CHAPTER FOUR

GENOME-WIDE FINGERPRINTING OF *EUCALYPTUS* SPECIES AND HYBRIDS USING DIVERSITY ARRAY TECHNOLOGY (DArT)

ABSTRACT

Species of *Eucalyptus* are used extensively in the establishment of commercial forest plantations in South Africa and other parts of the world. These plantations are commonly established using large numbers of different clones and hybrids. Deployment and management of nursery material can be complicated and tools to easily verify the identity of planting stock has become increasingly important. Our aim was to assess the value of Diversity Array Technology (DArT) as a diagnostic tool to identify *Eucalyptus* species and hybrids. For this purpose, a 384-probe prototype DArT chip was constructed by arraying DNA fragments of an interspecific F_2 hybrid backcross progeny set of *Eucalyptus grandis* and *Eucalyptus globulus* on glass slides. DArT fingerprints were obtained for six *Eucalyptus* species (*E. grandis*, *E. camaldulensis*, *E. tereticornis*, *E. globulus*, *E. smithii* and *E. nitens*) and five hybrids (*E. grandis* x *E. nitens*, *E. grandis* x *E. camaldulensis*, *E. grandis* x *E. globulus*, *E. grandis* x *E. tereticornis and E. grandis* x *E. smithii*) by hybridising labelled genomic representations of the individual trees to the 384-probe array. Microarray analysis showed that of the 384 DNA fragments on the chip, 142 (37%) were polymorphic among hybrids. Our results suggest that DArT is an efficient DNA marker technology to differentiate *Eucalyptus* species and F_1 hybrids and thus to more effectively manage plantation programmes.

INTRODUCTION

Eucalyptus species are commonly used to establish commercially important plantations in many parts of the world. The genus *Eucalyptus* includes more than 900 species that differ markedly in their value for forestry (Pryor and Johnson 1971; Chippendale 1988; Eldridge et al. 1994). The species of *Eucalyptus* that have been most commonly used for plantation forestry reside in three sections of the subgenus *Symphyomyrtus*, i.e. *Latoangulatae*, *Exsertaria* and *Maidenaria* (Brooker and Kleinig 1994; Steane et al. 2002). Species in these three sections have been used to produce superior F_1 hybrid progeny including hybrids with a variety of valuable traits such as disease tolerance, accelerated growth, wood quality and drought tolerance (Eldridge et al. 1994).

Many *Eucalyptus* hybrids used in South African plantations were naturally generated during the early stages of plantation forestry. During this period, *Eucalyptus* species were commonly planted in multi-species trials. Several generations of open pollination and natural selection produced *Eucalyptus* populations that were well adapted to the South African environment. A variety of such natural hybrids displaying favourable characteristics were consequently used in breeding programmes. As a result, many of the commercially grown South African *Eucalyptus* genotypes are believed to be natural hybrids, and not pure species (Eldridge et al. 1994, Steenkamp et al. 2003). It is thus important to know the species composition of improved *Eucalyptus* genotypes for further genetic improvement of such trees.

Eucalyptus species are relatively easy to hybridise, but interspecific hybridisation is most successful between species of the same section of a subgenus. Crosses between species in different sections of a subgenus often

result in hybrids with reduced viability and fertility (Griffin et al. 1988). Crosses may result in very low seed set and a high proportions of abnormal hybrid plants (Myburg et al. 2003). Such an example is the cross between *Eucalyptus grandis* (section *Latoangulatae*) and *Eucalyptus globulus* (section *Maidenaria*) (Griffin et al. 2000). From a management and developmental perspective, trees to be crossed could be chosen based on their genomic composition when producing interspecific F_1 hybrids (Myburg et al. 2003). Currently, a wide variety of natural and artificial hybrids are used in breeding programmes, and methods are needed to identify and discriminate pure species and hybrid genotypes. It would also be desirable to know what parts of genomes have been introgressed in advanced hybrid generations.

DNA-based molecular marker techniques provide a powerful approach for obtaining robust and unambiguous identification of breeding and planting stock. These techniques include restriction fragment length polymorphisms (RFLPs, Botstein et al. 1980), simple sequence repeats (SSRs, Weber and May 1989), random amplified polymorphic DNAs (RAPDs, Williams et al. 1990), and amplified fragment length polymorphisms (AFLPs, Vos et al. 1995). Despite the robustness and high throughput afforded by some of these methods, they require the amplification of individual markers, which have to be analysed serially by gel electrophoresis.

