GENETIC MAPPING OF QUANTITATIVELY INHERITED
ECONOMICALLY IMPORTANT TRAITS IN EUCALYPTUS
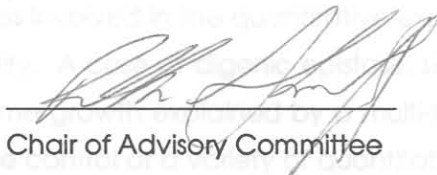
by

DARIO GRATTAPAGLIA

A thesis submitted to the Graduate Faculty of
North Carolina State University in partial fulfillment of the
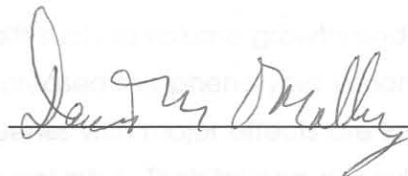requirements for the Degree of
Doctor of Philosophy
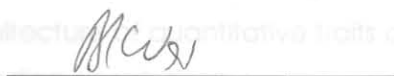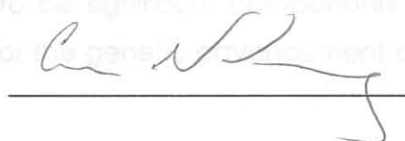
GENETICS and FORESTRY

Raleigh

1994

APPROVED BY:

_____
Chair of Advisory Committee

_____

_____

_____

ABSTRACT

GRATTAPAGLIA, DARIO. Genetic Mapping of Quantitatively Inherited Economically Important Traits in *Eucalyptus*. (Under the direction of Ronald R. Sederoff).

Genetic linkage maps of molecular markers offer a powerful tool to investigate the genetic architecture of quantitatively inherited traits and potentially assist in breeding, through the identification and manipulation of the individual genetic factors controlling such traits. Little is known about the genetic architecture of quantitative traits in natural populations of forest trees. In addition, tree breeding could be accelerated by marker assisted selection. The objective of the research was to detect and investigate the nature of discrete Quantitative Trait Loci (QTLs) controlling the variation in quantitatively inherited traits of economic importance in *Eucalyptus*. In a cross between undomesticated, highly heterozygous trees of *E. grandis* and *E. urophylla*, a ' pseudo-testcross' mapping strategy and RAPD (Random Amplified Polymorphic DNA) markers were used to construct two single-tree genetic linkage maps. This strategy is widely applicable to outbred heterozygous species. QTL mapping within-family was performed using two-generation pedigrees of full and half-sib families. This approach explores the within-family linkage disequilibrium and requires that the parent tree be heterozygous both at the QTL and linked marker. QTLs controlling traits related to the ability of vegetatively propagate trees were mapped in a full-sib cross. Effects of gene substitutions and multipoint estimates of the total phenotypic variation explained by the QTLs indicate that relatively large numbers of genes control the traits investigated, however major effect QTLs for rooting and micropropagation response were identified. A retrospective QTL mapping study was performed using an existing maternal open pollinated half-sib family of *Eucalyptus grandis*. Results indicate the existence of major genes involved in the quantitative expression of traits such as volume growth and wood quality. A case of digenic epistasis, significantly increased the phenotypic variance for volume growth explained by a multi-QTL model. Genes with major effects are involved in the control of a variety of quantitative traits in *Eucalyptus*.. Their linkage disequilibrium to markers should allow following these effects in generations of breeding and selection. Epistatic interactions amongst QTLs might prove to be significant components of the architecture of quantitative traits and important for the genetic advancement of tree breeding populations.

PERSONAL BIOGRAPHY

Dario Grattapaglia was born at the feet of the Alps in the city of Torino, Italy in a mild morning on June 18 1962. The future geneticist was born from an uncommon hybridization event at that time: a German mother and an Italian father, brought together by the international language Esperanto. Dario was born four years after his brother Guido and was the second and last son of Ursula and Giuseppe. He was raised in an international environment spending summers with his relatives in Germany and following his parents in Esperanto Congresses throughout Europe. Looking for a more ecological and meaningful life, Dario's parents moved to Brazil in 1974. The Grattapaglia's life style was radically changed when they went on to live in the inland high plains of Brazil close to the town of " High Paradise" - on parallel 15° in the Farm and School "Bona Espero" , " Good Hope" in Esperanto. To continue his studies Dario went to Brasilia where he completed high school in 1979. In 1980 he came to the U.S. as an exchange student and received a diploma from Teaneck High School, NJ. In 1981 Dario started college. Forest Engineering seemed to have the right blend of math and biology, Dario's favorite subjects. His fascination for Genetics, Evolution and Ecology led him to work as teaching assistant in these subjects and engage in research in breeding, conservation and biochemical systematic of native *Manihot* species. Graduated Forest Engineer in 1985, Dario went to work as a junior scientist in a newly established Plant Biotechnology Company in Campinas, Brazil. He worked with tissue culture, propagation and breeding of fruit and forest trees. It was during that time that he was attracted by the amazing growth and adaptation of eucalypt hybrids. In 1986 Dario was lucky enough to marry Tânia, a wonderful woman in every aspect. In 1990 he received a doctoral fellowship from the Brazilian Ministry of Education, left his job and went to North Carolina State University to pursue a Ph.D. in Genetics and Forestry with Ron Sederoff & Co. in the Forest Biotechnology Group. Three months before graduation, Dario together with Tânia, completed his most remarkable genetic experiment: their son Rafael, another hybrid as his dad, was born on February 3.

ACKNOWLEDGMENTS

I wish to thank my advisor Ron Sederoff for the genuine friendship, unconditional support and thoughtful guidance throughout this work. Special thanks go to him and all the committee members, David O' Malley, Bruce Weir and Gene Namkoong for always keeping my "big picture" perspective of genetics. Thank you to all the members and ex-members of the Forest Biotechnology Group at NCSU in particular my big buddy José Chaparro for showing me the " RAPD way" and for the animated discussions, Ben H. Liu for helping me with the intricacies of QTL mapping; Ross Whetten for the excellent instructions in molecular biology; Susan McCord for her patience and support, and the international mapping crowd, Christophe Plomion, Barbara Crane, Phil Wilcox, Henry Amerson and Denny Werner for their friendship and suggestions. Thanks to Zhao-Bang Zeng (NCSU Statistics) for valuable discussions and helpful power simulations; Toby Bradshaw for the great friendship and scientific interactions and my friend Scott Tingey for the encouragement.

This work could not have been possible without the great support of a large number of people in Brazil: Fernando Bertolucci for the sincere friendship, great discussions and unconditional help in several aspects of this project; Yara Ikemori and Edgard Campinhos for believing in my ideas back in 1990 and the encouragement throughout the project; my friend Ricardo Penchel and his crew for the huge work in the wood quality determinations; the field crews at Aracruz and São Mateus; all my friends at the University of Viçosa especially Eduardo Campinhos, Acelino Alfenas and his excellent group of people for the logistic support during my trip to Brazil.

TABLE OF CONTENTS

III. NUCLEAR DNA CONTENT OF COMMERCIALLY IMPORTANT *Eucalyptus* SPECIES AND HYBRIDS

IV. MULTIPLE APPLICATIONS OF RAPD MARKERS TO GENETIC ANALYSIS IN *Eucalyptus sp.*

V. GENETIC LINKAGE MAPS OF *Eucalyptus grandis* AND *E. urophylla* USING A PSEUDO-TESTCROSS MAPPING STRATEGY AND RAPD MARKERS

VI. GENETIC MAPPING OF QUANTITATIVE TRAIT LOCI CONTROLLING VEGETATIVE PROPAGATION IN *Eucalyptus grandis* AND *E. urophylla* USING A PSEUDO-TESTCROSS MAPPING STRATEGY AND RAPD MARKERS

VII. GENETIC MAPPING OF QUANTITATIVE TRAIT LOCI CONTROLLING GROWTH AND WOOD QUALITY TRAITS IN *Eucalyptus grandis* USING A MATERNAL HALF-SIB FAMILY AND RAPD MARKERS

LIST OF TABLES

CHAPTER VII

LIST OF FIGURES

CHAPTER VII

PREFACE

The Introduction of recombinant DNA technologies and the discovery of the Polymerase Chain reaction (PCR), have allowed a shift from genetic analysis based on the inference of genotypes from phenotypes, pioneered by Mendel, to methods based on the direct analysis of genetic variation at the level of DNA sequence. Genetic linkage maps of molecular markers offer a powerful tool to investigate the underlying genetic architecture of polygenic traits and potentially assist in breeding, by the identification and manipulation of the genetic factors controlling such traits on an individual basis.

The underlying difference between quantitative and qualitative characters is based on the magnitude of the effect of an allelic substitution at the genetic loci controlling the trait. If the ratio between the effect of a substitution and the total phenotypic variation is small, it is generally said that the trait is quantitative. Nevertheless, small allelic effects can also be observed even if there is a significant environmental influence. If, on the other hand, the effect of an allelic substitution at a single locus is large in relation to the total phenotypic variation, we say that the character is qualitative. Quantitative differences are usually, although not necessarily always, controlled by allelic substitutions at many loci. However, it is a basic premise of quantitative genetics that the inheritance of quantitative differences depends on genes subject to the same laws of transmission and having the same general properties as the genes whose transmission and properties are displayed by qualitative differences. The use of molecular markers for the study of quantitative inheritance essentially rests on this same premise inasmuch as it attempts to describe the continuous variation observed, in terms of its individual qualitative components.

The theory and techniques of quantitative genetics have been powerful tools for the study of polygenic inheritance. They have also been a fundamental tool for genetic progress in plant breeding. Nevertheless, the manipulation of complex traits continues to be a challenging task. Most of heritable traits of economic importance are the result of the joint action of several genes with unequal effects and variably influenced by the environment. The resulting phenotypes display a continuous variation instead of discrete phenotypic classes. Productivity, plant size and precocity are typical quantitatively inherited traits. For the vast majority of complex traits, very little information is available on their genetic architecture, i.e. the number, relative chromosomal location,

magnitude of effect and possible interactions of the genetic loci controlling the expression of the trait.

The main objective of the research presented herein was to shed some light on the nature of the discrete genetic units - Quantitative Trait Loci (QTLs) - controlling the observed variation in a number of quantitatively inherited traits of economic importance in *Eucalyptus.*. The first two chapters are introductory in nature and provide a current view on the existing knowledge in the area of molecular markers in plant genetics and breeding. Chapter I deals with the applications of molecular markers to plant breeding. It includes a brief history of the subject and a discussion of each application, focusing on the basic genetic principles involved, the experimental designs used for implementation, and on how the molecular information generated can be used by the breeder.

Several techniques of molecular biology are available today for the detection of genetic variability at the DNA level. These techniques allow the generation of a virtually unlimited number of molecular markers, which display Mendelian inheritance and cover whole genomes. The objective of Chapter II is to provide a general overview on the existing methods available for the generation of molecular markers specifically used for the genetic analysis in plant genetics and breeding. For each class of molecular marker we describe the origin, the genetic basis and the principles involved in detection. A comparative analysis of the advantages and limitations of each class of marker is then presented for applications in plant breeding. Finally, we describe some novel and extremely powerful techniques for the detection of DNA polymorphisms that open new perspectives for the analysis of complex genomes and potentially the isolation of economically important genes.

Chapters III trough VII represent original work dealing with genetic analysis of *Eucalyptus. Eucalyptus* (Myrtaceae) is a large genus of evergreen hardwood trees and shrubs that includes about 700 species. It is almost exclusively native to Australia, with a few species indigenous to adjacent islands. The genus *Eucalyptus* includes the most widely used tree species for plantation establishment in tropical and subtropical regions of the world. Today, eucalypts constitute the majority of the world's exotic hardwood forest and one of the main sources of woody biomass. Fast growth rates and a wide range in adaptability have contributed to the great interest that eucalypt species enjoy in many countries outside their native range. Besides the fast growth that allows for

shorter rotations, many species display wood properties that make them very suitable for fuel and charcoal production or pulp and paper manufacture.

Chapter III reports the nuclear DNA content estimates obtained by flow cytometry for a group of *Eucalyptus* species that includes those most widely planted throughout the world. In view of the importance of interspecific hybridization in the genus we also surveyed some of the most common hybrids used in plantation forestry. In Chapter IV, results are presented on a series of genetic analyses of individuals and populations carried out on *Eucalyptus* species using RAPD (Random Amplified Polymorphic DNA) genetic markers. These results describe the power, reliability, ease of use and multiplicity of applications of this new class of molecular markers for genetic studies in the genus *Eucalyptus* as well as in any forest tree species.

Chapter V describes the use of a "two-way pseudo-testcross" mapping strategy in combination with the RAPD assay to construct two moderate density genetic linkage maps for individual trees of *Eucalyptus*.. These are the first reported high coverage linkage maps for any species of *Eucalyptus* . The ability to quickly construct single-tree genetic linkage maps in any forest species opens the way for a shift from the paradigm of a species index map to the heterodox proposal of constructing several maps for individual trees of a population, therefore mitigating the problem of linkage equilibrium between marker and trait loci for the application of marker assisted strategies in tree breeding.

Chapter VI and VII focus specifically on the investigation of the genetic architecture of quantitative traits in *Eucalyptus*, using the linkage mapping information developed previously. QTL mapping was performed using two-generation pedigrees of full and half-sib families, exploring the within-family linkage disequilibrium. This approach requires that the parent tree be heterozygous both at the quantitative trait locus and linked marker locus, and results in individual-specific marker-trait associations .

In Chapter VI, the "pseudo-testcross" mapping strategy and RAPD markers are used to map QTLs controlling traits related to the ability of vegetatively propagate trees in *Eucalyptus*.. Results indicate that relatively large numbers of genes control the traits investigated, however major effect QTLs for rooting and micropropagation response were identified. Chapter VII reports the results of a retrospective QTL mapping study, using a maternal open pollinated half-sib family of *Eucalyptus grandis* at harvest age. This study also provides supporting evidence for the existence of QTLs of major effect involved in the quantitative expression of economically important traits related to forest

productivity in *Eucalyptus grandis*. Genetic analysis of continuous variation with molecular markers provides a powerful tool to study interactions amongst genes. Results in *Eucalyptus* suggest that epistasis might play an important role in the shaping of quantitative traits.

These findings shed some light on the architecture of quantitative traits in forest trees and may help direct future QTL mapping experiments in highly heterogeneous populations. Genes with major effects seem to be involved in the control of a significant proportion of the variation in a number of quantitative traits in *Eucalyptus*.. QTL analysis can now be contemplated in any available two-generation pedigree, opening the possibility of using existing families in retrospective QTL analyses. This approach should allow one to gather at once the quantitative data necessary to evaluate the potential of marker-assisted tree breeding on a case-by-case basis. However, several questions still need to be answered before marker assisted selection can be implemented at the operational level. We expect that our findings will foster research in a number of interesting and important issues related to QTL mapping such as the interaction of QTLs with environment and age, the stability of expression across generations and genetic backgrounds and the relative importance of epistasis in the advancement of populations through directional selection. Finally, the uncovering of major genes in essentially undomesticated populations, may have some important implications for understanding the basis of evolutionary change in adaptive traits in natural populations of forest trees.

1. INTRODUCTION

Plant breeding has been successfully practiced since the early days of human civilization. Genetic improvement can be achieved as long as: (1) there is accessible genetic variability, (2) the environment does not completely mask such variability and (3) selection and recombination of superior genotypes can be performed to establish the next generation. Traditionally, breeding progress has relied exclusively on the analysis of phenotypes. The success of this approach largely depends on the heritability of the trait involved. Polygenic inheritance, partial or complete dominance, environmental effects and the time necessary to complete a breeding cycle are some of the factors that frequently limit the efficiency of this approach.

Many of the limitations of phenotypic analysis could be alleviated by the direct identification of genotypes using a diagnostic system based on molecular markers that segregate with the traits of interest. Similarly, the use of molecular information in the analysis of genetic diversity and phylogenetic relationships of cultivated and undomesticated germplasm could facilitate the monitoring and expansion of the genetic base of breeding populations.

Plant breeding is carried out in a continuous and dynamic fashion. The product of a breeding program today, could become useful for a different objective tomorrow. Unpredictable changes in the ecological and economical environments, require managing breeding populations for multiple objectives (Namkoong *et al.* 1988). New traits are constantly being incorporated in cultivars or populations in different stages of the program. In this process, use has to be made of classical procedures of sexual recombination and, when possible, other means such as somatic hybridization, mutation breeding or genetic transformation. No single technique of genetic manipulation can resolve all the challenges of a breeding program. Plant breeding in the next decades will experience an increasing integration of biotechnology with classical techniques. In this context, molecular marker breeding technologies can contribute significantly to a better understanding of the specific crops and traits of interest and for the development of improved products.

This chapter deals with the applications of molecular markers to plant breeding. After a brief history of the subject, each application will be discussed, focusing initially on the basic genetic principles involved and the experimental designs typically used for implementation. Then, we will discuss how molecular information can be used by the

breeder. Each case will be illustrated with examples from the literature without attempting to provide an exhaustive review of the existing work in the area. The advantages and limitations of each type of molecular marker are discussed in a separate chapter. In the section dealing with the construction of genetic linkage maps and localization of regions controlling traits of interest, we present some emerging strategies for the conduction of such studies in outcrossed and essentially undomesticated forest and fruit tree species. The applications of molecular markers in basic studies of population and evolutionary genetics and genome organization have only been mentioned when justified in the context of plant breeding. An in depth discussion of this topic would constitute a further chapter. It is also beyond the scope of this chapter to discuss in detail the various statistical issues surrounding the analysis and interpretation of molecular breeding data.

2. HISTORICAL BACKGROUND

The twentieth century has experienced an indisputable revolution of the scientific thought in biology. The first year of this century witnessed the simultaneous rediscovery of Mendel's work by three independent people, De Vries, Tschermark and Correns (Sturtevant 1965). The development of population and quantitative genetics by Fisher, Wright and Haldane and their discussion of the evolutionary process, largely influenced the practice of plant breeding (see Provine 1987). The implementation of quantitative genetics methods by people such as Shull, Jones and East revolutionized the agricultural industry early in the century with the development of inbred lines and hybrid maize (Hallauer and Miranda 1988). The integration of several disciplines, particularly biochemistry, chemistry and microbiology culminated with the discovery of the DNA as the basic genetic material with the transformation experiments of Avery, MacLeod and MacCarty (1944) and by Hershey and Chase (1952). Subsequent work to decipher its structure and the multifaceted aspects of the genetic code, reported in the landmark papers by Watson and Crick (1953 a,b), resulted in the birth of molecular biology. In recent years, the continuous development of methods to clone and sequence nucleic acids, study gene expression, and analyze the enormous amounts of data generated, has offered unprecedented possibilities for the study of genetics of living organisms.

The use of genetic markers in plant breeding dates back to the beginning of the century when, in studies with peas, Bateson and Punnett (1905) indicated the possibility

of the existence of "reduplication" , later called genetic linkage, between genes controlling flower petal color and shape of pollen grain. J.B.S. Haldane would later prove the existence of genetic linkage in animals, but because of World War I, his work would only be published several years later (Haldane *et al.* 1915). In the meanwhile, the group led by Thomas Hunt Morgan was making great progress in this new area of genetics. Morgan was the first to clearly demonstrate an exception to one of Mendel's rules, that is that not all the genes segregate independently. Morgan postulated that such genes are linked in groups on the same string of material in which they reside, the chromosomes (Morgan 1910). One of his students, A. H. Sturtevant, pioneered the development of the first genetic linkage map in *Drosophila* (Sturtevant 1913). In plants, one of the first genetic maps to be developed was of maize, through the efforts of a prominent group of geneticists that included Emerson, Beadle, Rhoades and McClintock (Emerson *et al.* 1933). Tomato was the other plant species where early work was done in linkage mapping (MacArthur 1934).

The milestone for the use of markers in the practice of plant breeding was set by K. Sax (1923). He originally explored the association between a qualitative character, a marker, and a quantitative character for indirect selection, as he reported the association between seed coat color and seed size in common bean (*Phaseolus vulgaris*). Mather and Jinks (1982) summarized several cases where simply inherited markers were associated with continuously inherited variation. For example Rasmusson (1935) demonstrated linkage of flowering time in peas with a simply inherited gene for flower color. Everson and Schaller (1955) found morphological markers closely linked to a genomic segment affecting yield in barley. Extensive work in *Drosophila*, demonstrated the effects of individual chromosomes on quantitative traits (Mather and Harrison 1949). Thoday (1961) reiterated the possibility of using genetic markers for the selection of quantitative characters through the development of specific lines and segregating populations. He also suggested the use of a cross between homozygous lines differing in quantitative value as a general means to follow the effect of different genotypes on the expression of the quantitative character. These and other studies provided the framework of theory and observation on which the area of molecular breeding is based in present days. Some of the major contributors as pointed out by Stuber (1992) include Jayakar (1970), McMillan and Robertson (1974), Soller and Plotkin-Hazan (1977), Tanksley *et al.* (1982), Beckmann and Soller (1983, 1986) , Lebowitz *et al.*

(1987), Edwards *et al.* (1987) and Lander and Botstein (1989). Their specific contributions will be discussed later in this chapter.

The underlying difference between quantitative and qualitative characters is based on the magnitude of the effect of an allelic substitution at the genetic loci controlling the trait. If the ratio between the effect of a substitution and the total phenotypic variation is small, it is generally said that the trait is quantitative. Nevertheless, small allelic effects can also be observed even if there is a significant environmental influence. If, on the other hand, the effect of an allelic substitution at a single locus is large in relation to the total phenotypic variation, we say that the character is qualitative. Quantitative differences are usually, although not necessarily always, controlled by allelic substitutions at many loci. However, as pointed out by Falconer (1988) it is a basic premise of quantitative genetics that the inheritance of quantitative differences depends on genes subject to the same laws of transmission and having the same general properties as the genes whose transmission and properties are displayed by qualitative differences. The use of molecular markers for the study of quantitative inheritance and the practice of breeding for quantitative differences essentially rests on this same premise in that it attempts to describe the continuous variation observed in terms of its individual qualitative components.

The efficiency of marker assisted breeding procedures based on the joint segregation of the marker and the gene, is a function of their genetic map distance. Therefore the number of markers available for such studies becomes a critical variable. The advent of biochemical and molecular techniques based on the analysis of enzyme and DNA polymorphisms radically expanded the frontiers of this area of study as it mitigated the limitations of morphological phenotypes as markers both in terms of numbers available and their genetic properties.

Systematic dissection of quantitative traits using molecular markers was initially carried out using isozyme markers in tomato (Tanksley *et al.* 1982), and later in other crops such as maize oats and soybean (Stuber 1992). Several applications of molecular markers besides genetic analysis and improvement of economically important traits were immediately proposed including varietal identification, screening of genetic diversity in germplasm resources and protection of breeder's rights (Beckmann and Soller 1983, 1986; Tanksley 1983b). The rapid development of genetic linkage maps was quickly achieved for major crops such as maize (Helentjaris *et al.* 1986b) and tomato (Bernatzky and Tanksley 1986). Currently, linkage maps for many plant species are

available although at different levels of resolution (Table 2). Theoretical developments concerning experimental designs and biometrical tools have added more power for the use of molecular markers for the study and manipulation of quantitative traits (e.g. Soller *et al.* 1976; Edwards *et al.* 1987; Lander and Botstein 1989; Knapp *et al.* 1990; Zeng 1993a,b).

Today, the detection and mapping of genomic regions associated with quantitative traits has been achieved in a relatively large number of crops using different classes of molecular markers. The effective use of this approach in operational breeding programs is still justified on a case by case basis, and the potential of this technology in applied breeding has not yet been fully realized. The theoretical efficiencies of marker assisted selection and the cost/benefit relationships of this technology have been debated in the literature (Lande and Thompson 1990; Dudley 1993; Zehr *et al.* 1992). However, new and improved molecular marker technologies are constantly being developed that have radically improved our ability to generate molecular marker information (see chapter II). Some reviews are available on the subject of molecular markers in plant breeding, emphasizing genetic mapping of economically important traits (Michelmore and Hulbert 1987; Tanksley *et al.* 1989; Beckmann 1991; Stuber 1992; Dudley 1993).

3. APPLICATIONS OF MOLECULAR MARKERS TO PLANT BREEDING

In a review by Beckmann (1991) a list was presented covering the main applications of marker based methodologies in plant breeding. A modified and expanded version of that list (Table 1) outlines the discussion throughout this chapter.

Even before any association between one or more markers and any genetic locus of breeding interest is established, the availability of a battery of molecular markers allows one to undertake some important studies in a breeding program. The short term applications of molecular markers involve basically the identification and discrimination of genotypes. Down the analytical path, molecular markers allow one to quantify the existing genetic variability at the DNA sequence level, and correlate it with the variation in phenotypic expression in a procedure generally called genetic mapping. In the synthetic path, the molecular information acquired in the analytical phase is integrated to the classical methodologies of selection and recombination of genotypes as an additional tool to promote genetic progress. Finally, linkage maps of molecular markers have been important tools in physical mapping projects and have facilitated map-

based cloning approaches that could result in the isolation of useful genes for transgenic breeding.

Table 1. Main applications of molecular marker based methodologies in plant breeding.

