UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

ASSESSMENT AND DEVELOPMENT OF MICROARRAY-BASED DNA FINGERPRINTING IN *EUCALYPTUS GRANDIS* AND RELATED SPECIES

BY

SABINE LEZAR

Submitted in partial fulfilment of the requirements for the degree

PHILOSOPHIAE DOCTOR

Department of Genetics
in the Faculty of Natural and Agricultural Science
University of Pretoria
Pretoria

May 2005

SUPERVISOR: PROF. B.D. WINGFIELD

CO-SUPERVISORS: DR. A.A. MYBURG
PROF. D.K. BERGER
PROF. M.J. WINGFIELD

DECLARATION

I, the undersigned hereby declare that the thesis submitted herewith for the degree *Philosophiae Doctor* to the University of Pretoria, contains my own independent work and hitherto has not been submitted for any degree at any university or faculty.

Sabine Lezar

May 2005

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I wish to express my gratitude and sincere appreciation towards the following people and institutions:

- My promoters Profs. Brenda Wingfield, Zander Myburg, Dave Berger and Mike Wingfield for their guidance, comments and endless patience;
- The Department of Genetics and the Forestry and Agricultural Biotechnology Institute (FABI) for support and providing facilities to undertake my studies;
- The University of Pretoria; the National Research Foundation (NRF), the members of the Tree Protection Co-operative Programme (TPCP); the THRIP initiative for financial support;
- Vivienne Clarence for providing for my laboratory, as well as, emotional needs;
- My fellow students for their assistance whenever I needed it;
- My husband Andre for his love and emotional support to fulfil my dreams;
- My family and friends for inspiration and support;

PREFACE

Eucalyptus species are an essential component of commercial forest plantations in tropical and subtropical regions of the world. Major efforts are now directed towards breeding and selection of elite genotypes with tolerance to diseases such as Chrysoporthe canker caused by *Chrysoporthe austroafricana*. Currently, DNA-based genetic markers are the most widely used techniques for fingerprinting and identification of superior genotypes. These techniques, however, are limited by the resolution of the detection system, which commonly involves electrophoresis to separate fragments based on size. This limitation has prompted the investigation of microarray-based DNA fingerprinting presented in this thesis.

The first chapter of this thesis presents a review of literature regarding the current opportunities and challenges of microarray technology for disease resistance breeding in forest trees. This review is arranged in four sections. The first of these deals with the impact of disease on forest trees, with particular reference to *Chrysoporthe austroafricana*. This is followed by an overview of the DNA microarray technique. The third section is concerned mainly with the uses of microarray markers for genetic analysis of forest trees. The last section deals with the challenges of microarray data analysis and management.

Chapter Two of this thesis reports on the development and assessment of microarray-based DNA fingerprinting in *Eucalyptus grandis*. *Eucalyptus grandis*

clones were previously identified using random amplified polymorphic DNA (RAPD) and microsatellite markers. While these techniques are constrained by their reliance on gel or capillary electrophoresis, recently developed microarray technology is not constrained by these requirements and holds promise for parallel analysis of thousands of markers in plant genomes. The aim of the study was to generate a small genotyping array for *Eucalyptus* and to determine the usefulness of microarrays for fingerprinting a full-sib progeny of *Eucalyptus grandis*. For this purpose, we implemented the recently developed Diversity Array Technology (DArT) for *E. grandis*.

Chapter Three concerns the identification of molecular markers associated with *Chrysoporthe austroafricana* tolerance in *Eucalyptus* using bulk segregant analysis (BSA) and DarT. Currently, the most effective way to reduce the impact of Chrysoporthe canker on *Eucalyptus* trees is the selection of elite genotypes with best disease tolerance. These trees were selected by artificial inoculations with a virulent isolate of *Chr. austroafricana* and monitoring disease progress. However, trials to test disease tolerance of clonal hybrids using artificial inoculations are extremely time consuming. The work presented in this chapter was, therefore, designed to develop PCR markers converted from microarray markers for the fast and unambiguous screening of breeding stock for disease tolerance.

Chapter Four reports on the genome-wide fingerprinting of commerciallygrown *Eucalyptus* species and hybrids using DArT. The most widely used tools for

molecular diagnostics and genetic diversity studies in *Eucalyptus* are amplified fragment length polymorphism (AFLP), random amplified polymorphic DNA (RAPD) and microsatellite markers. Large volumes of information about these fingerprinting techniques are currently available. However, information pertaining to the efficiency of genome-wide fingerprinting of commercial species of *Eucalyptus* and hybrids using an AFLP-like complexity reduction method is not available yet.

This thesis presents a collection of studies conducted over four years that deal with microarray-based DNA fingerprinting in *Eucalyptus*. All studies were conducted independently and have been written as separate publishable units. Thus, some repetition between parts of chapters, which contain a progression of knowledge accumulated over a relatively long period of time, has been unavoidable. It is my sincere hope that the work presented in this thesis has contributed to our knowledge regarding the exploitation of DNA microarrays as a marker analysis method in *Eucalyptus* tree breeding programs.

SUMMARY

DNA microarray technology is a new and powerful technology that could substantially increase the speed of forest tree breeding programmes. This thesis represents a compilation of investigations that focus on the exploitation of DNA microarray technology for genetic marker analysis of *Eucalyptus* trees. The major focus of the studies presented in this thesis was on the assessment and development of microarray-based DNA fingerprinting in *Eucalyptus*.

A DNA chip for *Eucalyptus* was not available at start of the study. As a result of this study a 384-prototype chip was developed to evaluate the potential of microarrays for fingerprinting closely related *Eucalyptus* clones, species and hybrids. These studies show that microarrays are an efficient DNA marker technology for genome-wide fingerprinting of complex organisms for which no sequence data exist. However, cross-hybridisation and the lack of dedicated software products remain a challenge.

The 384-probe array developed in this study was subsequently employed for the detection of putative markers associated with tolerance to *Chrysoporthe austroafricana* in *Eucalyptus grandis*. Putative tolerance-associated markers were identified by bulk segregant analysis (BSA) and converted to cleaved amplified polymorphic sequence markers for further characterization in segregating *Eucalyptus* populations. BSA revealed a total of 109 scorable, polymorphic loci, of

which nine appeared to be associated with tolerance or susceptibility. Two DArT markers were converted to cleaved amplified polymorphic sequence (CAPS) markers, which discriminate susceptible and tolerant individuals. These PCR markers can be used for the rapid screening for disease tolerance in *Eucalyptus* planting and breeding stock.

The collection of studies included in this thesis demonstrated that DArT is an efficient DNA marker technology for genome-wide genotyping, particularly for application in less-studied plant genomes. Whole-genome profiling using DArT raises significant opportunities for tree breeding programmes and for future genome analysis of *Eucalyptus*.

CHAPTER ONE

LITERATURE REVIEW

DEVELOPMENT AND ASSESSMENT OF MICROARRAY-BASED DNA FINGERPRINTING IN *EUCALYPTUS GRANDIS*

INTRODUCTION

Forest trees are essential components of the natural landscape and of many ecological systems. They also form the foundation of a global, multi-billion dollar wood products industry. The economic importance of forest trees has contributed to a worldwide focus on the selection of breeding material with increased resistance or tolerance to pests and diseases. The first attempts to select for improved genotypes were based on morphological characteristics (Eldridge et al. 1994). However, most forest tree species are characterized by long generation intervals and much time is needed before assessment of disease resistance can be carried out (Grattapaglia 1994). This limitation has precluded the rapid identification and selection of disease resistant trees. A maximum of three to four generations of breeding have been completed in most commercial forest tree species. This has not provided sufficient opportunity to introgress disease resistance genes into germplasm that have otherwise only been selected for growth and form (Byrne et al. 1996). Breeders are in need of selection systems that will allow early identification of resistant germplasm and that will furthermore facilitate the introgression of resistance factors into elite or improved breeding material.

The discovery of isoforms and allelic variants of proteins (isozymes and allozymes) provided the first genetic markers based on molecular variation in plant and animal genomes (Markert and Moller 1959, Tanksley and Orton 1983, Weeden 1988). In the early 1970s, protein variants were first studied in trees by Bartels (1971) and

Bergmann (1971). Since then, much progress has been made in experimental genetic studies of forest tree species using isozyme analysis (Strauss and Conkle 1986, Niebling et al. 1987, Shirasi 1998). Isozyme markers have been useful for the identification of varieties (Ashari et al. 1989, Huang et al. 1994), recognition of economically important genes (Mackill et al. 1993), construction of genetic maps (Bernatzky and Tanksley 1986) and detection of genetic introgression (William and Mujeeb-Kazi 1993). However, isozyme markers are often dependent on the environmental conditions under which the plant is grown and the developmental stage of plant tissues used for protein extraction (Winter and Kahl 1995). In addition, relatively few proteins are amenable to isozyme analysis, which limits the genome coverage that can be achieved with these markers. These problems have largely been overcome with the development of DNA-based molecular markers, and high-throughput, automated genotyping technology.

A number of DNA-based molecular marker technologies have been developed in the past two decades. Such marker technologies include restriction fragment length polymorphisms (RFLPs, Botstein et al. 1980, Neale et al. 1989), simple sequence repeats (SSRs, Weber and May 1989, Jarne and Lagoda 1996, Goldstein and Schloetterer 1999), random amplified polymorphic DNA (RAPD, Williams et al. 1990, Rafalski et al. 1991) and amplified fragment length polymorphisms (AFLPs, Vos et al. 1995, Blears et al. 1998). These molecular markers have become fundamental tools for forest geneticists. They have assisted tree breeders in improving the efficiency of classical breeding techniques (Smith and Beavis 1996).

In most plant species, molecular markers have been particularly useful in manipulating disease resistance genes (Michelmore 1995, Okubara et al. 1995, Wilcox et al. 1996, Harkins et al. 1998). Currently, the major limitation of these molecular marker methods is their dependence on gel electrophoresis (Jaccoud et al. 2001). This hampers the processing of a large number of samples or markers in parallel. There are difficulties in precisely correlating bands with allelic variants, and only one or a few loci can be studied at a time.

Jaccoud et al. (2001) demonstrated that these limitations can readily be overcome by using microarrays. DNA microarray technology has the potential to greatly increase the level of automation and thoughput of molecular marker technology (Schuchhardt et al. 2000, Foster and Huber 2001). Hundreds or thousands of loci can be analysed in a short time. This significantly decreases the time required to identify disease resistance loci and develop markers for breeding purposes. The efficiency of this process facilitates the release of new varieties that display durable resistance. Furthermore, microarray-based breeding has the potential to allow whole-genome selection, which will enable tree breeders to manipulate the complete genomes of trees, while selecting for individual, high value loci.

The focus of this review is on the exploitation of DNA microarrays as a method for developing disease resistance markers to be used in tree breeding programmes, and particularly those related to Chrysoporthe canker caused by *Chrysoporthe austroafricana* in South Africa.

THE IMPACT OF DISEASE ON FOREST TREES IN SOUTH AFRICA, WITH PARTICULAR REFERENCE TO *CHRYSOPORTHE*

The forestry industry in South Africa depends on a small number of fast-growing species, most notably in the genera *Eucalyptus* and *Pinus* (Evans 1992). These species have been developed in South Africa due to their rapid growth and easy nursery propagation. In recent years, planting *Eucalyptus* species has become increasingly important (Directorate of National Forestry Planning 1987, Eldridge et al. 1994). This interest in eucalypts and the trend towards propagation of clones from cuttings has prompted concern regarding the role that diseases, such as those caused by *Chrysoporthe austroafricana* Gryzenhout & M. J. Wingf. (Gryzenhout et al. 2004), could have on the sustainability of the industry (Conradie et al. 1990, Wingfield et al. 1991).

Chrysoporthe species, previously united under the single name *Cryphonectria cubensis* (Bruner) Hodges (Gryzenhout et al. 2004), are serious canker pathogens of commercially grown *Eucalyptus* spp. in plantations (Boerboom & Maas 1970, Hodges et al. 1979, Hodges 1980, Florence et al. 1986, Wingfield et al. 1989, Wingfield 2003). DNA sequence comparisons showed that isolates from South East Asia, South America and South Africa represented distinct phylogenetic groups (Myburg et al. 2002, 2003). Specimens of *Cryphonectria cubensis* from South East Asia and South America could, however, not be separated or distinguished morphologically from the type specimen, originating from Cuba (Gryzenhout et al. 2004). For the present, these specimens are thus collectively transferred to

Chrysoporthe as a single species, *Chrysoporthe cubensis* (Bruner) Gryzenhout & M.J. Wingf. (Gryzenhout et al. 2004). Specimens previously treated as *Cryphonectria cubensis* from South Africa could be distinguished morphologically from those representing *Chrysoporthe cubensis* and was thus described as *Chrysoporthe austroafricana*.

Chrysoporthe spp. have severely limited the development of plantations of susceptible *Eucalyptus* species, where climatic conditions favour disease development (Alfenas et al. 1982, Sharma et al. 1985, Florence et al. 1986). Thus, it has become important to implement an effective disease management strategy for these pathogens. There are several means to accomplish this goal. The most effective means to reduce losses due to *Chr. cubensis* and *Chr. austroafricana* is to plant disease tolerant species or hybrid clones of *Eucalyptus* (Alfenas et al. 1983, Wingfield 1990). This approach has been shown to be effective in various parts of the world (Campinhos et al. 1983, Wingfield et al. 1990).

Various strategies have been used to select for trees with highest resistance to infection by both *Chr. cubensis* and *Chr. austroafricana*. A study conducted by Van Zyl and Wingfield (1999), assessed the relative susceptibility to *Chr. austroafricana (as C. cubensis)* by using the capacity of *Eucalyptus* clones to close wounds through callus production. Results indicated that tolerant clones close wounds significantly faster than susceptible trees. Another method currently used for resistance screening related to this group of pathogens is by artificial

inoculation (Ferreira et al. 1977, Alfenas et al. 1983, Van der Westhuizen 1992). However, it is also important to undertake disease screening in the areas where the clones will be commercially propagated, due to the genotype x environmental (G x E) interaction observed for disease susceptibility (Van Heerden and Wingfield 2002).

In the longer term, the impact of Chrysoporthe canker could potentially be reduced via biological control though hypovirulence. Hypovirulence is associated with the presence of double-stranded RNA (dsRNA) (van Alfen et al. 1975, Elliston 1985, Nuss 1992, van Heerden et al. 2001). Van Heerden showed that hypovirulence did not lead to a significant reduction in canker size, but did alter the morphology of the canker. Further, morphology of the transfected *Chr. austroafricana (*as *C. cubensis)* isolate is characterized by significantly smaller lesions than those associated with virulent, virus-free isolate (van Heerden et al. 2001).

Another exciting prospect to reduce the impact of Chrysoporthe canker is by using DNA microarray technology. The availability of DNA microarray technology will enhance the rate of tree improvement. Molecular markers tightly linked to resistance genes will allow assessment of the population and evolutionary genetics of resistance to *Chrysoporthe cubensis* and *Chr. austroafricana* on *Eucalyptus*. These markers could be indispensable for map-based approaches to clone resistance genes against Chrysoporthe canker. Furthermore, the availabililty of molecular markers allow for similarity studies between species involved in a tree breeding

programme. A successful hybridization programme, particularly in search of a hybrid species, does require genotyping of sample individuals from a population. Genotyping assists in the discovery of positive traits required from parents.

OVERVIEW OF DNA MICROARRAY-BASED MARKER ANALYSIS

TECHNICAL FOUNDATIONS

The field of array technology has evolved from Southern's key insight that hybridisation between nucleic acids (one of which is immobilized on a matrix) provides a core capability of molecular biology (Southern 1975, Kafatos et al. 1979, Southern et al. 1992). Southern blotting was only a small step towards filter-based screening of clone libraries, which introduced the employment of a pure, single, labelled oligonucleotide or polynucleotide species in the liquid phase and complex mixtures of polynucleotides attached to a solid support (Giege et al. 1998, Lockhart et al. 2000).

More relevant to microarrays is the method using gridded libraries, stored in microtitre plates and stamped onto filters in fixed positions (dot blotting). Each clone can be uniquely identified and information about it, accumulated (Southern 1975). The next advance was the automation and reduction in size of the dots on dot blots. This resulted in the large-scale exploitation of data emerging from hybridisation experiments (Lennon and Lehrach 1991, Southern et al. 1992). Several scientists explored expression analysis by hybridising mRNA to cDNA

libraries gridded on nylon filters (Jaakola et al. 2001, Voiblet et al. 2001). Although the ideas for such arrays were sound, the implementation had yet to be clarified.

Interest in array technologies developed rapidly after Ekins patented miniaturised assays for protein and DNA-RNA measurements in the mid 1980s (Ekins 1987, Ekins et al. 1989). Stanford University was the first institution to use DNA microarrays (Shena et al. 1995). Since then, application of the technology has grown rapidly. The importance is illustrated by ambitious genomic programmes using microarrays (Baldwin et al. 1999, Richmond et al. 2000) and by the establishment of core microarray facilities in the United States (http://www.rana.stanford.edu, http://www.nhgri.nih.gov, http://www.tigr.org), Australia (http://www.cambia.org, http://www.agrf.org), Germany (http://www.mips.bichem.mpg.de) and Switzerland (http://www.unil.ch). Lander et al. (1999) stated that "it is safe to predict that, not long after the turn of the century, researchers will be able to purchase standardised oligonucleotides and cDNA arrays containing the complete sets of 100 000 human and mouse genes" (Reymond et al. 2001).

MICROARRAY TECHNOLOGY

The key principle behind microarray technology is a hybridisation event. One binding partner, the target, is robotically deposited at indexed locations on a suitable surface such as a glass microscope slide (Lemieux 1998, Worley et al.

2000, Reymond et al. 2001). Targets can be of several types, including proteins (Zhu and Snyder 2001, MacBeath 2002, Sydor and Nock 2003), carbohydrates (Mellet et al. 2003) and tissues (Bubendorf et al. 2001, Hedvat et al. 2002). However, this review focuses only on two general methods for producing arrays: DNA-fragment based and oligonucleotide-based chips (Schena et al. 1995, Schena et al. 1996, Lockhart et al. 1996, Duggan et al. 1999). Detailed descriptions of both techniques can be found in a number of reviews, covering technical and comparative aspects of the technologies (Lipschutz et al. 1999, Lockhart et al. 2000, van Hal et al. 2000).

DNA fragment-based microarrays are generally applicable to non-model organisms, as they require only that a large library of DNA is available as a source of clones to be arrayed (Gibson 2002). Clones to be arrayed can originate from a variety of sources including anonymous genomic DNA or cDNA clones, EST clones, or DNA amplified from open reading frames (ORFs) found in sequenced genomes (Schena et al. 1995). Currently, it is feasible to array up to 10 000 DNA fragments/ 3.24cm^2 on the slide (http://cmgm.stanford.edu/pbrown).

Once produced, the microarrays are hybridised with fluorescently labelled probes (Marshall and Hodgson 1998, Ekins and Chu 1999, Aharoni and Vorst 2001). The fluorescently labelled samples allow the detection of individual species at a threshold of one part in 100 000 (Ruan et al. 1998) to 1 part in 500 000 (Schena et al. 1996). Since confocal laser scanning microscopy is an extremely sensitive method for fluorescence measurement at high resolution (Cheung et al. 1999,

Southern et al. 1999, Worley et al. 2000), both weak and strong signals can be monitored. The technology involved in the fabrication and analysis of DNA arrays has recently been extensively reviewed (Brazma and Vilo 2000, Haab et al. 2001, Planet et al. 2001, Foster and Huber 2001).

The alternate oligonucleotide technology, pioneered by Affymetrix GeneChips® (Pease et al. 1994, Lockhart et al. 1996, Lipschutz et al. 1999, Brown and Botstein 1999), uses light-directed synthesis for the construction of high-density DNA probe arrays (Lockhart et al. 1996, Bowtell 1999). Affymetrix GeneChips are short oligonucletide arrays and consist of up to 25-mer oligonucleotides per gene. The 25-mer oligonucleotides utilise perfect match and mismatch pairs that hybridise specifically or non-specifically (Gibson 2002, Hoheisel 1997). Arrays are hybridised to a single biotinylated amplified sample, and the intensity measure for each gene is computed by an algorithm that calculates the difference between the match and mismatch measurements and averages over each oligonucleotide, rather than comparing ratios (Gibson 2002). This technology is expensive, but has wide application for model organisms such as yeast and *Arabidopsis*.

MICROARRAY MARKERS FOR GENETIC ANALYSIS OF FOREST TREES

MICROARRAY MARKERS FOR DNA FINGERPRINTING

An often overlooked aspect of microarrays is that the sequence or even the origin of the probes does not need to be known to make interesting observations (Cheung et al. 1999). The complex profiles, consisting of thousands of individual observations can serve as a fingerprint. Fingerprints can be used for either the identification of diagnostic markers suitable for developing new PCR-based detection assays or for the characterization of population dynamics (genetic distance, similarity, phylogeny) of different genotypes. Experimentally, associating a particular genotype with a phenotype is an important step towards dissection of complex traits such as those associated with disease resistance. Opportunities and challenges of microarray fingerprinting are highlighted in this review.