Microarray technology provides a tool for the parallel processing of very large numbers of DNA fragments immobilized on a solid-state surface (Schena et al. 1995). To adopt microarray technology for fingerprinting and diversity studies, Jaccoud et al. (2001) reported the development of Diversity Array Technolgy (DArT™) in rice. Since no prior sequence information is needed for

DArT, this technique provides a useful analysis tool for previously uncharacterised genomes like *Eucalyptus*. Microarray-based genotyping as implemented in DArT allows the parallel analysis of thousands of marker loci by detection of selectively assayed genomic restriction fragments (Lezar et al. 2004).

The aim of this study was to determine levels of microarray-based restriction polymorphism among commercial species of *Eucalyptus* and hybrids of these species using DArT. The efficiency of microarray-based fingerprinting is expected to be higher in interspecific hybrids compared to within species, since interspecific hybrids should have increased levels of polymorphisms. A 384-marker DArT genotype chip was constructed to investigate this expectation for six *Eucalyptus* species and five hybrid combinations.

MATERIALS AND METHODS

Plant materials

Commercially-grown *Eucalyptus* species and hybrids were selected for this study (Figure 1). *Eucalyptus grandis*, *E. camaldulensis*, *E. globulus*, *E. nitens*, *E. tereticornis* and *E. smithii* species were grown from seed certified as true to type and has been collected from natural stands in Australia. The origin of each of the CSIRO seed sources can be obtained on the web at http://www.ffp.csiro.au/tigr/atscmain. Representative hybrids were *E. grandis* x *E. globulus* maintained by Forestral Oriental S.A. (FOSA, Uruguay), *E. grandis* x *camaldulensis*, *E. grandis* x *tereticornis* and *E. grandis* x *E. nitens* hybrids

obtained from Mondi Business Paper South Africa (Pietermaritzburg, South Africa), as well as *E. grandis* x *E. smithii* hybrids from SAPPI (South Africa). Genomic DNA was extracted individually from young leaves of two separate plants of each *Eucalyptus* species and hybrids using the method described by Murray and Thompson (1980).

Microarray analysis

A 384-probe microarray was constructed by arraying bulk of genomic DNA from eight individuals of an interspecific F_2 hybrid backcross progeny set of *Eucalyptus grandis* and *E. globulus* (Myburg et al. 2003) on glass slides. DNA fragments were prepared by restriction digestion, ligation and pre-amplification as described in the original AFLP protocol of Vos et al. (1995). The PCR pre-amplification reactions were performed in 30 µl containing 0.3 µM *Eco*RI (E + A) and *Mse*I (M + C) adapter primers (Vos et al. 1995), 0.2 mM of each dNTP, 0.6 U *Taq* polymerase, and 1 x reaction buffer (Roche Diagnostics GmbH). The PCR amplification consisted of 30 cycles of 94°C for 30 sec, 54°C for 30 sec, and 72°C for 1 min with a 1 sec per cycle increase; an initial step of 72°C for 30 sec, and a final extension step of 72°C for 2 min. Cloning, PCR amplification, array printing and processing of pre-amplified fragments was performed as described in Lezar et al. (2004). Eight replicates per fragment were arrayed on each slide.

DNA fingerprints were obtained by assaying the presence or absence of restriction fragments in pairs of trees on the 384-element DArT chip in 2-dye experiments (Cy3 or Cy5). Probe DNA preparation, labelling, hybridisation and

scanning were carried out as described by Lezar et al. (2004). One biological replicate was performed for each hybridisation, each replication consisting of a reverse labeling experiment from the same pair of trees. Signal intensity quantification was done with Array Vision 6.0 software (Imaging Research Inc., Molecular Dynamics, USA). A background-corrected spot intensity was obtained by subtracting the median local background of each spot from the median pixel intensity inside the spot. Background-corrected spots with signal intensities that varied more than 10% from the mean of the eight spot replicates on each slide were excluded from further analyses. The remaining intensity values were normalized across slides by global regression on the spot intensity data for the *E. smithii* tree, which was used as a reference for normalization of all spot intensity data. Spots with intensities that deviated more than 10% from the median spot intensity of a repeated slide were also excluded from further analyses.