=====

1. CLASSICAL BREEDING

- **Short term applications**
  - Identification of parental origin
  - Identification and protection of patented varieties and clones
  - Assignment of inbred lines to heterotic groups
  - Verification of line and hybrid purity
  - Verification of controlled crosses
  - Monitoring of outcrossing and selfing rates in forest trees seed orchards

- **Medium and long term applications**

- *Analytical path*
  - Construction of genetic linkage maps
  - Genetic mapping of Quantitative Trait Loci (QTL) of economical importance
  - Study of the architecture of quantitative traits (number, position, magnitude of effects and interactions of QTL's)
  - Exploration of homologous loci in other species or genera by syntenic or comparative mapping
  - Germplasm evaluation (diversity, classification, genetic distance and phylogeny)

- *Synthetic path*
  - Introgression of desired traits among cultivated and undomesticated germplasm
  - Guided selection and recombination of superior genotypes
  - Selection during development of inbred lines
  - Prediction of expected phenotypes
  - Indirect selection for traits difficult to evaluate (disease and pest resistance, environmental stresses, industrial performance traits)
  - Early selection of progeny in perennial species (within-family selection for clonal propagation, early parent selection)

2. TRANSGENIC BREEDING

- Fine structure mapping directed toward map-based cloning efforts of potential genes for transgenic manipulation

=====

The molecular marker literature includes an extensive number of studies which are analytical in nature, however very few cases have been reported where this technology has been effectively used in the generation of improved materials. The effective integration of molecular marker based approaches into operational breeding programs still represents a major challenge in this dynamically and rapidly evolving area of applied research.

3.1 Identification and verification of genotypes

Most individuals except clones, have a nucleotide sequence of DNA that make them unique and distinguishable from each other. The detection of such differences, either through enzyme or DNA based polymorphisms reveals a particular molecular pattern commonly called a "genetic fingerprint" that can be used for the unique identification and discrimination of individual genotypes. Tracking the inheritance of such marker patterns can also be used for the confirmation of the participation of a particular individual in the generation of progenies, an application commonly termed "paternity analysis, although it can also be used to verify the maternal seed parent.

In Citrus, a polyembryonic genus, isozymes have been used in breeding to distinguish zygotic embryos from apomyctic embryos derived from the nucellar tissue, and among the zygotic, those that were derived from the intended cross fertilization versus those arising through selfing (Soost et al. 1980). A similar procedure has been used to identify the genotype of rootstocks of trees that showed superior field performance in productivity and stress tolerance (Roose and Traugh 1988). Isozyme markers have also been usedto identify somatic hybrids in Citrus obtained through protoplast fusion (Grosser and Gmitter 1990). These studies were limited to wide crosses where the isoenzymatic polymorphism allowed genotype discrimination. DNA based markers offer a powerful tool for this kind of analysis, particularly RAPD (Random Amplified Polymorphic DNA) markers because they do not require prior probe development. Markers that are present in the putative father and absent in the mother are Indicative of the cross fertilized origin of the embryo. This strategy has been used for the confirmation of the interspecific and not parthenogenic originof dihaploid individuals in crosses of the genus Solanum (Waugh et al. 1992).

In forest genetics, isozyme markers have been extensively used to identify genotypes and infer patterns of hybridization and crossing among individuals (Adams 1983; Cheliak

et al. 1987). Outcrossing versus selfing rates in seed orchards can be estimated. The extent of recombination of selected individuals can be monitored to avoid excessive selfing or consanguineous matings that are generally detrimental in outcrossing species with high levels of genetic load. The analysis of multilocus genotypes at three to five isozyme loci for progeny arrays of 20 to 30 individuals for 20 to 30 open pollinated families is typically sufficient to estimate such rates under a mixed mating model. Furthermore, the use of individual specific alleles or rare alleles at the population level allows one to monitor: (1) the efficiency of supplemental mass pollination with superior pollen sources in seed orchards; (2) the contamination rates with foreign pollen to the orchard; (3) the fidelity of controlled crosses.

In breeding autogamous crops, markers can be used for the quantification of cross fertilization. In the production of hybrid seeds, involving open pollination between self-incompatible or male sterile lines, markers allow the estimation of contamination and seed quality control. All these applications are based on the ability to discriminate the paternal contribution to the seeds. The statistical power of discrimination is a function of the number of markers available and the allelic variability at each one of these loci. With the exception of cases involving rare alleles, complete genetic discrimination is rarely achieved with isozymes, in view of their limited number and low level of polymorphism. DNA based markers alleviate these limitations, however to different extents depending on their genetic properties. For example RAPD markers are limited in their power for paternity studies as well as for other studies that require the direct estimation of allelic and genotypic frequencies because they do not allow the discrimination of dominant homozygotes from heterozygotes.

By sampling several loci simultaneously, VNTR (Variable Number of Tandem Repeats) based markers (also called hypervariable loci), are particularly useful for the identification of genotypes. DNA fingerprinting using this class of markers has been used to distinguish rice cultivars (Dallas 1988); apple, blackberries and raspberries (Nybom *et al.* 1990); and autogamous and allogamous varieties of these species (Nybom and Schaal 1990a); to establish paternity in apple (Nybom and Schaal 1990b) and to identify dihaploid lines in *Beta vulgaris* (Schmidt *et al.* 1993). Though widely used for paternity tests in humans, the use of VNTR based markers in plants has been limited by the lack of adequate probes. Specific families of repeat elements in plants have been identified (see chapter II).

RAPD or AP-PCR have been used to identify and differentiate rice cultivars (Welsh and McClelland 1990), soybean (Caetano-Annoles *et al.* 1991a; Williams *et al.* 1992); barley (Weining and Langridge 1991); turf grass varieties (Caetano-Anolles *et al.* 1991b); *Brassica* (Hu and Quiros 1991); *Apium* (Yang and Quiros 1993); cocoa clones (Wilde *et al.* 1992); clones of *Eucalyptus* (Grattapaglia *et al.* 1992) and apple cultivars (Koller *et al.* 1992). In operational forest tree breeding, the identification of clones at the DNA level offers a very useful tool to resolve suspected mislabelings during the establishment of seed orchards and as a quality control system in operational clonal production plantations. RAPD markers have been used to evaluate the genetic stability of somatic embryos of *Picea mariana* (Isabel *et al.* 1993) and for the identification of somatic hybrids in potato (Baird *et al.* 1993).

AP-PCR has been used for the identification of hybrid parental lines in maize (Welsh *et al.* 1991). However, paternity determination in highly heterozygous species becomes difficult with RAPD markers. Due to the low information content per locus, dozens of markers with gene frequency for the presence of the band less or equal to 0.2 are necessary (P. Lewis personal communication; see also chapter 1). Co dominant markers such as RFLP's (Restriction Fragment Length Polymorphisms) and particularly those that display high levels of multiallelic polymorphism such as microsatellite based markers are powerful for studies that involve paternity assignment. However, the costs associated with the development of such markers are not justified exclusively for this application which is considered of less potential impact than genetic mapping.

In view of the high levels of polymorphism RAPD technology is particularly useful for identification of genotype-specific markers. RAPD fragments of this kind have been determined for tomato (Klein-Lankhorst *et al.* 1991b) *Brassica* (Quiros *et al.* 1991) and potato (Singsit and Ozias-Akins 1993). Species specific markers were obtained for tropical species of *Pinus*, native to Central America. In a procedure analogous to Bulk Segregant Analysis (Michelmore *et al.* 1991) for the targeted analysis of genetic linkage, equal amounts of genomic DNA of several individuals covering the geographical range of each *Pinus* species, were bulked into "Population Bulks" composing average genomes of the most representative DNA sequences in each species. With this strategy, RAPD markers unique to each species were identified and could be used to characterize the origin of particular remnant populations in terms of species composition, particularly in zones of hybrid swarms (Grattapaglia *et al.* 1992b).

The identification of a variety, line or hybrid at the DNA level has become an important tool to protect breeder's patent rights in countries where such laws exist and are enforced. DNA fingerprinting has been used in legal cases involving disputes of authorship (Smith 1989), and have been considered instrumental in supporting the continued investment in plant breeding by large companies in the seed business such as Pioneer HiBred with maize hybrid seeds (Smith and Chin 1992; Jondle 1992). In this context, four general criteria have to be satisfied by a molecular marker for varietal identification: distinctive intervarietal variation, minimal intravarietal variation, environmental stability and experimental reproducibility (Bailey 1983).

3.2 Survey of genetic diversity and distance in germplasm banks

Molecular markers provide a powerful way to generate a large amount of information on genetic diversity and phylogenetic relationships in the germplasm available to the breeder. In general, these data are obtained in a matrix format where a certain number of genotypes that can be varieties, lines or clones (rows), were genotyped for several hundred molecular markers that can be isozymes, RAPD, RFLP's, microsatellites etc. or combination of these. The large number of molecular markers available, generally allows an extensive sampling and coverage of the genome, without environmental influences and with high reproducibility. Most frequently, each marker is analyzed as a distinct and discrete phenotypic character, independent from the rest. The principle underlying the interpretation of this kind of data is simple: markers in common between genotypes represent genetic similarities, while markers not in common represent differences. Considering that different fragments could be allelic to each other and that linked markers segregate jointly, the premise of independence of characters is not entirely satisfied, and this become increasingly the case as more markers are used in the analysis. The violation of such premise, however, does not generally cause significant distortions in the results. Several statistical methods (parsimony, multivariate analysis, cluster analysis, etc.), coefficients of genetic distance or similarity (Nei's distance, Rogers, Jaccard etc.) and phenograms are used for the analysis and presentation of this kind of data. Software packages that implement these and several other related kinds of analysis are available on the market such as : PHYLIP (Phylogeny Inference Package) (J. Felsenstein, Dep. Genetics SK-50, U. Washington, Seattle WA 98195); PAUP (Phylogenetic Analysis Using Parsimony) (Swofford 1990); NTSYS

(F. James Rohl, Department of Ecology and Evolution, State University of New York Stony Brook, NY 11794) as well as some procedures of the SAS (Statistical Analysis Software) package (SAS Institute 1985).

Depending on the kind of molecular marker used, the interpretation of homology of fragments scored as being the same, may occasionally present some problems. For example, RFLP's are detected through hybridization with DNA probes. The fragments detected with such probes share a rather extensive but not always exact sequence homology. When RAPD markers are used, however, the sequence homology between the bands scored is limited to the terminal priming sites. The question in this case is whether the two RAPD fragments that are amplified by the same primer and have same length are homologous characters (originated form a common ancestor) or homoplastic (evolved independently in the population). The premise on which is carried out an analysis of genetic distance is that fragments of same electrophoretic mobility are in fact homologous in the sequences internal to the priming sites. This premise was verified by experiments where RAPD fragments were used as DNA probes back to blots of RAPD gels, where a positive hybridization signal indicated fragment relatedness of co migrating fragments in different individuals (Williams *et al.* 1992; Grattapaglia and Sederoff 1994). Furthermore, it is plausible that genetically close individuals have inherited a fragment derived from a common ancestor and not that the same fragment evolved independently (Tingey and DelTufo 1993). Thormann and Osborn (1992) and Thormann *et al.* (1993) verified that the results of genetic distance analysis with RFLP and RAPD markers were very similar among accessions within species of *Brassica*, but different at the interspecific level. An important technical detail in this respect is to ensure an efficient electrophoretic separation of the RAPD fragments to allow the detection of even small differences in the length, avoiding an erroneous interpretation of fragment similarity.

The molecular characterization of genetic diversity in the source of germplasm can help the breeder select progenitors of base populations to establish breeding programs. Such populations are established by crossing superior individuals or lines, frequently aiming at maximizing genetic distance and thus recombining genes or co adapted gene complexes in new combinations. Morpho-physiological traits can be used to establish such measures of genetic diversity (Rangel *et al.* 1990). However, molecular markers can generate a large number of additional characters that combined with phenotypic data provide a more complete picture for the grouping of genotypes and

planning of crosses. This approach is being used for the recombination of elite Eucalyptus clones to structure a composite elite breeding population (D. Grattapaglia and F. Bertolucci, personal communication). Detailed information on several phenotypic characteristics evaluated at the clone level is used in combination with some hundreds of mapped RAPD markers to guide matings for the production of progenies where new individuals will be selected for clonal propagation programs.

Molecular information of diversity and distance can assist in the enrichment of the genetic base as a breeding program proceeds. It can also be useful to evaluate the redundancy and specific deficiencies in germplasm banks and generate data on the efficiency of collection, maintenance and management strategies. (Phillips *et al.* 1993; Newbury and Ford-Lloyd 1993). Marker data can de utilized in the establishment of "core collections" to minimize the number of unnecessary accessions maintained in germplasm banks and to facilitate the access of breeders to such banks. (Frankel 1984; Brown 1989a,b). In recent years, the availability of RAPD markers has allowed analyses of genetic diversity and/or cultivar classification in several species for which such studies had previously been impossible due the lack of isozyme polymorphisms or RFLP probes. Some of these species include papaya (Stiles *et al.* 1993), cocoa (Wilde *et al.* 1992), garlic (Wilkie *et al.* 1993), *Microseris elegans* (VanHeusden and Bachmann 1992), celery (Yang and Quiros 1993), and peanut (Halward *et al.* 1992).

In maize breeding programs, hybrid performance can be predicted on the basis of the genetic relationship of the parental lines, and inbred lines can be allocated to heterotic groups. Earlier studies showed that genetic distance quantified by isozyme data was not a good predictor of grain yield or heterotic response (Frei *et al.* 1986), and that pedigree relationship was a better indicator (Smith and Smith 1989).

The advent of RFLP markers allowed a much more extensive sampling of the genome and a more detailed analysis of such an approach. Coefficients of determination on the order of 0.8 between genetic distance between lines measured with some tens of RFLP markers and heterotic response of the hybrids suggest that molecular similarity measures can be good predictors of hybrid performance (Smith *et al.* 1990; T. Osborn, personal communication). On the other hand, Dudley *et al.* (1991) verified that RFLP data agreed with pedigree information, but that the coefficient of genetic distance was not significantly correlated with hybrid productivities. In a later study, Dudley *et al.* (1992) used a measure of association of molecular markers with grain yield. The mean of the homozygous genotypes measured in the parental lines

were then highly correlated with the means of the hybrids. In this same line of study, it has also been verified and it is common practice today, the classification of maize lines in heterotic groups as well as the assignment to such groups of new lines with unknown origin (Godshalk *et al.* 1990; Livini *et al.* 1992; Melchinger *et al.* 1992).

Finally, molecular characterization also generates basic knowledge on the evolution of genomes, phylogenetic relationships and species diversification (Song *et al.* 1988; Second 1985; Skroch *et al.* 1992; Grattapaglia *et al.* 1992b; Phillips *et al.* 1993; Jung *et al.* 1993). This type of information can have important practical applications in the development of improved cultivars. For example, detailed phylogenetic data can assist in the introgression of genes through the selection of materials whose cross has better chances of resulting in fertile progenies. The degree of similarity between cultivars can still be used to reconstitute pedigrees and get insights on the development of a cultivar (Dweikat *et al.* 1993; Tinker *et al.* 1993). Other examples of applications of molecular characterization include : the maximization of heterotic response by increasing allelic diversity at each locus in polyploid organisms (Bonierbale *et al.* 1993), the analysis of the degree of recombination in populations (Tulsieram *et al.* 1992b) and the analysis of the extension of linkage drag around selected loci in backcross breeding (Young and Tanksley 1989).

3.3 Construction of genetic linkage maps

The development of linkage maps is often regarded as one of the more powerful applications of molecular markers in the genetic analysis of species and potentially in the practice of breeding. In the context of plant breeding, genetic maps allow: (1) extensive coverage of genomes for detailed genetic analysis; (2) the decomposition of complex genetic traits into its discrete Mendelian components; (3) the canalization of linkage information between markers and traits of interest into breeding strategies.

The availability of a large number of highly polymorphic and neutral genetic markers whose inheritance and segregation can be easily followed through generations has allowed the construction of genetic linkage maps for several plant species. In the last four years, the technological advances in the detection and analysis of genetic marker data has resulted in a unprecedented explosion of the field of genetic linkage analysis. Existing genetic maps for some major crops were refined and saturated with new markers. The greatest impact, however, consisted in the expansion of genetic

mapping studies to species where this was considered very difficult or impossible due to lack of appropriate populations and/or long generation times such as outcrossed forest and fruit trees.

The construction of a genetic linkage map basically involves the application of techniques of molecular biology to the original concepts of genetic inheritance. There are two basic requirements for the development of a genetic map: (1) sexual reproduction and generation of progenies; (2) large numbers of Mendelian markers. The methodology of map construction integrates a large number of techniques that may include the development of progenitor lines and segregating populations, the analysis of genotypes at marker loci through several different techniques, and the use of various statistical and computational tools for the estimation of recombination distance and order of marker loci.

Traditionally, the first step in genetic mapping has been the selection of progenitors for crosses with the objective of maximizing the probability of detecting genetic polymorphism at the DNA level. The ease with which polymorphic molecular markers are detected varies widely from species to species. As a general rule, autogamous species (selfing), essentially homozygous at all loci, are less polymorphic than preferentially allogamous (outcrossing), highly heterozygous species (Gale *et al.* 1990). For allogamous species such as maize and *Brassica,* the majority of RFLP probes are informative with one restriction enzyme or another, and frequently several alleles are detected.

On the other hand, to obtain a battery of polymorphic probes is a tedious process with autogamous species e.g. tomato and soybean (Helentjaris *et al.* 1985; Keim *et al.* 1989). Also, the closer the genetic distance between the different genotypes, more difficult it is to find genetic polymorphism. Therefore, lines without common ancestry, or the use of interspecific crosses have been frequently used to optimize the genotypic analysis (Bonierbale *et al.* 1988; Tanksley *et al.* 1988). Recently, however, molecular markers based on highly polymorphic simple sequence repeat sites (see chapter II), have, in principle, eliminated the problem of lack of polymorphism in selfing species (e.g. soybean, Morgante and Olivieri 1993). Finally, in the context of quantitative trait mapping, selection of progenitors becomes only a function of the contrasting phenotypes of the parents to evaluate the effects of allelic substitutions at loci that control traits of interest.

Once progenitors are selected, the generation of segregating progenies of several dozens of individuals is necessary to obtain a sufficient sample of meiotic events and estimate the frequency of recombination between segregating markers. The sample size required becomes several hundred for the analysis of Quantitative Trait Loci (QTL's) (see below). Traditionally, the pedigree used for the construction of genetic maps involves a cross between two inbred lines originally obtained through successive generations of selfing. The F_1 progeny are genotypically monomorphic. Either a selfing or backcrossing generation is performed to produce an F_2 or BC_1 generation, that segregates for the molecular marker genotypes. A polymorphic marker between the two inbred lines will have genotype (AA) in one line and (aa) in the other. The F_1 will all be heterozygous (Aa). In an F_2 produced by selfing the F_1 generation, the marker will segregate 1 (AA): 2 (Aa): 1(aa) for co dominant markers and 3 (A_) : 1 (aa) for a dominant one. In a backcross to the homozygous recessive line, tboth these markers will segregate 1 (AA): 1 (Aa). Alternatively, one can use a population of Recombinant Inbred Lines (RIL) derived from an F_2 through successive rounds of selfing (Burr et al. 1988; Reiter et al. 1992) or a population of doubled haploid lines produced by anther culture of F_1 individuals (Ferreira 1993). Both kinds of populations are analogous to a backcross in that a 1 (AA): 1(Aa) segregation ratio is expected.

As a segregating population is produced and a particular genetic mode of inheritance is expected, this mapping population is genotyped for a large number of markers. These markers had originally been selected because they are polymorphic (detect different alleles) between the parental inbreds. A matrix of data is generated where typically the rows correspond to the markers and the columns to the progeny individuals. A first statistical test ($\chi2$) is applied to each marker to test the null hypothesis of the expected segregation ratio, 1:1, 1:2:1 or 3:1. Significant $\chi2$ can indicate segregation distortions that can have a biological basis (e.g. lethals) or be merely the result of a small sample of individuals genotyped. Markers that pass this single-locus test, are then submitted to a second test of independence of segregation, or linkage. Basically this test is performed on all possible pairwise combinations of markers. Rejection of the null hypothesis - two markers segregating independently - is an indication of genetic linkage between the two markers.

The next step involves the estimation of genetic map distance, which is not a physical distance but reflects the frequency of genetic recombination events (crossovers) per meiosis per generation. This parameter is expressed in % recombination

or translated into centiMorgan units by using different mapping functions that take into consideration double crossovers and genetic interference (e.g. Haldane, Kosambi). In the simplest case, a backcross, genetic distance is the proportion between the observed number of recombinant genotypes and the total number of individuals genotyped; 10% recombination which roughly corresponds to 10 cM units corresponds to saying that there is a 10% probability of genetic recombination between the two markers during meiosis. The correlation between physical distance on the chromosome and the frequency of recombination between any two markers although seems intuitively high and linear, need not be so. There are regions of small physical distance where there is a high probability of genetic recombination, the so called "recombination hot spots". On the other hand, there are very long stretches of DNA where genetic recombination is suppressed due to the presence of heterochromatin, or around centromeres or in telomeric regions and therefore the genetic distance is small.

When all the markers are grouped in linkage groups and all two-point recombination distances are estimated, a genetic map can be constructed by hand in a classical procedure called two-point mapping where the relative position of the markers is determined and gradually a gene order is obtained. This procedure becomes tedious when hundreds of markers need to be ordered. Also each two-by-two distance has an associated error, which is a function of its magnitude and the sample size. This complicates the correct estimation of the relative position of the markers, and this becomes increasingly more difficult with closer markers. To mitigate these problems, maximum likelihood multipoint estimates of recombination frequency are typically used. An estimate of θ (recombination distance), that maximizes the probability of getting the observed data is obtained through iterative algorithms. Different locus ordering algorithms are then used to determine the relative position of the markers. A growing number of software packages have been developed and are available today to perform these procedures automatically in a more or less interactive fashion and with variable flexibility in terms of populations and mixtures of segregation ratios allowed. These include Linkage1 (Suiter et al. 1983), MapMaker (Lander et al. 1987), GMendel (Liu and Knapp 1992) and JoinMap (Stam 1993).

In the last ten years, partial genetic maps of morphological markers that took several decades to develop were rapidly saturated with molecular markers. Several species for which no information on genetic linkage was previously available, had

genetic maps constructed. Table 2 is a list of species for which genetic linkage maps were constructed using different kinds of pedigrees and classes of molecular markers.

Table 2. Plant species for which moderate density genetic linkage maps have been developed with different classes of molecular markers.

Common name	Species	Reference
Alfalfa	*Medicago sativa*	Brummer *et al.* 1993; Echt *et al.* 1993
Apple	*Malus domestica*	Lawson *et al.* 1992
Arabidopsis	*Arabidopsis thaliana*	Chang *et al.* 1988; Nam *et al.* 1989; Reiter *et al.* 1992
Barley	*Hordeum vulgare*	Shin *et al.* 1990; Kleinhofs *et al.* 1993
Bean	*Phaseolus vulgaris*	Vallejos *et al.* 1992; Nodari *et al.* 1993
Cauliflower	*Brassica oleracea*	Slocum *et al.* 1990; Landry *et al.* 1992
Citrus	*Citrus*	Durham *et al.* 1992
Colza	*Brassica napus*	Landry *et al.* 1991; Ferreira 1993
Eucalypt	*Eucalyptus grandis*	Grattapaglia and Sederoff 1992
	Eucalyptus globulus	Song and Cullis 1992
	Eucalyptus nitens	Moran *et al.* 1992
	Eucalyptus urophylla	Grattapaglia and Sederoff 1992
Faba bean	*Vicia faba*	Torres *et al.* 1993
Lentil	*Lens culinaris*	Havey and Muehlbauer 1989; Weeden *et al.* 1992
Lettuce	*Lactuca sativa*	Landry *et al.* 1987
Maize	*Zea mays*	Helentjaris *et al.* 1986b; Burr *et al.* 1988 Muehlbauer *et al.* 1991
Mustard	*Brassica rapa*	Song *et al.* 1991; Chyi *et al.* 1992
Pea	*Pisum sativum*	Ellis *et al.* 1992; Weeden *et al.* 1992
Peach	*Prunus persica*	Chaparro *et al.* 1994
Pepper	*Capsicum annuum*	Tanksley *et al.* 1988
Pine	*Pinus taeda*	Grattapaglia *et al.* 1991; Neale and Williams 1991; Neale and Sederoff 1991
	Pinus elliottii	Nance *et al.* 1992
	Pinus radiata	Moran *et al.* 1992
	Picea glauca	Tulsieram *et al.* 1992a
Poplar	*Populus trichocarpa*	Bradshaw and Stettler 1993
Potato	*Solanum tuberosum*	Bonierbale *et al.* 1988;

Table 2. Continued

		Gebhart *et al.* 1989;
		Tanksley *et al.* 1992
Rice	*Oryza sativa*	McCouch *et al.* 1988
Sorghum	*Sorghum bicolor*	Binelli *et al.* 1992
Soybean	*Glycine max*	Keim *et al.* 1990;
Sugar cane	*Saccharum spontaneum*	Al-Janabi *et al.* 1993;
		DaSilva *et al.* 1993
Tomato	*Lycopersicon esculentum*	
		Helentjaris *et al.* 1986b;
		Bernatzky and Tanksley, 1986;
		Gebhart *et al.* 1991;
		Tanksley *et al.* 1992
Vigna	*Vigna radiata*	Menancio-Hautea *et al.* 1993

The construction of genetic maps covering large portions of the genome, stimulates the acquisition of important information about the genetics of the species with potential applications into breeding. For germplasm surveys, mapping information can be used for selecting a battery of genetic markers evenly distributed throughout the genome, facilitating a better sampling of all genomic regions. This high genome coverage is essential for studies involving the detection of genetic linkage between markers and mono factorial as well as more complex quantitative traits of economical importance. The existence of a complete framework map of molecular markers, allows mapping morphological as well as cloned genes of known function, generally isolated through the use of mutants. However, the correlation between classical genetic maps and RFLP maps for model organisms such as maize and tomato has still been a difficult task. Specific mapping populations have to be created where both the morphological and molecular markers segregate. An alternative approach has been used in maize, for example, where some genetic loci that had been mapped on the classical map, were also mapped on the RFLP map by using probes derived from cloned genes corresponding to the morphological traits. Also correlations between the classical and RFLP map could be established by using special genetic stocks containing translocations (Helentjaris *et al.* 1986a; Wright *et al.* 1987). For the great majority of cultivated species, however, this type of study is impossible to perform because no classical genetic map of morphological markers and no well characterized mutants are available.

