima and *E. horistes*, the variation in oil production within populations is high and all populations have individuals with high oil yields. Heritability of oil yield in kochii is high ($h^2 = 0.83$) (Barton *et al.*, 1991). Breeding between high-yielding individuals from all populations would be expected to result in substantial increases in oil yield.

Acknowledgements

References

ABARE 1994. *Jojoba, Blue Mallee and Broombush: Market Assessment and Outlook*. Australian Bureau of Agricultural Research Report 94.9. Canberra, Australia.

ALDRICH, P. R. AND DOEBLEY, J. 1992. Restriction fragment variation in the nuclear and chloroplast genomes of cultivated and wild *Sorghum bicolor*. *Theor. Appl. Genet.*, **85**, 293–302.

BARTLE, J. R., CAMPBELL, C. AND WHITE, G. 1996. Can trees reverse land degradation? In: *Proceedings of the Australian Forest Growers Association*, pp. 68–75. Forest Growers Association, Mt Gambier.

BARTON, A. F. M. AND KNIGHT, A. R. 1997. High-cineole eucalyptus oils in degreasing applications. *Chemistry in Australia*, **64**, 4–6.

BARTON, A. F. M., COTTERILL, P. P. AND BROOKER, M. I. H. 1991. Heritability of cineole yield in *Eucalyptus kochii*, a mallee species endemic to Western Australia. *Silvae Genet.*, **40**, 37–38.

BOLAND, D. J., BROPHY, J. J. AND HOUSE, A. P. N. (eds) 1991. *Eucalyptus Leaf Oils*. Inkata Press, Melbourne.

BROOKER, M. I. H., BARTON, A. F. M., ROCKEL, B. A. AND TJANDRA, J. 1988. The cineole content and taxonomy of *Eucalyptus kochii* Maiden & Blakely and *E. plenissima* (Gardner) Brooker, with an appendix establishing these two taxa as subspecies. *Aust. J. Bot.*, **36**, 119–129.

BRUBAKER, C. L. AND WENDEL, J. 1994. Reevaluating the origin of domesticated cotton (*Gossypium hirsutum*; Malvaceae) using nuclear restriction fragment length polymorphisms (RFLPs). *Am. J. Bot.*, **81**, 1309–1326.

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

BYRNE, M., PARRISH, T. L. AND MORAN, G. F. 1998. Nuclear RFLP diversity in *Eucalyptus nitens*. *Heredity*, **81**, 225–232.

CHIPPENDALE, G. M. 1988. *Eucalyptus*, *Angophora* (Myrtaceae). In: *Flora of Australia*, vol. 19, pp. 245–255. Australian Government Publishing Service, Canberra.

FELSENSTEIN, J. 1993. PHYLIP (*Phylogeny inference package*) version 3.5c. Department of Genetics, The University of Washington, Seattle.

FERDOWSIAN, R., GEORGE, R. J., LEWIS, F., McFARLANE, D. J., SHORT, R. AND SPEED, R. 1996. The extent of dryland salinity in Western Australia. In: *Proceedings of the Fourth National Workshop on the Productive Use and Rehabilitation of Saline Lands*, pp. 89–98. Promaco Conventions Pty Ltd, Perth.

FJELLSTROM, R. G. AND PARFITT, D. E. 1994. Walnut (*Juglans* spp.) genetic diversity determined by restriction fragment length polymorphisms. *Genome*, **37**, 690–700.

LERCETEAU, E., ROBERT, T., PETIARD, V. AND CROUZILLAT, D. 1997. Evaluation of the extent of genetic variability among *Theobroma cacao* accessions using RAPD and RFLP markers. *Theor. Appl. Genet.*, **95**, 10–19.

LIU, Z. AND FURNIER, G. R. 1993. Comparison of allozyme, RFLP, and RAPD markers for revealing genetic variation within and between trembling aspen and bigtooth aspen. *Theor. Appl. Genet.*, **87**, 97–105.

MANTEL, N. 1967. The detection of disease clustering and a generalised regression approach. *Cancer Res.*, **27**, 209–220.

McGRATH, J. M. AND QUIROS, C. F. 1992. Genetic diversity at isozyme and RFLP loci in *Brassica campestris* as related to crop type and geographical origin. *Theor. Appl. Genet.*, **83**, 783–790.

NEI, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.

SLATKIN, M. 1987. Gene flow and the geographic structure of natural populations. *Science*, **236**, 787–792.

YEH, F. C., YANG, R.-C., BOYLE, B. J., YE, Z.-H. AND MAO, J. X. 1997. POPGENE, *the user-friendly shareware for population genetic analysis*. Molecular Biology and Biotechnology Centre, University of Alberta, Canada.