Normalized signal intensities were used to identify polymorphic fragments based on the bimodal distribution of their intensity values across individuals. Relative intensity values and frequency distributions were obtained as described by Lezar et al. (2004). Each polymorphic DArT marker identified was then scored for the presence (1) or absence (0) of the fragment in the genome representations of the *Eucalyptus* species and hybrids. Binary scores (presence: absence) were then used to determine the average distances among *Eucalyptus* species and hybrids with PAUP (Swofford 1998). The index of genetic similarity of Nei and Li (1979) was used to calculate pairwise genetic distances for all trees. Unweighted pair-group mean arithmetic analysis (UPGMA, Sheath and Sokal, 1973) was used for cluster analysis of the pairwise

distance matrix, which generated a dendrogram representing the genetic distances among the *Eucalyptus* species and hybrids.

RESULTS

Data quality

A 384-element chip was constructed from selectively amplified restriction fragments of an interspecific F_2 hybrid backcross progeny set of *Eucalyptus grandis* and *E. globulus* for the genome-wide fingerprinting of commercial species and hybrids of these species. Approximately 370 background-corrected spots (96.4%) had signal intensities that varied less than 10% from the mean of the eight spot replicates and were used for subsequent data analysis. Spots excluded from analysis (3.6%) showed inaccuracies in signal intensities. This can be ascribed to variability in the experimental process introducing inaccuracies in labelling, array hybridisation, signal detection and quantification. This approach allowed us to obtain repeatable scores, increasing the reliability of the data.

Proportion of polymorphic fragments useful for fingerprinting

The DArT fingerprinting chip was used to assay the presence or absence of microarray markers in the *Eucalyptus* species and hybrids. While many of the array features were common to all individuals (42%), or showed no hybridisation signal (19%), many were found to be polymorphic (39%) among

the species and hybrids represented in the study. Of the 384 polymorphic fragments, 142 (37%) were found to be polymorphic among hybrids. The hybrid combination *E. grandis* x *E. smithii* shared 56% of markers with the parental species, while *E. grandis* x *E. tereticornis*, *E. grandis* x *E. nitens*, *E. grandis* x *E. camaldulensis*, *E. grandis* x *E. globulus* shared 54%, 55%, 49% and 46% of the markers with the parental species, respectively.

In case of pure species, a total of 123 (32%) of DNA fragments were found to be polymorphic and were used for subsequent data analysis. The pure species *E. smithii* shared 48% with E. *grandis*, while E. *globulus*, *E. nitens*, *E. tereticornis* shared 46%, 47% and 52% of the markers with *E. grandis*, respectively. A unique microarray pattern was obtained for each *Eucalyptus* tree. All of the hybridization profiles allowed unambiguous discrimination of the individuals. Small randomly selected subsets of polymorphic DNA fragments determined that as few as six polymorphisms were sufficient to discriminate among species and seven among hybrids.

Cluster analysis

A total of 142 polymorphic DArT fragments were used for cluster analysis. On the basis of consistent binary data, we estimated the pair-wise distances by means of PAUP (Table 1). Between-individual distances are based on the mean character differences in the binary scores. As evident from the pair-wise distances, high interspecific polymorphisms are characteristic of the genus *Eucalyptus*. The coefficient of similarity (average differences) between species varied from 0.0347 to 0.3139. The hybrids have a similarity coefficient ranging from 0.0709 to 0.3160.

Cluster analysis of the pairwise distance matrix generated a dendrogram representing the relationships of hybridisation profiles among the *Eucalyptus* species and hybrids (Figure 2). With cluster analysis it was possible to distinguish among the six pure *Eucalyptus* species and the five hybrids included in this study. Biological replicates were all identical and replicate fingerprints all clustered as nearest neighbors (Figure 2). UPGMA analysis suggests that each section formed its own distinct cluster, with *Eucalyptus grandis* (section *Latoangulate*) being further separated. (Figure 2). The clustering of *E. grandis* with the hybrids can most likely be ascribed to the genetic composition of the hybrids used in this study. Most fragments were found to be common to the *Eucalyptus grandis* genotype since all hybrids were crosses between *E. grandis* and another pure species.