Finally, a genetic linkage map of a species, may be used to generate a large amount of information on the structure, organization and evolution of the genome based on patterns of inversions and translocations, gene duplications and distribution of gene families. Comparative or syntenic mapping using heterologous probes across related species has offered great opportunities to study the genomic structure from the perspective of gene homology and conservation of linkage and gene ordering relationships along the chromosomes (Weeden *et al.* 1992; Tanksley *et al.* 1988; Bonierbale *et al.* 1988; Slocum *et al.* 1990; Whitkus *et al.* 1992; Prince *et al.* 1993; Kianian and Quiros 1992). Extensive syntheny has been found for genes between distantly related species e.g in the grasses. This synteny will prove to be of enormous value in the identification and isolation of genes, just as the conservation of gene sequences is proving to be a major aid for identification of gene function. With this idea in mind, comparative mapping has been proposed as a strategy to generate a generalized reference map of genes with known products and functions for all plant species. A large international project is under development with this objective, using the small crucifer *Arabidopsis thaliana* as a model plant (Beckmann 1991). *Arabidopsis* has a small nuclear genome, it has highly saturated linkage maps, extensive physical maps and offers several advantages for genetic manipulation such as short generation cycles, availability of mutants and efficient plant transformation methods.

3.3.1 Construction of genetic linkage maps in cross fertilizing species

In the last three years, genetic mapping has experienced a significant expansion in highly heterozygous, preferentially outcrossing and essentially undomesticated species such as forest and fruit trees (Table 1). For such species, the pedigrees traditionally used for genetic mapping, involving inbred lines and either selfing or matings between relatives can not be easily obtained because of the typically high inbreeding depression associated with existing genetic load and the long time necessary to complete a generation. On the other hand, the heterozygous nature of these plants results in high levels of DNA sequence polymorphisms for the vast majority of species studies to date. Construction of genetic maps in trees was pioneered in loblolly pine (*Pinus taeda*) (Neale and Williams 1991) and poplar (*Populus trichocarpa*) (Bradshaw and Stettler 1993) using RFLP markers and three-generation pedigrees. In spite of the early advances, the majority of tree geneticists and breeders soon realized that RFLP

analysis was extremely laborious and that three-generation pedigrees were rare and not typical in breeding programs. The introduction of RAPD markers opened a new perspective in this area because, it not only facilitated molecular analysis, but also allowed the readoption or novel conception of mapping strategies more relevant and suitable to the genetic reality of such species (Grattapaglia *et al.* 1992c,1993; Grattapaglia and Sederoff 1992,1993).

A special feature of the genetic biology of conifers is the haploid megagametophyte tissue in the seeds, whose presumed function is to nourish the germinating embryo. Megagametophytes are mitotic derivatives of a single haploid megaspore and are derived from the same megaspore that gives rise to the maternal gamete. Therefore they are genetically equivalent to maternal gametes. If a tree is heterozygous at a given marker locus, half of the megagametophytes carry one allele and the other half the alternative allele. In the haploid mapping strategy one observes a 1:1 segregation ratio, equivalent to a back or testcross. The cosegregation of polymorphic allozymes in megagametophytes was originally proposed and used by Conkle (1981) to generate partial genetic maps with about two dozen loci in several conifer species. The limited number of genetic markers did not allow the construction of complete maps. RFLP marker analysis could not be used because they require 100 times more DNA than what is available in the megagametophyte tissue. RAPD markers on the other hand can be obtained in large numbers and because they are based on DNA amplification by PCR, rather than DNA hybridization, they only require very small amounts of DNA. The combination of RAPD markers and the haploid biology of conifers has allowed the direct analysis of segregation in gametes, and the rapid construction of genetic linkage maps in several conifer species. Grattapaglia *et al.* (1991) generated the first moderately saturated linkage map for a conifer comprising 191 RAPD markers for an elite clone of loblolly pine in a record two-month period. The same approach was immediately used by other groups and since then, several maps have been constructed for several individuals of various conifer species (Tulsieram *et al.* 1992a; Nance *et al.* 1992)

An analogous, but more generalized approach can be used for the construction of genetic maps in highly heterozygous, oucrossing species that do not have haploid biology, such as hardwood and fruit trees. For the vast majority of these species, it is difficult or impossible to produce or have access to three-generation pedigrees. On the other hand, crosses involving parents and progeny, i.e. two-generation pedigrees are

commonly available. The F_1 generation of such crosses is genetically heterogeneous and segregates for marker genotypes and phenotypic traits. An F_1 in the forest tree breeding concept is actually closer to an F_2 or backcross generation in an annual self fertilizing crop. Genetic markers segregate in different Mendelian ratios, depending upon the genotypes of the parents and the mating configuration (Ritter et al. 1990). For example if both parents are heterozygous for the same alleles, a 1:2:1 genotypic ratio for co dominant markers and 3:1 for dominant ones is expected. If the mating involves three or four alleles, as is often the case, a 1:1:1:1 ratio is expected for co dominant markers, however a more simple 1:1 ratio is expected, if only one dominant allele is detectable out of the four alleles in the mating configuration. Although less information per locus is acquired with dominant markers a more simple and straightforward analysis of a 1:1 ratio can be performed following essentially the same analysis used in the haploid mapping or in a backcross or testcross mating in annual crops. RAPD markers are particularly advantageous in this case because they are highly polymorphic and detect only one allele (presence of the amplified band) at a locus, while all the other alleles are detected as null. So essentially, the meiotic segregation of the two alleles (presence and absence) at a RAPD locus in one parent of the cross can be followed in exactly the same fashion as in the analysis of haploid megagametophytes, as long as the other parent tree has a RAPD null genotype at the same locus. Because the analysis of RAPD markers is quick and simple, it is possible to prescreen large numbers of oligonucleotide primers and select those that amplify markers segregating in the informative 1:1 configuration. This screening is typically done by analyzing the two parents and a sample of six or more F_1 progeny individuals in search of RAPD bands that are present in one of the parents, absent in the other and polymorphic, i.e. present and absent in the progeny sample. After genotyping several dozens of individuals for the preselected markers, two independent marker data sets are generated, one for each progenitor. These will in turn result in two independent genetic linkage maps by analyzing the co segregation of markers.

Markers for which only one allele is detectable have been called single dose polymorphisms or single dose restriction fragments when derived from RFLP analysis (Gebhardt et al. 1989; Ritter et al. 1990; Da Silva et al. 1993; Carlson et al. 1991; Sobral and Honeycutt, 1993). They have been efficiently used to map polyploid species such as sugar cane (Wu et al. 1992; Sobral and Honeycutt 1993; Al-Janabi et al. 1993). The name "pseudo-testcross" was suggested for this strategy because the 1:1 testcross

mating configuration of the markers is not known *a priori* as in a conventional testcross where the tester is homozygous recessive for the locus of interest. Rather, the configuration is inferred *a posteriori* after analyzing the parental origin and genetic segregation of the marker in the progeny of a cross between highly heterozygous parents with no prior genetic information. When this inference is done for both parents involved in the cross, the term "two-way pseudo-testcross" is more appropriately used (Grattapaglia and Sederoff 1994). This conceptually simple strategy combined with the polymorphism detection power or the RAPD assay, allowed the construction of the first linkage maps in *Eucalyptus grandis* and *E. urophylla* (Grattapaglia and Sederoff 1992, 1994) and has been widely used in other species such as *E. globulus* (Song and Cullis, 1992), apple (Lawson *et al.* 1992), grape (M. Lothi, personal communication), sugar cane (Sobral and Honeycutt, 1993; Al-Janabi *et al.* 1993) and conifer species (M. Carson personal communication and L. Pearson, personal communication).

The major advantage of the pseudo-testcross mapping strategy combined with RAPD markers is that it is a general strategy for the construction of genetic linkage maps in outbred forest trees as well as in any highly heterozygous living organisms. It can be immediately applied to these species without any prior genetic information. The only requirements are sexual reproduction between two individuals that results in the generation of a progeny set large enough to allow the reliable estimation of recombination frequencies between segregating markers. Its efficiency will be directly proportional to the level of genetic heterozygosity of the species under study, which is a function of the mating system, and the genetic divergence between the individuals crossed. In interspecific crosses, practically every arbitrary RAPD primer detects at least one informative pseudo-testcross marker, and frequently several. At the intraspecific level, the frequency is comparatively less, however, still high enough to make this strategy extremely efficient (Carlson *et al.* 1991; Grattapaglia, unpublished data; L. Pearson, personal communication). Finally, the possibility of constructing linkage maps using only two-generation pedigrees is of great relevance to tree breeding because it allows the use of existing populations. The ability to construct linkage maps for individual trees also mitigates the highly debated problem of linkage equilibrium between markers and trait alleles of interest for marker assisted breeding strategies (see below).

3.4 Genetic mapping of simply inherited traits

Associations between molecular markers and traits that display simple Mendelian inheritance have been established in several crop species. Isozymes, RFLP's and more recently RAPD's have been used to map genomic regions significantly associated with traits such as disease and pest resistance. From the theoretical standpoint, genetic mapping of such traits is quite simple, as more or less discrete phenotypic classes for the traits can be observed. The traditional mapping scheme involves crossing a susceptible line with a resistant one. The F_1 can be either susceptible, resistant or intermediate, depending on the pattern of interallelic interaction (recessive, dominant or incomplete dominance etc.). The F_1 is either selfed or backcrossed to one of the parents to obtain a population segregating for resistance and susceptibility. This population is subject to an inoculum pressure under controlled conditions and the phenotypes evaluated. When the phenotypes segregate according to a particular genetic model and the associated Mendelian expectations, the resistance trait is treated as another marker. Frequently, however, complications in this analysis arise due to incomplete penetrance (imperfect correspondence between genotype and phenotype), distorted segregation ratios due to the participation of more than one locus and epistatic interactions of loci. When only a few genes of large effects are involved, discrete phenotypic classes can still be determined. Typically subjective systems of disease severity scores are used and associations between markers and resistance can be found. The mapping procedure generally involves a first search for significance using one marker at every 25 or more centiMorgans, i.e. three to five markers per linkage group. As soon as a significant association is found, a larger number of markers in that region is surveyed, trying to establish a more precise location for the locus of interest.

A classical example of this approach was described in tomato (*Lycopersicon esculentum*) when a significant association was established between an isozyme marker locus (Alkaline Phosphatase 1) and a locus that controls resistance to the nematode *Meloidogyne incognita*. These two loci were found to be linked at a 1 cM distance (Medina-Filho 1980). Several tomato breeding programs have used this information to indirectly select nematode resistant plants based on the presence of a particular isozyme allele (Tanksley 1983a,b). Another example of strong association is between a locus that confers resistance to BYMV (Bean Yellow Mosaic Virus) and an allele at a plastid isozyme, PGM (Phospho-glucomutase) in pea (*Pisum sativum*) (Weeden *et al.* 1984). These two loci are at 2 cM apart. In both examples, the indirect selection on the isozyme allele avoids or reduces the manipulation of the pathogen and facilitates the

selection of resistant individuals. Genetic linkage between isozymes and simply inherited traits have also been detected in apple (*Malus*) for resistance to aphids, downy mildew and several self incompatibility alleles (Chavreau and Laurens 1987; Manganaris and Alston 1987). In barley an association has been detected between an isozyme locus and a major gene controlling differentiation of shoots in callus cultures (Komatsuda *et al.* 1993).

The use of RFLP and RAPD techniques has significantly increased the number of associations detected between markers and traits of economical interest, both with simple and complex inheritance (for a review see Melchinger 1990). The large number of markers allows a more complete sampling of the genome for significant linkages. RFLP's have been used to locate major genes involved in disease resistance in tomato (Young *et al.* 1988; Klein-Lankhorst *et al.* 1991a; Sarfatti *et al.* 1989), species of *Brassica* (Landry *et al.* 1992; Ferreira 1993), rice (Yu *et al.* 1991), soybean (Muehlbauer *et al.* 1991; Diers *et al.* 1992) and maize (Bentolila *et al.* 1991). Mutant lines have been used for the detection of an RFLP closely linked to the *ph1* gene that controls chromosome pairing in wheat (Gill and Gill 1991). A GA$_3$ insensitive dwarfing gene and growing habit were mapped with RFLP markers in rye (*Secale cereale*) (Plaschke *et al.* 1993). Since their introduction, RAPD markers have been extensively used for the purpose of mapping resistance genes in tomato (Klein-Lankhorst *et al.* 1991a,b; Martin *et al.* 1991), bean (Miklas *et al.* 1993; Haley *et al.* 1993) lettuce (Michelmore *et al.* 1991; Paran *et al.* 1991), *Beta vulgaris* (Uphoff and Wricke 1992) *Pinus lambertiana* (Devey *et al.* 1993), and *Pinus taeda* (P. Wilcox and H. Amerson, personal communication). Recently, RAPD markers linked to the two genes that control somatic embryogenesis in alfalfa have been identified (Yu and Pauls 1993).

Most of the studies where disease resistance loci were mapped, relied on a strategy involving Near Isogenic Lines (NIL) derived from sequential backcross breeding. Gene mapping using NIL is based on the principle by which a gene of interest is introduced in a cultivated line from a donor line. After several generations of backcrossing, the genome of the selected progenies becomes more and more similar to that of the recurrent parent with the exception of the DNA segments that contain the introgressed gene of interest. These segments, if genetically polymorphic at the DNA level in relation to the recurrent genome, can be used as a target to determine if a particular marker is located near the gene of interest. Near Isogenic Lines have been used to identify two RFLP probes very closely linked (0.4 cM) from the *Tm-2a* gene in tomato that confers

resistance to the Tomato Mosaic Virus (TMV) (Young *et al.* 1988). Similarly, a set of NIL derived from the cross of lines susceptible and resistant to *Pyricularia orizae* was used to identify two RFLP markers associated with the resistance genes (Yu *et al.* 1991). One of the markers, RFLP clone RF64, is 2.8 cM from a resistance locus and can potentially be used to select for resistant genotypes. That same marker is also strongly associated with a photoperiod sensitivity gene (MacKill *et al.* 1993). NIL were also used to map genes conferring resistance to *Fusarium oxysporum* and *Cladosporium fulvum* in tomato (Sarfatti *et al.* 1989); Van der Beek *et al.* 1992) and *Phythophthora megasperma* in soybean (Diers *et al.* 1992).

There are however two major limitations in using such an approach to identify linkages of breeding interest. The first one has to do with the creation of NIL. In the vast majority of plants this is impossible, especially in allogamous plants with long generations such as trees. When possible, it is still a time consuming process because many generations of backcrossing are necessary. The second limitation is that frequently many DNA segments of the donor genome are inadvertently introduced in the recurrent line (Young and Tanksley 1989). This results in the identification of polymorphic markers between the NIL's that in fact are not linked to any gene of interest. An alternative procedure which is also restricted to some species, involves the use of dihaploid lines derived from microspore culture of a single F_1 individual. Such an approach was used to identify DNA markers linked to loci for resistance to *Leptoshaeria maculans* and Albugo candida in colza (*Brassica napus*) (Ferreira 1993).

A powerful alternative technique to target genomic regions associated to simply inherited traits involves the analysis of pools of genomes that are fixed at the locus of interest. This strategy was originally proposed by Arnheim *et al.* (1985) for the identification of RFLP's in linkage disequilibrium with loci controlling diseases in humans as a result of natural selection. Markers linked to the locus of interest were identified based on their linkage disequilibrium with the disease trait in relation to the rest of the population. This idea was later adopted by Michelmore *et al.* (1991) that proposed its use for the rapid identification of markers in specific genomic regions using segregating populations, a technique called Bulk Segregant Analysis (BSA). Although any type of molecular marker can theoretically be used for this approach, RAPD markers are particularly efficient because large numbers of them can be quickly screened.

BSA consists in detecting differences between two bulked DNA samples derived from a segregating population from a single cross. These bulked samples are

constructed by pooling the DNA of individuals selected for having an identical genotype at the particular target genomic region and variable genotypes at regions unlinked to the target. At the DNA level, theoretically, the two samples will differ exclusively for the target region and will be monomorphic for all the other regions. The bulks are composed of DNA of the 10 to 20 individuals displaying extreme phenotypes at each end of the distribution of the trait of interest, for example a resistant bulk of the 10 to 20 more resistant phenotypes versus a susceptible one made up of a similar number of highly susceptible individuals. The DNA of the parental lines and the two bulks are analyzed with a large number of markers in search of markers that are polymorphic between the bulks, i.e. present in one but not in the other, and correctly originate from the resistant parental line. This preliminary evidence of genetic linkage is subsequently confirmed by a conventional co segregation analysis using a larger number of individuals from the population. Genetic linkage is then confirmed or rejected (false positive in the bulk screening frequently arise) and an estimate of genetic distance is obtained between the marker and target locus. Basically BSA is a short cut that uses a preliminary genomic selection to drastically reduce the time necessary to identify genetic linkage to major genes, quickly focusing the efforts and subsequent analysis on genomic regions associated with the trait. The genetic markers identified with this strategy are found to be within a 25 cM window around the target locus.

BSA has been originally used to locate genes associated with resistance to downy mildew (*Bremia lactucae*), an important pathogen in lettuce (*Lactuca sativa*) (Michelmore *et al.* 1991; Paran *et al.* 1991; Paran and Michelmore 1993). Recently it has been successfully used to map genetic loci conferring resistance to *Pseudomonas syringae* in tomato (Martin *et al.* 1991), *Puccinia graminis* in oat (*Avena sativa*) (Penner *et al.* 1993), *Uromyces phaseoli* in bean (Miklas *et al.* 1993; Haley *et al.* 1993), *Leptosphaeria maculans* in colza (Ferreira 1993), *Rhynchosporium secalis* in rye (Barua *et al.* 1993) and *Cronartium ribicola* in *Pinus lambertiana* (Devey *et al.* 1993).

As a tool to target specific genomic regions, BSA is very useful to identify markers in poorly covered regions of linkage maps (Michlemore *et al.* 1991). As a genetic map approaches saturation, the continued mapping of new markers becomes increasingly inefficient. To alleviate this problem, bulked samples can be constructed based on the polymorphism of the closest marker available to the target region (e.g. presence/absence of a RAPD marker or alternative alleles at an RFLP locus). New polymorphic markers between the bulks are identified and the cycle repeated again

based on the newly found marker. An iterative genetic walking procedure is therefore performed until sufficient coverage and marker density is achieved. This procedure has been termed 'local mapping' (Reiter *et al.* 1992) and was originally used to map a large number of markers specifically to one of the five *Arabidopsis* chromosomes. Finally, a pooling technique was also used to identify loci for other simply inherited traits. Giovannoni *et al.* (1991) used the pooling strategy based on known RFLP genotypes to identify markers linked to loci controlling fruit ripening (*non-ripening*) and pedicel abscission in tomato (*jointless*).

3.5 Genetic mapping of complex traits (Quantitative Trait Loci, QTL's)

Most of heritable traits of economic importance are the result of the joint action of several genes with unequal effects and variably influenced by the environment. These traits are called polygenic, quantitative or of complex inheritance. The resulting phenotypes display a continuous variation instead of discrete phenotypic classes. Productivity, plant size and precocity are typical quantitatively inherited traits. For the vast majority of complex traits, very little information is available on their genetic architecture. This latter term typically refers to the number, relative chromosomal location, magnitude of effect and possible interactions of the genetic loci controlling the expression of the trait.

The theory and techniques of quantitative genetics have been powerful tools for the study of polygenic inheritance. They have also been a fundamental tool for genetic progress in plant breeding. Nevertheless, the manipulation of complex traits continues to be a challenging task. Genetic linkage maps of molecular markers offers a powerful tool to investigate the genetic architecture of polygenic traits and potentially assist in the manipulation of the genetic factors controlling such traits on an individual basis, in procedures involving the selection and recombination of desired genotypes.

The determination of genetic linkage between markers and QTL's depends on the existence of linkage disequilibrium between the alleles at the marker locus and the QTL alleles. This disequilibrium generates quantitative effects associated with the marker that can be detected and estimated through appropriate statistical approaches. Linkage disequilibrium between two loci occurs when the two-locus genotypic frequencies significantly differ from the product of gene frequencies, denoting the existence of a non-random association between the two loci. Physical linkage between loci is only one

of the various causes of linkage disequilibrium. In mapping experiments, however, a significant association between marker and gene is usually interpreted as evidence of genetic linkage.

The classical QTL mapping strategy is based on pedigrees traditionally used for the construction of linkage maps. Two inbred lines, genetically divergent and preferentially extreme in the quantitative trait of interest are identified. These lines, homozygous for alternative alleles at the QTL's and polymorphic for a number of molecular markers, are crossed and a segregating population is generated from the F_1. Segregating populations can be of different kinds: F_2 intercross, backcross, recombinant inbred lines (RIL) or dihaploid lines. This type of pedigree where phenotypically extreme inbred lines are initially crossed is considered ideal because it maximizes linkage disequilibrium between markers and QTL's. A power to detect a QTL is a function of the magnitude of its effect on the quantitative trait, the heritability of the trait, the size of the segregating population and the genetic map distance (recombination frequency) between marker and QTL. Evidently the larger the effect, the heritability and the sample size, and the closer is the marker, the more powerful will be the efficiency of detection. QTL mapping experiments will therefore typically involve the evaluation of a few hundred individuals for the quantitative traits of interest and their genotyping for several dozens of molecular markers selected for being evenly spaced (10 to 30 cM) throughout the genome. A search is then carried out for significant associations between segregating markers and trait values

Various statistical methods can be applied to QTL analysis. A t test of difference between means is the most simple one, and it is generally considered to be robust especially for trait distributions departing from normality, or in cases where the genotypic classes involve a mixture of distributions due to recombination between marker and QTL (R. Doerge, personal communication). In a population where markers segregate 1:1, all the individuals can be classified in either one of two classes depending on their marker genotype. For each marker, a mean and a variance can be calculated for each genotypic class. A hypothesis test is constructed where the null hypothesis of no difference between the mean trait value of the two classes is tested using a t test. Alternatively, simple one-way ANOVA or linear regressions can be applied where the phenotypic trait is the dependent variable and each marker is fitted one at the time as the independent variable. A significant F test is an indication of genetic linkage. A regression approach was applied using isozyme data against several

quantitative traits in maize (Edwards *et al.* 1987, 1992; Stuber *et al.* 1987) tomato (Tanksley *et al.* 1982) as well as with RFLP data for these same crops (Nienhuis *et al.* 1987; Osborn *et al.* 1987; Tanksley and Hewitt 1988; Edwards *et al.* 1992). The use of a linear single marker analyses presents at least two limitations. The first one is the control of the experimentwise type I error rate, i.e. false positives (declaring a significant association where there is none). This can be controlled by using a higher significance threshold, compromising in statistical power (increase in type II error). The second problem is that linear models do not provide information as to the magnitude of effect of the QTL and its genetic distance to the marker. In other words a QTL of small effect but close to a marker and a QTL of large effect but further away from a marker cannot be discriminated.

As the result of recombination between marker and QTL, each of the genotypic marker classes will consist of a mixture of genotypes at the QTL. This mixture will produce a skewed distribution within each of the classes, the degree of skewness depending on the proportion of recombination between marker and QTL. The presence of recombinants in each marker class generates a difference in the within class variance. This means that in marker-QTL analyses, the shape of the within-marker genotype distribution carries important information in addition to the trait mean. These marker associated differences in shape of the distribution are inaccessible to linear ANOVA, but can be better exploited by maximum-likelihood methods. Such methods with a single marker approach were first applied to QTL mapping by Weller (1986, 1987) to try to separate main effects at the QTL and frequency of recombination between marker and QTL.

To further improve the power and precision of QTL analysis, current practice makes use of pairs of markers in procedures known as "marker brackets" (Soller and Beckmann 1983) or "interval mapping" when combined with maximum likelihood methods (Lander and Botstein 1989). Marker brackets, can disentangle the confounding of main effect and recombination found in quantitative effects associated with a single marker. The reason for this is that the difference between parental types for the marker bracket provides an estimate of the main effect of the QTL, virtually independent of recombination. Separating parental types and recombinant makes the parental types essentially monomorphic with respect to the QTL so that application of maximum likelihood methods based on distribution shape has little to offer to their analysis. The recombinant types, however, will be a mixture of QTL genotypes and therefore, besides

information on QTL location they will also contain information on the magnitude of effect of the QTL.