DNA microarrays as a diagnostic tool

DNA microarrays could potentially be used as tools to address multiple questions linked to species identification (Ye et al. 2001). The use of DNA microarrays for species identification and detection of rifampin resistance in microbes has for example, been described by Troesch et al. (1999). A total of 26 of the 27 species used in the study, as well as all rifampin-resistant mutants, were correctly identified. However, this approach can fail to fully reflect genetic potential of

many organisms. This failure could be overcome by comparative genomic methods with whole genome arrays of unknown sequences (Jaccoud et al. 2001).

For field applications, a portable system for sample preparation and oligonucleotide microarray analysis has been reported for microbial systems (Bavykin et al. 2001). This system has been developed by Nanogen (http://www.nanogen.com/technology). The company is addressing each step in the sample-to-result process on microfabricated chip-based devices (Sosnowski et al. 1997). This includes the integration of electronic cell separation, electronic sample transport, electronically accelerated hybridisation and electronic denaturation (Wang et al. 2000). The disposable LabChip has multiple interconnected reservoirs for samples, sizing ladder, sieving matrix and buffers (van de Goor 2003).

No portable system has yet been developed for DNA microarrays containing genomic regions with no sequence information. However, it is not difficult to envision the future construction of a DNA microarray that will contain unique rRNA, chloroplast DNA or other genomic DNA fragments of immobilised plant probes for identification purposes. These arrays could be useful in many applications. Borevitz et al. (2003) has already demonstrated repeatable and informative allelic variation scanning of the *Arabidopsis* genome. The same principle of the system could be used for the development for portable chips for the detection of resistant trees. However, there remains a lack of comparative data for

cDNA arrays versus long oligonucleotide arrays (Holloway et al. 2002). Care needs to be taken when changing from oligonucleotide arrays to cDNA arrays.

Genetic diversity studies

DNA microarray technology can also be used for genetic diversity studies. Genomic hybridisation of a whole genome array can be used to detect molecular variation between similar DNA regions in plants (Ye et al. 2001). The complex profiles, consisting of thousands of individual observations, can serve as 'fingerprints'. These fingerprints can be used to maintain genetic diversity within breeding populations with respect to disease.

Recently, Diversity Array Technology (DarT) has been used to assay for the presence (or amount) of a specific fragment in a representation derived from the total genomic DNA of rice (Jaccoud et al. 2001). DArT™ is a solid state fingerprinting technique based on AFLP and enables analysis of large numbers of marker loci without any DNA sequence information. Microarray-based genotyping as implemented in the DArT™ technique is a 2-dye approach and relies on the detection of DNA fragments in a complex mixture of selectively amplified restriction fragments. Reduction of complexity by selective amplification allows comparison of polymorphic fragments among genotypes. This is achieved by hybridising DNA to an array containing a large number of DNA fragments, derived from genomic representations of an organism. The study of Jaccoud et al. showed that DNA microarrays make it possible to resolve complex genomic samples into

respective components, offering a method to study genetic diversity at a genome-wide level.

At present, the major limitations of microarray technology as a marker system are cost and the difficulty of achieving both highly sensitive (proportion of polymorphism detected) and highly specific (proportion of true positives) marker identification. The ability of current efforts to satisfy the demand for one but not the other, resides in the difficulty of achieving the same hybridisation pattern for probes (Steinmetz 2000). It should also be noted that studies on genetic diversity based on molecular markers must be interpreted with caution. There are often low correlations with patterns of variation for disease resistance.

Inference of family relationships of mating systems

Inference of family relationships of forest trees has drawn increasing interest recently. This interest is a consequence of the development of hypervariable markers, such as microsatellites (Lexer et al. 1999) and AFLPs (Gerber et al. 1999, Wu et al. 2000). These highly polymorphic markers enable parentage analysis and thus the reconstruction of the mating system and/or gene flow (Dow and Ashley 1996, Ehm and Wagner 1996, Streiff et al. 1999). Furthermore, given additional genotyped family members, incorrect specification of relationship may be detected on the basis of apparent incompatibilities with Mendelian inheritance (Boehnke et al. 1997, Lexer et al. 1999). Within a family, unintentionally duplicated samples,

half sibs or unrelated pairs can be identified, if a high number of loci are considered (Goring et al. 1995, Streiff et al. 1999, Ziegenhagen et al. 1999).

To date, only microsatellite markers and AFLPs have successfully been adapted for segregation analysis of polymorphic bands (Wu et al. 2000). Since co-dominant (microsatellites) and dominant markers (AFLPs) can be used for the reconstruction of family relationships within forest tree populations, microarrays may be considered as an alternative. Microarrays provide a method to quickly genotype progeny at high resolution at the whole genome level (Borevitz et al. 2003). Parallel genotyping of a large number of markers permit linkage analysis. This, in turn, will be sufficient to allow accurate inference of relationships (Boehnke and Cox 1997).

Phylogenetic reconstruction

Fingerprinting data can be used for phylogenetic studies, if the genome studied fits the assumptions made by all methods of inference. This is, that organisms evolve mainly by drift and mutation under a bifurcating tree assumption (Swofford and Olsen 1991, Rzhetsky and Nei 1993, Allman and Rhodes 2004). Cho and Tiedje (2001) confirmed that this also applies to microarray data when using oligonucleotide arrays. These authors showed that cluster analysis of the hybridisation profiles revealed taxonomic relationships between *Pseudomonas* strains tested at species and strain level resolution. The results suggested that this approach is useful for the identification of bacteria as well as for determining the

phylogenetic relationships between the isolates. The overall topology of the dendrogram based on cluster analysis of similarity coefficient matrix was consistent with the phylogenetic tree obtained from 16S rDNA sequence data (Moore et al. 1996).

Oligonucleotide microarrays have been used to study phylogenetic relationships mainly in microbes and pathogens. However, Nouzava et al. (2001) used the DNA microarray technology to survey repetitive sequences. These authors devised a modified DNA-microarray-based technique to screen short fragment genomic libraries. The preliminary study showed that some of newly isolated repetitive sequences of *Vicia* species could be used as supportive molecular markers for phylogenetic reconstruction of this plant. A similar method can be used for microarray analysis of forest trees.

IDENTIFICATION OF MICROARRAYS LINKED TO SPECIFIC GENES OF INTEREST USING BSA

Since no near-isogenic lines (NILs) exist for forest trees, Bulked Segregant Analysis (BSA) provides an alternative method for identifying markers linked to resistance in forest trees (Michelmore et al. 1991, Weeden et al. 1994, Wilcox et al. 1996). The underlying principle of BSA is the grouping of informative individuals, so that a particular genomic region is studied against a randomised genetic background of unlinked loci. Informative individuals are representatives of

phenotypic extremes for a segregating trait and are bulked in two separate pools (Giovannoni et al. 1991, Lynch et al. 1997, Hill 1998). It is assumed that the individuals at the ends of the phenotypic distribution have opposing homozygous alleles for the trait in question. All the other loci forming the randomised genetic background are assumed to be in a heterozygous state due to the random contribution of these unselected alleles (Michelmore et al. 1991). Therefore, the polymorphisms identified are closely linked to the target locus (Asims 2002).

BSA can also be used to target multiple loci of highly heritable traits (Michelmore et al. 1991). This is especially important for disease resistance breeding in forest trees, because genetic variation for host resistance in endemic pathosystems typically appears continuous (Wilcox et al. 1996). With BSA, it is possible to determine the number of genes involved in resistance of a tree, or whether all trees harbour an inordinate number of susceptibility alleles (Mackay et al. 2000). At present, only a few discrete resistance genes are known in forest trees. Inheritance of disease resistance in forest trees has been commonly explained by polygenic models, where resistance is controlled by many genes, each with a small additive effect (Kinloch et al. 1970, Robinson 1987; von Weisenberg 1990, Kinloch and Walkinshaw 1991, Wilcox et al. 1996, Van Heerden et al. 2002).

BSA and three molecular techniques, namely restriction fragment length polymorphisms (RFLPs), amplified fragment length polymorphisms (AFLPs) and random amplified polymorphic DNA (RAPDs), have most commonly been used to

develop DNA-based molecular markers (Groover et al. 1994, Grattapaglia and Sederoff 1994, Bradshaw and Stettler 1995, Grattapaglia et al. 1995). However, the molecular techniques are constrained by their reliance on gel electrophoresis. This hampers the processing of a large number of samples in parallel (Smith and Beavis 1996). In contrast, the microarray technology holds the promise of parallel analysis of thousands of markers at the whole genome level. This will improve the candidate gene selection process. Further, microarrays may also be useful for BSA as the two bulks can be labelled with different dyes and incubated in a single microarray reaction. This would simplify comparisons considerably.

GENE MAPS AND QTL DETECTION

Genetic linkage maps can be used to locate genes affecting quantitative traits (Wayne and Mc Intyre 2002, Borevitz et al. 2003). Quantitative traits are usually controlled by many genes, termed quantitative trait loci (QTL). By using molecular markers linked to one or more QTL, information at the DNA level can be used for early selection. Molecular marker maps have been constructed for the major commercial genera of forest trees, such as eucalypts, pines and acacias. Updated information can be obtained at http://dendrome.uc.davis.edu/index.html (Groover et al. 1994, Bradshaw and Stettler 1995, Grattapaglia 1996).

Until recently, only microarray expression profile studies had been used to map genes (Winzeler et al. 1998, Cho et al. 1999, Brem et al. 2002, Wenzl et al. 2004).

However, high-density arrays can also be powerful new tools for parallel genotyping of a large number of markers of unknown sequence (Steinmetz et al. 2002). Informative spots obtained in this way can function as forensic fingerprints to assist selection in forest tree breeding (Samuel et al. 2003). The ability to use arrays for detection of variation lies in the specificity of DNA hybridisation, which allows the detection of polymorphisms within complex genomes. Polymorphisms are detected as DNA segments, which amplify from one parent but not the others, and can be used to construct genetic maps (Grattapaglia 1996).

Microarray technology, together with the large amount of sequence information, will allow marker-trait association studies in economically important forest tree species. Expressed gene sequences (ESTs), or full-genome sequences can be used to design probes for single nucleotide polymorphisms, or other types of polymorphisms. Such polymorphisms detected in segregating progeny can then be used to construct genetic maps. Borevitz et al. (2003) already demonstrated the use of microarrays to demarcate recombination events along chromosomes of recombinant inbred lines of *Arabidopsis*. Since chromosomal regions that have a measurable effect on economically important traits can be identified, microarrays hold the potential for the genome-wide selection of superior plant materials.

Although significant progress in marker genotyping with oligonucleotide arrays has been made in recent years, genotyping with microarrays is still in its infancy (Hardenbol et al. 2003, Borevitz et al. 2003, Wenzl et al. 2004). A great deal of

information must still be developed on the effective incorporation of markers. It is essential that clonally replicated experiments in several locations and genetic backgrounds are established (Schuchhardt et al. 2000). It is also important to understand that if a particular QTL allele is not detected in all genetic backgrounds, this does not imply that the locus is not present (Yanchuk 2002). Other better alleles at that locus occur in the population, which can be used for subsequent analysis (Byrne et al. 1997).

DEVELOPMENT OF SCARS

Sequence characterised amplified regions (SCARs) are also known as allele specific associated primers (ASAPs), or sequence tagged sites (STSs). A SCAR is a genomic DNA fragment at a single genetically defined locus that is identified by PCR amplification (Paran and Michelmore 1993). These markers are designed by cloning and subsequent sequencing a polymorphic fragment of interest. The unique nature of the marker system is determined by the sequence and spacing of the primer sequences, rather than by hybridisation. This makes the marker system ideal for routine screening of forest trees (Demeke and Adams 1994, Barreneche et al. 1998).

Current methods of SCAR-based marker discovery typically involve the detection of sequence variation using a marker technology. RAPD and AFLP markers are most commomly used to develop SCARs (Paran and Michelmore 1993; Adam-

Blondon et al. 1994) as they obviate the need for specific probe development, and are rapidly applied and technically simple. Although both marker systems have many advantages for SCAR development, they are restricted somewhat by their dominant nature, the sensitivity of the amplifications to precise reaction conditions and their dependence on gel electrophoresis. The necessity to run PCR products on a gel hampers the ability to process a large number of samples. This may be overcome by hybridisation-based microarrays, which would allow for the discrimination of allelic variation at several thousand loci in a single assay.

SCARs have successfully been used to obtain markers linked to brown spot needle blight resistance in longleaf pine (Nelson et al. 1994), and to the eastern filbert blight resistance genes in *Corylus avellana* (Mehlenbacher et al. 2004). These and other studies have shown that SCARs are reproducible, show low variability among different thermocyclers and when different DNA polymerises are used (Melotto et al. 1996, Schachermeyr et al. 1995). Therefore SCARs allow a level of reproducibility between laboratories that is independent of the method employed in the physical mapping. However, SCARs have no yet been developed from microarray markers.

MAP-BASED CLONING OF RESISTANCE GENES

The use of molecular markers has enabled geneticists to accelerate breeding programs (Shepherd et al. 1995, Xu et al. 1996, Yang et al. 1997). Using molecular

markers, tightly linked markers for particular genes of interest can be identified and cloned using map-based cloning procedures (Young 1990). Gene-flanking regions that contain at least one of two flanking markers are cloned and used to construct contigs spanning the genomic region defined by the molecular markers. By tagging economically important genes it is possible not only to select for presence of those genes in individuals, but also to isolate and introgress genes from one line to another.

Breeders expect marker-assisted selection to be useful, particularly in situations where trait heritability is low, typically in selection at the individual tree level (Grattapaglia 1996). With high heritability, the phenotype becomes a more efficient predictor of a genotype (Bradshaw and Grattapaglia 1994). However, an absolute requirement for the selection of traits based on map-based cloning of a gene is that markers must be closely associated with the gene (Winter and Kahl 1995, Feuillet et al. 1995).

Although map-based cloning has been successful in many species, the question that remains is whether there are real opportunities for incorporating microarrays as a genetic marker system in tree breeding programs. Microarrays might not be suitable for programs with small budgets and where significant gains can be achieved by implementing other marker systems or conventional selection procedures coupled to map-based cloning. The effective incorporation of

microarrays in breeding programems will require significant experimental work before they can be used as a marker system.

DATA ANALYSIS AND MANAGEMENT

During the course of the next few years, microarray technology is likely to become an integral part of molecular biology laboratories around the world (Kazan et al. 2001). Along with the rapid development of microarray technology, more data will be generated (Cole et al. 1998, Schena et al. 1998). The sensible storage, analysis, interpretation and management of this data will be a major challenge (Debouck and Goodfellow 1999, Wilson et al. 2002). Efforts to construct public repository databases for the management of information on the primary results of hybridisation and the construction of algorithms making it possible to examine the outputs from single and multiple experiments are clearly needed (Chen et al. 1997; Douglas et al. 1999).

Another requirement for microarray analyses is effective software for efficient handling of hybridisation-based DNA fingerprints or to extract binary scores from hybridisation data. The majority of available microarray software has been designed for expression profiling studies. For fingerprinting applications, the presence or absence of fragments (dominant scoring) or signal intensity (for co-dominant scoring) has to be determined to construct a fingerprint. Furthermore,

quality values must be assigned to each data point to evaluate the reliability of the combined fingerprint. Scoring markers remains tedious, and in many cases current programs fail to identify all fingerprints (Bassett et al. 1999).

At present, there is no standard solution for microarray data analysis and storage software. Successful interpretation of data will rely on integrating experimental data with external information resources, such as those encompassed by NCBI's Entrez system (Mc Entyre 1998, Schuler et al., 1996) and software available in public domains. Just as the nucleic acid and protein sequence databases depend on input from many groups, public microarray databases will similarly stimulate a level of analysis that is not possible with narrowly defined data sets. Patenting and trade marking should not be allowed to affect technology development, distribution and access (Duyk 2002). Co-operation seems to be a more likely option and has solved similar problems in the microelectronics industry before (Bassett et al. 1999).

CONCLUSIONS AND FUTURE DIRECTIONS

DNA microarray technology has a great number of potential applications that could be used in the improvement of forest trees. Geneticists have predominantly improved forest trees based on phenotypic characteristics. Due to the substantial time between seed germination and sexual maturity, trees have not benefited greatly from traditional breeding approaches. The availability of automated, highly

efficient, fast and productive technologies must clearly enhance the opportunities for, and the rate of tree improvement. This can be achieved by using DNA microarray technology. DNA microarrays allow genome-wide and high-throughput analysis for the identification of trait-linked markers available to the plant breeders in future, and therefore improve the efficiency of breeding programs (Borevitz et al. 2003). This will not only make loci and gene-specific markers available to the plant breeders, but will also increase our understanding of tree genomes for breeding purposes.

Despite remarkable progress, DNA microarrays are still in their early stages of development and this presents many challenges. One of the most challenging aspects relating to microarray technology in the next decade will deal with the flow of data generated (Bassett et al. 1999). Efforts to normalise and analyse data, as well as to centralise the information in public repository databases, are clearly needed. In addition, more effective software tools will be needed for the efficient handling of hybridisation-based DNA fingerprints or to extract binary scores from hybridisation data. The large amounts of data generated in microarray analyses often go beyond the research interests of the individual scientists that have produced them. Therefore, data should be shared by all researchers to ensure maximum impact of research. In future, researchers will most likely be required to submit their microarray data to a public depository before publication.

Although there are difficulties associated with data analysis and management of microarray results, useful applications are already emerging and others will undoubtedly follow. Innovative efforts, coupling fundamental biological and chemical sciences with technological advances in the field of micro-fabrication should lead to even more powerful devices that will accelerate the realisation of large-scale genetic testing (Wang, 2000). In the short term, we expect that DNA microarray technology will continue to be evaluated as a marker technology. DNA microarray technology and structural genomic tools, such as sequencing, will help to improve trees. Sequencing of plant genomes will allow for the saturation of existing forest tree maps. This will make loci and gene specific markers available to plant breeders in future.

In the long term, we expect that the ability to exploit and manipulate the genomes of forest trees will be possible. This is because of advances in the output of large-scale sequencing efforts and achievable advances in array technology (Lockhart et al., 2000). The combined use of microarray technology, sequence information, computational tools, integrated knowledge databases, biology, chemistry, physics, mathematics and genetics will increase the understanding of the mechanisms of disease. Locating, identifying and cataloguing genotypic differences will be the first steps in relating genetic variation to phenotypic variation in both normal and diseased states. However, the full potential of this technology as a marker system will depend on the number of laboratories that will use and evaluate the potential of microarray technology in theoretical and applied forest tree breeding. For

microarray technology to meet its full potential, it will need to become an integral part of the daily activities of the molecular biology laboratories (Mc Dowell et al. 2003).

The potential applications of DNA microarrays as a method for genetic marker analysis in forest tree breeding has been described in this review. It is likely that DNA marker technology will evolve rapidly in the coming years. Instruments for preparing microarrays and more effective software programs for analysing data from microarray experiments are continuously being improved. We thus expect that DNA microarray analysis as a method for genetic marker analysis will greatly assist forestry breeding programs.

REFERENCES

Adam-Blondon AF, Sevignac M, Dron M, Bannerot H (1994) A genetic map of common bean to localize specific resistance genes against anthracnose. Genome 37: 915 – 924.

Alfenas AC, Jeng R, Hubbes M (1983) Virulence of *Cryphonectria cubensis* on *Eucalyptus* species differing in resistance. Eur J Forest Path 13: 179 – 205.

Aharoni A, Vorst O (2001) DNA microarrays for functional plant genomics. Plant Mol Biol 48: 99 – 118.

Allman ES, Rhodes JA (2004) Phylogenetic invariants for stationary base composition. J Symbolic Computation, *in press*.

Ashari S, Aspinall D, Sedgeley M (1989) Identification and investigation of relationships of mandarin types using isozyme analysis. Scientia Hort 40: 305 – 315.

Asims MJ (2002) Present and future of quantitative trait locus analysis in plant breeding. Plant Breeding 121: 281 – 291.

Baldwin D, Crane V, Rice D (1999) A comparison of gel-based, nylon filter and microarray techniques to detect differential RNA expression in plants. Curr Opin Plant Biol 2: 96 – 103.

Barreneche T, Bodenes C, Lexer C, Trontin J F, Fluch S, Streiff R , Plomion C, Roussel G, Steinkellner H, Burg K, Favre JM, Glössl J and Kremer. A. (1998) A genetic linkage map of *Quercus robur* L. (pedunculate oak) based on RAPD, SCAR, microsatellite, minisatellite, isozyme and rDNA markers. Theor Appl Genet 97:1090-1103.

Bartels H (1971) Genetic control of multiple esterases from needles and macrogametophytes of *Picea abies*. Planta (Berl) 99: 283 – 289.

Bassett DE, Eisen MB, Boguski MS (1999) Gene expression information – it's all in your mine. Nature Genet. Supplement 21: 51 – 55.

Bavykin SG Akowski JP, Zakhariev VM, Barsky VE, Perov AN, Mirzabekov AD (2001) Portable system for microbial sample preparation and oligonucleotide microarray analysis. Appl Environ Microbiol 67: 922 – 928.

Bergmann F (1971) Genetische Untersuchungen bei *Picea abies* mit Hilfe der Isoenzym-Identifizierung. In:Moeglichkeiten fuer genetische Zertifizierung von Forstsaatgut. Allgemeine Forst-und Jagdzeitung 142: 278 – 280.

Bernatzky R, Tanksley SD (1986) Toward a saturated linkage map in tomato based on isozyme and random cDNA sequences. Genetics 112: 887 – 898.