DISCUSSION

Microarray-based genotyping of *Eucalyptus* species and hybrids will provide the opportunity for high-throughput, genome-wide fingerprinting of these important tree species. In this study, the main goal was to assess the applicability of this approach for molecular diagnostics and fingerprinting studies in commercially grown *Eucalyptus* species and hybrids. For this purpose, we used a small prototype genotyping array for *Eucalyptus* trees employing Diversity Array Technology. DArT detects DNA polymorphism by comparing the composition of genomic restriction products of different

genotypes through hybridisation to previously arrayed restriction fragments (Jaccoud et al. 2001). Our results suggest that DArT is an efficient DNA marker technology for genome-wide fingerprinting of *Eucalyptus* species and hybrids.

The advantages of DArT and the recent progress that has been made towards the application of microarray genotyping technology (Jaccoud et al. 2001, Borevitz et al. 2003, Lezar et al. 2004, Wenzl et al. 2004) have provided opportunities for the fingerprinting of *Eucalyptus* trees. However, the high cost of array-based genotyping remains a problem. This includes the cost of arrays, supplies and equipment. In some cases, the need for replicate experiments may also increase costs. In time, DArT should benefit from the development of improved computational and statistical approaches, as well as new array slides, formats and designs that allow cheaper assays. However, even at current costs, individual DArT assays are very competitive relative to other high-throughput technologies such as AFLP analysis.

The power of microarray-based DNA fingerprinting lies in its ability to compare different genomes at a large number of loci, in a single assay. In this context, direct comparison of hybridisation fingerprints may allow the accurate identification of individuals. We were able to generate highly repeatable hybridisation data. Hybridisation profiles were 95% identical to those obtained in replicate, but reverse labelling reactions, and the replicate fingerprints all clustered as nearest neighbours. Binary scores of the hybridisation profiles revealed that 32% of DArT markers were polymorphic among pure *Eucalyptus*

species, while the rate of polymorphism among hybrids (i.e. 3-way and 4-way genome comparisons) was 37%.

Previously, we used a pure-species DArT array to fingerprint *Eucalyptus grandis* genotypes (Lezar et al. 2004). Only 55 (15%) of 384 DArT loci were polymorphic and sufficiently robust for subsequent data analysis. In this study, 142 (37%) polymorphic loci were identified and were useful for data analysis. This probably reflects the higher amount of polymorphism in interspecific genome comparisons and the robustness afforded by a great proportion of strong presence/absence hybridisation signals. The prototype array constructed with DNA fragments of a F_2 hybrid backcross progeny set of *Eucalyptus grandis* and *Eucalyptus globulus* therefore proved to be an efficient genotyping platform.

In the longer term we are interested in developing molecular diagnostic procedures for identifying commercially grown *Eucalyptus* genotypes in South Africa. However, for developing such diagnostic methods, larger population studies of the *Eucalyptus* species are a prerequisite. Species-specific DArT fragments have to be identified from a sufficient number of individuals from a species and confirmed to be absent from other species. Since interspecific hybridisation of different species has become an important strategy for the improvement of *Eucalyptus*, microarrays should be useful diagnostic tools to confirm genotypes before crossing species and afterwards to confirm parentage. Although species-specific fragments still have to be identified to determine the

species composition of hybrids, this technology can already be used to track genomic regions inherited from parents in outcrossed pedigrees.

The results of this study illustrate the potential of DArT for identification of *Eucalyptus* trees in breeding programmes, particularly for application in less-studied plant genomes. DArT allowed us to easily differentiate pure species and hybrid individuals, which could be expected, as we had already demonstrated that a full-sib progeny could be distinguished (Lezar et al. 2004). Unique microarray fingerprints were identified for all genotypes.

Whole-genome profiling using DArT opens significant opportunities for tree breeding programmes and for future genome analysis of *Eucalyptus*. The *Eucalyptus* genome (*E. camaldulensis*) is currently being sequenced to 4x coverage at KAZUSA DNA Research Institute in Japan. A DArT array with 3000 – 4000 polymorphic features would be useful to construct very detailed genetic maps using *E. camaldulensis* as a parent in interspecific crosses. These maps could all be linked to the genome sequence by sequencing the polymorphic DArT fragments. This approach could also be used to anchor a physical map of the *Eucalyptus* genome to the genome sequence.

REFERENCES

Borevitz JO, Liang D, Plouffe D, Chang H-S, Zhu T, Weigel D, Berry CC, Winzeler EA, Cherry J (2003) Large-scale identification of single-feature polymorphisms in complex genomes. Genome Research 13: 513 – 523.