To express the significance level with which a QTL is detected, current methods use either a traditional p value or a LOD score. A LOD (logarithm of the likelihood of odds) score indicates in a log scale how much larger is the probability of the observed data to have arisen assuming the alternative hypothesis of the presence of a QTL than the null hypothesis of no QTL. A LOD= 3.0 indicates that this probability is roughly 1000 times larger. In general a significant threshold between 2 and 3 is used, depending on the number of markers used and the total map distance, so that the probability of falsely declaring a QTL (type I error) is kept below 0.05 in each interval individually. For example in tomato, with a genome comprising 12 linkage groups in 1100 cM and markers spaced at every 20 cM (about 55 markers) a LOD= 2.4 will provide an alpha level of 0.001 for each individual test (Lander and Botstein 1989).

The software package MapMaker-QTL (Lander *et al.* 1987) implements this type of QTL analysis. The output is given in numerical form by LOD scores at each map interval or user specified portion of interval. Alternatively these data can be presented graphically as LOD score profile scans for each linkage group where peaks indicate the presence and location of a QTL. Functions for the analysis of F_2 intercross, backcross and RIL are available. In the F_2 intercross model, options are available to fit different models of gene action at each QTL so that additive and dominance effects can be estimated. In the backcross situation, however, only the additive effect of an allelic substitution can be estimated. QTL's controlling several traits related to productivity have been identified in tomato (Paterson *et al.* 1988; Paterson *et al.* 1991), maize (Beavis *et al.* 1991) and wheat (Anderson *et al.* 1993).

Besides interval mapping implemented by MapMaker-QTL, other methods exists for the analysis of QTL's. The package QTLStat (Liu and Knapp 1992) uses a combination of non-linear regression and maximum likelihood interval mapping. Another commonly used procedure involves the detection of QTL's by interval mapping methods and subsequent use of SAS (Statistical Analysis Software) procedures to fit linear models with variable numbers of QTL's previously detected. This theoretically allows the determination of models that better explain the dependent phenotypic variable and allows for the study of epistatic interactions. This approach is frequently criticized because it assumes one QTL controlling the trait to find QTL's in the first step of the

analysis, and then tries to simultaneously fit the several QTL's found in the second step (B. Liu, personal communication).

The literature on different QTL mapping approaches is extensive and a more detailed discussion is beyond the scope of this chapter. Existing approaches suffer of a resolution problem when two or more QTL's are linked on the same chromosome. Difficulties arise in the correct estimation of QTL position and magnitude of effect. In this case, MapMaker-QTL allows one to fix one QTL at a particular position (thereby removing from the analysis the variation contributed by that QTL) and then scanning the linkage group again for the second peak. If the second peak appears again (with the same LOD score difference), it is considered a strong indication of a second linked QTL. Another statistical method was recently developed to separate multiple linked QTL's (Zeng 1993a,b). This method allows to analyze each map interval separately, conditional on different hypothesis and models constructed concerning the presence of QTL's at adjacent intervals.

The main difficulty in QTL mapping is the intrinsic fact that various genetic and environmental factors affect the final phenotypic expression. The commonly used experimental designs involve the measurements and genotyping of several hundred individuals to reach the desired resolution. For example Soller et $al.$ (1976) estimated that an F_2 population of approximately 2000 individuals would be necessary to detect QTL's with effect equal to 2-3% of the trait mean. In view of the limitation in the ability to genotype large numbers of individuals, progenies of some hundreds are typically used. Edwards et $al.$ (1987) in maize, used 187 F_2 individuals, and the individual effects detected with alpha 0.01 varied between 6 and 27.4% of the total variation in plant height trait components. Paterson et $al.$ (1988) used a backcross family of 237 individuals to explain around 50% of the total phenotypic variation in each one of three quantitative traits (fruit weight, soluble solids and pH). Between five and six QTL's were detected for each trait, and some mapped to overlapping positions.

In spite of the modest number of QTL mapping studies performed to date, in a limited number of species, some general features have emerged. Significant associations between markers and quantitative traits have been detected even with relatively small progeny sizes. Invariably, the architecture of the analyzed traits involves a few (3 to 7) major QTL's with large effects, that individually can account for 10 to 30% of the phenotypic variation, and a variable number of minor ones whose individual effect is between the detectable threshold (a few percent) and 10%. The cumulative

proportion of the variation explained by the mapped QTL's has varied between 30 and 70%, depending on several factors such as the specific cross analyzed (Beavis *et al.* 1991), the trait under consideration (Edwards *et al.* 1992), and the experimental design. i.e. the sample size and the map resolution in terms of number of markers (Edwards *et al.* 1987). In QTL analysis using segregating F_2 intercrosses, a remarkably higher proportion of the variation is accounted by dominant gene action at the QTL's when compared to standard biometrical procedures. Epistatic effects (QTL interactions), however, are typically non significant. It is also a common occurrence to identify QTL regions that simultaneously affect the expression of different traits (Paterson *et al.* 1988; Edwards *et al.* 1992). This could be due to (1) the different traits measured are in fact different expressions of the same trait; (2) real pleiotropic effects (the same QTL affecting more than one trait); (3) the presence of independently acting QTL's but closely linked on the same chromosome so that the individual resolution is impossible with the current marker density and statistical methods. Finally an interesting feature has been reported in maize, where the majority of the QTL's identified for plant height growth mapped on top or at close distance to known qualitative loci affecting growth. These results support the hypothesis of Robertson (1985) that qualitative loci are the same that affect quantitative traits, and also provide a strong biological validation of the statistically detected effects (Beavis *et al.* 1991).

Experimental mapping strategies have been suggested that allow a substantial reduction in the number of progeny that need to be genotyped for the markers, while keeping a similar statistical power of QTL detection. We will discuss two of these strategies that should substantially increase the efficency of QTL mapping experiments whenever growing and phenotyping additional progeny requires less effort than completely genotyping individuals for all the markers, which is typically the case in many species. The first one, called "selective genotyping", although originally proposed by Lander and Botstein (1989), had its basic principle previously described by Stuber *et al.* (1980). It is based on the fact that a significant difference in allelic frequency at the QTL's in the high and low extremes of the phenotypic distribution can be used as a test for linkage between marker and QTL. Lander and Botstein (1989) contend that progeny with phenotypes more than 1 standard deviation from the mean comprise about 33% of the total population but contribute about 81% of the total linkage information. By growing a population that was only about 25% larger and genotyping only these extreme progeny, the same total linkage information would be obtained from

genotyping only about 40% as many individuals. Therefore when adopting this strategy, an extended set of individuals is grown and evaluated for the quantitative traits, but only the extreme 5 to 20% of the total population is scored for the molecular markers.

As pointed out by Lander and Botstein (1989), standard linear regression methods cannot be used for selective genotyping since they would grossly overestimate the phenotypic effects of the QTL's due to the biased selection of progeny. Missing-data methods have to be used, provided that the phenotypes are recorded for all the progeny and the genotypes for the non extreme progeny are entered as missing data. In general, this strategy is interesting for studies involving the simultaneous segregation of only one or two independent traits. As more traits are included in the study, it becomes less and less efficient as new extremes have to be genotyped, which can eventually result in a larger number of marker assays than the complete genotyping of the original non-extended population (Darvasi and Soller 1992).

The second strategy involves some kind of replication of the individual genotypes to be evaluated in the experiment. This strategy basically reduces environmental variation, increasing the precision of the phenotypic evaluation and keeps constant the number of genotypic marker assays. Furthermore, genotype replication allows studies of QTL x Environment interaction and facilitates the exchange of genetic material among researchers. Some types of genotype replications commonly used include: F3 families or dihaploid lines in autogamous plants, progeny tests in outcrossed plants and animals (e.g. the 'granddaughter design' proposed by Weller et al. 1990) and vegetative propagules in species that can be propagated asexually (e.g. tissue culture, rooted cuttings etc.). If all the additive genetic variation can be accounted for by the markers, an additional replication of a clone increases the power of detection in a proportion equivalent to adding another progeny individual (Knapp and Bridges 1990). The use of clonal replication is particularly advantageous for traits of low heritability (0.1 to 0.2). Bradshaw and Foster (1992) estimated that in these cases, reductions of 50 to 70% in the number of individuals could be theoretically achieved, by using six clonal propagules per genotype, keeping the same statistical power of QTL detection.

The perspective of using molecular markers as an instrument for early selection in tree breeding has stimulated QTL mapping experiments in some of the main forest tree species. Three-generation pedigrees and RFLP markers were originally used for this purpose. Groover et al. (1993) identified QTL's controlling wood specific gravity in Pinus taeda. QTL's for phenological and growth related traits were also identified in an

interspecific cross of poplar (Bradshaw 1993). The kinds of pedigrees used in these studies, however, are not the ones typically generated in tree breeding programs. Two-generation pedigrees involving half-sib families obtained by polymix cross or open pollination, and full-sib families are commonly available or can be easily produced in a program. Alternative strategies for map construction were recently extended to QTL mapping in *Eucalyptus grandis* and *E. urophylla*. The pseudo-testcross analysis combined with clonal replication of genotypes, was used to investigate the genetic basis of vegetative propagation ability. Eleven genomic regions were detected accounting for 97% of the variation in *in vitro* micropropagation response, eight regions accounted for 71% of the variation in coppice sprouting ability and eight regions accounted for 70% of the adventitious rooting response. A second experiment involved QTL mapping of productivity related traits at harvest age (6.5 years) using an existing maternal half-sib family obtained by open pollination. Although only the maternal contribution was under scrutiny, four QTL's for volume growth (accounting for about 20% of the total within-family phenotypic variation) and four for wood specific gravity (36% of the variation) were mapped. The effect of an allelic substitution at these QTL's caused an upward shift of 0.3 to 0.66 phenotypic standard deviations. These results support the existence of major genes controlling quantitatively inherited, economically important traits with strong alleles of general combining ability effect (Grattapaglia and Sederoff 1993; Grattapaglia *et al.* 1994).

Strategies utilizing paternal half-sib families obtained by artificial insemination have been also applied for QTL mapping in cattle (Beever *et al.* 1990). The half-sib analysis depends on the ability to uniquely identify the marker alleles derived from the common progenitor, and detects exclusively the effect of QTL's in an heterozygous state. In conifers, the identification of the maternal genetic contribution can be easily obtained with the analysis of the haploid megagametophyte because it carries exactly the same genetic constitution as the fertilized ovule that gives rise to the embryo scored for the quantitative trait (Grattapaglia *et al.* 1992c). In most species haploid megegametophytes are not biologically available. A more general and equally effective alternative was used in the QTL mapping studies in Eucalyptus, where the high polymorphism detection power of the RAPD technology was used to preselect markers heterozygous in the common maternal progenitor and for which the plus allele (presence of the band) was absent or at very low frequencies in the pollen parents. With this set of markers, that displayed the expected 1:1 segregation in the half-sib

family, the maternal marker and associated QTL contribution could be uniquely identified (Grattapaglia and Sederoff 1993; Grattapaglia *et al.* 1994).

3.6 Marker assisted breeding

The introduction of DNA marker technologies and the ability to do detailed analysis of genetic linkage, have raised anew the debate on molecular marker assisted breeding strategies. While this possibility has been explored for simply inherited traits, the effective practice of marker assisted selection for quantitative traits still represents a very rarely tested hypothesis even in crops like maize and tomato where several QTL mapping studies have already been carried out.

An argument frequently raised by some is that the improvement of simply inherited traits does not require markers for selection, because the phenotypes can be easily scored. Although this argument is correct in principle, the use of molecular markers linked to genes for disease and pest resistance can have an important impact in the selection of resistant individuals. Indirect selection based on markers can be an important strategy when breeding for resistance to an exotic and not yet introduced pathogen in a region or country. As the handling of the pathogen or pest is prohibited by quarantine laws, marker assisted selection can effectively accelerate the generation of resistant materials. Another case would be when resistance is desired to two or more systemic diseases. One of the biggest problems in such cases is the handling and evaluation of the interactions of the pathogens infecting the same individual. Simultaneous infection frequently makes it complicated or impossible to simultaneously select genotypes with multiple resistances (Ferreira 1993). Finally, indirect marker assisted selection could be a useful tool in cases when the phenotypes are difficult to score or when there is a need to score the trait at an early age.

3.6.1 Backcross breeding

The use of molecular markers assisted backcrossing is probably the most concrete current application of this technology in plant breeding. Markers closely linked to the genes to be introgressed from a wild progenitor to a commercial line or from one line to another (backcross line conversion) are used to monitor and select for their presence of those genes in generations of backcrossing. At the same time, selection is also

performed based on molecular marker genotypes of the recurrent parent with the objective of recovering the recurrent genome. The concept of "graphical genotypes" (Young and Tanksley 1988) was introduced in this context where individuals were selected based on the proportion of recurrent genome as measured by molecular marker, significantly reducing the number of generations necessary for the development of new cultivars or lines.

Markers have been used for the introgression of viral resistance in tomato (Young and Tanksley 1989), tissue culture regeneration ability in maize (Armstrong *et al.* 1992) and have been proposed for gene introgression in domestic animals with the use of minisatellite markers (Hillel *et al.* 1990). Paterson *et al.* (1988) suggested that the use of markers could reduce by half the number of backcross generations. Young and Tanksley (1989) estimated that by selecting for the recurrent genome using RFLP markers, an introgressed segment could be reduced in two generations to a size that would require 100 generations of backcrossing without markers.

A common approach in marker aided backcross breeding is to select as early as possible and as strongly as possible on markers very close to the introgressed gene, generally without paying much attention to the non-carrier chromosome. However, Hospital *et al.* (1992) used computer simulations to show that this is not always the best strategy. They investigated the use of markers to accelerate the recovery of the recipient genome during introgression. Their results indicate that the segment surrounding the introgressed gene is better controlled by distant markers unless high selection intensity can be applied. Selection on this segment first can reduce the selection intensity available for selection on non-carrier chromosomes. They conclude that markers may lead to a gain in time of two generations, but that it does not seem possible to simultaneously optimize selection for both the carrier and non-carrier chromosomes without a compromise as to the efficiency of selection for each part of the genome.

Dudley (1993) correctly points out that marker assisted backcross is appropriate for traits controlled by a small number of loci. As the number of loci segregating for the trait increases, the number of backcross individuals that need to be grown to have a high probability of recovering a combination of favorable marker alleles at all loci also increases dramatically. Also, a larger proportion of unwanted donor genome will be carried by "linkage drag" at all the loci, making it more difficult to efficiently select against it. The use of flanking markers has the advantage of providing a greater

precision during selection, however it reduces the probability of finding all the favorable marker alleles in one individual. Tanksley *et al.* (1989) predicted that the use of molecular markers to access and transfer genes from exotic germplasm to commercial varieties would rapidly become the most significant contribution of this technology in breeding of cultivated crops. A combination of marker assisted backcross line conversion and embryo rescue was used to deploy a new maize hybrid carrying a particular disease resistance in a record time of one year (M. Ragot, Ciba-Geigy, personal communication).

3.6.2 Development of inbred lines

Marker assisted selection can be used during inbreeding in an analogous manner to the backcross, i.e. by reducing the number of generations of selfing or consanguineous matings necessary to achieve homozygosity (fixation) at the desired QTL loci. Simultaneously, selection against recessive detrimental alleles (negative QTL's) at overall fitness QTL's could also be accomplished which could significantly contribute to the selection of better lines for seed production purposes. The efficiency of such a procedure will again depend on the number of QTL's that need to be selected, and the frequency of recombination between marker and QTL. Murray *et al.* (1988) demonstrated the potential utility of this procedure for the selection of lines with resistance to maize dwarf mosaic virus. They identified seven genomic regions associated with the resistance. Then they used this linkage information to account for 95% of the variation in resistance among lines selfed out of the F3 generation, resulting in the selection of lines at an early stage of inbreeding.

3.6.3 Improvement of quantitative traits

Potentially, the greatest impact of marker assisted breeding techniques is expected for the improvement of complex polygenic traits. Indirect selection based on markers has to be evaluated on a case by case basis considering simultaneously the selection intensity, heritability, genetic correlations, length of a breeding cycle (selection and recombination) and the relative cost of each alternative. The efficiency of selection for traits of low heritability can be significantly improved with the use of markers whose heritability approaches 1. On the other hand, the detection of QTL's for such complex

traits is intrinsically more challenging under the assumptions of quantitative genetics, where a large number of genes of small and equal effect are assumed to be controlling the trait. The analysis of a very large number of individuals becomes necessary to achieve sufficient statistical power of detection.

Lande and Thompson (1990) analyzed the efficiency of marker assisted selection in the improvement of quantitative traits. They proposed the use of selection indices where the sum of the additive effects at the QTL's would be used as a molecular score which would in turn be combined with phenotypic information at the individual level and at the family mean level. The efficiency of this approach would depend on the heritability of the character, the proportion of the additive variance associated with the marker loci and the selection scheme. When the proportion of the additive variance explained by markers exceeds the heritability of the character, selection on the markers alone is more efficient than selection on the individual phenotype (Smith 1967). In a recent review, Dudley (1993) discussed other approaches that have been suggested for using molecular marker/QTL linkage data in the practice of selection for quantitative traits.

Lande and Thompson (1990) recognized three limitations for the potential utility of markers in applied breeding: (1) the number or markers loci necessary to detect significant associations with QTL's; (2) sample sizes needed to detect QTL's for traits of low heritability and (3) sampling error in the estimation of relative weights in the selection index combining molecular and phenotypic information. On the other hand they also indicate that (1) hybridization between divergent lines is a powerful mechanism for generating linkage disequilibria (associations) between markers and trait loci; (2) that the number of marker loci scored can be greatly reduced in subsequent generations following hybridization, by neglecting those that initially were not significantly associated with the QTL's; (3) the number of individuals needed to detect substantial additive genetic variance associated with markers ranges from a few hundred to a few thousands depending on the heritability and the number of QTL's involved and (4) that the loss of efficiency due to incorrect estimation of index weights will be quite small when similar sample sizes are used to estimate these parameters. Although they recognize that many details on the use of marker assisted selection still need research, they conclude that their results support the fact that molecular markers can be used to achieve substantial increases in the efficiency of artificial selection. Finally, they recognize that due to the sample sizes required, better technologies in the assays of

molecular markers are needed to accomplish such work, and predicted that such technologies would be developed in the near future. Since then, new and improved technologies have come about that allow a much more rapid acquisition of marker data (see chapter II).

Other factors to be considered when contemplating marker assisted breeding include: (1) the repeatibility of the linkage information between marker and QTL across families; (2) QTL x Environment interaction; (3) simultaneous selection for multiple traits. Experimental results in maize and tomato are controversial in relation to the first two factors. Beavis *et al.* (1991) found a few QTL in common among four different maize populations. On the other hand, Lee *et al.* (1991) found seven QTL's in common for resistance to european corn borer, of which four mapped to regions previously identified in independent experiments. Stuber *et al.* (1992) found little evidence of QTL x Environment interaction for grain yield, a trait that typically shows significant genotype by environment (GxE) interaction. However Paterson *et al.* (1991) verified that out of 29 QTL's identified for several traits in tomato, only four were common across three divergent environments, 10 across two environments and 15 expressed specifically in each environment.

Significant QTLx Environment interactions are not particularly surprising because GxE is routinely found in breeding experiments and varies with different traits. Therefore, this type of interaction is not a peculiarity of marker assisted breeding, but rather it has to be taken into account in any form of artificial selection. In an analogous manner, the simultaneous improvement of multiple traits using markers faces the same challenges presented in conventional breeding and requires similar choices of priorities of one trait over another. However, the access that markers provide to the individual components of each trait and the knowledge of the expected magnitude of effect associated with each one of them will certainly allow for increased efficiency both when GxE is strong and when multiple trait breeding is the objective.

The experimental results of marker assisted selection for quantitative traits are still limited and controversial (Dudley 1993). For grain yield in maize, Stuber and Edwards (1986) originally compared selection with 15 isozyme markers with phenotypic selection and concluded that the gain was the same. Johnson (1991) suggested that selection of S4 lines for yield based on markers and S_2 testcross performance was more efficient than selection based on testcross performance alone. Stromberg (1992) verified that selection with markers among F_2 plants and subsequent selection among S_4 plants

within lines derived from the selected F_2 plants was as efficient as selecting based on F_2 testcross means.

3.6.4 Breeding of perennial forest and fruit species

The use of molecular markers as a supplemental tool in breeding long lived species is a very appealing possibility. When taking into account the time necessary to complete a generation of breeding, an efficient way to carry out early selection could result in a substantial increase in genetic gain per unit time. This perspective makes the genetic improvement of perennial species the applied area where marker assisted breeding technologies have the highest chances of making a significant impact in productivity.

Recently the first experimental results in the area demonstrated that QTL's can be detected controlling economically important quantitative traits in some of the major tree species (Groover *et al.* 1993; Bradshaw 1993; Grattapaglia and Sederoff 1993; Grattapaglia *et al.* 1994). The identification of some major effect QTL's explaining considerable proportions of the total variation in all the traits and species, suggests that the architecture of quantitative traits such as volume growth and wood specific gravity in trees have a genetic structure similar to those of polygenic traits investigated in annual crops. Also, QTL mapping is possible on two-generation pedigrees commonly available in breeding programs. This opens the possibility of using existing families in retrospective QTL analyses (Grattapaglia *et al.* 1993).

The main argument traditionally raised against the feasibility of marker assisted breeding in forest and fruit species is the expected absence of linkage disequilibria between marker and trait loci (Beckmann and Soller 1986; Strauss *et al.* 1992). Population genetics theory maintains the view that in large random mating populations the alleles at QTL's and alleles at marker loci would be randomly associated in different individuals, so that it would be impossible to establish significant marker/QTL associations at the population level. Breeding populations of annual crop plants on the other hand have characteristics that generate high levels of linkage disequilibrium and facilitate the detection of significant associations. These include: restricted genetic base; autogamous reproductive habit; or when naturally allogamous the availability of inbred lines and pedigrees where significant amounts of linkage disequilibrium can be generated.

The assumption of linkage equilibrium is based on data coming from natural populations of forest trees, which more closely resemble the idealized population model. Also they are based on a few and sparsely distributed isozyme loci, typically physically unlinked. However, breeding populations of forest trees, are significantly more reduced in size, are frequently structured in subpopulations, are under artificial selection and often involve high degrees of distant hybridizations both between species or provenances in their initial establishment. Genetic drift, selection and hybridization are all powerful forces to create significant linkage disequilibrium between loci. Also, the number of markers that can be assayed at the DNA level are much larger than isozymes, and consequently physical linkages between loci can be detected at a much higher efficiency and closer distances. The assumption of lack of disequilibrium might not be a realistic one in the context of breeding populations with more powerful marker technologies available.

However, even if the assumption of lack of disequilibrium proves true, an alternative and powerful approach is possible today. This approach involves the construction of genetic linkage maps of molecular markers for individual genotypes. Close linkages established between markers and QTL's at the individual level could be followed for several subsequent generations of selection and recombination. This possibility could never have been considered before because the marker techniques that existed until a few years back presented a major technological barrier for the rapid acquisition of marker data. Furthermore, several forest geneticists followed the orthodox concept of a unique reference linkage map for the species, derived from the current thinking in the annual crop QTL mapping arena, where breeding is based on a very restricted genetic base. The introduction of RAPD markers (Williams *et al.* 1990) and the construction of single-tree linkage maps in conifers (Grattapaglia *et al.* 1991; Tulsieram *et al.* 1992a) and angiosperms (Grattapaglia and Sederoff 1992) has provoked a reevaluation of the perspectives of marker assisted breeding in forest trees, and the suggestion of potential strategies for its implementation (Grattapaglia *et al.* 1992c,1993).

This reevaluation is based on the heterodox proposal of constructing linkage maps for every individual tree of an elite breeding population. The use of small elite populations of between 4 and 30 individuals is becoming more common in forest tree breeding, especially in more advanced programs, both in the form of sublines (McKeand and Bridgwater 1992), nucleus breeding (Cotterill *et al.* 1988) or superlines (White *et al.* 1993). These strategies involve a larger concentrartion of efforts in selection

and recombination in these elite groups, aiming at rapid gains. The integration of more sophisticated selection technologies would therefore be appropriate to these new breeding strategies. QTL mapping in each individual tree would be done by analyzing the performance of their progeny. Large progeny sizes and many marker assays would be required in the initial detection step to achieve sufficient power for detection. In subsequent generations, however, the number of markers would be substantially reduced as only those particular marker segments containing the QTL's of interest would be tracked. Progeny sizes could then vary, not so much for statistical power purposes but more for increasing the probability of recovering genotypes with the correct QTL allele profiles. The size will depend on the objectives of selection, especially at the within-family level.

An important aspect of such a proposal is that not only would individual maps be constructed but also individual QTL's would be identified, i.e. no population level QTL's are assumed. Although such "population level" or general quantitative trait loci (QTL) should exist, their relative importance in the overall level of genetic variation in quantitative traits in forest trees is still unclear. The identification and manipulation of QTL's specific to individual trees might emerge as being more important for the advancement of quantitative traits by marker assisted breeding. For a similar phenotypic expression, it is likely that different QTL's can act in different ways in different individuals, depending on the genetic background and the kinds of selective pressures that the individuals have been subject to. Furthermore, linkage relationships between markers and QTL's vary. Experimental evidences of inconsistency of QTL expression across populations have been reported in maize and tomato (Beavis *et al.* 1991; Tanksley and Hewitt 1988) suggesting that population level QTL's should be seen with caution.