Blears MJ, de Grandis SA, Lee H, Trevors JT (1998) Amplified fragment length polymorphism (AFLP): a review of the procedure and its applications. J I ndust Microbiol and Biotechnol 21: 99 – 114.

Boehnke M, Cox NJ (1997) Accurate inference of relationships in sib-pair linkage studies. Am J Hum Genet 61: 423 – 429.

Boerboom JHA and Maas PWT (1970) Canker of *Eucalyptus grandis* and *Eucalyptus saligna* in Surinam caused by *Endothia havanensis*. Turrialba 20: 94 – 99.

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

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

Bowtell DDL (1999) Options available – from start to finish – for obtaining expression data by microarray. Nature Genet Supplement 21: 25 – 32.

Bradshaw HD, Grattapaglia D (1994) QTL mapping in interspecific hybrids of forest trees. Forest Genetics 1: 191 – 196.

Bradshaw HD, Stettler RF (1995) Molecular genetics of growth and development in *Populus*, IV: Mapping QTLs with large effects on growth, form and phenology traits in a forest tree. Genetics 139: 963 – 973.

Brazma A, Vilo J (2000). Gene expression data analysis. FEBS Lett. 480: 17 – 24.

Brem RB, Yvert G, Clinton R, Kruglyak I (2002) Genetic detection of transcriptional budding yeast. Science 296: 752 – 755.

Brown PO, Botstein D (1999) Exploring the new world of the genome with DNA microarrays. Nature Genet Supplement 21: 33 – 37.

Bruner RF (1916) A new species of *Endothia*. Mycologia 8: 239 – 242.

Bubendorf L, Nocito A, Moch H, Sauter G (2001) Tissue microarray (TMA) technology: miniaturized pathology archives for high-throughput *in situ* studies. J Pathol 195: 72 – 79.

Byrne M, Murell JC, Owen JV, Kriedemann P, WilLiams ER, Moran GF (1997) Identification and mode of action of quantitative trait loci affecting seedling height and leaf area in *Eucalyptus nitens*. Theor Appl Genet 94: 647 – 681.

Campinhos E, Ikemori YK (1983) Mass production of *Eucalyptus* spp. by root cuttings. Silvicultura 8: 770 – 775.

Chen Y, Dougherty ER, Bittner ML (1997) Ratio-based decisions and the quantitative analysis of cDNA microarray images. J Biomed Optics 2: 364 – 374.

Cheung VG, Morley M, Aguilar F, Massimi A, Kucherlapati R, Childs G (1999) Making and reading microarrays. Nature Genet 21: 15 – 19.

Cho RJ, Mindrinos M, Richards DR, Sapolsky RJ, Anderson M, Drenkard E, Drewdney J, Reuber TL, stammers M, Federspiel N, et al. (1999) Genome-wide mapping with biallelic markers in *Aradopsis thaliana*. Nature Genet 23:203 – 207.

Cho JC and Tiedje JM (2001) Bacterial species determination from DNA-DNA hybridisation by using genome fragments and DNA microarrays. Appl Environ Microbiol 67: 3677 – 3682.

Cole KA, Krizman DB, Emmort-Buck MR (1999) The genetics of cancer – a 3D model. Nature Genet 21: 38 – 41.

Conradie E, Swart WJ, Wingfield MJ (1990) *Cryphonectria* canker of *Eucalyptus*, an important disease in plantation forestry in South Africa. SA Forestry J 152: 43 – 49.

Debouck C, Goodfellow B (1999) DNA microarrays in drug discovery and development. Nature Genet Supplement 21: 48 – 50.

Demeke T and Adams RP (1994) The use of PCR-RAPD analysis in plant taxonomy and evolution. In: PCR Technology: Current Innovations (eds) Griffin HD and Griffin AM, chapter 21. CRC Press, Boca Raton, Florida.

Department of Environmental Affairs (1987) Directorate of National Forestry Planning. Commercial timber resources and roundwood-processing in South Africa 1985/1986.

Douglas EB Jr, Eisen MB, Boguski M (1999) Gene expression informatics – it's all in your mine. Nature Genet Supplement 21: 51 – 55.

Dow BD, Ashley MV (1996) Clusteranalyse und Diskriminanzanalyse. Gustav Fischer Verlag, Stuttgart.

Dow BD, Ashley MV (1996) Microsatellite analysis of seed dispersal and parentage of saplings in bur oak *Quercus macrocarpa*. Mol Ecol 5: 615 – 627.

Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM (1999) Expression profiling using cDNA microarrays. Nature Genet Supplement 21: 10 – 14.

Duyk GM (2002) Sharper tools and simpler methods. Nature Genet Supplement 32: 465 – 468.

Ehm MG, Wagner M (1996) Test statistic to detect errors in sib-pair relationships. Am J Hum Genet Supplement 59: A217.

Ekins RP (1987) US Patent Application 8803000.

Ekins RP, Chu F and Micallef J (1989) High specific activity chemiluminescent and fluorescent markers: their potential application to high sensitivity and 'multianalyte' immunoassays. J Biolum Chemilum 4: 59-78.

Ekins R, Chu FW (1999) Microarrays: their origins and applications. Trends Biotechnol 17: 217 –218.

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

Elliston JE (1985) Further evidence for two cytoplasmic hypovirulence agents in a strain of *Endothia parasitica* from western Michigan. Phytopathology 75:1405 - 1413.

Evans J (1992) Plantation forestry in the tropics. Oxford, UK, Clarendon Press, pp 432.

Feuillet C, Messmer M, Schachermeyer G, Keller B (1995) Genetic and physical characterisation of the Lr1 leaf rust resistance locus in wheat (*Triticum aestivum* L.). Mol Genet 248: 553 – 562.

Ferreira FA, Reis MS, Alfenas AC, Hodges CS (1977) Avaliacao da resistencia de *Eucalyptus* spp. ao cancro causado por *Diaporthe cubensis* Bruner. Fitopatol Brasil 2: 225 – 241.

Florence EJM, Sharma JK, Mohanan C (1986) Stem canker disease of *Eucalyptus* caused by *Cryphonectria cubensis* in Kerala. Kerala Forest Research Institute Scientific Paper 66: 384 –387.

Foster WR, Huber RM (2001) Current themes in microarray experimental design and analysis. DDT 7: 290 – 292.

Gerber S, Mariette S, Streiff R, Bodenes C, Kremer A (1999) Comparison of microsatellites and AFLP markers for parentage analysis. In: Gillet EM (ed) Which DNA marker for which purpose? Final Compendium of the Research Project 'Development, optimisation and validation of molecular tools for assessment of biodiversity in forest trees' in the European Union DGXII Biotechnology FW IV Research Programme 'Molecular Tools for Biodiversity'.

Gibson G (2002) Microarrays in ecology and evolution: a preview. Mol Ecol 11: 17 – 24.

Gibson IAS (1981) A canker disease of *Eucalyptus* new to Africa. FAO, Forest Genet Res Info 10: 23 – 24.

Giege P, Konthur Z, Walter G, Brennicke A (1998) An ordered *Arabidopsis thaliana* mitochondrial cDNA library on high-density filters allows rapid systematic analysis of plant gene expression: a pilot study. Plant J 15: 721 – 726.

Giovannoni JJ, Wing RA, Ganal MD, Tanksley SD (1991) Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. Nucleic Acids Research 19: 6533 – 6558.

Goldstein DB, and Schlötterer C (1999) Microsatellites: Evolution and Applications. Oxford University Press, New York.

Goring HHH, Ott J (1995) Verification of sib relationship without knowledge of parental genotypes. Am J Hum Genet Supplement 57: A192.

Grattapaglia D, Sederoff RR (1994) Genetic Linkage Maps of *Eucalyptus grandis* and *Eucalyptus urophylla* Using a Pseudo-Testcross: Mapping Strategy and RAPD Markers. Genetics 137: 1121 – 1137.

Grattapaglia D, Bertolucci FL, Sederoff RR (1995) Genetic mapping of quantitative trait loci controlling vegetative propagation in *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross mapping strategy and RAPD markers. Theor Appl Genet 90: 933 – 947.

Grattapaglia D, Bertolucci FL, Penchel R, Sederoff RR (1996) Genetic mapping of quantitative trait loci controlling growth and wood quality traits in *Eucalyptus grandis* using a maternal half-sib family and RAPD markers. Genetics 138: 1205 – 1214.

Groover A, Devey M, Fiddler T, Lee J, Megraw T, Mitchell-Olds T, Sherman B, Vujcic C, Williams C, Neale D (1994) Identification of quantitative trait loci

influencing wood specific gravity in an outbred pedigree of loblolly pine. Genetics 138: 1293 – 1300.

Gryzenhout M, Myburg H, van der Merwe NA, Wingfield BD, Wingfield MJ (2004) *Chrysoporthe*, a new genus to accommodate *Cryphonectria cubensis*. Studies in Mycol 50: 119-142.

Haab BB, Dunham MJ, Brown PO (2001) Protein microarrays for highly parallel detection and quantification of specific proteins and antibodies in complex solutions. Genome Biol 2: 4.1 – 4.13.

Hardenbol P, Baner J, Jain M, Nilsson M, Namsaraev EA, Karlin-Neumann GA, Fakhrai-Rad H, Ronaghi M, Willis TD, Landegren U, Davis RW (2003) Multiplexed genotyping with sequence-tagged inversion probes. Nature Biotech 21: 673 – 678.

Harkins DM, Johnson GM, Skaggs PA, Mix AD, Dupper GE, Devey ME, Kinlock BB Jr., Neale DB (1998) Saturation mapping of a major gene for resistance to white pine blister rust in sugar pine. Theor Appl Genet 97: 1355 – 1360.

Hedvat CV, Hedge a, Chaganti RS, Chen B, Qin J, Filippe DA, Nimer SD, Teruya-Feldstein J (2002) Application of tissue array technology for the study of non-Hodgkin's and Hodgkin's lymphoma. Hum Pathol 33: 968 – 974.

Hill WG (1998) A note on the theory of artificial selection in finite populations and application to QTL detection by bulk segregant analysis. Genet Research (Cambridge) 72: 55 – 58.

Hodges CS and Reis MS (1974) Identificacae do funfo causado de cancro de *Eucalyptus* spp. no Brasil. Brasil Forest 5:19.

Hodges CS, Geary TF, Cordell CE (1979) The occurrence of *Diaporthe cubensis* on *Eucalyptus* in Florida, Hawaii and Puerto Rico. Plant Disease Reporter 63: 216 – 220.

Hodges CS (1980) The taxonomy of *Diaporthe cubensis*. Mycologia 72: 542 – 548.

Hoheisel JD (1997) Oligomer-chip technology. Trends Biotechnol 15: 465 – 469.

Holloway AJ, van der Laar RK, Tothill RW, Bowtell DDL (2002) Options available – from start to finish – for obtaining data from DNA microarrays. Nature Genet Supplement 32: 481 – 489.

Huang H, Dane F, Norton JD (1994) Genetic analysis of 11 polymorphic isozyme loci in chestnut species and characterization of chestnut cultivars by multi-locus allozyme genotypes. J Am Soc Hort Sci 199: 840 – 849.

Jaakola L, Pirttila AM, Hohtola A (2001) cDNA Blotting Offers an Alternative Method for Gene Expression Studies. Plant Mol Biol 19: 125–128.

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

Jarne P, Lagoda PJL (1996) Microsatellites, from molecules to populations and back. TREE 11: 424- 429.

Kafatos FC, Jenes CW, Efstratiadis A (1979) Determination of nucleic acid sequence homologies and relative concentrations by a dot blot hybridisation procedure. Nucleic Acids Res 24: 1541 – 1552.

Kazan K, Schenk PM, Wilson I, Manners JM (2001) DNA microarrays: new tools in the analysis of plant defence responses. Mol Plant Pathol 2: 177 – 185.

Kinloch BB, Parks GK, Fowler CW (1970) White pine blister rust: simply inherited resistance in sugar pine. Science 167: 193 – 195.

Kinloch BB, Walkinson CH (1991) Resistance to fusiform rust in southern pines: How is it inherited? In: Proceedings of the International Union of Forestry Research Oragnizations Rusts of Pine Working Party Conference. Natl Tech Info *Service*, Springfield VA. Publication #44 of the Southern Forest Tree Improv. Comm., pp. 403 – 411.

Lander ES (1999) Array of hope. Nature Genet 21: 3 – 4.

Lennon GG, Lehrach H (1991) Hybridization analyses of arrayed cDNA libraries. Trends Genet 7: 314 – 317.

Lemieux B, Aharoni A, Schena M (1998) Overview of DNA chips technology. Mol Breeding 4: 227 – 289.

Lexer C, Heinze B, Gerber S, Steinkellner H, Ziegenhagen B, Kremer A, Gloessl J (1999) In: Gillet EM (ed). Which DNA Marker for which purpose? Final Compendium of the Research Project 'Development, optimisation and validation of molecular tools for assessment of biodiversity in forest trees' in the European Union DGXII Biotechnology FW IV Research Programe 'Molecular Tools for Biodiversity'. Chapter 6.

Lipschutz RJ, Fodor SP, Gingeras TR, Lockhart DJ (1999) High density synthetic oligonucleotide arrays. Nature Genet 21: 20 – 24.

Lockhart, DJ, Dong, H., Byrne, MC, Follettie, MT, Gallo, MV, Chee, MS, Mittmann M., Want, C., Kobayashi, M., Horton, H. & Brown, EL (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. Nature Biotech 14: 1675 – 1680.

Lockhart DJ, Winzeler EA (2000) Genomics, gene expression and DNA arrays. Nature 405: 827 – 836.

Lynch M, Walsh B (1997) Genetics and analysis of quantitative traits. Sinauer Associates, Inc. MA.

Mackay IJ, Caligari PDS (2000) Efficiencies of F_2 and Backcross Generations for Bulked Segregant Analysis using dominant markers. Crop Sci 40: 626 – 630.

Mackill DJ, Salam MA, Wang ZY, Tanksley SD (1993) A major photoperiod-sensitivity gene tagged with RFLP and isozyme markers in rice. Theor Appl Genet 85: 536 – 540.

Markert CL and Moller F (1959) Multiple forms of enzymes: Tissue, ontogenetic, and species-specific patterns. Proc Nat Acad Sci USA 5: 753 – 763.

Marshall A, Hodgson J (1998) DNA chips: an array of possibilities. Nature Biotech 16: 27 – 31.

Mc Beath G (2002) Protein microarrays and proteomics. Nature Genet Supplement 32: 526 – 532.

Mc Dowell JM, Woffenden J. (2003) Plant disease resistance genes: recent insight and potential applications. Trends Biotech 21: 178 – 183.

Mc Entyre J (1998) Linking up with Entrez. Trends Genet 14: 39 – 40.

Mehlenbacher SA, Brown RN, Davis JW, Chen H, Bassil NV, Smith DC, Kubisiak TL (2004) RAPD markers linked to eastern filbert blight resistance in *Corylus avellana*. Theor Appl Genet 108: 651 – 656.

Mellet CO, Fernandez JM (2003) Carbohydrate Microarrays. Chem Bio Chem 3: 819 – 822.

Melotto M, Afanador L, Kelly JD (1996) Development of a SCAR marker linked to the *I* gene in common bean. Genome 39: 1216 – 1219.

Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci USA 88: 9828 – 9832.

Michelmore RW (1995) Isolation of disease resistance genes from crop plants. Curr Opin Biotech 6:145-152.

Moore ERB, Mau M, Arnscheidt A, EC Böttger, Hutson RA, Collins MD, van de Peer Y, de Wachter R, Timmis KN (1996) The determination and comparison of the 16S rRNA gene sequences of species of the genus *Pseudomonas* (sensu stricto) and estimation of the natural intrageneric relationship. Syst Appl Microbiol 19: 478 - 492.

Myburg H, Gryzenhout M, Wingfield BD, Wingfield MJ (2002) β-tubulin and Histone *H3* gene sequences distinguish *Cryphonectria cubensis* from South Africa, Asia and South America. Can J Bot 80: 590 - 596.

Myburg H, Gryzenhout M, Wingfield BD, Wingfield MJ (2003) Conspecificity of *Endothia eugeniae* and *Cryphonectria cubensis*: A re-evaluation based on morphology and DNA sequence data. Mycoscience 104: 187 - 196.

Neale DB, Tauer CG, Gorzo DM, Jermstad KD (1989) Restriction fragment length polymorphism mapping of loblolly pine: methods, applications and limitations. In: Proceedings of the 20th Southern Forest Tree Improvement Conference. Charleston, SSC, pp 363 – 372.

Nelson CD, Kubisiak TL, Stine M, Nance WL (1994) A Genetic Linkage Map of Longleaf Pine (*Pinus pallus*tris Mill.) Based on Random Amplified Polymorphic DNAs. J Heredity 85: 433 – 439.

Niebling CR, Johnson K, Gerhold HD (1987) Electrophoretic analysis of genetic linkage in scots pine (Pinus sylvestris L.). Biochem Genet 25: 803 - 814.

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

Nuss DL (1992) Biological control of chestnut blight: an example of virus-mediated attenuation of fungal pathogenisis. Microbiol Reviews 56: 561 – 576.

Old KM, Murray DIL, Kile GA, Simpson J, Malafant K (1986) The pathology of fungi isolated from eucalypt cankers in south-eastern Australia. Australian Forest Res 16: 21 – 36.

Okubara P, Anderson P, Michelmore RW (1995) Mutation analysis of genes for resistance to downy mildew in lettuce. Genetics 137:867 – 874.

Paran I, Michelmore RW (1993) Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. Theor Appl Genet 85: 985 – 993.

Pease AC, Solas D, Sullivan EJ, Cronin MT, Holmes CP, Fodor SA (1994) Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proc Natl Acad Sci USA 91: 5022 – 5026.

Planet PJ, DeSalle R, Siddall M, Bael T, Neil Sarkar I, Stanley SE (2001) Systematic Analysis of DNA Microarray Data: Ordering and Interpreting Patterns of Gene Expression. Genome Research 11:1149 - 1155.

Rafalski JA, Tingey SV, Williams JGK (1991) RAPD markers – a new technology for genetic mapping and plant breeding. Agric Biotech News and Info 3: 645 – 648.

Reymond P (2001) DNA microarrays and plant defence. Plant Physiol Biochem 39: 313 – 321.

Richmond T, Sommerville S (2000) Chasing the dream: plant EST microarrays. Curr Opin Plant Biol 3: 108 – 116.

Robinson RA (1987) Host management in crop pathosystems. *Mac Millan, New York.*

Ruan Y, Gilmore J, Conner T (1998) Towards *Arabidopsis* genome analysis: monitoring expression profiles of 1400 genes using cDNA microarrays. Plant J 15: 821 – 833.

Rzhetsky A, Nei M. (1993) Theoretical foundation of the minimum-evolution method of phylogenetic inference. Mol Biol Evol. 10: 1073 – 1095.

Samuel P, Kay H, Kay S (2003) Gene arrays are not just for measuring gene expression. Trends in Plant Science 8: 413 – 415.

Schachermeyr G, Messmer MM, Feiullet C, Winzeler H, Winzeler M, Keller B (1995) Identification of molecular markers linked to the *Agropyron elongatum*-based leaf rust resistance gene *Lr 24* in wheat. Theor Appl Genet 90: 982 – 990.

Sharma JK, Mahanan C, Florence EJM (1985). Occurrence of Cryphonectria canker disease of *Eucalyptus* in Keralda, India. Annals of Appl Biol 106: 265 – 276.

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

Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW (1996). Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. Proc Natl Acad Sci USA 93: 10614 – 10619.

Schena M, Heller RA, Theriault TP, Konrad K, Lachenmeier E, Davis RW (1998) Microarrays: biotechnologys' discovery platform for functional genomics. Trends Biotech 16: 301 – 306.

Schuchhardt J, Beule D, Malik A, Wolski E, Eickhoff HL, Herzel H (2000) Normalization strategies for cDNA microarrays. Nucleic Acids Res 28: e47.

Schuler GD, Epstein JA, Ohkawa H, Kans JA (1996) Entrez: molecular biology database and retrieval system. Methods Enzymol 266: 141 – 162.

Shepherd H, Chaparro J, Dal G, Jefferson L, Duong H, Vogel H, Walsh J, Gibbings M, Teasdale R (1995) Mapping insect resistance and essential oil traits in a

tropical *Eucalyptus* hybrid. Proceedings of the CRC/ IUFRO Conf.- *Eucalyptus* plantation: improving fibre yield and quality. Hobart, Australia 420 – 423.

Shirasi S (1998) Linkage relationships among allozyme loci in Japanese black pine, *Pinus thunbergii* Parl. Silvae Genetica 37: 60 – 66.

Smith S, Beavis W (1996) Molecular marker assisted breeding in a company environment. In: The impact of Plant Molecular Genetics, ed. B.W.S. Sobral, chap 15. Birkhauser, Boston.

Sosnowski RG, Tu E, Butler W, O'Connell J, Heller MJ (1997) Rapid determination of single base mismatch mutations in DNA hybrids by direct electric field control. Proc Natl Acad Sci USA 94: 1119 – 1123.

Southern EM (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98: 503 – 517.

Southern EM et al. (1992) Analysing and comparing nucleic acid sequences by hybridisation to arrays of oligonucleotides: evaluation using experimental models. Genomics 13: 1008 – 1017.