Botstein D, White R, Skilnick M, Davis R (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32: 314 – 331.

Brooker MIH, Kleinig DA (1994) Field guide to Eucalypts. Inkata Press, Australia.

Chippendale GM (1988) *Eucalyptus*, *Angophora* (*Myrtaceae*). Flora of Australia 19. Australian Government Publishing Service, Canberra, 542pp.

Eldridge K, Davidson J, Hardwood C, van Wyk G (1994) Eucalypt Domestication and Breeding. Clarendon Press, Oxford, 288pp.

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

Griffin AR, Harbard J, Centurion C, Santini P (2000) Breeding *Eucalyptus grandis* x *globulus* and other interspecific hybrids with high inviability – problem analysis and experience at Shell Forestry Projects in Uruguay and Chile. In: Dungey, H.S., Dieters, M.J., Nikles, D.G. (eds). April 2000: Hybrid breeding and genetics of forest trees. Proc. QFRI/ CRC – SPF Symposium, 9 – 14 April 2000, Noosa, Queensland, Australia, Department of Primary Industries, Brisbane, pp. 1- 13.

Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity Arrays: a solid state technology for sequence independent genotyping. Nucleic Acids Research 29: e25.

Lezar S, Myburg AA, Berger DK, Wingfield MJ, Wingfield BD (2004) Development and assessment of microarray-based DNA fingerprinting in *Eucalyptus grandis*. Theor Appl Genet 109:1329-1336.

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Research 8: 4321 – 4325.

Myburg AA, Griffin AR, Sederoff RR, Whette RW (2003) Comparative genetic linkage maps of *Eucalyptus grandis*, *Eucalyptus globulus* and their F_1 hybrid based on a double pseudo-backcross mapping approach. Theor Appl Genet 107: 1028 – 1042.

Nei M, Li W-H (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Natl Acad Sci USA 76: 5269 – 5273.

Nouzová M, Neumann P, Navrátilová A, Galbraith DW, Macas J (2001) Microarray-based survey of repetitive genomic sequences in *Vicia* spp. Plant Molecular Biology 45:229-244.

Pryor LD, Johnson LAS (1971) Classification of the Eucalypts. Australian National University Press, Canberra, 102pp.

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270: 467 – 470.

Sneath PHA, Sokal RR (1973) Numerical Taxonomy. Freeman W.H., San Francisco.

Steane DA, Nicolle D, McKinnon GE, Vaillancourt RE, Potts BM (2002) Higher-level relationships among the eucalypts are resolved by ITS-sequence data. Aust Syst Bot 15:49 – 62.

Steenkamp ET, van der Nest MA, Wingfield MJ, Wingfield BD (2003) Detection of hybrids in commercially propagated Eucalyptus using 5S rDNA sequence. Forest Genet 10:195-205.

Swofford DL (1998) PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland Massachusetts.

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Fritjers A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23: 4407 – 4414.

Weber J, May P (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 44: 388 – 396.

Wenzl P, Carling J, Kudrna D, Jaccoud D, Hyttner E, Kleinhofs A, Kilian A (2004) Diversity Array Technology (DarT) for whole-genome profiling of barley. Proc Natl Acad Sci USA 101: 9915 - 9920.

Williams J, Kubelik A, Livak K, Rafalski J, Tingey S (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research 18: 6531 – 6535.

Winzeler EA, Richards DR, Conway AR, Goldstein AL, Kalman S, McCullough MJ, McCusker JH, Stevens DA, Wodicka L, Lockhart DJ et al. 1998: Direct allelic variation scanning of the yeast genome. Science 281: 1194 – 1197.

Table 1. Average differences among six *Eucalyptus* species and five hybrids studied