However, when marker/trait associations have been established at the individual level, substantial linkage disequilibria are expected to be maintained between selectively neutral loci with recombination distance $r < 1/T$ where T is the number of generations of random mating following the initial hybridization (Kimura and Ohta 1971). Also genetic drift in a randomly mating population of size Ne is expected to produce substantial associations between loci with recombination rates $r < 1/(4Ne)$ (Hill and Robertson 1968). So, apart from the effects of selection which will be operating to establish associations, if we assume for example 3 generations of random mating since the establishment of a tree breeding populations based on hybridizations, even sparsely

dense maps (average distance of 30% recombination) should be useful for detecting associations. Also, regardless of hybridization, in small breeding subpopulations of 5 individuals substantial linkage disequilibria could be expected at distances around 5% recombination. Therefore, given that close linkages are established between markers and QTL in the mapping phase, the decay of marker-trait associations with time, would not be of immediate concern in the context of the long generations of tree breeding.

As the associations between markers and QTL's are established in each individual, two possibilities exist for the use of this information: (1) retrospective selection of progenitors based on some kind of selection index that combines phenotypic information at the family level and the QTL profile of each potential progenitor. The objective would be to select parents to synthesize new crosses where a larger number of favorable alleles could be recovered in single progeny genotypes; (2) prospective individual tree selection at high intensity within families, using the added precision of QTL mapping information. In species where non-additive genetic effects are important, and it is possible to capture such effects through vegetative propagation, (e.g. *Eucalyptus*, poplar, sweetgum) the QTL mapping information could become especially important to increase selection precision of superior individuals.

A significant impact of the QTL mapping information could be realized in early selection, both in the context of recurrent selection in small elite populations, as well as for the selection of individuals for vegetative propagation. In recurrent selection, selected individuals could be recombined more rapidly to produce the next generation, potentially increasing the genetic gain per unit time. In the selection of clones, the possibility of practicing an intense preliminary selection at a very juvenile stage, would circumvent the common problem of loss of regeneration or adventitious rooting ability with physiological phase change. Preliminary selection of individuals would be carried out based on QTL allele composition, substantially reducing the number of selected clones. Selected individuals could then be micropropagated or kept as juvenile hedges immediately. In a first stage of this practice, results of clonal trials in the field would be needed to corroborate the preliminary selection. In later stages of the program, once validation and prediction experiments had been accomplished, a strict selection on QTL's could be implemented and selected individuals could be immediately deployed as clones.

Another interesting option would be indirect selection for traits that are difficult to evaluate, such as adventitious rooting response or wood properties (e.g. cellulose pulp

yield, fiber characteristics). These traits require lengthy and costly procedures for phenotypic evaluation in greenhouse or laboratory. Similarly indirect/early selection for fruit quality traits or disease and pest resistance would have great impact in fruit crop breeding.

Finally, molecular markers could be very useful as an aid in the development of inbred lines in forest trees. Such a strategy was considered to be impractical in forest tree improvement because of the low vigor and poor seed set of inbreds and the resultant drastic reduction in the size of the breeding population (Franklin 1969). However the promotion of different levels of inbreeding has recently been introduced into advanced breeding programs as a supplemental mechanism to achieve fixation of favorable alleles and removal of genetic load (McKeand and Bridgwater 1992; M. Carson, personal communication). The experience in maize demonstrated that the production of inbred lines is a very lengthy and difficult task, however extremely high productivities can be realized in the hybrids. The phenomenal growth of several interspecific poplar and eucalypt hybrids suggests that dominant gene actions is important in the expression of phenotypic superiority. The availability of partially endogamous lines could allow a better exploitation of such effects. Crosses between unrelated S2 individuals (second generation selfs) of *Eucalyptus grandis* in Brazil, have shown remarkable growth and uniformity (R. Vencovsky, personal communication). Molecular markers tightly linked to favorable alleles would be used to monitor the inheritance and fixation of favorable alleles in generations of selfing or consanguineous matings and to select desired genotypes. At the same time, genomic regions displaying segregation distortions could allow the identification of detrimental lethals or semi-lethals at the seedling stage, facilitating their early removal from the population.

3.7 Map-based or positional cloning

The identification and cloning of economically important genes is still a very difficult task. The classic paradigm of reverse genetics involves the identification of a gene product, typically an enzyme, its purification to homogeneity, the production of a specific monoclonal antibody to the protein and the use of this antibody to screen a complementary DNA (cDNA) expression library to identify the cDNA clone that encodes the protein. Based on the sequence of the cDNA, a complete genomic clone is isolated including upstream *cis* regulatory regions. When an appropriate assay system is

available, transcription factors involved in the regulation of the gene can potentially be identified and also cloned. This procedure is elegant, however it is limited to those few cases where the relevant gene products are known. When the exact gene product of interest is not known or the phenotypic expression does not involve *a priori* a defined protein, another strategy can be used. This consist in targeting genes that encode the enzymes involved in the biosynthesis of compounds that either affect the characteristic or are structurally closely related to those compounds. Additionally, the specific *cis* and *trans* regulatory elements are needed for the correct tissue or developmental stage specific expression. For example genes that encode for the enzymes involved in the biosynthesis of specific fatty acids could be targeted with the objective of manipulating oil contents in seeds. From the breeding standpoint, such an approach is based on a rather strong assumption that once the gene(s) is cloned and introduced in a cultivated variety, its expression will effectively result in a phenotypic change, and more importantly in the desired direction. This occasionally works, when the trait is simple enough that one or a few genes can result in a significant phenotypic effect. Oligo and polygenic traits have not yet been manipulated in this manner, and this perspective seems rather difficult.

An alternative to isolate and clone genes in cases where the gene product is not known, is the use of map-based cloning strategies also called positional cloning. In this approach, the actual phenotypic effect of a gene is verified *a priori* through the measurement of the trait in segregating families. The approximate map position of the QTL in relation to molecular markers is then determined by linkage analysis. The following step is typically called fine structure mapping, and aims at identifying markers very closely linked (< 1 cM) to the mapped phenotypic effect (QTL) of interest. Different chromosomal segments that contain such markers are identified and their overlapping regions determined using all the available markers in that region. Phenotypic effects of each chromosomal segment are determined by QTL analysis in segregating progenies. Effects that are shared by different segments are attributed to QTL's in overlapping regions of those segments. Exclusive effects of a segment are attributed to exclusive portions of that segment. The resolution of this kind of mapping strategy depends on the number of molecular markers and on the availability of large numbers of progeny individuals to increase the probability of recovering informative recombination events in the segments of interest. Fine structure mapping has been accomplished in tomato, where segments of only 3 cM in length were identified containing QTL's for fruit quality

traits (Paterson *et al.* 1990). The tomato genetic map is now highly saturated with markers spaced at an average 1.2 cM (Tanksley *et al.* 1992), facilitating this kind of approach. However, 1 cM of map distance still corresponds to some hundred kilobase pairs of DNA, containing several genes. In *Eucalyptus* for example, it has been estimated that 1 cM corresponds on the average to 600 to 700 Kb of DNA (Grattapaglia and Bradshaw 1994)

To move from a fine structure map around the locus of interest to the actual gene is still a very challenging task today. Chromosome walking involves the use of large clones of genomic DNA (cosmid clones generally) that contain overlapping DNA sequences. The procedure depends on obtaining a small segment of DNA from one end of the first clone and using this segment to rescreen the library to obtain new clones, and so on until an overlapping set of clones is obtained spanning the region between the two markers flanking the locus of interest. Chromosome jumping techniques have also been used to accelerate this procedure. Special libraries are prepared where cloned fragments contain sequences that are widely separated on the chromosome. The jumping library can be screened with a probe and positive clones should contain a DNA fragment from a region the length of the jump from the probe (Watson *et al.* 1992). A powerful alternative technique that is currently being adopted is the STS approach (Sequence Tagged Sites). A STS is a unique sequence from a known location that can be amplified by PCR. The unique primer sequences provide a common language among laboratories facilitating large scale map-based cloning efforts. STS can be used for ordering clones by searching for clones common to two or more STS's. (Olson *et al.* 1989)

However, gene mapping and analysis based on overlapping cosmids is labor intensive because so many clones are needed to encompass segments of interest. Yeast Artificial Chromosomes (YAC's) offer a powerful alternative for map-based cloning experiments because fragments of several hundred kilobases and up to 1 megabases can be cloned and screened. So potentially one or two overlapping clones can be found in a YAC library that span the whole region of interest between the two flanking markers. However, even when a clone of DNA is finally isolated that contains the locus of interest, the next challenge is to identify the gene of interest among the several genes that are likely to be present along the segment. A particularly useful strategy is the use of mutants that lack the phenotypic function of interest, in combination with a plant transformation system. Subclones of the segment identified as containing the gene are

transformed and expressed into the mutant plant and restoration or complementation of function is evaluated. This procedure allows narrowing down to particular DNA sequences that precisely contain the gene desired. In the model plant *Arabidopsis thaliana*, it has been used for map-based cloning of a gene controlling biosynthesis of fatty acids (Arondel *et al.* 1992). This strategy, however, has still a limited applicability in crop plants since it requires YAC libraries, mutants and efficient plant transformation systems.

4. CONCLUSIONS

Detailed genetic analysis at the DNA level is becoming increasingly efficient. Large amounts of information about plant genomes is being generated for several cultivated species. The prospects are good for an even more dramatic explosion of knowledge in this area in the next decades. Numerous opportunities will be available for a more efficient use and management of the existing genetic variability both in the cultivated and undomesticated germplasm resources. The identification of genomic regions controlling important traits is possible not only in species traditionally amenable to mapping studies but also to essentially undomesticated species with high levels of heterozygosity such as forest and fruit trees. The impact of this technology could prove far greater for these species in view of the long generation times necessary for breeding.

The literature of genetic mapping using molecular markers includes a large number of analytical studies. Adequate experimental designs and methods for statistical analysis are now available. The technological limitations for the rapid acquisition of genotypic data for large numbers of polymorphic markers in large families have been significantly mitigated by PCR based marker technologies. On the other hand, both theory and experimental data on the implementation of marker assisted technologies for the generation of improved plant products, are still very limited.

The major challenge facing researchers in the area of molecular breeding is how to integrate the knowledge acquired through the use of marker data, in the effective practice of plant breeding. In the area of marker assisted selection, several aspects still need intensive investigation, such as: (1) the stability of QTL expression in variable genetic backgrounds and environments; (2) the final phenotypic effect following the accumulation of supposedly favorable alleles at different QTL's; (3) the potential exploration of epistatic interaction among QTL's for increased and continuous gains

through selection; (4) methodologies of practical integration of QTL mapping information in selection indices; (5) comparative estimates of realized gains through selection with and without QTL mapping information.

The practical application of this technology will depend on a cost/benefit analysis, on a case by case basis, taking into consideration the biological properties of the crop, the specific traits that need to be improved and the populations and marker techniques available. Since the initial experiments by Sax in 1923, the great promise still remains for marker assisted selection. Its greatest impact is predicted for traits of low heritability or difficult evaluation, in situations where undesirable genetic correlations exists between traits, and especially in the practice of early selection in perennial crops. The physical isolation and manipulation of individual genes that control quantitative traits of commercial importance still seems to be a distant objective, however theoretically attainable. It will be interesting to see how soon this will be realized.

5. LITERATURE CITED

Adams, W.T 1983 Application of isozymes in tree breeding. In: Isozymes in plant genetics and breeding. Part A. Tanksley, S.D. and Orton, T.J. (eds.) Elsevier Science Publishers, Amsterdam pp. 60-64.

Al-Janabi, S.M.; Honeycutt, R.J.; McClelland, M. and Sobral, B.W.S. 1993 A genetic linkage map of Saccharum spontaneum L. 'SES 208'. Genetics 134:1249-1260.

Anderson, J.A.; Sorrells, M.E. and Tanksley, S.D. 1993 RFLP analysis of genomic regions associated with resistance to preharvest sprouting in wheat. Crop Sci. 33:453-459.

Armstrong, C.L.; Romero-Severson, J. and Hodges, T.K. 1992 Improved tissue culture response of an elite maize inbred through backcross breeding, and identification of chromosomal regions important for regeneration by RFLP analysis. Theor. Appl. Genet. 84:755-762.

Arnheim, N.; Starnge, C. and Erlich, H. 1985 Use of pooled DNA samples to detect linkage disequilibrium of polymorphic restriction fragments and human disease: studies of the HLA class II loci. Proc. Natl. Acad. Sci. USA 82:6970-6974.

Arondel, V.; Lemieux, B.; Hwang, I.; Gibson, S.; Goodmann, H.M. and Somerville, C.R. 1992 Map-based cloning of a gene controlling omega-3 fatty acid desaturation in Arabidopsis. Science 258:1353-1355.

Avery, O.T.; MacLeod, C.M. and MacCarty, M. 1944 Studies on the chemical nature of the substance inducing transformation of pneumococcal types. J. Exp. med. 79:137-158.

Bailey, D.C. 1983 Isozymic variation and plant breeders rights. In: Isozymes in plant genetics and breeding. Part A. Tanksley, S.D. and Orton, T.J. (eds.) Elsevier Science Publishers, Amsterdam pp. 425-441.

Bailey, D.C. 1983. Isozymic variation and plant breeders rights. In: Tanksley S.D. and Orton T.J. ed Isozymes in plant genetics and breeding. Part A. p. 425-441. Amsterdam Elsevier.

Baird, E.; Cooper-Bland, S.; Waugh, R.; DeMaine, M. and Powell, W. 1993 Molecular characterization of inter- and intraspecific somatic hybrids of potato using randomly amplified polymorphic DNA (RAPD) markers. Mol. Gen. Genet. (in press).

Barua, U.M.; Chalmers, K.J.; Hackett, C.A.; Thomas, W.T.B.; Powell, W. and Waugh, R. 1993 Identification of RAPD markers linked to a *Rhynchosporium secalis* resistance locus in barley using near-isogenic lines and bulked segregant analysis. Heredity 71:177-184.

Bateson, W. and Punnett, R.C. 1905 Experimental studies in the physiology of heredity. Reports to Evol. Comm. Royal Soc. 2:1-131.

Beavis, W.D.; Grant, D.; Albertsen, M. and Fincher, R. 1991 Quantitative trait loci for plant height in four maize populations and their associations with qualitative genetic loci. Theor. Appl. Genet. 83:141-145.

Beckmann, J.S. 1991 Genomic genetics and plant genetic improvement. In: Gene-mapping techniques and applications. Schook, L.B.; Lewin, H.A. and McLaren D.G. (eds.) Marcel Dekker Inc., New York pp. 201-230.

Beckmann, J.S. and Soller, M. 1983 Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. Theor. Appl. Genet. 67:35-43.

Beckmann, J.S. and Soller, M. 1986 Restriction fragment length polymorphisms and genetic improvement of agricultural species. Euphytica 35:111-124.

Beever, J.E.; George, P.D.; Fernando, R.L.; Stormont, C.J. and Lewin, H.A. 1990 Associations between genetic markers and growth and carcass traits in a paternal half-sib family of Angus cattle. Animal Sci. 68:337-344.

Bentolila, S.; Guitton, C.; Bouvet, N.; Sailland, A.; Nykaza, S. and Freyssinet, G. 1991 Identification of an RFLP marker tightly linked to the *Ht1* gene in maize. Theor. Appl. Genet. 82:393-398.

Bernatzky, R. and Tanksley, S.D. 1986 Toward a saturated linkage map in tomato based on isozymes and random cDNA sequences. Genetics 112:887-898.

Binelli, G.; Gianfrancheschi, L.; Pe, M.E.; Taramino, B.; Busso, C.; Stenhouse, J. and Ottaviano, E. 1992 Similarity of maize and sorghum genomes as revealed by maize RFLP probes. Theor. Appl. Genet. 84:10-16.

Bonierbale, M.W.; Plaisted, R.L. and Tanksley, S.D. 1988 RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics 120:1095-1103.

Bonierbale, M.W.; Plaisted, R.L. and Tanksley, S.D. 1993 A test of the maximum heterozygosity hypothesis using molecular markers in tetraploid potatoes. Theor. Appl. Genet. 86:481-491.

Botstein, D.; White, R.L.; Skolnick, M. and Davis, R.W. 1980 Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32:314-331.

Bradshaw, H.D.Jr. and Foster, G.S. 1992 Marker-aided selection and propagation systems in trees: advantages of cloning for studying quantitative inheritance. Can. J. For. Res. 22:1044-1049.

Bradshaw, H.D. Jr. 1993 Genome mapping in *Populus*. Poplar Molecular Network 1:1-3.

Bradshaw, H.D.Jr. and Stettler, R.F. 1993 Molecular genetics of growth and development in Populus. I. Triploidy in hybrid poplar. Theor. Appl. Genet. 86:301-307.

Brown, A.H.D. 1989a The case for core collections. In: The use of plant genetic resources. Brown, A.H.D.; Frankel, O.H.; Marshall, d.R. and Williams, J.T. (eds.). Cambridge University Press, cambridge, pp. 136-156.

Brown, A.H.D. 1989b Core collections: a practical approach to genetic resources management. Genome 31:818-824 .

Brummer, E.C.; Bouton, J.H. and Kochert, G. 1993 Development of an RFLP map in diploid alfalfa. Theor. Appl. Genet. 86:329-332.

Burr, B.; Evola, S.V.; Burr, F.A. and Beckmann, J.S. 1983 The application of restriction fragment length polymorphism to plant breeding. In: Genetic engineering: principles and methods. Setlow, J.K. and Hollaender, A. (eds.). Plenum Press, Vol. 5, pp. 45-59.

Burr, B.; Burr, F.A.; Thompson, K.H.; Albertson, M. and Stuber, C.W. 1988 Gene mapping with recombinant inbreds in maize. Genetics 118:519-526.

Caetano-Anólles, G.; Bassam, B.J. and Gresshoff, P.M. 1991a High resolution DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. Biotechnology 9:553-557.

Caetano-Annóles, G.; Bassam, B.J. and Gresshoff P.M. 1991b DNA Amplification Fingerprinting: a strategy for genome analysis. Plant Mol. Biol. Rep. 4:294-307.

Carlson, J.E.; Tulsieram, L.K.; Galubitz, J.C.; Luk, V.W.K.; Kauffeldt, C. and Rutledge, R. 1991 Segregation of random amplified DNA markers in F_1 progeny of conifers. Theor. Appl. Genet. 83:194-200.

Chang, C.; Bowman, J.L.; DeJohn, A.W.; Lander, E.S. and Meyerowitz, E.M. 1988 Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. (USA) 6856-6860.

Chaparro, J.X.; Werner, D.J.; O'Malley, D.M. and Sederoff, R. 1994 Targeted mapping and linkage analysis of morphological, isozyme and RAPD markers in peach. Theor. Appl. Genet. (in press).

Cheliak, W.M.; Yeh, F.C.H. and Pitel, J.A. 1987 use of electrophoresis in tree improvement programs. For. Chron. 63:89-96.

Chevreau, E. and Laurens, F. 1987 The pattern of inheritance in apple: further results from leaf isozyme analysis. Theor. Appl. genet. 75:90-95.

Chyi, Y.-S.; Hoenecke, M.E. and Semyk, J.L. 1992 A genetic linkage map of restriction fragment length polymorphism loci for *Brassica rapa* (syn. *campestris*). Genome 35:746-757.

Conkle, M.T. 1981 Isozyme variation and linkage in six conifer species. In: Proceedings of the Symposium on Isozymes in North American Forest Trees and Forest Insects. Pacific Southwest Exp. St. USDA Genet. tech. Rep. PSW-48. pp. 11-17.

Cotterill, P.P.; Dean, C.A.; Cameron, J.N. and Brindbergs, M.L. 1988 Nucleus breeding: a new strategy for rapid improvement under clonal forestry. In: Proceedings of IUFRO International Conference "Breeding tropical trees" Section 2.02-08 Pattaya, Thailand.

Da Silva, J.A.G.; Sorrels, M.E.; Burnquist, W.L. and Tanksley, S.D. 1993 RFLP linkage map and genome analysis of *Saccharum spontaneum*. Genome 36:782-791.

Dallas, J.F. 1988 Detection of DNA "fingerprints" of cultivated rice by hybridization with a human minisatellite DNA probe. Proc. Nat. Acad. Sci. (USA) 85:6831-6835.

Darvasi, A. and Soller, M. 1992 Selective genotyping for determination of linkage between marker locus and a quantitative trait locus. Theor. Appl. Genet. 85:353-359.

Devey, M.E.; Delfino-Mix, A.; Donaldson, D.; Kinloch, B.B.Jr. and Neale, D.B. 1993 Efficient mapping of a gene for resistance to white pine blister rust in sugar pine. (submitted).

Diers, B.W.; Mansur, L.; Imsande, J. and Shoemaker, R.C. 1992 Mapping Phytophthora resistance loci in soybean with restriction fragment length polymorphism markers. Crop Sci. 32:377-383.

Dudley, J.W.; Saghai Maroof, M.A. and Rufener, G.K. 1991 Molecular markers and grouping of parents in maize breeding programs. Crop Sci. 31:718-723.

Dudley, J.W.; Saghai Maroof, M.A. and Rufener, G.K. 1992 Molecular marker information and selection of parents in corn breeding programs. Crop Sci. 32:301-304.

Dudley, J.W. 1993 Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. Crp. Sci. 33:660-668.

Durham, R.E.; Liou, P.C.; Gmitter Jr, F.G. and Moore, G.A. 1992 Linkage of restriction fragment length polymorphisms and isozymes in *Citrus*. Theor. Appl. Genet. 84:39-48.

Dweikat, I.; Mackenzie, S.; Levy, M. and Ohm, H. 1993 Pedigree assessment using RAPD-DGGE in cereal crop species. Theor. Appl. Genet. 85:497-505.

Echt, C.S.; Kidwell, K.K.; Osborn, T.C.; Knapp, S.J. and McCoy, T.J. 1993 Linkage mapping in diploid alfalfa (*Medicago sativa*). Genome (in press).

Edwards, M.D.; Stuber, C.W. and Wendel, J.F. 1987 Molecular-marker-facilitated investigations of quantitative-trait loci in maize. I - Numbers, genomic distribution and types of gene action. Genetics 116:113-125.

Edwards, M.D.; Helentjaris, T.; Wright, S. and Stuber, C.W. 1992 Molecular-marker facilitated investigations of quantitative trait loci in maize. 4 Analysis based on genome saturation with isozyme and restriction fragment length polymorphism markers. Theor. Appl. Genet. 83:765-774.

Ellis, T.H.N.; Turner, L.; Hellens, R.P.; Lee, D.; Harker, C.L.; Enard, C.; Domoney, C. and Davies, D.R. 1992 Linkage maps in pea. Genetics 130:649-663.

Emerson, R.A.; Beadle, G.W. and Fraser, A.C. 1935 A summary of linkage studies in maize. Cornell Univ. Agr. Exp. Sta. Memoir 180.

Falconer, D.S. 1988 Introduction to quantitative genetics. New York, Longman.

Ferreira, M.E. 1993 Linkage mapping of molecular markers and loci associated with flower induction and disease resistance in *Brassica napus*. Ph.D. thesis. University of Wisconsin, Madison, WI, USA.

Frankel, O.H. 1984 Genetic perspectives of germplasm conservation. In: Arber, W.; Limensee, K.; Peacock, W.J. and Starlinger, P. (eds.) Genetic manipulation: impact on man and society. Cambridge Univ. Press., Cambridge, pp. 161-170.

Franklin, E.C. 1969 Inbreeding as a means of genetic improvement of loblolly pine. Proceeding of the. X Southern Forest Tree Improvement Conference, Houston (TX) pp.107-115.

Frei, O.M.; Stuber, C.W. and Goodman M.M. 1986 Uses of allozymes as genetic markers for predicting performance in maize single-cross hybrids. Crop Sci. 26:37-42.

Gale, M.D.; Chao, S. and Sharp, P.J. 1990 RFLP mapping in wheat - progress and problems. In: Gene manipulation in plant improvement. Vol. 2. Gustafson J.P. (ed). Plenum Press, New York, pp. 353-364.

Gebhart, C.; Ritter, E.; Barone, A.; Debener, T.; Walkemeier, B.; Schachtschabel, U.; Kaufmann, H.; Thompson, R.D.; Bonierbale, M.W.; Ganal, M.W.; Tanksley, S.D. and

Salamini, F. 1991 RFLP maps of potato and their alignment with the homeologous tomato genome. Theor. Appl. Genet. 83:49-57.

Gebhart, C.; Ritter, E.; Debener, T.; Schachtschabel, U.; Walkemeier, B.; Uhrig, H. and Salamini, F. 1989 RFLP analysis and linkage mapping in *Solanum tuberosum*. Theor. Appl. Genet. 78:65-75.

Gill, K.S. and Gill, B.S. 1991 A DNA fragment mapped within the submicroscopic deletion of Ph1, a chromosome pairing regulator gene in polyploid wheat. Genetics 129:257-259.

Giovannoni, J.J.; Wing, R.A.; Ganal, M.W. and Tanksley, S.D. 1991 Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. Nucl. Acid. Res. 19:6553-6558.

Godshalk, E.B.; Lee, M. and Lamkey, K.R. 1990 Relationship of restriction fragment length polymorphisms to single-cross hybrid performance of maize. Theor. Appl. Genet. 80:273-280.

Grattapaglia, D.; Wilcox, P.; Chaparro, J.X.; O'Malley, D.; McCord, S.; Whetten, R.; McIntyre, L. and Sederoff, R. 1991 A RAPD map of loblolly pine in 60 days. Third International Congress of Intl. Soc. for Plant Mol. Biol. abs. 2224.