Southern E, Mir K, Shchepinov M (1999) Molecular interactions on microarrays. Nature Genet 21: 5 – 9.

Steinmetz LM, Davis RW (2000) High-density arrays and insights into genome function. Biotechnol Genet Eng Rev 17: 109-146.

Steinmetz LM, Sinha H, Richards DR, Spiegelman JI, Oefner PJ, McCusker JH, Davis RW (2002) Dissecting the architecture of a quantitative trait locus in yeast. Nature 416: 326 – 330.

Strauss SH and Conkle MT (1986) Segregation,, linkage, and diversity of allozymes in knobcone pine. Theor Appl Genet 72: 483 – 493.

Streiff R, Duncousso A, Lexer C, Steinkellner H, Gloessl J, Kremer A (1999) Pollen dispersal inferred from paternity analysis in a mixed oak stand of *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. Mol Ecol 7: 317 – 328.

Swofford DL, Olsen GJ, Waddell PJ, Hissis DM (1996) Phylogenetic inference. In Molecular Systematics, D. M. Hilliss, C. Moritz, and B. K. Mable, Eds. Sunderland, MA: Sinauer Associates, pp 407-514.

Sydor JR, Nock S (2003) Protein expression profiling arrays: tools for the multiplexed high-throughput analysis of proteins. Proteome Sci 1: e3.

Tanksley SD, Orton TJ (1983) Isozymes in plant genetics and breeding. Parts 1A and 1B. Elsevier, Amsterdam.

Troesch A, Nguyen H, Miyada CG, Desvarenne S, Gingeras TR, Kaplan PM, Cros P, Mabilat C (1999) Mycobacterium species identification and rifampin resistance testing with high-density DNA probe arrays. J Clin Microbiol 37: 49 – 55.

Van Alfen NK, Jaynes RA, S. L. Anagnostakis SL, Day PR (1975) Chestnut blight: biological control by transmissible hypovirulence in *Endothia parasitica*. Science 189:890-891.

Van de Goor (2003) The principle and promise of Labchip technology. Pharma Genomics 5:16 – 18.

Van der Westhuizen IP, Wingfield MJ, Kemp GHJ, Swart WJ (1992) Comparative susceptibility of *Eucalyptus grandis* clones and hybrids to *Cryphonectria cubensis*. Phytophylactica 24: 107 (abstract).

Van Hal NL, Vorst O, van Houwelingen AM, Kok EJ, Peijnenburg A, Aharoni A, van Tunen AJ, Keijer, J. (2000) The application of DNA microarrays in gene expression analysis. J Biotechnol 78: 271 – 280.

Van Heerden SW, Wingfield MJ (2001) Genetic diversity of *Cryphonectria cubensis* isolates in South Africa. Mycological Res 105: 94 – 99.

Van Heerden SW, Geletka LM, Preisig O, Nuss DL, Wingfield BD, Wingfield MJ (2001) Characterization of South African *Cryphonectria cubensis* isolates infected with *C. parasitica* hypovirus. Phytopathology 91: 628 – 632.

Van Heerden SW, Wingfield MJ (2002) Effect of environment on the response of *Eucalyptus clones* to inoculation with *Cryphonectria cubensis*. Forest Pathol 32: 395 – 402.

Van Zyl LM, Wingfield MJ (1999) Wound response of *Eucalyptus* clones after inoculation with *Cryphonectria cubensis*. Eur J For Pathol. 29: 161 – 167.

Voiblet C, Duplessis S, Encelot N, Martin F (2001) Identification of symbiosis-regulated genes in *Eucalyptus globulus-Pisolithus tinctorius* ectomycorrhiza by differential hybridization of arrayed cDNAs. Plant Science 25: 181.

Von Weisenberg K (1990) Silva Femica 24:129 – 139.

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 Res 23: 4407 – 4414.

Wang R, Guegler K, LaBrie S T, Crawford N M (2000) Genomic analysis of a nutrient response in *Arabidopsis* reveales diverse expression patterns and novel

metabolic and potential regulatory genes induced by nitrate. Plant Cell 12: 1491 – 1510.

Wayne ML, Mc Intyre (2002) Combining mapping and arraying: An approach to cadidate gene identification. Proc Natl Acad Sci USA 99: 14903 – 14906.

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.

Weeden NF, Timmerman GM, Lu J (1994) Identifying and mapping genes of economic importance. Euphytica 73 191 – 198.

Weeden NF, Zamir D, Tadmore Y (1988) Application of isozyme analysis in pulse crops. In: Summerfield R (ed.). World Crops. Cool Season Food Legumes: Amsterdam Nijhoff pp 979 – 987.

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.

Wilcox PL, Amerson HV, Kuhlman EG, Lui B-H, O'Malley DM, Sederoff RR (1996) Detection of a major gene for resistance to fusiform rust disease in loblolly pine by genomic mapping. Proc Natl Acad Sci USA 93: 3859 – 3864.

William MDHM, Mujeeb-Kazi (1993) *Thinopyrum bessarabicum*: biochemical and cytological markers for the detection of genetic introgression in its hybrid derivatives with *Triticum aestivum* L. Theor Appl Genet 86: 365 – 370.

Williams J, Kubelik A, Livak K, Rafalski J, Tingey S (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 16: 6531 – 6535.

Wilson DL, Buckley MJ, Helliwell CA, Wilson IW (2002) New normalization methods for cDNA microarray data. Bioinformatics 19: 1325 – 1332.

Wingfield MJ, Swart WJ, Abear B (1989) First record of *Cryphonectria* canker of *Eucalyptus* in South Africa. Phytophylactica 21: 311 – 313.

Wingfield MJ (1990). Current status and future prospects of forest pathology in South Africa. SA J of Science 86: 60 – 62.

Wingfield MJ, Swart WJ, Kemp GHJ (1991) Pathology considerations in clonal propagation of *Eucalyptus* with special reference to the South African situation In: Proceedings of the 1991 IUFRO Symposium – Intensive Forestry, The Role of *Eucalyptus* pp. 811 – 820.

Wingfield MJ (2003) Daniel McAlpine Memorial Lecture. Increasing threat of diseases to exotic plantation forests in the Southern Hemisphere: lessons from Cryphonectria canker. Australasian Plant Path 23: 133-139.

Winter P, Kahl G (1995) Molecular marker technologies for plant improvement. World J Microb & Biotech 11: 438 – 448.

Winzeler EA, Richards DR, Conway AR, Goldstein AL, Kalman S, Mc Cullough MJ, Mc Cusker JH, Stevens DA, Wodicka L, Lockhart DJ, et al. (1998) Direct allelic variation scanning of the yeast genome. Science 281: 752 – 755.

Worley J, Bechtol K, Penn S, Roach D, Hanzel D, Trounstine M, Barker D (2000) A Systems Approach to Fabricating and Analyzing DNA Microarrays. In: Microarray Biochip Technology, Eaton Publishing, pp 65 – 85.

Wu RL, Han YF, Hu JJ, Fang JJ, Li L, Li ML, Zeng ZB (2000) An integrated genetic map of *Populus deltoids* based on amplified fragment length polymorphisms. Theor Appl Genet 100: 1249 – 1256.

Xu H, Bakalinskey AT (1996). Identification of grape (*Vitis*) rootstocks using sequence characterized amplified region DNA markers. Hort Science 31: 267 – 268.

Yanchuk A (2002). The role and implications of biotechnology in forestry. Forest Genetic Res 30: 18 – 22.

Yang HY, Korban SS, Krueger J, Schmidt H (1997) The use of bulked segregant analysis to identify a molecular marker linked to a scab resistance gene in apple. Euphytica 94: 175 – 182.

Ye RW, Wang T, Bedzyk L, Croker KM (2001) Applications of DNA microarrays in microbial systems. J of Microbiol Methods 47: 257 – 272.

Young ND (1990) Potential applications of map-based cloning to plant pathology. *Phys.* Mol Plant Path 37: 81 – 94.

Zhu H and Snyder M (2001) Protein arrays and microarrays. Curr Opinion Chem Biol. 5: 40 – 45.

Ziegenhagen B, Kuhlenkamp V, Brettschneider R, Scholz F, Stephan BR, Degen B (1999) DIG-labelled AFLPs in oaks – A DNA marker for reconstruction of full or half sib family relationships. In: Gillet EM (ed) Which DNA marker for which purpose? Final Compendium of the Research Project 'Development, optimisation and validation of molecular tools for assessment of biodiversity in forest trees' in the European Union DGXII Biotechnology FW IV Research Programme 'Molecular Tools for Biodiversity'.

CHAPTER TWO

DEVELOPMENT AND ASSESSMENT OF MICROARRAY-BASED DNA FINGERPRINTING IN *EUCALYPTUS GRANDIS*

Published: *Theoretical and Applied Genetics* (2004) 109: 1329-1336

ABSTRACT

Development of improved *Eucalyptus* genotypes involves the routine identification of breeding stock and superior clones. Currently, microsatellites and random amplified polymorphic DNA (RAPD) markers are the most widely used DNA-based techniques for fingerprinting of these trees. While these techniques have provided rapid and powerful fingerprinting assays, they are constrained by their reliance on gel or capillary electrophoresis, and therefore, relatively low throughput of fragment analysis. In contrast, recently developed microarray technology holds the promise of parallel analysis of thousands of markers in plant genomes. The aim of this study was to develop a DNA fingerprinting chip for *Eucalyptus grandis*, and to investigate its usefulness for fingerprinting of eucalypt trees. A DarT-prototype chip was prepared using a partial genomic library from total genomic DNA of 24 *Eucalyptus grandis* trees, of which 22 were full siblings. A total of 384 cloned genomic fragments were individually amplified and arrayed onto glass slides. DNA fingerprints were obtained for 17 individuals by hybridising labeled genome representations of the individual trees to the 384-element chip. Polymorphic DNA fragments were identified by evaluating the binary distribution of their background-corrected signal intensities across full-sib individuals. Among 384 DNA fragments on the chip, 104 (27%) were found to be polymorphic. Hybridisation of these polymorphic fragments was highly repeatable ($R^2 > 0.91$) within the *E. grandis* individuals and they allowed us to identify all 17 full-sib individuals. Our results suggest that DNA microarrays can be used to effectively fingerprint large numbers of closely related *Eucalyptus* trees.

INTRODUCTION

Eucalyptus spp. are widely planted as exotics in many tropical and subtropical regions of the world (Eldridge et al. 1993). Since many of these plantations are commonly developed using vegetative propagation, the routine identification of clones and selection of elite genotypes has become increasingly important. Until recently, tree breeders have had to rely on detailed pedigree information and careful labeling to identify individual trees in breeding programs. However, incorrect identification is common and poses a major problem in forestry operations (Keil et al. 1994). DNA-based molecular markers have provided a solution to this problem. Several studies have thus shown that individual genotypes can be discriminated using molecular markers (Epplen et al. 1991; Nybom1991; Weising et al. 1991).

A variety of molecular marker techniques can be used for DNA fingerprinting. 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 high throughput afforded by some of these methods, they are all constrained by their dependence on gel electrophoresis. This hampers the processing of a large number of samples or markers in parallel (Smith and Beavis 1996). Furthermore, several of these methods require processing with many independent restriction enzymes or probes to achieve low error rates.

Originally designed for analysis of gene expression, DNA microarrays permit the parallel processing of large numbers of DNA fragments immobilised on a solid-state surface (Schena et al. 1995). To adopt microarray technology for fingerprinting and diversity studies, Jaccoud et al. 2001 recently reported the development of Diversity Array Technology (DArT™) in rice, while Borevitz et al. 2003 reported the use of oligonucleotide arrays to detect and genotype single feature polymorphisms (SFP) in *Arabidopsis*. No oligonucleotide arrays are available for *Eucalyptus* and therefore, the Diversity Array technique is the only microarray-based genotyping method that would be applicable for these trees. DArT™ is a solid state fingerprinting technique based on AFLP and enables analysis of large numbers of marker loci without any DNA sequence information. Microarray-based genotyping as implemented in the DArT™ technique is a 2-dye approach and relies on the detection of DNA fragments in a complex mixture of selectively amplified restriction fragments. Reduction of complexity by selective amplification allows comparison of polymorphic fragments among genotypes. This is achieved by hybridising DNA to an array containing a large number of DNA fragments, derived from genomic representations of an organism. However, plant genomes contain large amounts of highly repetitive DNA sequences and it is not clear how this feature might affect the rigor of hybridisation-based fingerprinting.

The aim of this study was to develop a prototype microarray chip to evaluate the potential of DNA microarrays for fingerprinting closely related *Eucalyptus* clones. In this study, the reproducibility of microarray hybridisation

profiles in *Eucalyptus grandis* was evaluated and recommendations for using this technology in plantation forestry were provided.

MATERIALS AND METHODS

Plant material and DNA extraction

A total of 15 full-sib progeny of *E. grandis* clone ZG14 (Mondi Forests, South Africa) were fingerprinted in this study. Clone ZG14 was used in a controlled cross with *E. grandis* clone TAG-S (Mondi Forests), from which 22 cloned progeny (clones 44D, 32A, 67D, 36E, 31C, 62D, 74C, 53B, 12C, 17C, 10D, 28D, 18C, 60D, 30B, 4D, 13C, 44C, 17D, 74C, 16C and 56E) were selected for the generation of a genomic representation of the whole full-sib family (described below). Genomic DNA was extracted from tree ZG14 and one ramet of each tree as described by Murray and Thompson (1980). The second parent tree (TAG-S) was lost during the early stages of this study and plant material was not available for it. A DNA sample was, therefore, obtained from tree TAG-5, a putative sibling relative of TAG-S.

Generation of genome representations

The method used for preparation of genome representations (Figure 1) was essentially the same as that described by Jaccoud et al. (2001). DNA samples were

pooled from 23 trees (144 ng DNA in total from 22 full-sib progeny and parental tree ZG14). The DNA in the pool was digested with 20 U *Pst*I using buffer H (Roche Diagnostics GmbH, Mannheim, Germany) in a reaction volume of 50 µL. The reactions were incubated at 37°C for 3 h and the restriction enzyme removed using an equal volume of phenol:chloroform. The DNA fragments were then precipitated with 100% EtOH and 100 mM NaCl. The precipitated DNA was washed with 70% ETOH and resuspended in 20 µl deionized water to a final concentration of 30 ng/µl.

Purified DNA was ligated to *Pst*I-specific adapters (Jaccoud et al. 2001) in a total volume of 30 µl at 10°C, overnight. The ligation mixture consisted of 1X ligation buffer, 2 U T_4 DNA ligase (Roche Diagnostics GmbH), 10 ng/µL l BSA (Amersham Biosciences, Piscataway, USA), 1.0 mM ATP (Amersham Biosciences) and 10 µM *Pst*I adapters. After ligation, 0.2 mM EDTA was added and the samples were heated at 70°C for 5 min to inactivate the ligase. The mixture was then diluted to 100 µl with water and 2 µl used as a template in a subsequent selective PCR reaction.

The PCR was performed in 50 µl containing 0.8 µM PCR primer (adapter +T), 0.25 mM of each dNTP, 1 U *Taq* polymerase, and 1 x reaction buffer (Roche Diagnostics GmbH). The PCR amplification consisted of 30 cycles of 94°C for 30 sec, 53°C for 45 sec, and 72°C for 1 min, with an initial denaturation step of 94°C for 5 min, and a final extension step of 72°C for 8 min.

Cloning, PCR amplification and sequencing of genomic fragments from representations

The amplified products were inserted into the PCR 2.1-TOPO vector using a T/A cloning kit (Invitrogen, Carlsbad, California, USA). After transforming *Escherichia coli* TOP 10F' host cells with ligation products, single colonies were grown overnight at 37°C in LB medium containing 50 µg/ml ampicillin. Recombinant *E. coli* clones were diluted in 1 vol of 50% glycerol and stored at -80°C. From each culture, 10 µl were transferred to 10 µl water and boiled for 10 min to disrupt the cells and release plasmid DNA into the growth medium. A 1 µl aliquot of this solution was used in a 100 µL PCR reaction with M13 forward (-20) and M13 reverse primers (Invitrogen). The reaction mix contained 1X PCR buffer, 1 U *Taq* polymerase (Roche Diagnostics GmbH), 0.25 mM of each dNTP, and 0.4 µM of each primer. The PCR amplification consisted of 30 cycles of 94°C for 30 sec, 53°C for 30 sec, and 72°C for 1 min, with an initial denaturation step of 95°C for 5 min, and a final extension step of 72°C for 7 min. Aliquots of the PCR products were separated on a 1.4% agarose gel for quality control. The remainder of each sample was then precipitated in 90% ethanol and 0.9 mM NaAc (pH 5.2) to exclude low molecular weight fragments. The precipitate was collected by centrifugation at 3600 x *g* for 30 min. Pellets were washed in 70% ethanol, dried, and then resuspended in deionized water at ~ 250 ng/µl.

Out of 384 amplified clones, forty random clones were sequenced. The insert sequences were subjected to similarity searches in GenBank using BLASTN and BLASTX. BLAST alignments were used to estimate the number of repetitive

clones in the library that could result in cross-hybridisation or uninformative spots on the array.

Array printing and processing

Equal volumes (10 µL each) of purified PCR product and 100% DMSO were transferred into a 384-well plate (Amersham Pharmacia Biotech). Eight replicates per fragment were arrayed on each slide at 250 µm spacing onto Vapour Phase Coated Glass Slides (Amersham Pharmacia Biotech) using a Molecular Dynamics Gen III spotter at the African Centre for Gene Technologies (ACGT) Microarray Facility, University of Pretoria, Pretoria, South Africa (http://fabinet.up.ac.za/microarray). Following printing, the slides were allowed to dry at 45-50% relative humidity overnight. Spotted DNA was then bound to the slides by UV-crosslinking at 250 mJ and baking at 80°C for 2 h.

Preparation of labeled probes

For microarray hybridisations, genome representations from parent tree ZG14 and 15 full-sib progeny were used. Tree TAG-5, the putative relative of parent TAG-S, was also included. Probe DNA from individual plants was prepared by restriction enzyme digestion of genomic DNA (144 ng per tree), ligation of restriction fragments to adapters, and subsequent amplification following the protocol described above. Amplicons were precipitated in one volume isopropanol to remove excess dNTPs. Labeling of the amplified fragments was carried out using the Klenow fragment of DNA Polymerase I (Roche Diagnostics GmbH). Each

labeling reaction contained 5 µg amplified DNA, 1.8 mM dNTP mix (0.3 mM of dATP, dGTP, dCTP each, 0.8 mM of dTTP, 0.1 mM Cy3-dUTP (Amersham Biosciences, Buckinghamshire, UK), 1 x hexanucleotide mix (Roche Diagnostics, GmbH) and 8 U Klenow enzyme (Roche Diagnostics GmbH). The reaction was incubated at 37°C overnight. After labeling, the DNA was column-purified (QIAquick PCR purification Kit, Qiagen GmbH, Germany).

Hybridisation and washing

Microarray slides were pre-hybridised for 20 min at 60°C in a solution containing 3.5 x SSC, 0.2 % SDS and 1% BSA (Roche Diagnostics GmbH). Slides were rinsed three times in deionized water and dried with N_2 gas. The Cy3-labeled probe was then dissolved in hybridisation solution containing 50% formamide (SIGMA), 25% 2 x hybridisation buffer (Amersham Pharmacia Biotech), and 25% deionized water. The mixture was denatured at 92°C for 5 min and quickly cooled on ice. The denatured probe (approximately 35 µl) was pipetted directly onto the microarray surface and covered with a glass coverslip (24 mm x 60 mm, No.1, Marienfeld, Germany). Slides were placed in a custom made hybridisation chamber (N. B. Engineering Works, Pretoria, South Africa) and incubated for 16 -18h in a 42°C water bath.

After hybridisation, slides were washed once in 1 x SSC, 0.2% SDS at 37°C for 4 min, twice in 0.1 x SSC, 0.2% SDS at 37°C for 4 min, twice in 0.1 x SSC at room temperature for 1 min, and then rinsed in deionised water for 2 seconds. Slides were dried using N_2-gas.

Scanning, image processing and data analysis

Slides were scanned using a GenePix 4000B Scanner (Molecular Dynamics, USA). The mean pixel intensity within each spot and the local background the spot were determined using Array Vision 6.0 software (Imaging Research Inc., Molecular Dynamics, USA). All signal intensities were background corrected. Abnormal spots (e.g. high background, dust, irregularities, etc.) were manually flagged for removal. Anomalous spots not detected through manual inspection were removed if the signal intensity of such spot was greater than 10% of the mean of the eight replicates on each slide. The mean background-corrected spot intensity of the remaining replicates of each DNA fragment was used in subsequent data analyses.

The single dye (Cy3) data were normalised across slides by regression on the spot intensity data for tree ZG14, which was used as a reference for normalisation of all progeny data. The normalised data were then converted into $\log_2$ intensity values.