Eucalyptus Individual	Average differences[a]										
	E. smithii	E. smithii x E. grandis	E. globulus	E. globulus x E. grandis	E. grandis	E. grandis x E. tereticornis	E tereticornis	E. grandis x E. camaldulensis	E. camaldulensis	E. grandis x E. nitens	E. nitens
E smithii	-										
E smithii x E. grandis	0.07640316	-									
E globulus	0 03524108	0 09995972	-								
E. globulus x E. grandis	0.29838815	0.3103627	0 26757753	-							
E. grandis	0.26988026	0.09484581	0.31392452	0.40193033	-						
E. grandis x E tericornis	0.15636763	0 14809948	0 12190829	0.23796162	0.14496805	-					
E. tereticornis	0.06311531	0.12731433	0.0773048	0.23580265	0 25481093	0 12736213	-				
E grandis x E. camaldulensis	0.13080375	0.09323128	0.16472898	0.27078599	0.11667355	0.16643387	0.10851608	-			
E. camaldulensis	0.06108693	0 11066774	0.09463277	0 23434879	0.2183733	0.15757513	0.03785607	0.08884408	-		
E. grandis x E. nitens	0 10711877	0.07089831	0.10239429	0.26757753	0.08575626	0.12915635	0.13608161	0.10136519	0.1282395	-	
E. nitens	0.06426094	0.09930894	0.03470733	0.24555765	0.20522882	0.1444647	0.08654109	0.14105862	0 09856221	0.06665345	-

[a] Average differences were determined with PAUP (Swofford 1998), using the binary score for each of the individuals studied. Binary scores were obtained with ArrayVision and Microsoft Excel.