Grattapaglia, D., O'Malley, D.M., Sederoff, R.R. 1992a. Multiple applications of RAPD markers to genetic analysis of *Eucalyptus* sp In: Proceedings of IUFRO International Conference "Breeding tropical trees" Section 2.02-08 Cali, Colombia, pp. 436-450.

Grattapaglia, D., O'Malley, D.M., Dvorak, W. 1992b Phylogenetic analysis of Central American and Mexican pines using RAPD markers on bulked DNA samples. In: Proceedings of IUFRO International Conference "Breeding tropical trees" Section 2.02-08 Cali, Colombia, pp. 132-147.

Grattapaglia, D., Chaparro, J., Wilcox, P., MCCord, S., Werner, D., Amerson, H., MCKeand, S., Bridgwater, F., Whetten, R., O'Malley, D., Sederoff, R. 1992c. Mapping in woody plants with RAPD markers: applications to breeding in forestry and horticulture. In: Proceedings of the Symposium on Applications of RAPD Technology to Plant Breeding. CSSA/ASHS/AGA. pp. 37-40.

Grattapaglia, D. and Sederoff, R.R. 1992. Pseudo-testcross mapping strategy in forest trees: single tree RAPD maps of *Eucalyptus grandis* and *E. urophylla*. Abstracts of the International Conference on the Plant Genome I, San Diego, CA poster # 47.

Grattapaglia, D. and Sederoff, R.R. 1993. QTL mapping in *Eucalyptus* using pseudo-testcross RAPD maps, half and full-sib families . Proceedings of the XXII Southern Forest Tree Improvement Conference, Atlanta GA, p. 487.

Grattapaglia, D., Chaparro, J., Wilcox, P., MCCord, S., Crane, B., Amerson, H., Werner, D., Liu, B.H., O'Malley, D.; Whetten R., MCKeand, S., Goldfarb, B., Greenwood, M., Kuhlman, G., Bridgwater, F. and Sederoff, R. 1993. Application of genetic markers to tree breeding. In: Proceedings of the XXII Southern Forest Tree Improvement Conference, Atlanta GA , pp. 452-463.

Grattapaglia D. and Bradshaw, H.D., Jr. 1994 Nuclear DNA content of commercially important *Eucalyptus* species and hybrids. Can. J. For. Res. (in press).

Grattapaglia, D. and Sederoff, R.R. 1994 Genetic linkage maps of *Eucalyptus grandis* and *E. urophylla.* using a pseudo-testcross strategy and RAPD markers. Genetics (in press).

Grattapaglia, D., Bertolucci, F.L., Penchel, R. and Sederoff, R.R. 1994. Genetic mapping of QTL's controlling volume growth, vegetative propagation and wood quality traits in *Eucalyptus grandis.* and *E. urophylla.* Abstracts of Plant Genome II, The International Conference on the Plant Genome. San Diego CA. poster # 71.

Groover, A.T.; Devey, M.E.; Fiddler, T.A.; Lee, J.M.; Megraw, R.A.; Mitchell-Olds, T.; Sherman, B.K.; Vujcic, S.L.; Williams, C.G. and Neale, D.B. 1993 Genetic mapping of quantitative trait loci influencing wood specific gravity in loblolly pine (*Pinus taeda*). In: Proceedings of the XXII Southern Forest Tree Improvement Conference, Atlanta GA (in press).

Grosser, J.W. and Gmitter, F.G. 1990 Somatic hybridization of *Citrus* with wild relatives for germplasm enhancement and cultivar development. HortScience 25:147-151.

Haldane, J.B.S.; Sprunt, A.D. and Haldane, N.M. 1915 Reduplication in mice. J. Genet. 5:133.

Haley, S.D.; Mikals, P.N.; Stavely, J.R.; Byrum, J. and Kelly, J.D. 1993 Identification of RAPD markers linked to a major rust resistance gene block in common bean. Theor. Appl. Genet. 86:505-512.

Hallauer, A.R. and Miranda, J.B. 1988 Quantitative genetics in maize breeding. 2nd edition. Iowa State University Press, Ames, USA.

Halward, T.; Stalker, T.; LaRue, E. and Kochert, G. 1992 Use of single-primer DNA amplifications in genetic studies of peanut (*Arachis hypogaea* L.) Plant Mol. Biol. 18:315-325.

Havey, M.J. and Muehlbauer, F.J. 1989 Linkages between restriction fragment length, isozyme, and morphological markers in lentil. Theor. Appl. Genet. 77:395-401.

Helentjaris, T.; King, G.; Slocum, M.; Siedenstrang, C. and Wegman, S. 1985 Restriction fragment length polymorphisms as probes for plant diversity and their development as tools for applied plant breeding. Plant Mol. Biol. 5:109-118.

Helentjaris, T.; Weber, D.F. and Wright, S. 1986a Use of monosomics to map cloned DNA fragments in maize. Proc. Natl. Acad. Sci. USA 83:6035-6039.

Helentjaris, T.; Slocum, M.; Wright, S.; Schaefer, A. and Nienhuis, J. 1986b Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphisms. Theor. Appl. Genet. 72:761-769.

Hershey, A.D. and Chase, M. 1952 Independent functions of viral protein and nucleic acid in growth of bacteriophage. J. Gen. Physiol. 36:39-56.

Hill, W.G. and Robertson, A. 1968 Linkage disequilibrium in finite populations. Theor. Appl. Genet. 38:226-231.

Hospital, F.; Chevalet, C. and Mulsant, P. 1992 Using markers in gene introgression breeding programs. genetics 132: 1199-1210.

Hillel, J.; Schaap, T.; haberfeld, A.; Jeffreys, A.J.; Plotzky, Y.; Cahaner, A. and lavi, U. 1990 DNA fingerprints applied to gene introgression in breeding programs. genetics 124:783-789.

Hu, J. and Quiros, C.F. 1991 Identification of broccoli and cauliflower cultivars with RAPD markers. Plant Cell Rep. 10:505-511.

Isabel, N.; Tremblay, L.; Michaud, M.; Tremblay, F.M. and Bousquet, J. 1993 RAPDs as an aid to evaluate the genetic integrity of somatic embryogenesis-derived populations of Picea mariana (Mill.)B.S.P. Theor. Appl. Genet. 86:81-87.

Johnson, G.R. 1991 RFLP assisted early generation selection for advanced generation testcross performance. In: Proceedings 27th Illinois Corn Breeders School, Champaign.

Jondle, R.J. 1992 Legal aspects of varietal protection using molecular markers. In: Proceedings of the Symposium on Applications of RAPD Technology to Plant Breeding. CSSA/ASHS/AGA. pp 50-52.

Jung, C.; Pillen, K.; Frese, L.; Fahr, S. and Melchinger, A.E. 1993 Phylogenetic relationships between cultivated and wild species of the genus Beta revealed by DNA 'fingerprinting'. Theor. Appl. Genet. 86:449-457.

Keim, P.; Shoemaker, R.C. and palmer, R.G 1989 Restriction fragment length polymorphism diversity in soybean. Theor. Appl. Genet. 77:786-792.

Keim, P.; Diers, B.W.; Olson, T.C. and Shoemaker, R.C. 1990 RFLP mapping in soybean: association between marker loci and variation in quantitative traits. Genetics 126:735-742.

Kianian, S.F. and Quiros, C.F. 1992 Generation of a Brassica oleracea composite RFLP map: linkage arrangements among various populations and evolutionary implications. Theor. Appl. Genet. 84:544-554.

Kimura, M. and Ohta, T. 1971 Theoretical aspects of population genetics. Princeton University Press, Princeton NJ.

Klein-Lankhorst, R.; Rietveld, P.; Verkerk, R.; Weide, R.; Gebhardt, C.; Koornneef, M. and Zabel, P. 1991a RFLP markers linked to rrot knot nematode resistance gene Mi in tomato. Theor. Appl. Genet. 81:661-667.

Klein-Lankhorst, R.M.; Vermunt, A.; Weide, R.; Liharska, T. and Zabel, P. 1991b Isolation of molecular markers for tomato (*Lycopersicon esculentum*) using random amplified polymorphic DNA (RAPD). Theor. Appl. Genet. 83:108-114.

Kleinhofs, A.; Kilian, A.; Saghai Maroof, M.A.; Biyashev, R.M.; Hayes, P.; Chen, F.Q.; Lapitan, N.; Fenwick, A.; Blake, T.K.; Kanazin, V.; Ananiev, E.; Dahleen, L.; Kurdna, D.; Bollinger, J.; Knapp, S.J.; Liu, B.; Sorrels, M.; Heun, M.; Franckowiak, J.D.; Hoffman, D.; Skadsen, R. and Steffenson, B.J. 1993 A molecular, isozyme and morphological map of the barley (*Hordeum vulgare*) genome. Theor. Appl. Genet. 86:705-712.

Knapp, S.J. and Bridges, W.C. 1990 Using molecular markers to estimate quantitative trair locus parameters: power and genetic variances for unreplicated and replicated progeny. Genetics 126:769-777.

Knapp, S.J.; Bridges, W.C. and Birkes, D. 1990 Mapping quantitative trait loci using molecular marker linkage maps. Theor. Appl. Genet. 79:583-592.

Koller, B.; Lehmann, A.; McDermott, J.M. and Gessler, C. 1993 Identification of apple cultivars using RAPD markers. Theor. Appl. Genet. 85:901-904.

Komatsuda, T.; Annaka, T. and Oka, S. 1993 Genetic mapping of a quantitative trait locus (QTL) that enhances the shoot differentation rate in *Hordeum vulgare* L. Theor. Appl. Genet. 86:713-720.

Lande, R. and Thompson, R. 1990 Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743-756.

Lander, E.S. and Botstein, D. 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185-199.

Lander, E.S.; Green, P.; Abrahamson, J.; Baarlow, A.; Daly, M.J.; Lincoln, S.E. and Newburg, L. 1987 MapMaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174-181.

Landry, B.; Hubert, N.; Etoh, T.; Harada, J.J. and Lincoln, S.E. 1991 A genetic map for *Brassica napus* based on restriction fragment length polymorphisms detected with expressed DNA sequences. Genome 34:543-552.

Landry, B.S.; Hubert, N.; Crete, R.; Chiang, M.S.; Lincoln, S.E. and Etoh, T. 1992 A genetic map for *Brassica oleracea* based on RFLP markers detected with expressed DNA sequences and mapping of resistance genes to race 2 of *Plasmodiophora brassicae* (Woronin). Genome 35:409-420.

Landry, B.S.; Kesseli, R.V.; Farrara, B. and Michelmore, R.W. 1987 A genetic map of lettuce (*Lactuca sativa* L.) with restriction fragment length polymorphism, isozyme, disease resistance and morphological markers. Genetics 116:331-337.

Lawson, D.; Hemmat, M. and Weeden, N. 1992 Identification of genes influencing morphological and physiological traits in apple. Abstracts of the International Conference on the Plant Genome I, San Diego, CA poster # 148.

Lee, M.; Melchinger, A.E. and Guthrie, W.D. 1991 Molecular marker analysis of host-plant resistance to European corn borer in corn. In: Proceedings 27th Illinois Corn Breeders School, Champaign, IL.

Liu, B.H. and Knapp, S.J. 1992 GMendel 2.0, a software for gene mapping. Oregon State University.

Liu, B.H. and Knapp, S.J. 1992 QTLStat 1.0, a software for mapping complex trait using non-linear models. Oregon State University.

Livini, C.; Ajmone-Marson, P.; Melchinger, A.E.; Messmer, M.M. and Motto, M. 1992 Genetic diversity of maize inbred lines within and among heterotic groups revealed by RFLPs. Theor. Appl. Genet. 84:17-25.

MacArthur, J.W. 1934 Linkage groups in tomato. J. Genet. 29:123-133.

Mackill, D.J.; Salam, M.A.; Wang, Z.Y. and Tanksley, S.D. 1993 A major photoperiod-sensitivity gene tagged with RFLP and isozyme markers in rice. Theor. Appl. Genet. 85:536-540.

Manganaris, A.G. and Alston, F.H. 1987 Inheritance and linkage relationships of glutamate transaminase isoenzymes in apple. 1. The gene GOT-1, a marker for the S incompatibility locus. Theor. Appl. Genet. 74:154-161.

Martin, G.B.; Williams, J.G.K. and Tanksley, S.D. 1991 Rapid identification of markers linked to a Pseudomonas resistance gene in tomato by using random primers and near-isogenic lines. Proc. Natl. Acad. Sci (USA) 88:2336-2340.

McCouch, S.R.; Kochert, G.; Yu, Z.H.; Wang, Z.Y.; Khush, G.S.; Coffman, W.R. and Tanksley, S.D. 1988 Molecular mapping of rice chromosomes. Theor. Appl. Genet. 76:815-829.

McKeand, S.E. and Bridgwater, F.E. 1992 Third-generation breeding strategy for the North Carolina State University-industry cooperative tree improvement program. In: Proceedings of IUFRO International Conference "Breeding tropical trees" Section 2.02-08 Cali, Colombia (in press).

Medina-Filho, H.P. 1980 Linkage of Aps-1, Mi and other markers on chromosome 6. Rep. Tomato Genet. Coop 30:26-28.

Melchinger, A.E. 1990 Use of molecular markers in breeding for oligogenic disease resistance. Plant Breed. 104-1-19.

Melchinger, A.E.; Boppenmaier, J.; Dhillon, B.S.; Pollmer, W.G. and Herrmann, R.G. 1992 Genetic diversity of RFLPs in European maize inbreds. II - Relation to performance of hybrids within versus between heterotic groups for forage traits. Theor. Appl. Genet. 84:672-681.

Menancio-Hautea, D.; Fatokun, C.A.; Kumar, L.; Danesh, D.; and Young, N.D. 1993 Comparative genome analysis of munbean (*Vigna radiata* L. Wilczek) and cowpea (*V. unguiculata* L. Walpers) using RFLP mapping data. Theor. Appl. Genet. 86:797-810.

Michelmore, R.W. and Hulbert, S.H. 1987 Molecular markers for genetic analysis of phytopathogenic fungi. Ann. Rev. Phytopathol. 25:383-404.

Michelmore, R.W.; Paran, I. and Kesseli, R.V. 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.

Miklas, P.N.; Stavely, J.R. and Kelly, J.D. 1993 Identification and potential use of a molecular marker for rust resistance in common bean. Theor. Appl. Genet. 85:745-749.

Moran , G.F., Bell, J.C., Byrne, M., Devey, M.E. and Smith, D.1992. Molecular marker maps in *Pinus radiata* and *Eucalyptus nitens* for application in tree breeding programs. Abstracts of the International Conference on the Plant Genome I, San Diego, CA poster # 93.

Morgan, T.H. 1910 Sex limited inheritance in *Drosophila*. Science 32:120-122.

Morgante, M. and Olivieri, A.M. 1993 PCR-amplified microsatellites as markers in plant genetics. The Plant J. 3:175-182.

Muehlbauer, G.J.; Staswick, P.E.; Specht, J.E.; Graef, G.L.; Shoemaker, R.C. and Keim, P. 1991 RFLP mapping using near-isogenic lines in the soybean (*Glycine max* (L.) Merr.). Theor. Appl. Genet. 81:189-198.

Murray, M.G.; Ma, Y.; Romero-Severson, J.; West, D.P. and Cramer, J.H. 1988 Restriction fragment length polymorphisms: what are they and how can breeders use them? In: Proceedings 43rd Annual Corn and Sorghum Research Conference. Am. Seed Trade Assoc. 43:72-87.

Nam, H.G.; Giraudat, J.; Boer, B. den; Moonan, F.; Loos, W.D.B.; Hauge, B.M. and Goodman, H.M. 1989 Restriction fragment length polymorphism linkage map of *Arabidopsis thaliana*. The Plant Cell 1:699-705.

Namkoong, G.; Kang, H.C. and Brouard, J.S. 1988 Tree breeding: principles and strategies. Monographs on Theoretical and Applied Genetics. Springer-Verlag, New York.

Nance, W.L.; Nelson, C.D.; Stine, M.; Kubisiak, T.L.; VanBujtenen, J.P.; Kong, X. and Doudrick, R.L. 1992 A RAPD based genetic linkage map of slash pine (*Pinus elliottii* Englm. var. *elliotii*) with comparision to other RAPD based maps. Abstracts of the International Conference on the Plant Genome I, San Diego, CA .poster # 11.

Neale, D.B. and Williams, C.G. 1991 Restriction fragment length polymorphism mapping in conifers and applications to forest genetics and tree improvement. Can. J. For. Res. 21:545-554.

Neale, D. and Sederoff, R. 1991 Genome mapping in pine takes shape. Probe 1:1-3.

Newbury, H.J. and Ford-Lloyd, B.V. 1993 The use of RAPD for assessing variation in plants. Plant Growth Regul. 12:43-51.

Nienhuis, J.; Helentjaris, T.; Slocum, M.; Ruggero, B. and Schaefer, A. 1987 Restriction fragment length polymorphism analysis of *loci* associated with insect resistnace in tomato. Crop Sci. 27:797-803.

Nodari, R.O.; Tsai, S.M.; Gilbertson, R.L. and Gepts, P. 1993 Towards an integrated linkage map of common bean. II. Development of an RFLP-based linkage map. Theor. Appl. Genet. 85:513-520.

Nybom, H.; Rogstad, S.H. and Schaal, B.A. 1990 Genetic variation detected by use of the M13 "DNA fingerprint" probe in *Malus, Prunus* and *Rubus (Rosaceae)*. Theor. Appl. Genet. 79:153-156. .

Nybom, H. and Schaal, B.A. 1990a DNA "fingerprints" reveal genotypic distribution in natural populations of blackberries and raspberries (*Rubus* L. *Rosaceae*) Am. J. Bot. 77:883-889.

Nybom, H. and Schaal, B.A. 1990b DNA "fingerprints" aplied to paternity analysis in apples (*Malus x domestica*). Theor. Appl. Genet. 79:763-768.

Osborn, T.C.; Alexander, D.C. and Williams, B. 1987 Identification of restriction fragment length polymorphisms linked to genes controlling soluble solids content in tomato fruit. Theor. Appl. Genet. 73:350-356.

Olson, M.; Hood, L.; Cantor, C. and Doststein, D. 1989 A common language for physical mapping of the human genome. Science 254:1434-1435.

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

Paran, I.; Kesseli, R. and Michelmore, R. 1991 Identification of restriction fragment length polymorphism and random amplified polymorphic DNA markers linked to downy mildew resistance genes in lettuce, using near-isogenic lines. Genome 34:1021-1027.

Paterson, A.H.; Lander, E.S.; Hewitt, J.D.; Peterson, S.; Lincoln, S.E. and Tanksley, S.D. 1988 Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721-726.

Paterson, A.H.; DeVerna, J.W.; Lanini, B. and Tanksley, S.D. 1990 Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecies cross of tomato. Genetics 124:735-742.

Paterson, A.H.; Damon, S.; Hewitt, J.D.; Zamir, D.; Rabinowitch, H.D.; Lincoln, S.E.; Lander, E.S. and Tanksley, S.D. 1991 Mendelian factors underlying quantitative traits in tomato: comparison across species, generations and environments. Genetics 127:181-197.

Penner, G.A.; Chong, J.; Levesque-Lemay, M.; Molnar, S.J. and Fedak, G. 1993 Identification of a RAPD marker linked to the oat stem rust gene Pg3. Theor. Appl. Genet. 85:702-705.

Phillips, T.D.; Murphy, J.P. and Goodman, M.M. 1993 Isozyme variation in germplasm accessions of the wild oat *Avena sterilis* L. Theor. Appl. Genet. 86;54-64.

Plaschke, J.; Börner, A.; Xie, D.X.; Koebner, R.M.D.; Schlegel, R. and Gale, M.D. 1993 RFLP mapping of genes affecting plant height and growth habit in rye. Theor. Appl. Genet. 85:1049-1054.

Prince, J.P.; Pochard, E. and Tanksley, S.D 1993 Construction of a molecular linkage map of pepper and a comparision of synteny with tomato. Genome 36:404-417.

Provine, W. B. 1987 The origins of theoretical population genetics. The University of Chicago Press, Chicago and London.

Quiros, C.F.; Hu, J.; This, P.; Chevre, A.M. and Delseny, M. 1991 Development and chromosomal localization of genome-specific markers by polymerase cahin reaction in *Brassica*. Theor. Appl. Genet. 82:627-632.

Rangel, P.H.N.; Cruz, C.D.; Vencovsky, R. and Ferreira, R.P. 1990 Selection of local lowland rice cultivars based on multivariate genetic divergence. Brazil. J. Genetics 14(2):437-453.

Reiter, R.S.; Williams, J.G.K.; Feldmann, K.A.; Rafalski, J.A.; Tingey, S. V. and Scolnik, P.A. 1992 Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random amplified polymorphic DNAs. Proc. Natl. Acad. Sci. USA 89:1477-1481.

Ritter, E.; Gebhardt, C. and Salamini, F. 1990 Estimation of recombination frequencies and construction of linkage maps from crosses between heterozygous parents. Genetics 125:645-654.

Robertson, D.S. 1985 A possible technique for isolating genomic DNA for quantitative traits in plants. J. Theor. Biol. 117-1-10.

Roose, M.L. and Traugh, S.N. 1988 Identification and performance of *Citrus* trees on nucellar and zygotic rootstocks. J. Am. Soc. Hort. Sci.:113:100-105.

Sarfatti, M.; Katan, J.; Fluhr, R.; Zamir, D. 1989 An RFLP marker in tomato linked to the *Fusarium oxysporum* resistance gene *I2*. Theor. Appl. Genet. 78:755-759.

Sax, K. 1923 The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics 8:552-560.

Schmidt, T.; Boblenz, K.; Metzlaff, M.; Kaemmer, D.; Weising, K. and Kahl, G. 1993 DNA fingerprinting in sugar beet (*Beta vulgaris*) - identification of double-haploid breeding lines. Theor. Appl. Genet. 85:653-657.

Second, G. 1985 Evolutionary relationships in the *sativa* group of *Oryza* based on isozyme data. Genet. Sel. Evol. 17:89-114.

Shin, J.S.; Chao, S.; Corpuz, L. and Blake, T. 1990 A partial map of the barley genome incorporating restriction fragment length polymorphism, polymerase chain reaction, isozyme, and morphological marker loci. Genome 33:803-810.

Singsit, C. and Ozias-Akins, P. 1993 Genetic variation in monoploids of diploid potatoes and detection of clone-specific random amplified polymorphic DNA markers. Plant Cell Reports 12:144-148.

Skroch, P.; Tivang, J. and Nienhuis, J. 1992 Analysis of genetic relationships using RAPD marker data. In: Proceedings of the Symposium on Applications of RAPD Technology to Plant Breeding. CSSA/ASHS/AGA. pp. 26-30.

Slocum, M.K.; Figdore, S.S.; Kennard, W.C.; Suzuki, J.Y. and Osborn, T.C. 1990 Linkage arrangement of restriction fragment length polymorphism loci in *Brassica oleracea*. Theor. Appl. Genet. 80:57-64.

Smith, J.S.C. 1989 The characterization and assessment of genetic diversity among maize (*Zea mays* L.) hybrids that are widely grown in France: Chromatographic data and isozymic data. Euphytica 43:73-85.

Smith, J.S.C and Smith, O.S. 1989 The description and assessment of distances between inbred lines of maize. II. The utility of morphological, biochemical and genetic descriptors and a scheme for the testing of distinctivness between inbred lines. Maydica 34:151-161.

Smith, O.S.; Smith, J.S.C.; Bowen, S.L.; Tenborg, R.A. and Wall, S.J. 1990 Similarities among a group of elite maize inbreds as measured by pedigree, F_1 grain yield, grain yield, heterosis and RFLP's. Theor. Appl. Genet. 80:833-840.

Smith, S. and Chin. E. 1992 The utility of random primer-mediated profiles, RFLP's and other technologies to provide useful data for varietal protection. In: Proceedings of the Symposium on Applications of RAPD Technology to Plant Breeding. CSSA/ASHS/AGA. pp. 46-49.

Smith, C. 1967 Improvement of metric traits through specific genetic loci. Anim. Prod. 9:349-358.

Sobral, B.W.S. and Honeycutt, R.J. 1993 High output genetic mapping of polyploids using PCR-generated markers. Theor. Appl. Genet. 86:105-112.

Soller, M.; Brody, T. and Genizi, A. 1976 On the power of experimental designs for the detection of linkage between marker *loci* and quantitative *loci* in crosses between inbred lines. Theor. Appl. Genet. 47:35-39.

Soller, M. and Beckmann, J.S. 1983 Genetic polymorphisms in varietal identification and genetic improvement. Theor. Appl. Genet. 67:25-33.

Song, Y. and Cullis, C.A. 1992. RFLP and RAPD mapping of Eucalyptus globulus. Abstracts of the International Conference on the Plant Genome I, San Diego, CA poster # 133.

Song, K.M.; Osbom, T.C. and Williams, P.H. 1988 Brassica taxonomy based on nuclear restriction gragment length polymorphisms (RFLPs). I - Genome evolution of diploid and amphidiploid species. Theor. Appl. Genet. 75:784-794.

Song, K.M.; Suzuki, J.Y.; Slocum, M.K.; Williams, P.H. and Osbom, T.C. 1991 A linkage map of Brassica rapa (syn. campestris) based on restriction fragment length polymorphism loci. Theor. Appl. Genet. 82:296-304.