Identification of polymorphic fragments

Polymorphic DNA fragments were identified in Microsoft Excel based on the bimodal distribution of their normalised intensity values across slides, consistent with their segregation as dominant PCR-based testcross (Aa:aa = 1:1) or intercross (A:aa = 3:1) markers. Relative intensity values were obtained by scaling the signal intensities to that of the DNA fragment with the highest intensity value across

slides (set to 1.0). The ranked spot intensities were plotted for each DNA fragment, and identification of DNA fragments with bimodal distribution was based on the presence of two clearly defined intensity classes with mean relative intensity values differing by at least 0.5. A binary scoring table of polymorphic spots was developed for all the *Eucalyptus* trees analyzed. The data for all the polymorphic spots were used to calculate the relative "distances" between the hybridisation profiles of individual *Eucalyptus* trees using Spearman correlation and hierarchical clustering (CLUSTER, available at http://rana.lbl.gov/). The clustering results were visualized with TREEVIEW (Eisen et al. 1998).

Verification of DNA polymorphisms

Two of the DNA polymorphisms detected in the array experiment were analysed by Southern hybridisation. *Pst*I digested total genomic DNA of nine individual trees were resolved on agarose gels and transferred to nylon membranes. Probes representing two of the polymorphic DNA fragments were labeled and hybridised to the *Pst*I digested DNA on the nylon membranes using the DIG High Prime DNA Labeling and Detection Starter Kit I (Roche Diagnostics GmbH, Germany).

Reproducibility of DNA microarray fingerprints

Tests were done on the reproducibility of hybridisation profiles starting from independently prepared genome representations, and that of stripping and re-hybridisation of the same slides. Repeated stripping and re-hybridisation of slides

allows for multiple rounds of hybridisation on the same slides. To test the reproducibility of the hybridisation fingerprints obtained from stripped slides, slides were treated using the protocol of Dolan et al. (2001) with minor modifications. Used slides were immersed four times in stripping buffer (2.5 mM Na_2HPO_4, 0.1% SDS) at 95°C for 25 s. Slides were then washed in deionised water at room temperature for 2 seconds and dried using N_2-gas. Stripped slides were scanned to verify that all signal had been removed. The stripped slides were then used for a repeat of the same hybridisation as before, but with independently labeled DNA. Data analysis was performed as described above. Independent replicates were also prepared from fresh leaf samples of the genome representations of tree ZG14. These genome representations were labeled and hybridised to new slides. Signal intensity values of the replicate hybridisations were plotted against each other in Microsoft Excel.

RESULTS

DNA Microarray Analysis

To consider the potential use of microarrays for fingerprinting *Eucalyptus* clones, a prototype DNA microarray chip was constructed with selectively amplified restriction fragments of pooled genomic DNA of an *E. grandis* full-sib family. The technique used to generate a genome representation of the full-sib family and of

each *Eucalyptus* tree employs the principle of AFLP (Vos et al. 1995). The complexity of each genomic DNA sample was reduced 16-fold by using +1/+1 selective nucleotides for PCR amplification of genomic restriction fragments. PCR amplicons prepared in this way ranged from 0.2 to 1.5 kb with an average insert size of 700 bp. Sequencing of 40 of the cloned PCR products revealed that there was a low proportion (17%) of "repeat" clones (i.e. clones with microsatellite or other simple repeat sequences, or multiple copies of the same genomic DNA fragment) in the *Eucalyptus* library generated (data not shown).

Proportion of polymorphic fragments useful for fingerprinting

To determine the proportion of polymorphic DNA fragments on the fingerprinting chip, tree ZG14 and 15 full-sib progeny were used in single dye experiments (Figure 2). While many of the array features were common (momomorphic) to all individuals (58%), or showed no hybridisation signal (15%), many (104 or 27%) were clearly polymorphic among individuals. However, only 55 of these spots (15%) were selected for further analyses. The analysis was limited to these 55 spots because clear threshold values (difference of 0.5 in relative intensity between two intensity classes) could be assigned for them (Figure 3A) and they were easily convertible into a binary scoring table (results not shown). In contrast, non-polymorphic spots, including both clearly monomorphic loci and loci that were not possible to score as either monomorphic or polymorphic (Figure 3B), exhibited a greater proportion of high relative intensity values. This can be attributed to the fact that monomorphic loci share the same signal intensities. Polymorphic spots for

which no clear threshold values could be assigned are responsible for the lower relative intensity values.

The CLUSTER software programme allowed us to visualise the relationships of the hybridisation profiles using TreeView (Eisen et al. 1998, Figure 4). The branching orders of duplicate experiments were all identical and duplicate experiments clustered as nearest neighbors. However, depending on which similarity metric setting was used, the overall branching order varied substantially. Since the Spearman correlation analysis provides a more conservative and reliable estimation of the relationship between hybridisation profiles (Murray et al. 2001), this correlation was used for data analysis. The dendrogram generated merely provides a means to visualize the relationship of fingerprints and should not be seen as representative of genetic relationships between the full-sib progeny.

All of the hybridisation profiles were unique and allowed unambiguous discrimination of the full-sib individuals. The probability of obtaining a particular 55- locus fingerprint is 2.7×10^{-17}, assuming no linkage among polymorphic spots. This provides an upper estimate of the discriminating power of our data. Randomly selected small subsets of polymorphic DNA fragments were used to determine that as few as 7 polymorphisms were sufficient to discriminate among full-sib progeny.

Reproducibility

To assess the reproducibility of the experimental procedure, replicate experiments were performed for nine individuals (Figure 2 and Figure 4). Signal intensities of the experimental replicates exhibited regression coefficients (R^2) ranging from 0.90

to 0.93 (Table 1). These are considered to reflect acceptable levels of reproducibility for microarray analysis (Hertzberg et al. 2001). These values were compared to the repeatability of binary scores obtained from the same hybridisations. Binary scores of replicate experiments were on average 1.5% higher than regression coefficients.

The regression of the hybridisation (normalized signal intensity) data obtained from two different sources of DNA (Figure 5) for the parent ZG14 revealed a linear regression coefficient (R^2) of 0.91. This was not significantly different from the regression coefficient obtained for the experimental replicates of the same tree ($R^2 > 0.93$), suggesting that independent DNA sampling did not introduce much additional experimental variance.

Validation of DNA polymorphisms

Two polymorphic DNA fragments (no 227 and 229) were analyzed by Southern hybridisation. When probe 227 of the genomic library was hybridised to a blot of the representations, trees 44D, 32A, 67D, TAG-S, 36E and 30B produced a band of 300bp in size, while a band of 430bp was detected for the other genotypes. The genomic Southern blot of probe 229 resulted in a band of 500bp in the case of trees 28C, ZG14, 60D, 17C, 10D, and a band of 300bp in size for the other genotypes. These RFLP banding patterns were converted to absence/presence of a band. These RFLPs were consistent with the bimodal hybridisation pattern observed for these two probes in the microarray experiment (Table 2).

Stripping and re-use of slides

Coefficients of determination, which are a measure of the correlation between two variables (experiments), were observed to be higher than 0.90 in replicate hybridisation experiments on stripped slides (data included in Table 1). This confirmed that re-used slides resulted in reproducible data. Although the signal intensities decreased on average by 10% after each successive hybridisation (Figure 2), spot signal intensities remained detectable and were quantifiable.

DISCUSSION

In this study we have shown that microarray technology can be used for genome-wide fingerprinting of closely related *Eucalyptus* trees. Several features of the DNA microarray technology make it attractive for this purpose. The DNA for hybridisation is prepared by selective PCR amplification of short restriction fragments. This means that <250 ng of total genomic DNA provides essentially unlimited starting material for future genotyping of the same trees. This technique, like AFLP analysis, also allows genomic fingerprinting of organisms such as *Eucalyptus* tree species with no prior DNA sequence information (Jaccoud et al. 2001). Most importantly, analysis of the polymorphic fragments is not restricted by the need for gel electrophoresis, and thousands of polymorphic loci in each tree genome, can potentially be analysed in a single assay. Gel electrophoresis in

contrast is limited in throughput and suffers from difficulties in precisely matching allelic variants of the same size on different gels (Ticknor et al. 2001).

Despite the recent progress that has been made towards the application of microarray technology for DNA fingerprinting and high-throughput genotyping in plants (Jaccoud et al. 2001, Borevitz et al. 2003), cross-hybridisation remains a problem. The highly repetitive DNA content of plant genomes undoubtedly results in cross-hybridisation of DNA fragments to printed probe DNA. This increases the overall spot intensity of many probes, and it masks potential polymorphisms. It has been demonstrated that small regions of similarity can lead to cross-hybridisation on oligonucleotide microarrays. Kane et al. (2000) found that in 50-mer oligonucleotide arrays, cross-hybridisation occurred between fragments of relatively low sequence similarity. This has also been observed on microarrays with PCR-based probes (Wren et al. 2002). In general, cross-hybridisation of many different genomic fragments will result in the conversion of polymorphic probes into monomorphic probes. However, a much more serious problem is presented by background segregation of a small number of strongly cross-hybridising fragments, which will result in mixed hybridisation patterns and incorrect marker phenotypes. This problem can be detected at the locus level in segregating progeny, but not in population or fingerprinting studies.

The prototype microarray chip developed in this study for fingerprinting *Eucalyptus* clones allowed for the discrimination among full sib progeny and thus

very closely related individuals. The hybridisation profiles obtained for *Eucalyptus grandis* individuals were highly repeatable ($R^2 > 0.9$), and allowed us to identify distinct intensity classes (bimodal intensity distributions) for 55 (14.3%) of the 384 printed probes (Figure 3). An additional 49 of the probes showed bimodal intensity distributions, but the overlap between the two intensity classes for these probes was inordinately great to easily assign them to presence or absence classes. The total proportion of bimodal probes (27%) and polymorphisms (14.3%) that could be scored was somewhat lower than the rate of polymorphisms often reported for gel-based AFLP markers in outcrossed *Eucalyptus* pedigrees (up to 50%, Myburg et al. 2003). The lower rate of scorable polymorphisms is most probably the result of cross-hybridisation obscuring polymorphic features. This is in addition to the "normal" inaccuracies introduced during labeling and hybridisation.

In an outcrossed pedigree, the majority of restriction fragment polymorphisms would be expected to segregate in testcross configuration (Aa:aa = 1:1), while a smaller proportion are expected to segregate in intercross configuration (AA:Aa:aa = 1:2:1 or 3:1). The majority of fragments will segregate as testcross fragments since a higher heterozygosity is expected in an outcrossing pedigree. Our pedigree set (15 full-sibs) was not sufficiently large to reliably distinguish between intercross and testcross segregation patterns, or to determine whether these fragments can be scored in a dosage dependent (co-dominant) fashion on microarrays. Therefore, the bimodal intensity distribution shown in Figure 3A probably contains a mixture of testcross and intercross fragments, which may explain the wide and high.

Signal intensity differences among genotypes can be compared across arrays using either single-dye or two-dye colour detection. The Diversity Array technique as described by Jaccoud et al. (2001) represents a two-dye approach. Differences among genotypes (presence or absence of fragments) are detected by comparing the Cy3 signal of each array element to the Cy5 signal of a reference (another genome representation, or a labeled vector fragment). Polymorphic spots show a bimodal distribution of log ratios relative to the reference. The use of a vector-based reference therefore provides an internal standard for each spot and a way to control for differences in the amount of DNA spotted on each array. However, if the same amount of DNA is spotted in each position across arrays, as can be expected for spots printed with the same pin, the value of the reference channel has to be balanced against the additional cost of labeling. Significant variation in printing across arrays was not observed, and therefore used a single-dye approach and normalized signal intensities rather than signal ratios. The normalised intensities were used to identify polymorphic spots based on their bimodal frequency distribution across individuals.

Reproducibility is essential in genotyping and fingerprinting. We tested for reproducibility of fingerprinting profiles at the experimental and biological level and found that the R^2 of normalized mean signal intensities was always higher than 0.90 in duplicate experiments, even when different sources of genomic DNA were used). The observed variability in signal intensities of 6 – 9 % between replicates of the same individual (in different labeling and hybridisation reactions) can be

ascribed to variability in the experimental process. Spot variability probably resulted from inaccuracies introduced in labeling, array hybridisations, signal detection and quantification, or low hybridisation signal. A higher frequency of errors at lower signal intensities was also observed due to signals being close to the background noise (Hertzberg et al. 2001). In comparison to the mean signal intensities, the (dominant) binary scores obtained from the same hybridisation data were more repeatable (>95%). This was due to the fact that correct scores could still be obtained when signal intensities varied within signal intensity classes, and due to the low occurrence of spots that varied sufficiently to be erroneously classified. In addition, the repeatability of the hybridisation profiles based on the 55 scored polymorphic probes was on average approximately 1.5% higher than that based on the full set of 384 probes (data not shown).

The power of microarray-based fingerprinting lies in its ability to compare different genomes at large numbers of loci, in a single assay. In this context, direct comparison of signal intensity profiles may allow accurate identification of individuals, if proper normalization procedures are followed. Our results suggest that binary scores based on underlying hybridisation patterns are only marginally more repeatable than the hybridisation data. However, binary (or ideally co-dominant) scores are required to determine allelic frequencies in target populations and in order to calculate probabilities of misidentification for forensic purposes. Binary scores are also required for linkage analysis in mapping pedigrees. Although the use of this technology for linkage mapping still remains to be tested,

our results suggest that the technology is useful for rapid genome-wide comparison of closely related germplasm. This study showed that the branching orders of replicate hybridisation fingerprints were all identical and replicate fingerprints all clustered as nearest neighbors. This allowed for the unambiguous identification of *Eucalyptus grandis* individuals and the identification of two unknown samples (included as blind test samples).

Microarray-based fingerprints may allow the identification of genomic regions shared between related individuals, or identification of genomic regions inherited from specific parents in outcrossed pedigrees. Borevitz et al. (2003) recently demonstrated the use of oligonucleotide probes to demarcate recombination events along chromosomes of recombinant inbred lines of *Arabidopsis*. In our case, map information is not available, but in the future the internal sequences of probes will be useful to link polymorphisms to a genome sequence when that becomes available for *Eucalyptus*. The clustering of probes into columns according to levels of similarity based on their hybridisation (or segregation) patterns across individuals suggests the presence of major linkage groups. This approach may allow ordering of polymorphic markers if the population size is increased adequately.

No dedicated software products are currently available to define hybridisation-based DNA fingerprints or to extract binary scores from hybridisation data. The majority of available microarray software is designed for two-color expression profiling studies. For single-color fingerprinting applications, such as the one used

in this study, the presence or absence of fragments (dominant scoring) or signal intensity (for co-dominant scoring) has to be determined to construct a fingerprint, and quality values need to be assigned to each data point to evaluate the reliability of the combined fingerprint. Kingsley et al. (2002) used the APEX (automated peak extraction) algorithm to measure spot intensities, and to determine whether a spot is "on" or "off". This algorithm has advantages over the software used in the present study, and should be considered for future work.

The long-term objective of the research presented in this study is to develop a larger array, or set of arrays, with informative probes that can be used for genome-wide fingerprinting of most commercially planted *Eucalyptus* tree species. This will require multiple rounds of selection of polymorphic probes within *E. grandis*, and selection of polymorphic probes in other species or interspecific mapping pedigrees. Such an array of polymorphic probes will be useful to saturate existing genetic linkage maps, and may also allow comparative mapping of many eucalypt genomes. Future fingerprinting arrays may be based on oligonucleotides residing in genes (Borevitz et al. 2003) or on genomic restriction fragments such as those cloned in this study.

REFERENCES

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

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

Dolan PL, Wu Y, Ista LK, Metzenberg RL, Nelson MA, Lopez GP (2001) Robust and efficient synthetic method for forming DNA microarrays. Nucleic Acids Research 29(21): e107.

Eisen MB, Spellmann PT, Brown, PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci, USA 95: 14863 – 14868.

Eldridge KJ, Davidson J, Harwood C, van Wyk G (1993) Eucalypt Domestication and Breeding. Oxford Univ. Press, Oxford.

Epplen JT, Ammer H, Epplen C, Kammerbauer C, Mitreiter R, Roewer L, Schwaiger W, Steinle V, Zischler H, Albert E, Andreas A, Beuermann B, Meyer W, Buitkamp J, Nanda I, Schmid M, Nuernberg P, Pena SDJ, Poeche H, Sprecher W, Schartl M, Weising K, Yassouridis A (1991). Oligonucleotide fingerprinting using single repeat motifs: a convenient, ubiquitously applicable method to detect hypervariablity for multiple purposes. In: Burke T, Dolf G, Jeffreys AJ, Wolff R (eds) DNA fingerprinting approaches and applications, Birkhaeuser Verlag, Basel, pp 50 – 69.

Hertzberg M, Sievertzon M, Aspeborg H, Nilsson P, Sandberg G, Lundeberg J (2001) cDNA microarray analysis of small plant tissue samples using a cDNA tag target amplification protocol. The Plant Journal 25(4): 1 – 9.

Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity Arrays: a solid state technology for sequence independent genotyping. Nucleic Acids Research 29(4) : e25.

Kane MD et al. (2000) Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. Nucleic Acids Research 28: 4552 – 4557.

Keil M, Griffin AR (1994) Use of random amplified polymorphic DNA (RAPD) markers in the discrimination and verification of genotypes in Eucalyptus. Theor Appl Genet 89: 442 – 450.

Kingsley MT, Straub TM, Call DR, Daly DS, Wunschel SC, Chandler DP (2002) Fingerprinting closely related Xanthomonas pathovars with random nonamer oligonucleotide microarrays. Applied and Environmental Microbiology 68(12): 6361-6370.

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

Murray AE, Lies D, Li G, Nealson K, Zhou J, Tiedje JM (2001) DNA/DNA hybridization to microarrays reveals gene-specific differences between closely related microbial genomes. Proc Natl Acad Sci 98 (17): 9853-9858.

Myburg AA, Griffin AR, Sederoff RR, Whetten RW (2003) Comparative genetic linkage maps of *E. grandis, E. globulus* and their F_1 hybrid based on a double pseudo-backcross mapping approach. Theor Appl Genet (in press).

Nybom H (1991) Applications of DNA fingerprinting in plant breeding. In: Burke T, Dolf G, Jeffreys AJ, Wolff R (eds) DNA fingrprinting approaches and applications, Birkhaeuser Verlag, Basel, pp. 294 – 331.

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

Smith S, Beavis W (1996) Molecular marker assisted breeding in a company environment. In: The Impact of Plant Molecular Genetics, ed. B.W.S. Sobral, chap15. Birkhauser, Boston.

Ticknor LO, Kolsto A-B, Hill KK, Keim P, Laker MT, Tonks M, Jackson PJ (2001) Fluorescent amplified fragment length polymorphism analysis of Norwegian *Bacillus cereus* and *Bacillus thuringiensis* soil isolates. Appl Environ Microbiol 67: 4863 – 4873.

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 polymearse chain reaction. Am J Hum Genet 44: 388 – 396.

Weising K, Ramser J, Kaemmer D, Kahl G, Epplen JT (1991) Oligonucleotide fingerprinting in plant and fungi. In: Burke T, Dolf G, Jeffreys AJ, Wolff R (eds) DNA fingrprinting approaches and applications, Birkhaeuser Verlag, Basel, pp 312 – 329.

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.

Wren JD, Kulkarni A, Joslin J, Butow RA, Garner HR (2002) Cross-Hybridization on PCR-Spotted Microarrays. IEEE Engineering in Medicine and Biology. March/ April 71 – 75.

Table 1. Repeatability of hybridization profiles and binary scores. R^2 values are based on two separate labeling reactions and hybridisations starting from a single genome representation of each individual.

Eucalyptus individual no	R^2 [a] Hybridization profile	Repeatability of binary scores [b]
ZG14 (parent tree)	93.53	98.18
ZG14 (parent tree – biological replicate)[c]	91.47	94.55
TAG-5 (relative)	91.72	96.37
18C	92.34	96.37
28C	91.79	94.55
53B	93.95	98.18
36E	92.52	96.37
30B	93.09	96.37
67D	91.86	94.55
74C	93.97	98.18

[a] Based on the spot intensities in two replicate experiments of all 384 features on the array.

[b] Proportion of polymorphic probes (55 total) with same binary score across experimental replicates [1-(number of misscores/55)]x 100%

[c] For tree ZG14, in addition to a direct experimental replicate, an independently obtained DNA sample and genome representation was used as biological replicate.

Table 2. Hybridization patterns of individual RFLP alleles and microarray features. Hybridization patterns were only determined for replicated individuals (see Table 1).

		ZG14	74C	18C	28C	53B	36E	30B	67D	TAG-5
Probe 227	RFLP allele **(300 bp)**	_	_	_	_	+	+	_	+	+
	Microarray hybridisations	_	_	_	_	+	+	_	+	+
Probe 229	RFLP allele **(350 bp)**	_	_	+	+	+	+	+	+	+
	Microarray hybridisations	_	_	+	+	+	+	+	+	+

Figure 1. Schematic representation of the microarray-based genotyping method used in this study. Note that a pooled DNA sample was used to prepare the genome representation that was printed on the array. Also, the length of the selectively amplified restriction fragments determine the number of incorporated dye molecules per fragment, and therefore the average intensity of the corresponding spot on the array.