Figure 1. Schematic representation of commercially-grown *Eucalyptus* species and hybrids used in this study.

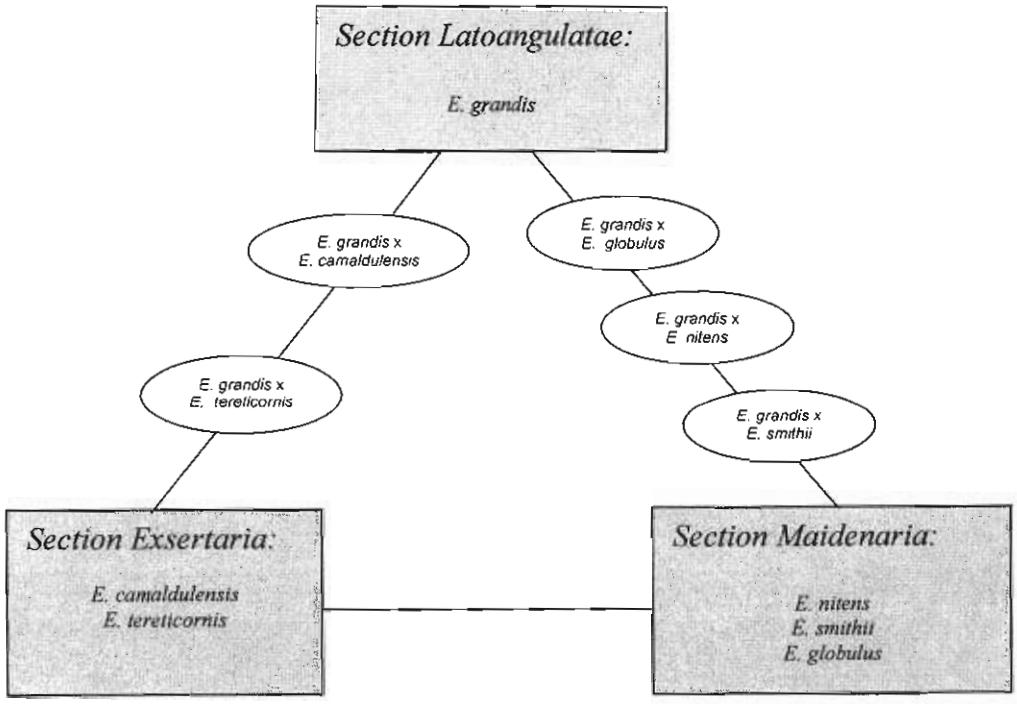
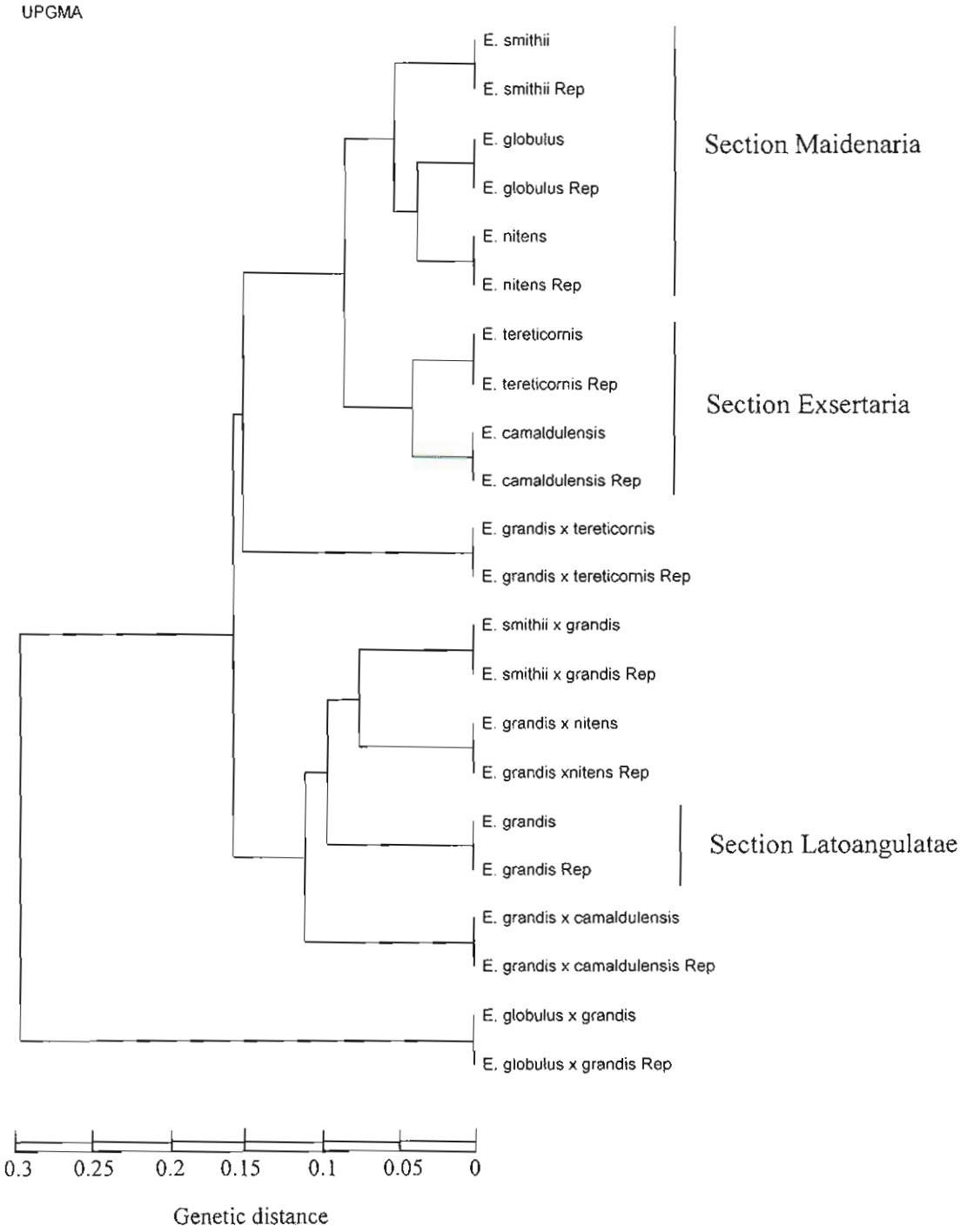

Section Latoangulatae:
E. grandis
E. grandis x E. camaldulensis
E. grandis x E. globulus
E. grandis x E. nitens
E. grandis x E. tereticornis
E. grandis x E. smithii
Section Exsertaria:
E. camaldulensis
E. tereticornis
Section Maidenaria:
E. nitens
E. smithii
E. globulus

Figure 2. Dendrogram generated by UPGMA cluster analysis of pair-wise distance data for six *Eucalyptus* species and five hybrids. Eleven of the hybridizations were performed in replicate (indicated as REP).


UPGMA
E. smithii
E. smithii Rep
E. globulus
E. globulus Rep
E. nitens
E. nitens Rep
Section Maidenaria
E. tereticornis
E. tereticornis Rep
E. camaldulensis
E. camaldulensis Rep
Section Exsertaria
E. grandis x tereticornis
E. grandis x tereticornis Rep
E. smithii x grandis
E. smithii x grandis Rep
E. grandis x nitens
E. grandis xnitens Rep
E. grandis
E. grandis Rep
Section Latoangulatae
E. grandis x camaldulensis
E. grandis x camaldulensis Rep
E. globulus x grandis
E. globulus x grandis Rep
0.3
0.25
0.2
0.15
0.1
0.05
0
Genetic distance