Soost, R.K.; Williams, T.E and Torres, A.M. 1980 Identification of nucellar and zygotic seedlings of Citrus with leaf isozymes. HortScience 15:728-729.

Southern, E.M. 1975 Detection of specific sequences among DNA fragments separated by electrophoresis. J. Mol. Biol. 98:503-517.

Stam, P. 1993 Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. The Plant J. 3:739-744.

Stiles, J.I.; Lemme, C.; Sondur, S.; Morshidi, M.B. and Manshardt, R. 1993 Using randomly amplified polymorphic DNA for evaluating genetic relationships among papaya cultivars. Theor. Appl. Genet. 85:697-701.

Stromberg, L.D. 1992 Incorporating molecular markers into applied maize breeding. Ph.D. diss. University of Illinois. (diss abstr. 92:5591b) Urbana.

Strauss, S.H.; Lande, R. and Namkoong, G. 1992 Limitations of molecular-marker aided selection in forest tree breeding. Can. J. For. Res. 22:1050-1061.

Stuber, C.W.; Mull, R.H.; Goodman, M.M.; Shaffer, H.E. and Weir, B.S. 1980 Allozyme frequency changes associated with selection for increased grain yiel in maize (Zea mays L.). Genetics 95:225-236.

Stuber, C.W. and Edwards, M.D. 1986 Genotypic selection for improvement of quantitative traits in com using molecular marker loci. In: Proceedings 41st Annual Com and Sorghum Research Conference. Am. Seed Trade Assoc. 41:40-83.

Stuber, C.W.; Edwards, M.D. and Wendel, J.F. 1987 Molecular-marker facilitated investigations of quantitative trait loci in maize. II Factors influencing yield and its component traits. Crop Sci. 27:639-648.

Stuber, C.W.1992a Biochemical and molecular markers in plant breeding. In: Plant breeding reviews Vol. 9. Dudley, J.W.; Hallauer, A.R. and Ryder, M. (eds) John Wiley and Sons, Inc.

Stuber, C.W.; Lincoln, S.E.; Wolff, D.W.; Helentjaris, T. and Lander, E.S. 1992b Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics 132:823-839.

Sturtevant, A.H. 1913 The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. J. Exp. Zoology 14:43-59.

Sturtevant, A.H. 1965 A history of genetics. Harper and Row, Pubishers, New York.

Suiter, K.A.; Wendell, J.F. and Case, J.S. 1983 Linkage -1. J. Hered. 74:203-204.

Swofford, D.L. 1990 PAUP: Phylogenetic Analysis Using Parsimony, version 3.0. Illinois Natl. Hist. Surv., Champaign, IL.

Takaiwa, F.; Oono, K.; Lida, Y. and Sugiura, M. 1985 The complete nucleotide sequence of a rice 25S rRNA gene. Gene 37:255-259.

Tanksley, S.D.; Medina-Filho, H. and Rick, C.M. 1982 Use of naturally occurring enzyme variation to detect and map genes controlling quantitative traits in an interspecific backcross of tomato. Heredity 49:11-25.

Tanksley, S.D. 1983a Gene mapping. In: Isozymes in plant genetics and breeding, part A. Tanksley, S.D. and Orton, T.J. (eds). Elsevier Science Publishers B.V., Amsterdam.

Tanksley, S.D. 1983b Molecular markers in plant breeding. Plant Mol. Biol. Report 1(1):3-8.

Tanksley, S.D. and Hewitt, J.D. 1988 Use of molecular markers in breeding for soluble solids in tomato - a re-examination. Theor. Appl. Genet. 75:811-823.

Tanksley, S.D.; Bernatzky, R.; Lapitan, N.L. and Prince, J.P. 1988 Conservation of gene repertoire but not gene order in pepper and tomato. Proc. Natl. Acad. Sci. (USA) 85:6419-6423.

Tanksley, S.D.; Young, N.D.; Paterson, A.H. and Bonierble, M.W. 1989 RFLP mapping in plant breeding: new tools for an old science. Bio/Technology 7:257-264.

Tanksley, S.D.; Ganal, M.W.; Prince, J.P.; de Vicente, M.C.; Bonierbale, M.W.; Broun, P.; Fulton, T.M.; Giovannoni, J.J.; Grandillo, S.; Martin, G.B.; Messeguer, R.; Miller, J.C.; Miller, L.; Paterson, A.H.; Pineda, O.; Roder, M.S.; Wing, R.A.; Wu, W. and Young, N.D. 1992 High density molecular linkage maps of the tomato and potato genomes. Genetics 132:1141-1160.

Thoday, J.M. 1961 Location of polygenes. Nature 191:368-370.

Thormann, C.E. and Osborn, T.C. 1992 Use of RAPD and RFLP markers for germplasm evaluation. In: Proceedings of the Symposium on Applications of RAPD Technology to Plant Breeding. CSSA/ASHS/AGA. pp. 9-11.

Thormann, C.E.; Ferreira, M.E.; Camargo, L.E.A.; Tivang, J.G. and Osborn, T.C. 1993 Comparison of genetic relationship estimates among cruciferous species based on RFLP and RAPD markers. Theor. Appl. Genet. (in press).

Tingey, S.V. and DelTufo J.P. 1993 Genetic analysis with Random Amplified Polymorphic DNA markers. Plant Physiol. 101:349-352.

Tinker, N.A.; Fortin, M.G. and Mather, D.E. 1993 Random amplified polymorphic DNA and pedigree relationships in spring barley. Theor. Appl. Genet. 85:976-984.

Torres, A.M.; Weeden, N.F. and Martin, A. 1993 Linkage among isozyme, RFLP and RAPD markers in *Vicia faba*. Theor. Appl. Genet. 85:937-945.

Tulsieram, L.K.; Glaubitz, J.C.; Kiss, G. and Carlson, J. 1992a Single tree genetic linkage mapping using haploid DNA from megagametophytes. Bio/Technology 10:686:690.

Tulsieram, L.; Comptom, W.A.; Morris, R.; Thomas-Compton, M. and Eskridge, K. 1992b Analysis of genetic recombination in maize populations using molecular markers. Theor. Appl. Genet. 84:65-72.

Uphoff, H. and Wricke, G. 1992 Random Amplified Polymorphic DNA (RAPD) markers in sugar beet (*Beta vulgaris* L) - mapping the genes for nematode resistance and hypocotyl colour. Plant Breeding - Zeitschrift Fur Pflanzenzuchtung 109:168-171.-.

Vallejos, C.E.; Sakiyama, N.S. and Chase, C.D. 1992 A molecular marker-based linkage map of *Phaseolus vulgaris* L. Genetics 131:733-740.

Van der Beek, J.G.; Verkerk, R.; Zabel, P. and Lindhout, P. 1992 Mapping strategy for resistance genes in tomato based on RFLPs between cultivars: *Cf9* (resistance to *Cladosporium fulvum*) on chromosome 1. Theor. Appl. Genet. 84:106-112.

VanHeusden, A.W. and Bachmann, K. 1992 Genotype relationships in *Microseris elegans* (*Asteraceae, Lactucaceae*) revealed by DNA amplification from arbitrary primers (RAPDs). Pl. Syst. Evol. 179:221-233.

Watson, J.D. and Crick, F.H.C. 1953a Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. nature 171:737-738.

Watson, J.D. and Crick, F.H.C. 1953b Genetical implications of the structure of deoxyribonucleic acid. Nature: 171:964-967.

Watson, J.D.; Gilman, M.; Witkowski, J. and Zoller, M. 1992 Recombinant DNA. 2nd edition, Scientific American Books, Freeman and Co. New York.

Waugh, R.; Baird, E. and Powell, W. 1992 The use of RAPD markers for the detection of gene introgression in potato. Plant Cell Rep. 11:466-469.

Weeden, N.F.; Provvidenti, R. and Marx, G.A. 1984 An isozyme marker for resistance to bean yellow mosaic virus in *Pisum sativum*. The Journal of Heredity 75:411-412.

Weeden, N.F.; Muehlbauer, F.J. and Ladizinsky, G. 1992 Extensive conservation of linkage relationships between pea and lentil genetic maps. Journal of Heredity 83:123-129.

Weining, S. and Langridge, P. 1991 Identification and mapping of polymorphisms in cereals based on the polymerase chain reaction. Theor. Appl. Genet. 82:209-216.

Weller, J.I. 1986 Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. Biometrics 42:627-640.

Weller, J.I. 1987 Mapping and analysis of quantitative trait loci in Lycopersicon (tomato) with the aid of genetic markers using approximate maximum likelihood methods. Heredity 59:413-421.

Weller, J.I.; Kashi, Y.and Soller, M. 1990 Power of 'daughter' and 'granddaughter' designs for mapping of quantitative traits in dairy cattle using genetic markers. J. Dairy Sci. 73:2525-2532.

Welsh, J. and McClelland, M. 1990 Fingerprinting genomes using PCR with arbitrary primers Nucleic Acids Res. 18:7213-7218.

Welsh, J.; Honeycutt, R.J.; McClelland, M. and Sobral, B.W.S. 1991 Parentage determination in maiza hybrids using the arbitrarily primed polymerase chain reaction (AP-PCR). Theor. Appl. Genet. 82:473-476.

White, T.L.; Hodge, G.R. and Powell, G.L. 1993 An advanced-generation tree improvement plan for slash pine in the southeastern united states. Silvae Genetica 42: 359-371

Whitkus, R.; Doebley, J. and Lee, M. 1992 Comparative genome mapping of Sorghum and Maize. Genetics 132:1119-1130.

Wilde, J.; Waugh, R. and Powell, W. 1992 Genetic fingerprinting of Theobroma clones using randomly amplified polymorphic DNA markers. Theor. Appl. Genet. 83:871-877.

Wilkie, S.E.; Isaac, P.G. and Slater, R.J. 1993 Random amplified polymorphic DNA (RAPD) markers for genetic analysis in Allium. Theor. Appl. Genet. 86:497-504.

Williams, J.G.; Kubelik, A.R.; Livak, K.J.; Rafalski, L.A. and Tingey, S.V. 1990 DNA polymorphism amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 18:6531-6535.

Williams, J.G.K.; Rafalski, J.A. and Tingey, S.V. 1992 genetic analysis using RAPD markers. Methods Enzymol. (in press).

Wright, S.; Helentjarls, T. and Kikuchi, Y. 1987 RFLP mapping of cloned genes in maize. Maize Genet. Coop. Newsl. 61:89-90.

Wu, K.K.; Burnquist, W.; Sorrells, M.E.; Tew, T.L.; Moore, P.H. and Tanksley, S.D. 1992 The detection and estimation of linkage in polyploids using single-dose restriction fragments. Theor. Appl. Genet. 83:294-300.

Yang, X. and Quiros, C. 1993 Identification and classification of celery cultivars with RAPD markers. Theor. Appl. Genet. 86:205-212.

Young, N.D. and Tanksley, S.D. 1988 Restriction fragement length polymorphism maps and the concept of graphical genotypes. Theor. Appl. Genet. 77:95-101.

Young, N.D.; Zamir, D.; Ganal, M.W. and Tanksley, S.D. 1988 Use of isogenic lines and simultaneous probing to identify DNA markers tightly linked to the *Tm-2a* gene in tomato. Genetics 120:579-585.

Young, N.D. and Tanksley, S.D. 1989 RFLP analysis of the size of chromosomal segments retained around the *Tm-2* locus of tomato during backcross breeding. Theor. Appl. Genet. 77:353-359.

Yu, K. and Pauls, K.P. 1993 Identification of a RAPD marker associated with somatic embryogenesis in alfalfa. Plant Mol. Biol. 22:269-277.

Yu, Z.H.; Mackill, D.J.; Bonman, J.M. and Tanksley, S.D. 1991 Tagging genes for blast resistance in rice via linkage to RFLP markers. Theor. Appl. Genet. 81:471-476.

Zehr, B.E.; Dudley, J.W. and Chojecki, J. 1992 Some practical considerations for using RFLP markers to aid in selection during inbreeding of maize. Theor. Appl. Genet. 84:704-708.

Zeng, S. 1993a Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc. Natl. Acad. Sci. USA 90:10972-10976.

Zeng, S. 1993b Precision mapping of quantitative trait loci. Genetics 136:1457-1468.

1. INTRODUCTION

The term 'molecular marker' in plant genetics and breeding, includes any molecular phenotype detected as a distinct protein or segment of DNA or RNA. The molecular marker can be anonymous or not in sequence and function. It may correspond to a genetically expressed sequence or not and it merely functions as a reference point for a site in the genome. When this molecular marker behaves according to Mendel's basic law of inheritance of a simple factor, it is additionally defined as a genetic marker. An important distinction that has to be made is that because a molecular marker is physically a DNA molecule or it is functionally close to it, it does not necessarily mean that it operates and can be used as a genetic marker. For example, some DNA sequences are modified during development by methylation with resulting changes in expression or sensitivity to restriction enzymes used for their detection.

The introduction of recombinant DNA technologies and the discovery of the Polymerase Chain reaction (PCR), have caused a shift from genetic analysis based on the inference of genotypes from phenotypes, pioneered by Mendel, to genetic methods based on the direct analysis of variation at the level of DNA sequence. Beckmann (1988) defined this as a shift from Mendelian Genetics to Genomic Genetics. Several techniques of molecular biology are available today for the detection of genetic variability at the DNA level, and the analysis of genetic polymorphism. These techniques allow the generation of a virtually unlimited number of molecular markers, which display Mendelian inheritance and cover whole genomes. These markers can be used for many applications in genetic studies and the practice of plant breeding.

The technological development in the area of molecular markers has been extremely rapid and dynamic. The study of plant genomes has taken great advantage of the advances that come from the Human Genome Project. More efficient and accessible methods are constantly being developed and implemented. More powerful statistical methods and computational capabilities follow, allowing the management of the large amounts of data. For these reasons, any attempt of an exhaustive review of this subject would rapidly become out of date. The objective of this chapter is to provide a general overview on the existing methods available for the generation of molecular markers specifically used for the genetic analysis in plant genetics and breeding. For each class of molecular marker we describe the origin, the genetic basis

and the principles involved in detection. A comparative analysis of the advantages and limitations of each class of marker is then described for applications in plant breeding. Details and examples of such applications will be made in a separate chapter. Finally, we describe some novel and extremely powerful techniques for the detection of DNA polymorphisms that open new perspectives for the analysis of complex genomes and potentially the isolation of economically important genes.

2. MORPHOLOGICAL AND MOLECULAR MARKERS

Prior to the 1960's, markers used in plant genetics and breeding were those derived from genes controlling discrete phenotypes of easy visual identification such as dwarfism, chlorophyll deficiencies, flower, seed or leaf color and morphology. These morphological markers contributed significantly to the theoretical development of genetic segregation and linkage and constituted the first versions of genetic maps. Due to their limited number, however, these markers were rarely physically close to economically important genes, and therefore not of use in the practice of plant breeding. Furthermore, the availability of such markers was essentially restricted to a few plant species used as model systems for the study of inheritance, such as pea, maize and tomato.

This picture began to change in the 1960's with the development of molecular markers based on isozymes polymorphisms. The number of genetic markers was increased by at least one order of magnitude and more importantly their application included potentially any plant species of interest. The advent of the modern techniques of molecular biology brought about several methods for the detection of DNA polymorphisms. In the 1970's the discovery and use of restriction enzymes and DNA hybridization, allowed the analysis of Restriction Fragment Length Polymorphisms (RFLP's). More recently, following the discovery of the PCR (Polymerase Chain reaction) process by which segments of DNA can be amplified using a thermostable DNA polymerase (Mullis and Faloona 1987; Saiki *et al.* 1988) several classes of molecular markers based on this technology have been described. These techniques, combined to advanced cloning and sequencing procedures, have allowed a rapid accumulation of information on the structure of eukaryotic genomes and the discovery of diverse classes of repetitive DNA such as mini and microsatellites, which constitute a rich source of genetic polymorphisms. Today, taking all these classes of markers together, a virtually

unlimited number of highly variable (polymorphic) genetic markers can be obtained and explored in virtually any living organism.

3. ADVANTAGES OF MOLECULAR MARKERS AS GENETIC MARKERS

Some of the main advantages of molecular markers over morphological markers include:

(1) a large number of genetic marker loci can be followed in segregating populations, offering the possibility of genetic mapping. The polymorphism information content (PIC) of such marker loci is generally high, while morphological markers have typically a low PIC. In the past, great effort was needed to construct genetic maps based on morphological markers, because several crosses were necessary in order to bring together and study the joint segregation of markers originally found in different lines;

(2) molecular markers are generally neutral in relation to phenotypic effects with minimal or no epistatic or pleiotropic effects. Morphological markers frequently restrict the normal plant development (albinos, dwarfs etc.) and show epistatic effects complicating the correct identification of genotypes;

(3) molecular markers are generally co dominant, containing more information per locus than morphological markers, which typically show dominant or recessive inheritance (Tanksley 1983a,b; Beckman and Soller 1983; Burr et al. 1983; Stuber 1992).

Besides these aspects, morphological markers can only be identified at the whole plant level. Biochemical or molecular markers, however, can be genotyped from samples of tissues or cells at any developmental stage, provided that enough protein or DNA is made available. This aspect offers the possibility to accelerate breeding and recombination of desired individuals, reducing the time necessary to complete a breeding cycle and considerably increasing the efficiency of a breeding program. This becomes particularly relevant when breeding long lived woody perennials such as forest and fruit trees (see chapter I).

In the following discussion, we will describe the basic principles involved in the generation and detection of each class of molecular marker. Some examples will be used to illustrate these principles. We also present a comparative analysis of the main characteristics of the four main classes of molecular markers used in plant genetics and breeding (Table 1). Further details are given throughout the text. Some of the marker

classes discussed in the text were not included in the comparative analysis due to their basic similarity to markers already covered in the analysis and to their very limited use to date in plant breeding.

4. ISOZYME AND ALLOZYME MARKERS

Isozymes constitute a group of multiple molecular forms of the same enzyme that exist in a species as a result of more than one gene encoding for the enzymes (Moss 1982). Since the description of starch gel electrophoresis (Smithies 1955), the histochemical visualization of proteins (Hunter and Markert 1957) and the classical studies of population genetics using isozyme markers (Lewontin and Hubby 1966), techniques of isozyme analysis have been applied to a large number of organisms to answer the most diverse genetic questions. In the last twenty years, a large number of studies in population and evolutionary genetics have used isozyme techniques to estimate the levels and understand the structure of the genetic variability in natural populations, study genetic flow, hybridization and species dispersion as well as to carry out phylogenetic analyses. In plant breeding, isozyme markers have been used to detect linkage to mono and oligogenic traits, varietal identification and germplasm evaluation (for reviews see Tanksley and Orton 1983). In forest genetics, isozyme markers have been extensively used making important contributions in breeding practices (see Adams 1983; Cheliak *et al.* 1987).

4.1 Genetic basis of isozyme markers

Isozymes display the same catalytic activity, however they can have different kinetic properties and be separated by biochemical procedures. This means that isozymes of the same group differ in their amino acid sequence which may result in a different secondary, tertiary or quartenary protein structure. The number of isozymes for a particular enzyme can be related to the number of subcellular compartments where the same catalytic reaction takes place (Gottlieb 1982).

The basic assumption made when using isozyme analysis is that differences in mobility along an electric field are the result of differences at the level of the DNA segment that encodes for such enzymes. Thus, if the banding patterns of two individuals differ, it is generally assumed that such differences have a genetic basis and are

inherited (Murphy *et al.* 1990). The genetic control of isozymes occurs by several genes, which can be alleles at a same locus or be located in different loci. Isozymes encoded by allelic genes are also defined as allozymes. Murphy *et al.* (1990) recognize two general forms of protein data. One is derived from isozymes which are all functionally similar forms of enzymes, incuding all polymers of subunits produced by different genetic loci or by different alleles at the same locus. The other data set consists of allozymes, which are variants of polypeptides representing different allelic alternatives of the same gene locus. To simplify, we will use the more general term "isozyme" to refer to both these kinds of data.

The expression of isozymes is generally co dominant, i.e. both alleles at a locus in a diploid individual are expressed and can be visualized. To interpret the banding patterns obtained, it is important to have previous knowledge about the number of subunits of the enzyme. For example, monomeric enzymes are formed by only one polypeptide, while dimeric enzymes by two. Heterozygotes for a dimeric enzyme, besides the two bands corresponding to the two polypeptides, also show a third intermediary band on the gel, product of the conjugation of the two polypeptide subunits. It is common that more than one genetic locus is resolved on the same gel, and that the migration of bands of each locus is visualized in different zones of the gel. Occasionally, subunits of an enzyme can be encoded by distinct genetic loci, complicating the genetic interpretation of resulting zymograms.

4.2 Detection of isozymes

Isozyme detection involves basically three steps: protein extraction from the appropriate plant tissue, separation of these proteins by electrophoresis and histochemical staining of the gel to allow the visualization of bands. The majority of plant tissues can, in principle, be used for isozyme analysis. Once the adequate tissue has been identified both from the genetic (e.g. ploidy), physiological (e.g. age) and practical (e.g. collection and storage) standpoints, it is ground in the presence of a buffer that allows protein extraction and maintenance of their catalytic activity, while preventing oxidation of associated phenolic compounds. The protein extract is then separated by gel electrophoresis. Different gel matrices (starch, agarose, polyacrylamide) and several buffer systems can be used for this purpose. The

appropriate choice of tissue and the optimization of a buffer system for a particular isozyme and a particular species are critical variables for the success of this analysis.

Following electrophoresis, the isozymes are visualized by specific histochemical stains, that supply substrates to the enzymes. The observed banding patterns (zymogram) result from the catalysis of a biochemical reaction by the isozyme present at that position in the gel, allowing therefore the visualization of a band.

4.3 Advantages of isozyme markers

For most of the isozymes, their genetic control is well known. This allows for genetic inferences directly from the observed banding patterns. Isozymes generally provide ample genetic information for several applications. The methods involved are technically accessible and inexpensive to perform. In spite of the limited number (see below), several loci can be rapidly and simultaneously analyzed. For example, the most widely used starch gel systems allow horizontal slicing of the gel in up to 10 slices where one or more loci is analyzed in each slice. In spite of the current "DNA trend" in the field of genetic analysis, the practicality of isozymes will continue to make them a very useful class of molecular markers when an extensive sampling of the genome is not required.

Isozyme alleles are typically co dominant, i.e. heterozygotes and homozygotes genotypes at a locus can be easily distinguished. This allows direct estimation of genetic parameters such as genotype frequencies, gene frequencies and from these several measures of diversity and heterozygosity. Deviations from Hardy-Weinberg equilibrium at one locus or gametic linkage disequilibrium at several loci can be tested with isozyme data (Weir 1990).

4.4 Limitations of isozyme markers

When the investigation requires a more ample coverage of the genome, such as the case of genetic mapping, isozymes present two basic limitations: (1) the number of loci that can be resolved and analyzed; (2) the number of alleles per locus, i.e. the level of genetic polymorphisms detectable at each locus (or Polymorphism Information Content, PIC). Even if we considered that the total potential number of isozyme loci that can be resolved in an organism is around 100 (Murphy *et al.* 1990), only a limited fraction, typically between 10 and 30, can actually be resolved in a particular species.

This level of resolution does not permit the genome coverage necessary for the construction of genetic linkage maps. Evidently, with only a partial genome coverage, any attempt of establishing significant associations between markers and genes that control desirable traits, particularly of polygenic inheritance, becomes limited and difficult. Although isozymes are considered as selectively neutral markers, the level of enzymatic polymorphism has a limit from the evolutionary standpoint, besides the intrinsic limitation in the detection technique. Thus, besides the limited number of loci, the absence of genetic polymorphism at some or several of these loci is frequently observed, seriously limiting the power of genetic analysis.

Other limitations of isozymes as genetic markers are: (1) pos-translational modifications of the enzymes resulting in the so called conformational isozymes, or multiple forms of the products of the same gene that differ in secondary or tertiary structures and appear as distinct phenotypes on the gel; (2) specificity of some isozyme forms to some particular tissues, which can become a temporal limitation for the analysis; (3) isozyme polymorphisms in response to environmental conditions or physiological states of the plant; (4) difficulties in the interpretation of zymograms when isozymes with identical electrophoretic mobilities are products of different loci of the same enzymatic system. If allelic variation exists at such "isoloci", it can be very difficult or impossible to determine which alleles belong to which loci (see Murphy *et al.* 1990 for a review).

5. RESTRICTION FRAGMENT LENGTH POLYMORPHISMS (RFLP'S)

The acronym RFLP's describes the polymorphisms observed in the length of the fragments of DNA produced by cleavage with restriction enzymes, and visualization by DNA hybridization with probes containing homologous sequences. Restriction enzymes are enzymes that cut double stranded DNA molecules at specific sites defined by 4 to 8 base pairs. More than 500 restriction endonucleases have been isolated from bacteria, and several tens of those are commercially available. The natural function of restriction enzymes is to protect the bacteria from foreign DNA. Methylation of the bacterial chromosomal DNA at the recognized site, prevents cleavage. The exogenous DNA, not containing the specific pattern of DNA methylation is recognized and digested.