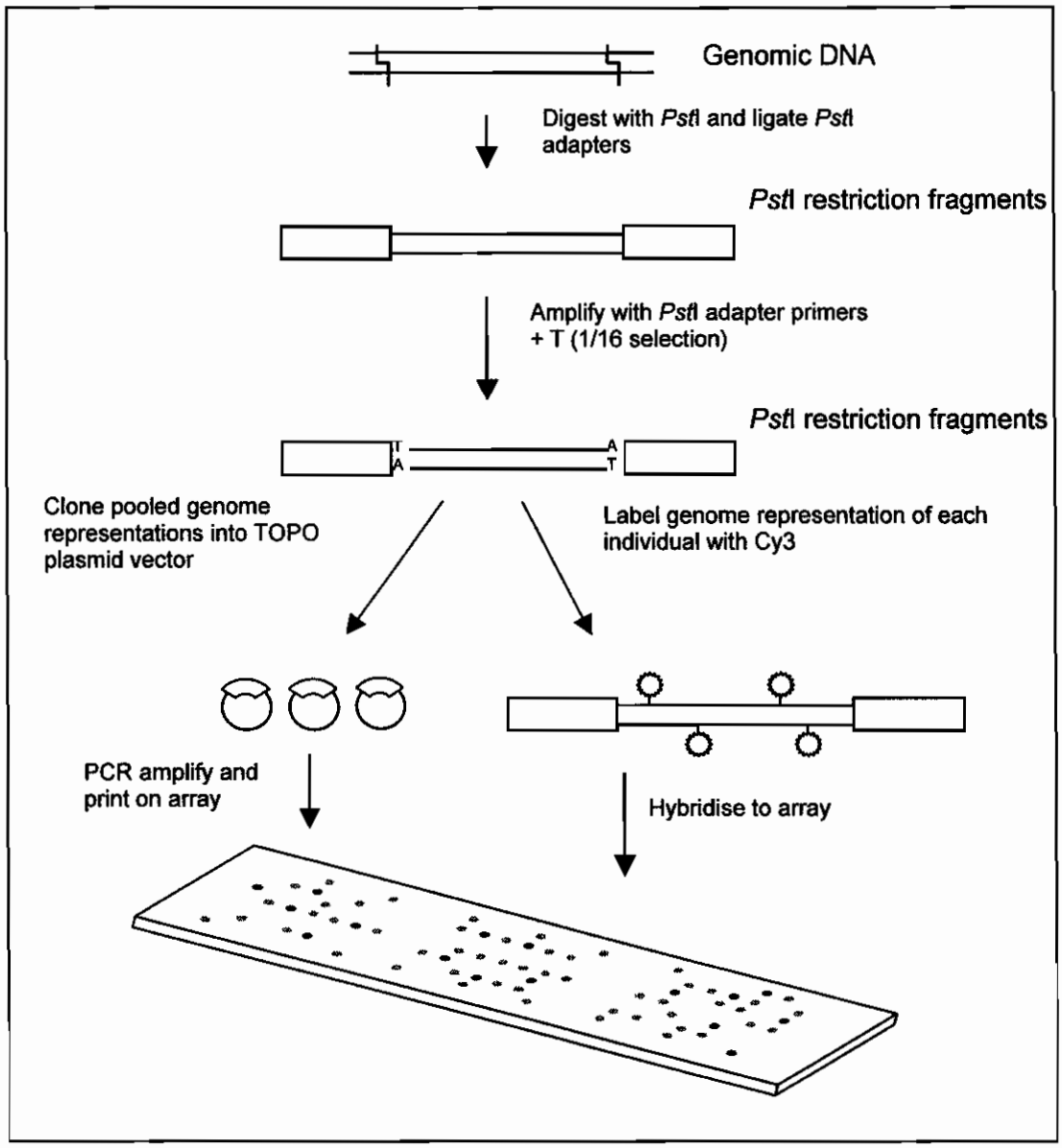

Genomic DNA
Digest with *Pst*I and ligate *Pst*I adapters
*Pst*I restriction fragments
Amplify with *Pst*I adapter primers + T (1/16 selection)
*Pst*I restriction fragments
Clone pooled genome representations into TOPO plasmid vector
Label genome representation of each individual with Cy3
PCR amplify and print on array
Hybridise to array

Figure 2. Microarray hybridization patterns of two different *Eucalyptus* individuals on the same section of the slide. Each column represents four replicates of the same spot. (A) Hybridization fingerprint of *Eucalyptus* individual 67D and (B) parent ZG14. C and D are hybridization of the same individuals on replicate, stripped arrays.

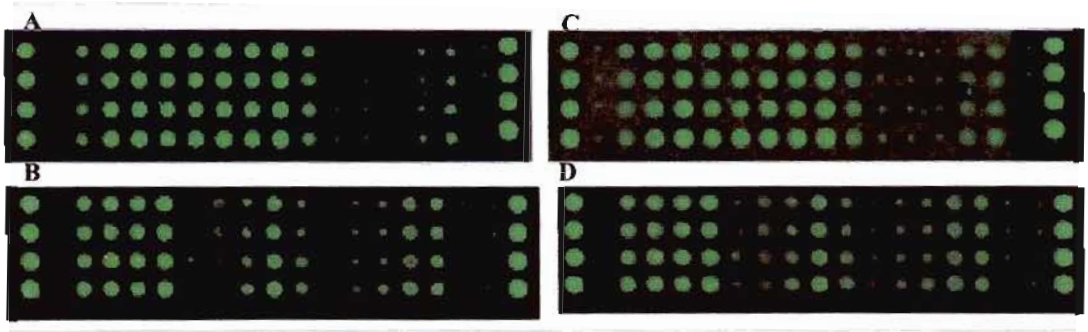
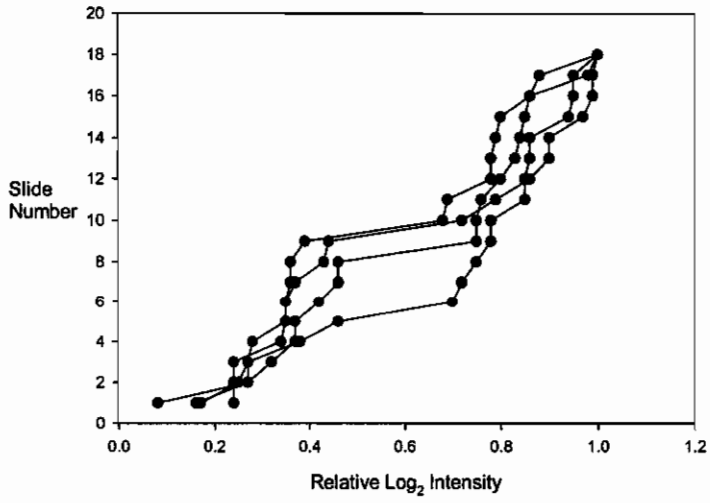


Figure 3. Examples of signal intensity distributions of log-transformed hybridisation data among 17 *Eucalyptus* individuals. (A) Distribution of relative (normalized) log intensities of four random polymorphic fragments that show a clear bimodal distribution across slides. (B) Non-polymorphic spots show a unimodal distribution.

A


Slide Number
0
2
4
6
8
10
12
14
16
18
20
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Relative Log2 Intensity

B

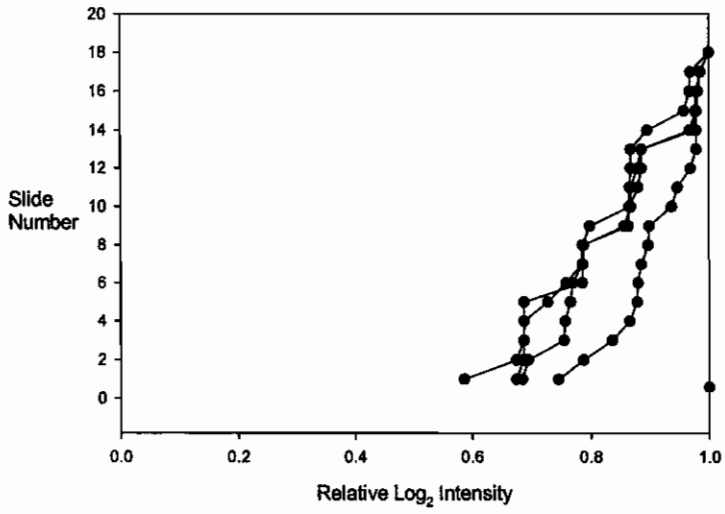

Slide Number
0
2
4
6
8
10
12
14
16
18
20
0.0
0.2
0.4
0.6
0.8
1.0
Relative Log2 Intensity

Figure 4. TreeView (Eisen et al. 1998) representation of relationships of hybridization profiles among 17 *Eucalyptus* individuals based on microarray analysis with the 384-probe array. Columns represent hybridization profiles of individuals (or replicates) and rows represent the mean log intensities for labeled DNA/DNA hybridisations across individuals. *Red* and *green* bars indicate high and low mean log intensity values, *black* bars indicate intermediate values and *grey* bars show missing data. Nine of the hybridisations were performed in replicate (indicated as REP). The replicate for ZG14 is a biological replicate, i.e. starting from independently obtained leaf samples of the same tree.

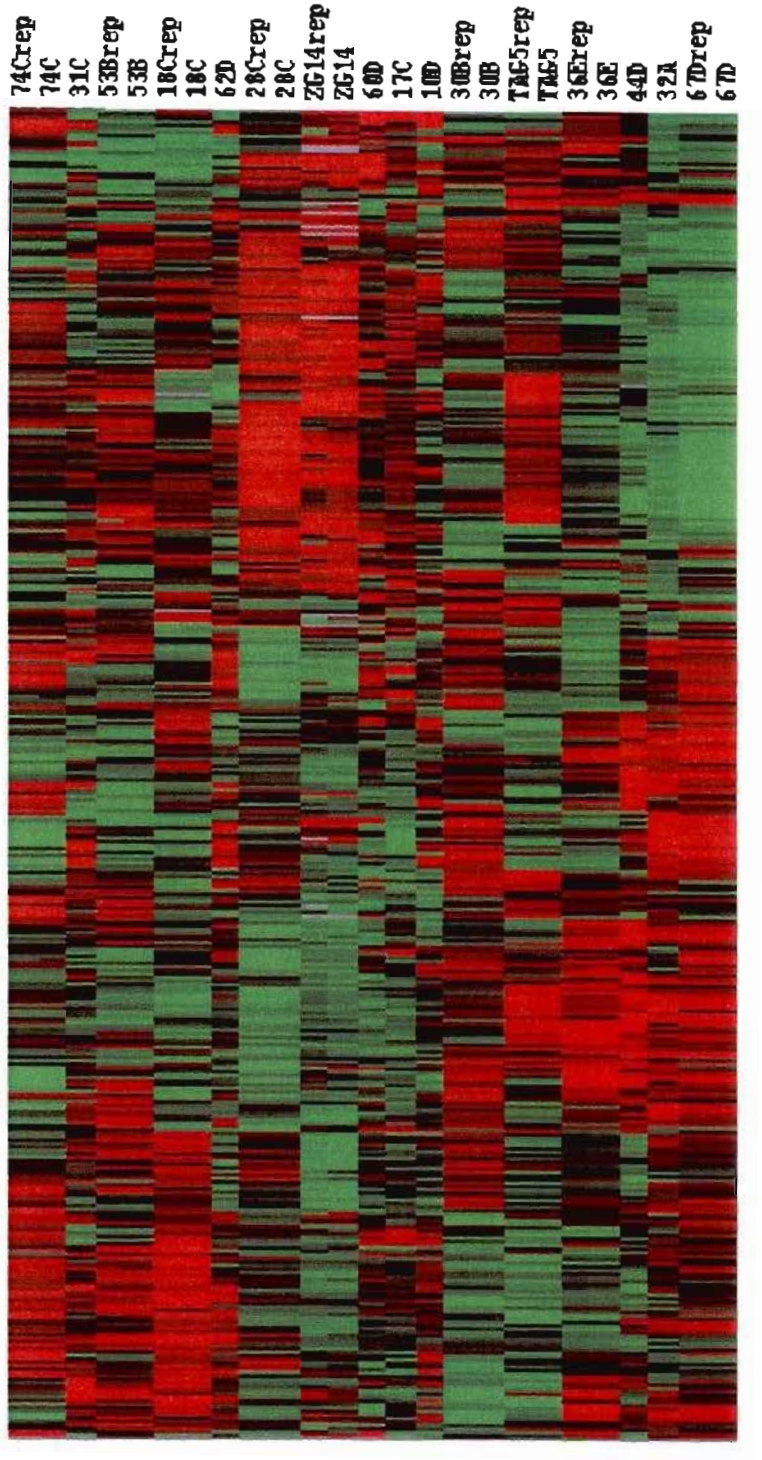
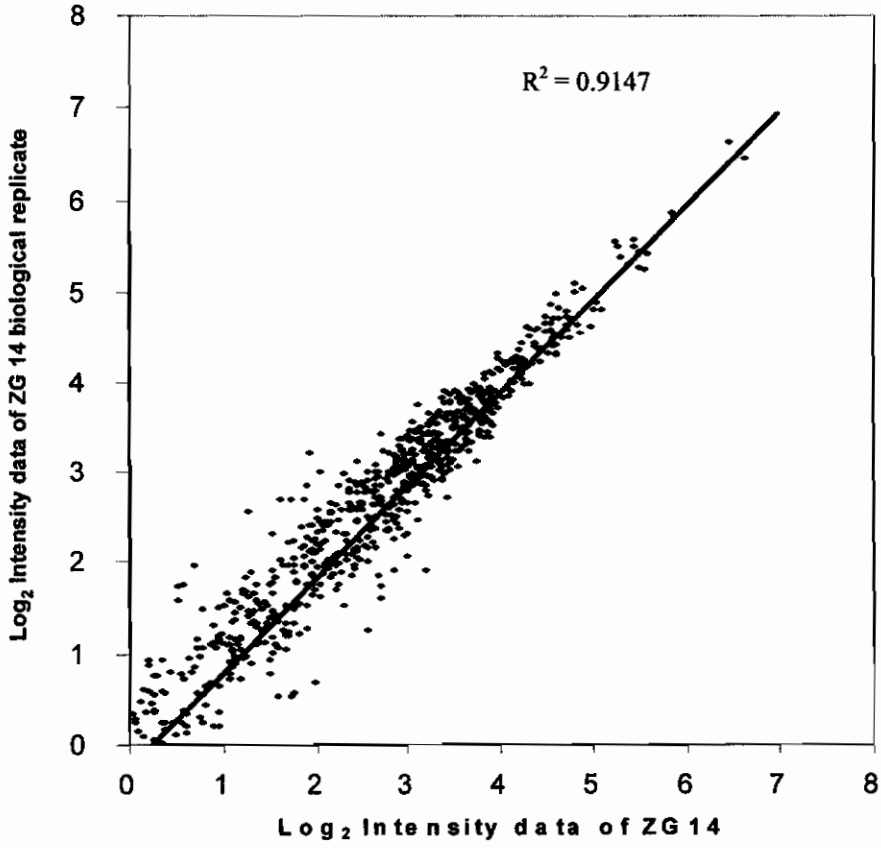

74Crep
74C
31C
53Brep
53B
18Crep
18C
62D
28Crep
28C
ZG14rep
ZG14
60D
17C
10D
30Brep
30B
TM5rep
TM5
36Erep
36E
44D
32A
67Drep
67D

Figure 5. Log plot of the microarray hybridisation signals of *Eucalyptus* individual (ZG14). The signal intensity obtained with ZG14 (*x*axis) was plotted against its biological replicate.


$R^2 = 0.9147$
Log_2 Intensity data of ZG 14 biological replicate
Log_2 Intensity data of ZG 14
0
1
2
3
4
5
6
7
8

CHAPTER THREE

IDENTIFICATION OF MICROARRAY MARKERS ASSOCIATED WITH *CHRYSOPORTHE AUSTROAFRICANA* TOLERANCE IN *EUCALYPTUS* USING DIVERSITY ARRAY TECHNOLOGY (DArT) AND BULK SEGREGANT ANALYSIS

ABSTRACT

Eucalyptus spp. are important plantation trees in South Africa and other parts of the world. The canker fungus *Chrysoporthe austroafricana,* previously known as *Cryphonectria cubensis in South Africa,* has resulted in serious damage, particularly to clonal plantations. Selection of disease tolerant *Eucalytpus* planting stock is thus a priority of the South African forestry industry. The aim of this study was to develop DNA-based molecular markers for Chrysoporthe tolerance in *Eucalyptus grandis* using a bulk segregant analysis (BSA) approach and Diversity Array Technology (DArT). A genotyping microarray had previously been prepared by arraying a partial library of genomic restriction fragments of 24 *Eucalyptus grandis* trees onto a glass slide. The multi-locus marker genotypes of tolerant and susceptible bulks, as well as individual genotypes of the progeny, were determined using a 384-feature prototype genotyping array. BSA revealed a total of 109 scorable, polymorphic loci, of which nine appeared to be associated with tolerance or susceptibility. Two DArT markers were converted to cleaved amplified polymorphic sequence (CAPS) markers, which discriminate susceptible and tolerant individuals. This study shows that DArT-based genotyping in combination with BSA provides a powerful approach for the detection of markers associated with disease tolerance in forest trees, and it should be equally applicable in other crop species.

INTRODUCTION

Chrysoporthe cubensis, previously known as *Cryphonectria cubensis* (Gryzenhout et al. 2004), is one of the most important pathogens of plantation-grown *Eucalyptus* spp. (Hodges et al. 1979, Hodges 1980, Florence et al. 1986, Wingfield et al. 1989). The fungus causes a serious stem canker disease and it is widely accepted to have influenced changes to vegetative propagation of *Eucalyptus* spp. for plantation development (Wingfield et al. 2003) in areas where climatic conditions favor the development of the disease (Hodges et al. 1979, Alfenas et al. 1982, Sharma et al. 1985b, Florence et al. 1986).

A fungus very similar to and originally thought to be the same as *Chr. cubensis* (Gryzenhout et al. 2004) causes a serious stem canker disease of *Eucalytpus* in South Africa (Gibson 1981, Hodges et al. 1986, Micales et al. 1987, Wingfield et al. 1989, Roux et al. 1999, 2003). Specimens previously treated as *Cryphonectria cubensis* from South Africa could, however, be distinguished morphologically from those representing *Chrysoporthe cubensis* and have thus been described as *Chrysoporthe austroafricana* (Gryzenhout et al. 2004). The discovery of *Chrysoporthe austroafricana* and the associated canker disease on *Eucalyptus* in South Africa (Wingfield et al. 1989) has caused concern regarding the impact that the disease could have on the success of *Eucalyptus* plantations in this country.

A substantial part of the forestry industry in South Africa depends on the wide-scale planting of *Eucalyptus* clones. The result is large areas of genetically uniform stands that are seriously threatened by pests and diseases (Wingfield 1990). Thus, it has become important for forestry companies both in South Africa and elsewhere in the world to implement an effective disease management strategy for these pathogens (Wingfield 2003).

Deployment of *Eucalyptus* planting stock, tolerant to canker caused by *Chr. austroafricana,* is considered the only feasible means of controlling the disease in affected areas (van Zyl 1990, Wingfield 2003, van Heerden et al. 2003). Trials to test disease tolerance of clonal hybrids using traditional breeding methods are time-consuming and costly. In addition, efficient selection of tolerant clones is possible only after trees are more than one year old (van der Westhuizen et al. 1992). Rapid, yet reliable, screening methods for disease susceptibility or tolerance are thus needed to identify and deploy tolerant genotypes.

Polymerase chain reaction (PCR)-based marker systems have been available for more than a decade. Among these, RAPD (Williams et al. 1990), AFLP (Vos et al. 1995) and SSR (Weber and May 1989) have been used extensively for fingerprinting, gene-tagging and mapping (Brondani et al. 2002, Myburg et al. 2003). These marker systems are technically accessible and rapidly provide polymorphic markers with universal reagents and assay protocols. However, they require gel electrophoresis for marker genotyping. Common limitations of gel electrophoresis are the limited number of detectable polymorphisms (up to 100 on the highest resolution acrylamide gels), and

difficulties in the precise correlation of bands with allelic variants resolved on different gels (Jaccoud et al. 2001).

Recently, methods have been developed for genotyping very large numbers of marker loci in parallel based on microarray hybridization of genomic DNA fragments. Jaccoud et al. (2001) reported the development of Diversity Array Technology (DArT™). This technique relies on the detection of DNA fragments in a complex mixture of selectively amplified restriction fragments and was recently used to develop the first genetic maps in a plant species (barley) based on microarray markers (Wenzl et al. 2004). Borevitz et al. (2003) reported the use of oligonucleotide arrays for repeatable and informative detection of allelic variation in the *Arabidopsis* genome. Since no oligonucleotide arrays are available for *Eucalyptus*, DArT is the only microarray-based genotyping method currently applicable in these species. This technique provides a tool for the parallel analysis of thousands of restriction fragment-based marker loci in the *Eucalyptus* genome, and we have used it for genome-wide fingerprinting of closely related *Eucalyptus* trees (Lezar et al. 2004).

In this study, DArT was used as a tool to identify markers associated with tolerance to *Chr. austroafricana* in *Eucalyptus*. Putative tolerance-associated markers were identified by bulk segregant analysis (BSA; Michelmore et al. 1991) and converted to cleaved amplified polymorphic sequence (CAPS, Konieczny et al. 1993) markers for further characterization in segregating *Eucalyptus* populations.

MATERIALS AND METHODS

Plant material and DNA extraction

A full-sib pedigree of *Eucalyptus grandis* was used to identify the DArT markers. This pedigree consisted of progeny from a controlled cross of the susceptible *E. grandis* clone ZG14 (Mondi Business Paper South Africa) and a single tolerant *E. grandis* tree TAG-S (Mondi). Genomic DNA of twenty-two clonally propagated F_1 progeny and the two parents were used for the generation of a genomic representation of the whole full-sib family (Lezar et al. 2004). Genomic DNA was extracted from young leaves as described by Murray and Thompson (1980). A DNA sample was also obtained from TAG5 (Mondi), a putative sibling relative of TAG-S, because tree TAG-S was lost during the early stages of this study.

Canker assessment

A total of 210 F_1 progeny (two clonal ramets per F_1 individual) and the two parental genotypes (ramets of clones ZG14 and TAG5) were planted in a commercial trial near Kwambonambi, Kwazulu-Natal, South Africa in 1997. The clonal replicates were planted in a randomized block design and maintained using standard silvicultural procedures. All parent and progeny trees in the trial were inoculated with one isolate (CMW 2113) of *Chr. austroafricana*. This isolate was previously shown to be highly virulent in a population of isolates (van Heerden et

al. 2001). The inoculum was prepared by culturing the fungus on 2% Malt Extract Agar (MEA; 20 g Malt extract/ 1000 ml water) plates. For inoculations, a cambial disc was removed from the main stems at a height of approximately 140 cm using a 20-mm diameter cork borer. A disc taken from the actively growing margin of a culture of *Chr. austroafricana* was placed in each wound with the mycelium facing towards the inside. To reduce desiccation, wounds were sealed with masking tape. Lesion lengths were measured 11 weeks after inoculation. The data collected from this trial were used to identify 11 disease tolerant and 10 susceptible F_1 progeny (Figure 1) that had the most extreme phenotypes and exhibited consistent phenotypes in three consecutive years of disease assessment. Genomic DNA was obtained from disease-free clonal ramets of these F_1 progeny and used for microarray analysis.

Microarray analysis

A 384-probe prototype DArT microarray was constructed by arraying genomic restriction fragments derived from total genomic DNA of 22 F_1 progeny and two parents onto glass slides (Jaccoud et al. 2001, Lezar et al. 2004). The DArT fragments were prepared by digesting a mixture of genomic DNA of all 24 trees with the restriction enzyme *Pst*I followed by ligation of enzyme-specific adapters to the restriction fragments. The genome complexity was reduced by selective amplification using adaptor primers ending in a single selective nucleotide (Lezar

et al. 2004). Eight replicates of the 384 DArT fragments were arrayed on each slide.

For BSA, two bulk genome representations were constructed by mixing equal amounts of genomic DNA of the 10 susceptible and 11 tolerant F_1 genotypes (from the phenotypic extremes, Figure 1), respectively, followed by restriction digestion and reduction of genome complexity as described above. A multi-locus DArT genotype was generated for each bulk genome representation by hybridizing differentially labeled (Cy3 or Cy5) genome representations of the two bulks to the 384-probe DArT genotyping array. Probe DNA preparation, labeling and hybridization were carried out following the published DArT procedure (Jaccoud et al. 2001) with modifications described by Lezar et al. (2004). One technical replicate of each hybridization was performed, each replication consisting of a reverse labeling experiment. In addition, the whole experiment was repeated with fresh DNA samples (i.e. four microarray slides were used in total). DArT genotypes were generated for the individual F_1 progeny included in the bulks and for the parents in a recent DArT fingerprinting study (Lezar et al. 2004). Array Vision 6.0 software (Imaging Research Inc., Molecular Dynamics, USA) was used to quantify the signal intensities of the array features.

For BSA, DArT fragments that differentiated the DNA pools derived from the tolerant and susceptible F_1 progeny were identified. Signal intensity ratios between the susceptible and tolerant bulks were obtained by dividing the normalized average signal intensity (across four reverse labeling replicates and eight spot replicates) of the tolerant bulk by that of the susceptible bulk. The $\log_2$

fold change and significance of the difference in bulk means were computed in SAS (version 8.2, Cary, NC, USA) for each of the 384 DArT features as described elsewhere (Okinaka et al. 2002). A Bonferroni multiple testing threshold (0.05/384 = 0.00013) was used to obtain an experiment-wise significance threshold of α = 0.05.

For analysis of the individual F_1 progeny and parents, the signal intensities of the parents and F_1 individuals were background corrected and normalized across slides by regression on the spot intensity data for parent ZG14 (Lezar et al. 2004). The normalized signal intensity values were used for identification of polymorphisms among F_1 individuals (Lezar et al. 2004).

DArT markers that differed significantly between the two bulks were evaluated in each of the 21 F_1 progeny constituting the bulks and in parent ZG14 using the single-dye (Cy3) experiments reported earlier (Lezar et al. 2004). Polymorphic DArT fragments were identified in Microsoft Excel based on the bimodal distribution of their intensity values across slides, consistent with their segregation as dominant PCR-based testcross (Aa:aa = 1:1) or intercross (A:aa = 3:1) markers. Each polymorphic DArT marker was scored for the presence (AA or Aa) or absence (aa) of the fragment in the genome representations of F_1 progeny and parent ZG14. The observed segregation ratios in the F_1 progeny were used to select a subset of markers with significant signal intensity differences between the two bulks and that were exclusively present in tolerant or susceptible F_1 progeny.

Sequencing of DArT markers

Nine DArT fragments that were putatively associated with tolerance were sequenced and subjected to similarity searches to reveal the possible identity of each tolerance marker. The cloned DNA fragments were recovered from *Escherichia coli* TOP10F' host cells harboring the genomic library. Recombinant *E. coli* clones were grown overnight at 37°C in LB medium containing 50 µg/ml ampicillin. A 1.0 µL aliquot of this solution was used for PCR with the M13 forward (-20) and M13 reverse primers (Invitrogen) as described by Lezar et al. (2004). Amplified fragments were then column-purified (QIAquick PCR purification Kit, QIAGEN GmbH) for sequencing on an ABI 3100 DNA sequencer (BigDye v3.2, Applied Biosystems). All fragment sequences were subjected to similarity searches using BLASTN and BLASTX.

CAPS marker design and analysis

For all nine sequenced DArT fragments, genome walking was carried out in the forward direction (across the 3'end of the cloned fragment) and the reverse direction (across the 5'end of the fragment) in two susceptible and two resistant individuals as described by Siebert et al. (1995). Each successfully cloned genome walking product was evaluated for the presence of a *Pst*I restriction site plus approximately 200 nucleotides beyond the restriction site. Eight pairs of CAPS

primers (Table 1) were designed using the Primer Designer 4 package (Version 4.2, Scientific and Educational Software, Cary, NC). The CAPS fragments were amplified in 20 µl containing 0.4 µM of each forward and reverse primer, 0.25 mM of each dNTP, 1X reaction buffer (Roche Diagnostics GmbH), 0.8 U *Taq* polymerase (Roche), and 6 ng template DNA. The PCR amplification consisted of 10 cycles of denaturation at 94°C for 30 sec, primer annealing at 63°C (reduced by 1°C per cycle) for 30 sec, and primer extension at 72°C for 1 min; 25 cycles of denaturation at 94°C for 20 sec, primer annealing at 53°C for 30 sec, and primer extension at 72°C for 1 min; with an initial denaturation of 94°C for 2 min, and a final extension of 68°C for 15 min. Following amplification, 12 µl of the reaction mix were transferred to a new tube and digested with 5 U of *Pst*I. Digested amplification products were resolved in 2.5% agarose gels. The marker phenotype of each CAPS marker was determined in parent ZG14, the 21 F_1 progeny used for BSA and in the two bulks. The CAPS markers were further tested in an additional 21 susceptible and 20 tolerant F_1 progeny of the same cross that also exhibited reliable disease phenotypes.

RESULTS

Canker assessment

After 11 weeks, inoculations on the parental and F_1 trees resulted in stem lesions of variable length. The F_1 progeny differed significantly in lesion length producing a phenotypic distribution with non-overlapping extremes (Figure 1). This result was expected since the parent trees were known to differ in tolerance. Only the most susceptible and most tolerant trees with consistent data over three consecutive years were selected for further analysis using microarrays.

Identification of DArT markers associated with tolerance

Sixty-two of the 384 array features showed some amount of differentiation between the contrasting bulks. Out of the 62 polymorphic features identified, 23 (6.0%) had significantly higher signal intensities in the tolerant bulk, while 39 (10.2%) had significantly higher intensities in the susceptible bulk at the Bonferroni-corrected threshold. Putative DArT markers differentiating the two bulks were evaluated in the 21 individuals included in the bulks (data available in Lezar et al. 2004, plus four additional individuals).

The CLUSTER and TreeView software (Eisen et al. 1998) allowed visualization of the relationships of the bulk and individual hybridization profiles (Figure 2). Two unique sub-clusters were identified that were associated with

canker tolerance. Sub-cluster I contained five DArT markers that were absent in all tolerant and present in all susceptible individuals (Figure 2), while sub-cluster II contained four markers that were absent in all susceptible and present in all tolerant *Eucalyptus* individuals. Therefore, only 9 (2.3% of 384) of the 62 polymorphic features were present in all individuals of one bulk and absent in all individuals of the other. They were consequently targeted for CAPS marker development (Table 2).

Sequencing of microarray markers associated with tolerance

Similarity searches based on the sequenced DArT fragments revealed that four fragments (78, 167, 189 and 341) were not homologous to any sequences in public databases. However, one fragment (4) showed weak similarity to *Arabidopsis* disease tolerance protein RPP8. Fragment 39 showed similarity to *Lotus japonicus* genomic DNA, while fragments 73 and 312 showed similarity to an *Arabidopsis thaliana* CTP-synthase-like protein and *Solanum tuberosum* mRNA for putative external rotenone-insensitive NADH dehydrogenase respectively. All E-values showed weak similarities.

CAPS marker development

Sequence data obtained from the cloned genome walking fragments revealed that for eight of the nine DArT markers, there were base-pair or indel polymorphisms

among susceptible and tolerant F_1 progeny within the *Pst*I restriction sites at the 3' end of the original DArT fragments. One fragment had a polymorphism in the 5'*Pst*I restriction site. Specific primers were designed to amplify DNA fragments containing the *Pst*I restriction site plus at least 200 nucleotides beyond the site (Table 1). We were able only to amplify eight of the nine CAPS fragments (results not shown). Two of the eight primer sets were successfully used in CAPS marker assays. Primers F178 and R178 amplified a monomorphic band of 600 bp in all parents an F_1 progeny, as expected from the sequence data. After restriction digestion of the fragment with *Pst*I, it was possible to differentiate between susceptible and tolerant individuals (Figure 3 A). A 600 bp undigested band was present in all susceptible F_1 individuals, whereas fragments of 500 bp and 100 bp were observed in all tolerant F_1 individuals. Three susceptible individuals (45, 47, 50) were heterozygous for the CAPS marker. Amplification of DNA with primers F312 and R312 resulted in a monomorphic band of 700 bp. Subsequent restriction digestion with *Pst*I produced a 600 bp and 100 bp band in the 10 susceptible individuals and a 700 bp undigested band in the 11 disease tolerant individuals (Figure 3B).

Primers for DArT markers 352, 4, 189, 73, 167, 39 were also tested in susceptible and tolerant individuals, but the PCR products were found not to be useful for CAPS analysis. The primers 4, 39 did not identify single loci, but produced multiple bands. To overcome this problem, annealing temperatures were increased. However, this resulted in loss of amplification. Primers 352, 189, 167,

73 amplified monomorphic bands, but showed no polymorphisms between individuals after restriction digestion when run on agarose gels (results not shown).

DISCUSSION

In this study, bulk segregant analysis (BSA, Michelmore and Paran, 1991) and DArT enabled us to identify restriction polymorphisms associated with tolerance to stem canker caused by *Chrysoporthe austroafricana* in a segregating full-sib family of *E. grandis* trees. A DArT genotyping microarray was previously prepared by selective PCR amplification of short restriction fragments derived from total genomic DNA of 24 *Eucalyptus grandis* trees onto a glass slide (Lezar et al. 2004). The multi-locus marker genotypes of tolerant and susceptible bulks, as well as individual genotypes of the progeny, were determined using a 384-feature prototype DArT genotyping array. Direct comparison of signal intensities revealed that of a total of 109 scorable, polymorphic loci, nine appeared to be associated with tolerance or susceptibility.

The combination of BSA and DArT is an efficient approach to rapidly identify genomic markers linked to traits of interest. Up to 10 000 DArT fragments can be printed on a single array, which allows very high genome coverage for gene-tagging in segregating families, or even in natural populations. The observed number of putatively associated markers detected with BSA (nine) corresponded

well with that predicted for one or two major tolerance genes. If we assume that linkage can be detected within a third of a chromosome, then 11.6 (384/33) of the markers could be expected to be linked to a single tolerance locus. Of the 384 DArT loci assayed, nine (2.3%) were found to have very different signal intensities in the resistant and susceptible bulks.

Several features of DArT make it attractive for genome-wide fingerprinting of *Eucalyptus* clones for the identification of trait-linked markers. This technique, like AFLP, allows genomic fingerprinting of organisms such as *Eucalyptus* with no prior DNA sequence information (Jaccoud et al. 2001, Wenzl et al. 2004). More importantly, the DNA for hybridization is prepared by amplification of short restriction fragments derived from throughout the genome using generic adapter primers. This eliminates the need for single PCR amplifications with different primers followed by restriction digestion as is used in CAPS (Paran et al. 1993) or PCR-RFLP approaches (Williamson et al. 1994), and allows a very high multiplexing ratio. The technology is, however, fairly expensive. This limits its usefulness for routine marker analysis of large numbers of trees. However, the fact that all DArT fragments are cloned before printing, allows the very quick conversion of DArT markers to simple, co-dominant PCR-RFLP based markers. PCR-RFLP based markers are less expensive to use for routine screening and marker-assisted breeding and more informative than DArT markers at the single marker level.

In this study, BSA revealed a total of 109 scorable, polymorphic loci, of which nine appeared to be associated with tolerance or susceptibility. We were able to successfully convert two of the nine DArT polymorphisms to co-dominant CAPS markers. These co-dominant markers provide a valuable tool for the accurate assessment of the genotype at the linked locus in mostly heterozygous forest trees. The two CAPS markers co-segregated perfectly with the DArT markers from which they were derived. However, failure of the enzyme *Pst*I to digest the amplified fragment would result in the incorrect scoring of genotypes. The activity of the enzyme should, therefore, be monitored by including known resistant and susceptible controls in marker assays. The remaining seven of the putatively associated DArT markers were not successfully converted to CAPS markers. We were not able to amplify one of the CAPS fragments, while the primers designed for two of these markers did not identify single loci. The other four CAPS markers that amplified the same size fragment from all individuals failed to digest with *Pst*I. We are investigating the design of the primers and identity of the PCR fragments.

The power of DArT fingerprinting combined with BSA lies in its ability to track the segregation of specific genomic regions inherited from the two parents in outcrossed pedigrees. Borevitz et al. (2003) reported the use of oligonucleotide arrays to demarcate recombination events along chromosomes of recombinant inbred lines of *Arabidopsis*. Unfortunately, genetic map information is not available for the cross of *E. grandis* parents used in this study and internal sequences of probes could not be used to link polymorphisms to a genome

sequence. Nonetheless, we found that the clustering of the DArT markers according to similar hybridization patterns across individuals suggested the presence of major linkage groups. Clustering of DArT markers allowed us to identify two unique subclusters that corresponded to Chrysoporthe canker tolerance or susceptibility.

The approach used in this study to identify microarray markers associated with tolerance to stem canker in *Eucalyptus* caused by *Chr. austroafricana*, should be applicable to other traits and other plant species. The long-term objectives of the research presented in this study is to analyze linkage relationships in a larger set of F_1 progeny and to associate the two markers identified with tolerance in *Eucalyptus grandis* breeding populations at a genome-wide level.

REFERENCES

Alfenas AC, Hubbes M, Couto L (1982) Effect of phenolic compounds from *Eucalyptus* on the mycelial growth and conidial germination of *Cryphonectria cubensis*. Canadian J Bot 60: 2535-2541.

Alfenas AC, Jeng R, Hubbes M (1983) Virulence of *Cryphonectria cubensis* on *Eucalyptus* species differing in resistance. Eur J Forest Path 13: 179-205.

Brondani RPV, Brondani C, Grattapaglia D (2002) Towards a genus-wide reference linkage map for *Eucalyptus* based exclusively on highly informative microsatellite markers. Mol Genet Genomics 267: 338-347.

Florence EJM, Sharma JK, Mohanan C (1986) Stem canker disease of *Eucalytpus* caused by *Cryphonectria cubensis* in Kerala. Kerala Forest Research Institute Scientific Paper 66: 384-387.

Gibson IAS (1981) A canker disease of *Eucalyptus* new to Africa. FAO, Forest Genetics Resources Info 10: 23-24.

Gryzenhout M, Myburg H, van der Merwe NA, Wingfield BD, Wingfield MJ (2004). *Crysoporthe*, a new genus to accommodate *Cryphonectria cubensis*. *Studies in Mycol* 50: 119-142.

Hodges CS, Geary TF, Cordell CE (1979) The occurrence of *Diaporthe cubensis* on *Eucalyptus* in Florida, Hawaii and Puerto Rico. Plant Disease Rep 63: 216-220.

Hodges CS, Alfenas AC, Ferreira FA (1986) The conspecificity of *Chrysoporthe cubensis* and *Endothia eugeniae*. Mycologia 78: 343-350.

Konieczny A, Ausubel F (1993) A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR based markers. Plant J 4:403–410.

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.

Micales JA, Stipes RJ, Bonde MR (1987) On the conspecificity of *Endothia eugeniae* and *Cryphonectria cubensis*. Mycologia 79: 700-720.

Michelmore RW, Paran L, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions using segregating populations. Proc Acad. Sci USA 88: 9828-9832.

Myburg AA, Griffin AR, Sederoff RR, Whetten 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.

Okinaka Y, Yang C-H, Perna NT, Keen NT (2002) Microarray profiling of *Erwinia chrysanthemi* 3937 genes that are regulated during plant infection. Mol. Plant-Microbe Interact 15: 619-629.

Old KM, Murray DIL, Kile GA, Simpson J, Malafant K (1986) The pathology of fungi isolated from eucalypt cankers in south-eastern Australia. Australian Forest Res 16: 21 – 36.

Paran I, Michelmore RW (1993) Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. Theor Appl Genet 85: 985-993.

Roux J, Wingfield MJ, Coutinho TA, Bouillett JP, Leigh P (1999) Diseases of plantation *Eucalyptus* in the Republic of Congo. SA J Science 96: 454-456.

Roux J, Myburg H, Wingfield BD, Wingfield MJ (2003) Biological and phylogenetic analyses suggest that two *Cryphonectria* spp. cause cankers of *Eucalyptus* in Africa. Plant Disease 87: 1329-1332.

Sharma JK, Mohanan C, Florence EJM (1985a) Disease survey in nurseries and plantations of forest tree species grown in Kerala. Research Report 36, Kerala Forest Research Institute, India.

Sharma JK, Mahanan C, Florence EJM (1985b) Occurrence of *Cryphonectria canker* of *Eucalyptus* in Kerala, India. Annals Appl Biol 106: 265-276.

Siebert PD, Chenchik A, Kellogg DE, Lukyanov KA, Lukyanov SA. (1995) An improved method for walking in uncloned genomic DNA. Nucleic Acids Res 23:1087–1088.

Van der Westhuizen IP, Wingfield MJ, Kemp GHJ, Swart WJ (1992) Comparative susceptibility of *Eucalyptus grandis* clones and hybrids to *Cryphonectria cubensis*. In: Proceedings of the 30th Congress of the South African Society for Plant Pathology, Cintsa, East London, 23 – 26 January, South Africa.

Van Heerden SW, Wingfield MJ (2001) Genetic diversity of *Cryphonectria cubensis* isolates in South Africa. Mycol Res 105: 94–99.

Van Heerden SW, Wingfield MJ (2002) Effect of environment on the response of *Eucalyptus clones* to inoculation with *Cryphonectria cubensis*. Forest Pathol 32: 395 – 402.

Van Zyl LM, Wingfield MJ (1999) Wound response of *Eucalyptus* clones after inoculation with *Cryphonectria cubensis*. Eur J Forest Pathol 29: 161 – 167.

Vos P, Hogers R, Blecker M, Reyans 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 Res 23: 4407-4414.

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(26): 9915-9920.

Williamson V M, H o J-Y, W u F F, M iller N , K aloshian I (1994) A P CR-based marker tightly linked to the nematode resistance gene, *Mi*, in tomato. Theor Appl Genet 87: 757-763.

Wingfield MJ, Swart WJ, Abear BJ (1989) First record of Cryphonectria canker of *Eucalyptus* in South Africa. Phytophylactica 21: 311-313.

Wingfield MJ (1990) Current status and future prospects of forest pathology in South Africa. S A J Science 86: 60-62.

Wingfield M J (2003) Daniel M cAlpine M emorial Lecture. Increasing threat o f diseases to exotic plantation forests in the Southern Hemisphere: lessons from Cryphonectria canker. Australasian Plant Path 23: 133-139.

Table 1. List of primers developed for evaluation of CAPS markers

Spot Number	Primer Name[a]	Sequence 5' 3'
312	**F312**	CATGGTGGGTGTATAGT
312	**R312**	GGTTCCGAAGCTTGAGTAG
178	**F178**	CCAGCTGCTGATGGAT
178	**R178**	TTACCTCTCTCGGCATCT
352	**F352**	TGGGCTCTTCTTCTCC
352	**R352**	CGGCGGAAGGGTAAAA
4	**F4**	CTGCAGTGGGGAAAG
4	**R4**	CTAAGACGCGTTGCACGA
73	**F73**	CAGTGATGCCAGTGTA
73	**R73**	ATCCGCCGGAAGGGTCAAA
189	**F189**	TGCAGTACGGAGGCTTGT
189	**R189**	CCGAGGATCTTGAATCTC
167	**F167**	TCCAGTGCAGTTGCCAACA
167	**R167**	GAGCAGTAACCGTTGAC
39	**F39**	CTGCCTGATGCAGACA
39	**R39**	TACTGTCGGACGGGCAGAT

[a] The forward primers (F) were used for genome walking. The reverse primers (R) were designed based on the genome walking products.

Table 2. Putative markers for tolerance to *Eucalyptus* stem canker caused by *Chr. austroafricana*, based on signal intensity ratios and segregation patterns in F_1 progeny

MARKER IDENTIFICATION SPOT NO	Ratio (Tolerant v. Susceptible)	*P*-value	Individual Data	
			Tolerant	Susceptible
4	2.402	0.00007	11	10
39	2.257	0.00001	11	10
73	2.033	0.00005	11	10
78	0.451	0.00001	10	11
167	0.380	0.00000	10	11
189	0.422	0.00000	10	11
312	0.472	0.00000	10	11
341	3.757	0.00000	11	10
352	4.955	0.00000	11	10

Figure 1. Mean lesion length (mm) of 61 F_1 progeny with consistent disease phenotypes after artificial inoculation with *Chrysoporthe austroafricana* for 3 consecutive years (1997-1999). The results are based on one replicate per tree and three annual measurements. Standard deviations are indicated by error bars. Additional trees genotyped with CAPS markers are indicated by *.

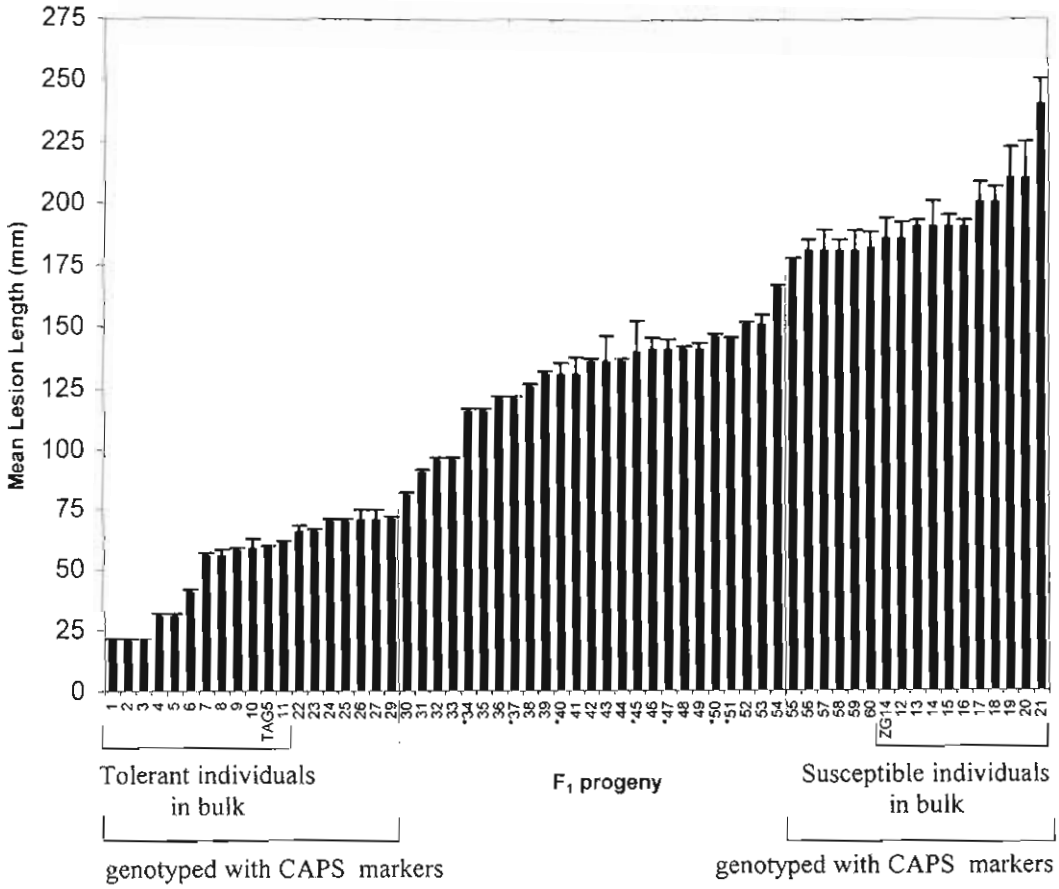

Mean Lesion Length (mm)
275
250
225
200
175
150
125
100
75
50
25
0
TAG5
ZG14
Tolerant individuals in bulk
genotyped with CAPS markers
F1 progeny
Susceptible individuals in bulk
genotyped with CAPS markers

Figure 2. TreeView (Eisen et al. 1998) representation of clustering of hybridization profiles among 21 *Eucalyptus* individuals based on microarray analysis with the 384-probe array. Columns represent hybridization profiles of individuals (or replicates) and rows represent the mean log intensities for labeled DNA/DNA hybridisations across the parent ZG14 and the individuals of the full-sib progeny. *Green* bars and *red* bars indicate high and low mean log intensity values and *black* bars indicate intermediate values. The numbers I and II represent sub-clusters identified associated with canker tolerance. Sub-cluster I contains DArT markers that were absent in all tolerant and present in all susceptible individuals, while sub-cluster II contained markers that were absent in all susceptible and present in all tolerant *Eucalyptus* individuals.

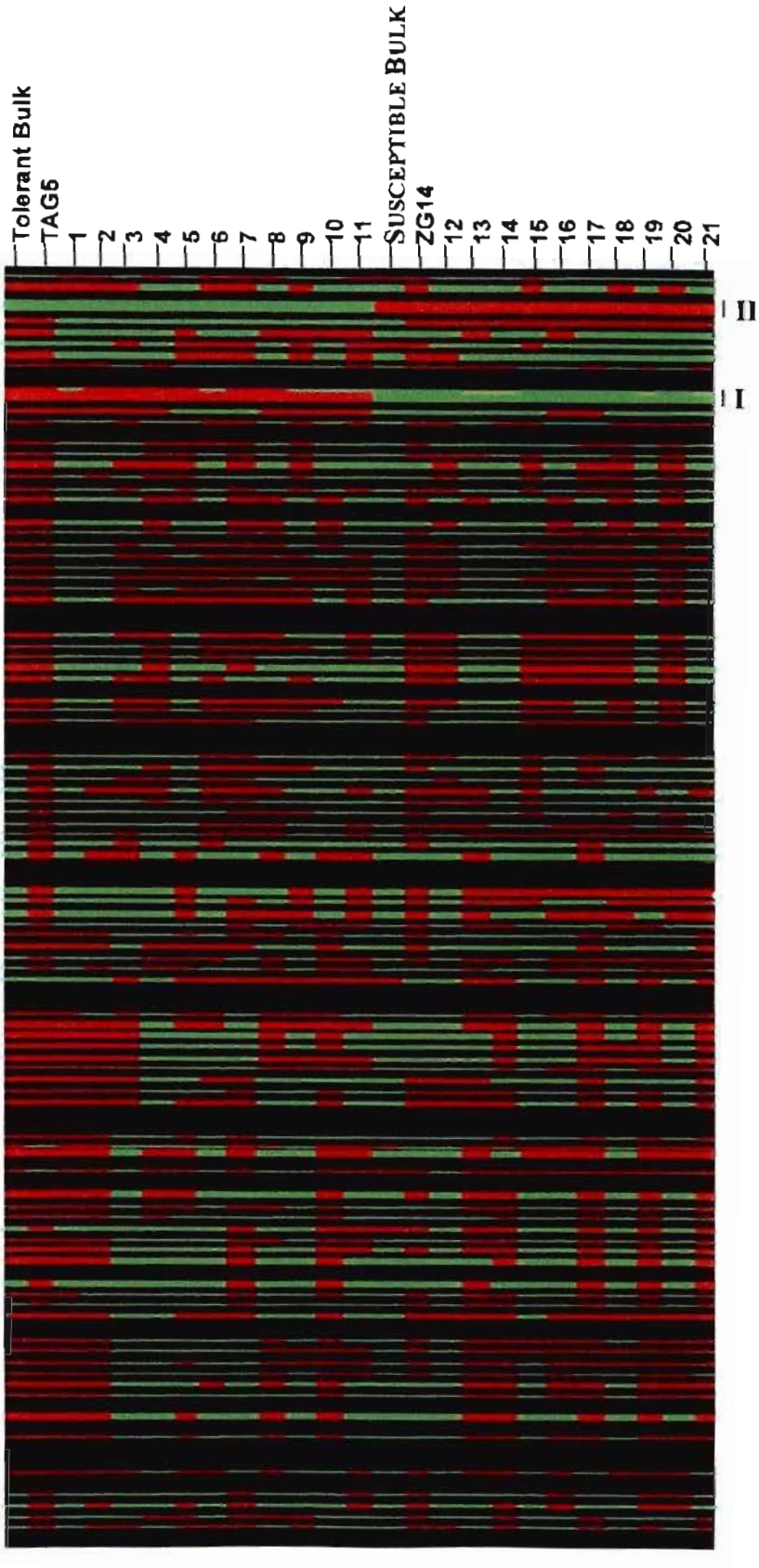
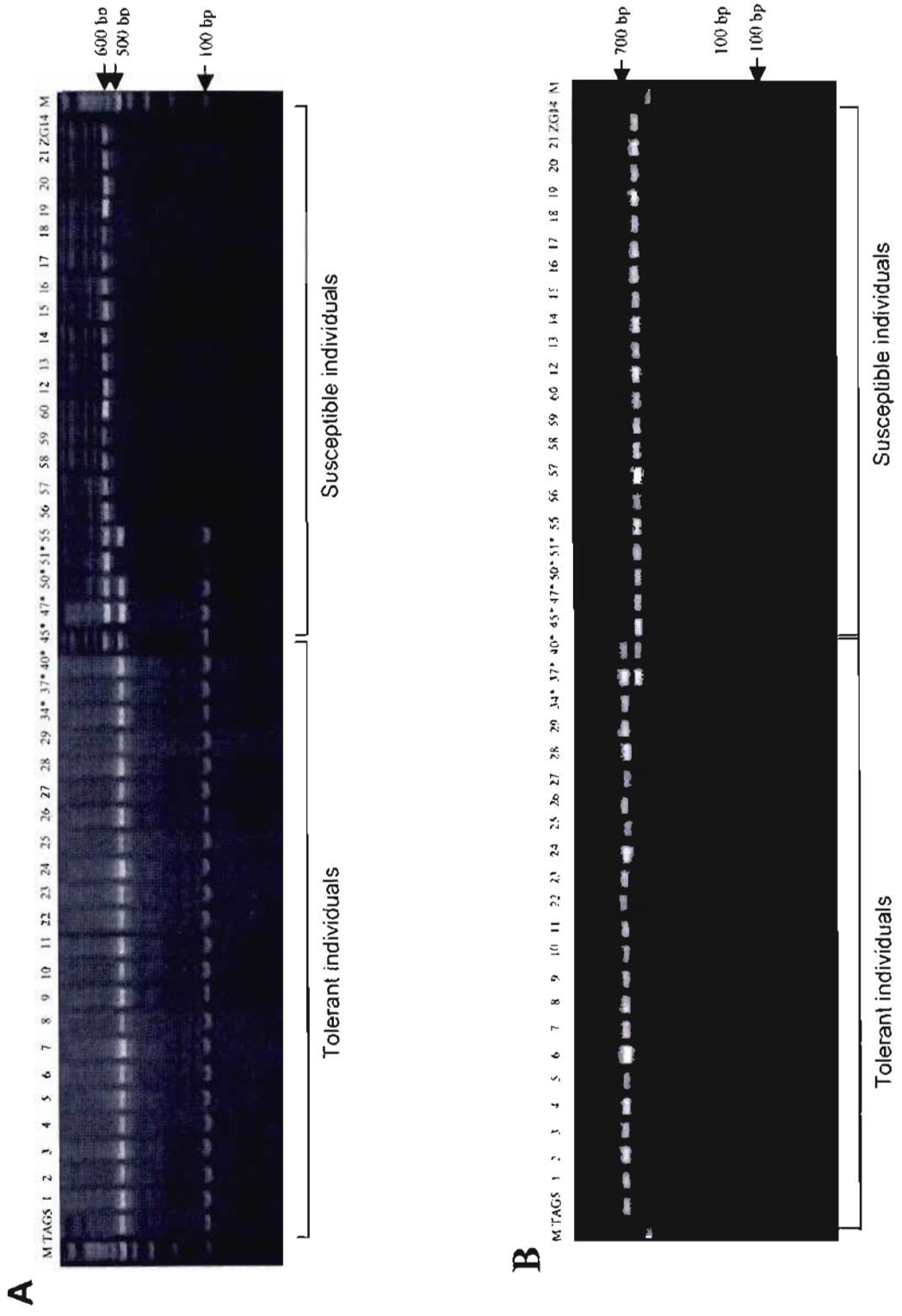


Figure 3. Electrophoretic pattern of PCR-amplified DNA products digested with *Pst*I enzyme among F_1 derivatives of *Eucalyptus grandis*. (A) DNA fragments amplified with primers F78 and R78. Trees TAG5 – 40: F_1 derivatives of parent ZG14 and TAG-S are all tolerant, whereas trees 45 – ZG14 are susceptible genotypes. (B) DNA fragments amplified with primers F312 and R312. Trees TAG5 – 40 are tolerant genotypes and lanes 45 – ZG14 are susceptible genotypes. F_1 progeny with intermediate disease phenotypes after artificial inoculation with *Chrysoporthe austroafricana* are indicated by *.


A
600 bp
500 bp
100 bp
M TAGS 1 2 3 4 5 6 7 8 9 10 11 22 23 24 25 26 27 28 29 34* 37* 40* 45* 47* 50* 51* 55 56 57 58 59 60 12 13 14 15 16 17 18 19 20 21 ZG14 M
Susceptible individuals
Tolerant individuals
B
700 bp
100 bp
100 bp
M TAGS 1 2 3 4 5 6 7 8 9 10 11 22 23 24 25 26 27 28 29 34* 37* 40* 45* 47* 50* 51* 55 56 57 58 59 60 12 13 14 15 16 17 18 19 20 21 ZG14 M
Susceptible individuals
Tolerant individuals

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.

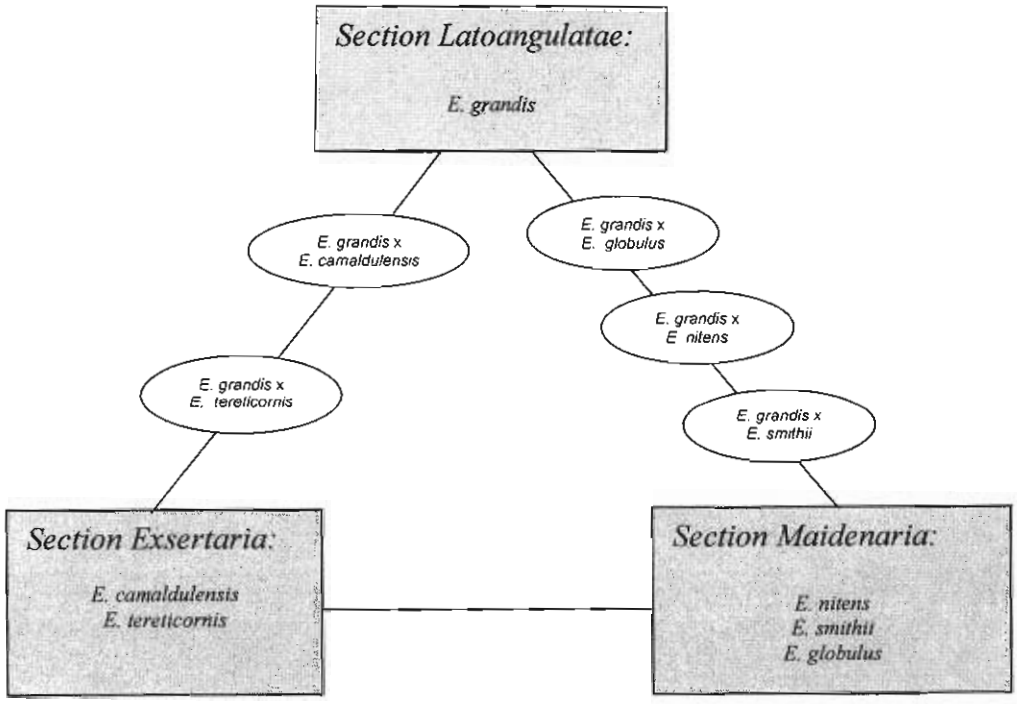
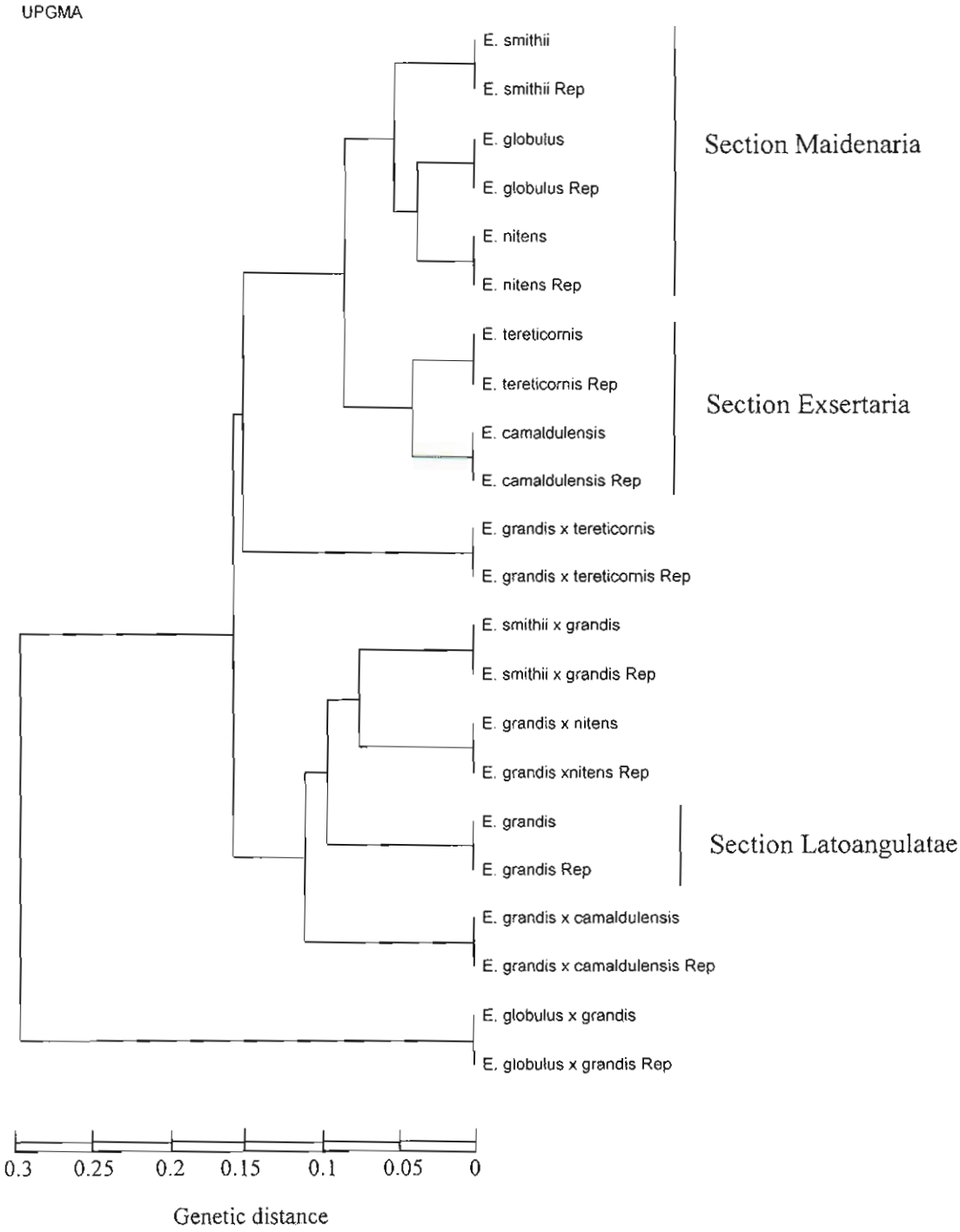

Section Latoangulatae:
E. grandis
E. grandis x E. camaldulensis
E. grandis x E. globulus
E. grandis x E. nitens
E. grandis x E. tereticornis
E. grandis x E. smithii
Section Exsertaria:
E. camaldulensis
E. tereticornis
Section Maidenaria:
E. nitens
E. smithii
E. globulus

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