RFLP's were originally used for the detection of mutations in viral DNA (Grodzicker *et al.* 1974).The use of RFLP's as genetic markers with the potential to saturate the

genome for the construction of linkage maps was initially proposed for the human genome (Botstein et al. 1980; Wyman and White 1980). The use of RFLP's in plant breeding was suggested thereafter by Beckmann and Soller (1983) and Burr et al. (1983). RFLP's constitute today the most widely used class of molecular marker in plant genetics and breeding. Its applications range from germplasm identification and evaluation to the construction of linkage maps to monitor the inheritance of Quantitative Trait Loci (QTL's) (for reviews see Helentjaris et al. 1985; Tanksley et al. 1989; Beckmann 1991; Neale and Williams 1991; Stuber 1992).

5.1 Genetic basis and detection of RFLP's

RFLP's occur because the DNA of distinct individuals differ in the presence or absence of specific 4 to 8 base pairs sites recognized by restriction enzymes. Alternatively, the DNA sequence of two genotypes can differ as a result of insertions, deletions or other rearrangements that alter the relative distance between any two restriction sites. The RFLP detection technique involves several steps. Initially the DNA of the individuals that one wishes to analyze is extracted and subjected to digestion with a particular enzyme. The DNA sequence is cut into a large number of fragments. Differences in the DNA sequence between individuals (DNA polymorphisms) results in different populations of fragments, which are size fractionated by agarose gel electrophoresis. Due to the large number of fragments produced in eukaryotic genomes, a smear of fragments is observed on the gel. Discrete fragments cannot be discerned and therefore differences in length of the fragments between individuals cannot be visualized directly on the gel.

To perform the detection of RFLP's, the fragments separated by electrophoresis are transferred and immobilized by capillary or vacuum transfer onto nylon or nitrocellulose membranes in a process of blotting developed by Southern (1975). The membranes are often called "Southern Blots". The identification of polymorphic fragment between individuals is done by hybridizing small cloned DNA fragments called labeled probes (see below), to homologous sequences of the fragments immobilized onto the membrane. Probe labeling usually involves the incorporation of nucleotides containing radioactive phosphorous (^{32}P) or sulfur (^{35}S) isotopes. Following probe hybridization, the membrane is exposed to an X-ray film in a process called autoradiography, resulting in dark bands or spots on the film that constitute the genetic

marker data. These bands result from the sensitivity of the film to beta particles emitted by the radioactivity incorporated to the probe. The probe hybridizes exclusively to fragments containing homologous sequences, and only where the probe hybridizes, a signal is obtained on the film.

Alternatively, non-radioactive detection systems use modified nucleotides containing groups such as digoxigenin or biotin, that are later recognized by antibody-enzyme conjugates. In the chemiluminescence system, for example, the antibody recognizes the digoxigenin group on the hybridized probe, and the conjugated enzyme, typically an alkaline phosphatase, catalyzes a chemiluminescent reaction involving a substrate that, when degraded by the enzyme, emits light photons. These photons play the same role as the beta particles in sensitizing the film.

If two individuals display fragments that differ in length following restriction digestion, such fragments will migrate to different positions on the gel and also be immobilized at different positions on the membrane. Upon probe hybridization and detection, bands in distinct positions will be observed on the X-ray film, characterizing a so called RFLP. If the Mendelian segregation and inheritance is confirmed by analyzing parents and progeny, the particular RFLP is a defined locus and can be used as a genetic marker. The membrane prepared at the beginning of the procedure can be reused several times (up to 10) for hybridization with different probes, provided that the previous probe is adequately removed and the immobilized DNA is not degraded. Each probe typically provides one polymorphic marker, and occasionally detects more than one.

5.2 Production of probes for RFLP analysis

The DNA clones to be used as probes can be obtained in different ways, the most common being: (1) by reverse transcription of messenger RNA of the organism under study, producing a complementary DNA (cDNA) library of clones corresponding to expressed sequences of the genome; (2) random genomic DNA sequences, in the form of a genomic library that includes both expressed and non expressed sequences; (3) from known clones of specific genes, either by PCR amplification or synthetic oligonucleotide construction.

Typically, once the library of fragments to be used as probes is obtained, a necessary step before any analysis can be carried out is the selection of clones to be

used as probes. In this screening procedure, the objective is to select those clones that contain exclusively single or low copy sequences, that is, that do not contain repetitive sequences present in the genome. Clones containing such repetitive DNA, will hybridize to all the fragments immobilized on the membrane that contain such repetitive elements. After autoradiography, this pattern of hybridization will show large numbers of bands or continuous smears that cannot be interpreted or followed as genetic markers. On the other hand, probes containing moderately repetitive elements become very useful for the objective of obtaining unique fingerprints of genomes (see section 6). Libraries of cDNA fragments are intrinsically made up of a large majority of low to single copy clones. Genomic libraries are not, however they can be constructed in specific ways as to enrich them in low copy sequences. In general both types of libraries supply useful probes with similar amounts of detectable polymorphism.

5.3 Advantages of RFLP markers

RFLP's present the main advantage of potentially covering the whole genome of interest, depending on the type of library used to obtain the probes. The use of RFLP markers substantially increases the probability of finding significant associations between markers and genetic loci of breeding interest. It was postulated that a single segregating population is sufficient to study a large number of segregating traits through the use of markers distributed throughout the genome, given that such traits are segregating in the population under scrutiny (Tanksley et al. 1982; Soller and Beckmann 1983).

RFLP markers typically display co dominant inheritance, and therefore a larger amount of genetic information is obtained per locus when compared to dominant markers. Co dominant markers allow the analysis of intralocus interactions between alleles at genetic loci of interest as well as epistatic interactions between alleles at different loci. RFLP's typically sample genetic variation in coding regions and in a smaller scale in non-coding regions of the genome. This is mainly a result of the selection process of useful low copy probes as described previously. Unlike isozymes, the potential number of RFLP markers is almost unlimited and the polymorphism information content per locus is much larger. This is because several restriction enzymes can be used which, combined to a large number of probes (cDNA and genomic clones) can detect several alleles at a marker locus and several marker loci in a genome. Finally, as any other DNA

based markers, RFLP's share the advantage of the high stability of DNA which can be extracted from any tissue, at any stage of development and can be stored and reused over long periods of time.

5.4 Limitations of RFLP markers

Considering that the vast majority of experiments in plant breeding involves the analysis of hundreds or even thousands of individuals for several tens or hundred of markers, it is critical that the technique used to assess genotypes be extremely efficient in the acquisition of data and that it can be automated. In other words, the technique has to be "breeder-friendly". RFLP analysis presents an important limitation in this respect because it involves several steps that are labor intensive and time consuming.

Another limitation exists when a library of useful probes is not immediately available when a project is started. When this is the case, libraries have to be constructed and clones screened for probes. Only after this time consuming stage, which typically takes several months or years of work, will any type of molecular analysis be possible with RFLP's. Probe libraries are available for several of the main crops such as maize, tomato, rice, wheat, soybean, bean, potato and brassicas. This is not yet the case, however, for minor crops, as well as for the vast majority of forest and fruit trees.

The use of RFLP's requires personnel with skills in recombinant DNA technology and a more sophisticated lab setup than that necessary for isozyme work. When phosphorous isotopes are used in the detection process, appropriate facilities have to be built and approved by federal agencies and safety regulations have to be followed for the management of radioactive waste. All these aspects taken together make it difficult, for example, the transfer of this technology directly to breeding stations and its use by breeders.

6. HYPERVARIABLE MINISATELLITES BASED MARKERS

A large proportion of eukaryotic genomes is composed of repetitive DNA sequences. This proportion varies: for example 30% of the DNA in *Drosophila* is repetitive, while 70% in tobacco (Lewin 1990). Several classes of repeated DNA elements have been described and characterized in different species, from mammals to plants. These differ in the number and composition of nucleotides and are found clustered or

dispersed throughout the nuclear genome forming structurally highly variable sites. From the standpoint of molecular markers, these repeated sequences offer the possibility to simultaneously sample a relatively large number of polymorphic genetic loci distributed throughout the genome.

Hypervariable sequences called VNTR's (Variable Number of Tandem Repeats) are the most commonly used molecular markers involving repetitive elements. A hypervariable locus, also called a minisatellite locus, is composed of a variable number of repeat elements arranged in tandem. These repeats have between 15 and 100 base pairs and are repeated up to 50 times at each hypervariable locus. These minisatellites sites are dispersed in the genome, constituting several loci in different chromosomes and a detectable proportion of the genome which will vary depending on the length of the elements and the number of repeat units. Different loci display different number of repeats and there are different families of minisatellites. The first hypervariable region was isolated by chance by Wyman and White (1980) from a human genomic library. The first minisatellite was formally described for the human genome by Jeffreys et al. (1985a). It was a 33 base pairs sequence repeated four times, derived from a human myoglobin intron. Since then, similar sequences have been identified and cloned from other organisms including several plant species (Wu and Wu 1987; Zhao et al. 1989; Schmidt 1991; Thomas et al. 1993).

The name minisatellite comes from the fact that these repeated sequences form a satellite peak distinct from the main genomic DNA peak along a cesium chloride gradient, because they contain a GC content that is different from the average of the whole genome (Lewin 1990). The names Hypervariable loci, VNTR loci and minisatellites are used interchangeably in the literature to refer to this class of molecular markers. In plant breeding, these markers have been used mainly for varietal identification, analysis of genetic diversity among individuals and paternity studies (Dallas 1988; Nybom and Hall 1991; Rogstad et al. 1988; Gepts et al. 1992; Broun et al. 1992). The genetic basis of molecular markers based on VNTR loci is such that its use for mapping and linkage studies is limited due to difficulties in the genetic interpretation of alleles and loci. In spite of this limitation, its use has been suggested to monitor introgression in backcross breeding programs (Hillel et al. 1990).

6.1 Genetic basis and detection of hypervariable loci

The principles underlying the generation and detection of VNTR based markers are essentially the same as those used for RFLP's. In fact, technically speaking all VNTR polymorphisms are by definition RFLP's. In VNTR analysis, in the same manner as in RFLP analysis, the DNA of individuals is digested with restriction enzymes, size fractionated, transferred to a membrane and the detection is performed by hybridization with a DNA probe. When the restriction enzyme cuts in regions adjacent to the hypervariable locus, the RFLP's produced will vary according to the number of repeat units present in the minisatellite. This rich source of DNA sequence polymorphism is probably the result of unequal crossing over or slippage during DNA replication (Jeffreys *et al.* 1985a).

The basic difference between RFLP and VNTR based markers resides in the type of probe used in the polymorphism detection step. In RFLP's, probes homologous to single copy regions are used, detecting one locus at the time, occasionally more than one. VNTR's polymorphisms on the other hand are detected with probes which are or contain core sequences homologous to the repeat elements. In this way, all the hypervariable loci that contain an homologous repeat element are detected simultaneously. Thus, instead of a simple banding pattern having one band for homozygotes and two for heterozygotes, a complex banding pattern is obtained where alleles and loci cannot be easily discerned, and typically each band is treated as a separate entity.

An interesting aspect of this technique is that several of the probes used in the detection of hypervariable loci can be used across an ample spectrum of organisms, from viruses to plants and mammals. A typical example is the 15 base pairs core sequence found in the M13 *E. coli* phage (a commonly used cloning vector). This sequence has been widely used to detect hypervariable polymorphisms in humans, other mammals as well as plants including angiosperms and gymnosperms (Vassart *et al.* 1987; Rogstad *et al.* 1988). In a similar manner a minisatellite identified in the human genome has been used fro the detection of polymorphisms in rice and rubber tree (*Hevea brasiliensis*) (Dallas 1988; Besse *et al.* 1993). When specific probes need to be developed, a genomic library is constructed and the screening aims at selecting moderately repetitive clones, a different objective than the RFLP library screening. Sequencing and isolation of monomers of these repeats are isolated and used as probes (Wu and Wu 1987; Schmidt 1991).

6.2 Advantages and limitations of hypervariable marker loci

As VNTR based markers are technically equivalent to RFLP's, they share the same advantages and limitations discussed previously for RFLP based markers. Due to the basic difference in the probes used in the detection, VNTR based markers not only explore the fragment length polymorphism at each locus, but take advantage of the polymorphism in the number and distribution of such loci throughout the genome, sampling several loci simultaneously in the same assay. This complex banding pattern produced corresponds to all the alleles at all the hypervariable loci detected by the probe and displays typically 10 to 40 discernible bands.

The complexity of such patterns, reflecting the high polymorphism content of such loci, results in the fact that virtually every individual has a unique banding profile, reminiscent of a bar code. From this observation the revolutionary concept of genetic fingerprint was suggested, for which the VNTR technique was the first to be introduced in humans (Jeffreys *et al.* 1985b). On the other hand, however, due to the complexity of the banding patterns observed, it is practically impossible to interpret which bands are allelic and derive from the same loci and thus determine genotypes to test hypotheses about Mendelian segregation and linkage relationships. For such reasons, hypervariable loci are very powerful tools for fingerprinting and paternity studies, however they are of limited use for genetic mapping and linkage studies.

A modification of the original VNTR technique was proposed by Nakamura *et al.* (1987). This involves the detection of hypervariable loci one at the time using as probes synthetic oligonucleotides derived from consensus sequences of known VNTR's. Although more labor intensive, this technique allows the identification of alleles and genotype determination. Essentially this modification results in a return to the simple banding patterns obtained by RFLP, however exploring the high PIC content of hypervariable loci, i.e. several alleles at a given locus. Although interesting, this proposal is limited to those cases where prior information is available about what probe to use. Furthermore, at least in human genomes it has been shown that hypervariable loci tend to be clustered in proterminal regions of chromosomes, and thus do not provide an unbiased coverage of the genome for mapping applications (Royle *et al.* 1988).

7. POLYMERASE CHAIN REACTION (PCR)

The technique of Polymerase Chain Reaction (PCR) was conceived by Kary Mullis in mid 1980's (Mullis and Faloona 1987; Saiki *et al.* 1985). Since its introduction, this technology has revolutionized the way in which DNA analysis and manipulation has been carried out in research and diagnostics. Several traditional cloning methods have been accelerated or even bypassed by PCR. A large number of variations on the PCR theme have been described in recent years, revealing an impressive creativity on the part of molecular biologists, and allowing studies that had been impossible before. The ease, quickness, versatility and sensitivity of the method makes it particularly powerful for studies involving the analysis of large numbers of individuals of any living organism. The PCR technique has rapidly reached widespread use in the most diverse areas of biology. For recent reviews about PCR refer to White *et al.* (1989), Arnheim *et al.* (1990) and Erlich *et al.* (1991).

7.1 Basic principles of the Polymerase Chain Reaction

PCR is a powerful technique that involves the enzymatic synthesis "in vitro" of millions of copies of a specific segment of DNA. This reaction is based on the annealing and enzymatic extension of a pair of short oligonucleotide primers that bracket the target DNA sequence. These primers are designed in such a way that their nucleotide sequences are complementary to specific sequences that flank the target segment. A PCR cycle involves three steps: denaturation, annealing and extension. The template DNA is denatured by raising the temperature to 92-95 $^{\circ}$C. In the annealing step, the temperature is rapidly reduced to 35 to 60 $^{\circ}$C depending on the length and GC content of the primers, allowing the hybridization of each primer to their complementary regions. The temperature is then raised again, this time to 72 $^{\circ}$C, which is the optimal temperature for the thermostable DNA polymerase to catalyze extension from the 3' end of the primer. This extension involves the addition of nucleotides using the target sequence as template, resulting in the synthesis of a new copy of the target sequence in each cycle. The cycle is repeated for 20 to 40 times. The amplification process follows a geometric progression, as the number of target sequences doubles at every cycle. Thus in only 20 cycles, more than one million times the original amount of target sequence is produced. This rate of amplification allows starting the process with minute amounts of template DNA (on the order of 10^{-6} picograms) and end up with large amounts of DNA of a specific segment of interest.

Originally, a DNA polymerase of *E. coli* was used to catalyze the process. It had to be added fresh at every cycle because its activity was lost every time the temperature was raised to denature the template DNA. A major technological breakthrough was achieved when Saiki *et al.* (1988) isolated a thermostable DNA polymerase from the thermophylic bacteria *Thermus acquaticus* (*Taq* polymerase), allowing the complete automation of the PCR process. Its optimal extension temperature is 72 °C and it can keep its activity for some hours at 95 °C.

8. RANDOM AMPLIFIED POLYMORPHIC DNA (RAPD) MARKERS

The advent of the PCR method, allowed the production of large amounts of specific DNA sequences that could be visualized directly in an electrophoretic gel with ethidium bromide staining of DNA. More importantly it represented the first step toward automation of genetic analysis. However as a source of molecular markers, PCR by itself still presented a serious limitation. The design of primers still required prior sequence information on flanking regions to the target. For some highly conserved genes, sequence information could be transferred across species. However in most cases the development of PCR based markers depended on cloning and sequencing efforts, thereby limiting its immediate application as a way to obtain large numbers of genetic markers.

A simple but great innovation was introduced in 1990 with the idea of using short primers of arbitrary sequence to direct the PCR, eliminating the need of prior sequence information. This technique was independently developed by two groups in the U.S. Williams *et al.* (1990) at Dupont, named and patented the process using the captivating name RAPD (Random Amplified Polymorphic DNA). They developed and demonstrated the use of the technique to obtain large numbers of DNA polymorphisms that were inherited as genetic markers and could be used for mapping and linkage studies. Welsh and McClelland (1990) used the technically more correct name Arbitrarily Primed PCR, because the primers have arbitrary sequence and the amplification is not random. They proposed the use of such markers for genomic fingerprinting using higher resolution polyacrylamide gels. A third group also independently developed the same idea, but a delay in the publication process did not grant them the public recognition for the discovery (P.M. Gresshoff, personal communication). The name DAF (DNA

Amplification Fingerprinting) for the detection of AFLP (Amplification Fragment Length Polymorphisms) was then proposed for the method (Caetano-Annóles *et al.* 1991a).

Irrespective of the acronym used and the slight variations in the methods, all the applications proposed were of great relevance in the study and practice of plant genetics and breeding. A new perspective was opened for the genotypic analysis of individuals and populations. Since its description, the use of RAPD markers in plant genetics and breeding has seen a rapid and extensive diffusion. Some of the applications have included: clonal and varietal fingerprinting, analysis of genetic diversity in natural populations, construction of genetic maps and localization of important genes (for recent reviews see Hedrick 1992; Rafalski *et al.* 1991; Caetano-Annóles *et al.* 1991b; Williams *et al.* 1992; Tingey and Deltufo 1992).

8.1 Genetic basis of RAPD markers

RAPD amplification is basically a variation of the PCR protocol, with two distinctive characteristics: (1) it uses one primer only instead of a pair of primers; (2) this primer is typically short (10 to 17 bases) and has arbitrary sequence, such that its target region is anonymous. When two sites complementary to the primer sequence are adjacent (< 4000 base pairs) and in an inverted orientation (technically a pair of short inverted repeats), the amplification between these sites takes place following the same procedure as in a standard PCR. As a result of the large quantity of DNA produced, the amplified fragment can be directly visualized as a band on an agarose gel by ethidium bromide staining. Alternatively high resolution polyacrylamide gels, coupled to isotope incorporation during PCR and autoradiography or silver staining can be used to increase the sensitivity of the method. Typically each primer directs the synthesis of several DNA fragments simultaneously in different regions of the genome, so that several bands of different sizes are observed on the gel, whose size depends on the distance between the pair of priming sites. Welsh and McClelland (1991) suggested the use of multiplexing pairs of arbitrary primers to increase the number of detectable polymorphisms.

The probability with which a RAPD site occurs in a genome of a certain size can be estimated, such that a linear relationship is expected between the sequence complexity of the genome and the number of bands observed. Experimental data have shown, however, that the number of bands is relatively independent from the

genome complexity. In other words, the number of amplified bands from a simpler genome such as bacteria, is similar to that amplified from more complex genomes such as soybean (Williams *et al.* 1990). These results indicate that the final products of the RAPD reaction depend more on the competition between different RAPD sites than on the number of sites, and imply that not all the fragments amplified are the result of a perfect match between the primer and the template DNA. A more specific complementarity at the 3' end of the primer, where the extension begins, seems to be more critical than the pairing at the 5' end (Williams *et al.* 1990).

The molecular nature of RAPD polymorphisms is not entirely known. However, experimental evidences indicate that single base pair differences are sufficient to cause a primer template mismatch, preventing amplification (Williams *et al.* 1990). Other sources of polymorphisms may include priming sites deletions, or insertions that change the distance bewteen the priming sites. The genetic polymorphism detected by the RAPD assay is binary in nature, i.e. a fragment is amplified (presence of a band on the gel) or not (absence of the band).

RAPD markers usually behave as dominant genetic markers. Dominance in this case does not refer to the classical meaning in terms of intralocus interaction between alleles, rather purely from the standpoint of relationship between genotype and phenotype. When a RAPD band is observed on the gel, it is not possible to distinguish whether that fragment is derived from the amplification from one or two allelic dosages. That is, in a diploid homozygous individual (AA) for a RAPD locus, amplification is carried out from two copies of the RAPD allele (A). In a heterozygous individual (Aa) for the same locus, the allele (A) is amplified and the allele (a) is not. The RAPD fragment detection does not have enough quantitative sensitivity to discriminate between these two cases, and a band of identical intensity will be observed on the gel in both cases. So, while the homozygous recessive genotyoe (aa) is identified as the absence of the band (null genotype), the homozygous and heterozygous genotypes have the same molecular phenotype, i.e. the presence of a band. This behavior is equivalent to that of a morphological dominant marker. The RAPD assay detects only one allele at each locus. The absence of the band corresponds to all the other genotypes represented by undetectable alleles at that locus that cannot be amplified. These are all pooled into the same class (null genotype). When haploid tissue is available for the genetic assay, this general problem is circumvented because haplotypes can be determined directly, such as the case of haploid megagametophytes in conifers (see chapter I).

8.2 Advantages of RAPD markers

The basic technological difference between the RFLP and RAPD assays is that while RFLP is based on DNA hybridization, RAPD is based on DNA amplification. This difference results in a number of practical advantages that can be summarized in two attributes: simplicity and speed. Data acquisition is several order of magnitude faster. For example RAPD was four to six times more technically efficient that RFLP's when mapping polymorphisms linked to disease resistance, and 10 times more efficient in terms of time and labor (Paran et al. 1991). This is basically the result of the detection procedure which is done directly by visualizing the bands on the gel, eliminating all the steps involving Sothern blotting, probe preparation and autoradiography.

The RAPD technique does not require the prior development of a library of probes for the organism of interest. The same set of oligonucleotide primers can be used for any organism. Because it is not based on DNA hybridization, it does not require the handling of any type of radioactive isotope. Another important advantage is the minimal amounts of DNA that are necessary for the genotypic analysis of an individual. For each RFLP data point, microgram quantities of DNA are needed. RAPD requires only nanogram amounts of DNA, i.e. three orders of magnitude less. This substantially facilitates and accelerates DNA extraction procedures. These properties allowed for example the construction of a linkage map with 200 markers for *Pinus taeda* in 6 person/month (Grattapaglia et al. 1991) and *Arabidopsis thaliana* in 4 person/month (Reiter et al. 1992).

As for RFLP based markers, the RAPD assay allows the generation of a large number of polymorphic DNA markers distributed throughout the genome. RAPD markers, however, offer the possibility to sample genomic regions containing repetitive DNA, because the primers used are arbitrary in sequence while RFLP probes are pre-selected for low copy regions. RAPD has shown to be more sensitive in the detection of DNA polymorphisms than RFLP. For example Foolad et al. (1993) verified that only 16% of the RFLP probes tested revealed polymorphisms at the intraspecific level in tomato, while 63% of the RAPD primers tested detected at least one polymorphism. Similar results have been observed in soybean (S.V. Tingey, personal communication). The sensitivity in polymorphism detection has also been shown to be very powerful for the establishment of genetic fingerprints (Welsh and McClelland 1990). RAPD also offers an alternative to

cloning of random genomic fragments which is extremely fast and does not require vectors and host manipulations. The RAPD fragments can be isolated directly from the gel, maintained in the form of a genomic library "in vitro" and reamplified whenever necessary by PCR with the original arbitrary primers. These fragments can be used for example as RFLP or VNTR probes depending on their internal sequence content in terms of copy number.

The cost of the RAPD technique is lower than RFLP in terms of cost per data point. If one includes the cost of probe library development, the cost for RAPD would be much lower. This difference in cost results mainly from the difference in labor costs and to a lesser extent to differences in reagents and supplies needed. However it is important to point out that substantially fewer reagents are necessary to conduct RAPD analysis than RFLP and isozymes. This is certainly an advantage for the utilization of this technology in developing countries like Brazil, where reagent supply is limited.

RAPD associates the technical simplicity of direct visualization of marker data of isozymes, with the numbers and resolution power of DNA based markers. The use of RAPD markers does not require experience in molecular biology, nor high tech facilities. It is a very accessible technology that can be immediately transferred to breeding stations. In the context of the typical breeding experiments involving hundreds or thousands of individuals, RAPD is the best technology there is today to efficiently carry out detailed genetic analysis for a large number of markers. It also offers a more realistic chance of automating this analysis. The binary nature of the RAPD polymorphism, if on one hand implies in a lower information content per locus (since heterozygotes cannot be distinguished), on the other hand it is more appropriate for the framework of a digital electronic data acquisition system.

8.3 Limitations of RAPD markers

The main limiting feature of RAPD markers is their low genetic information content per locus, since only one allele is amplified, while all other alleles are detected as nulls. This limitation is commonly treated as dominant behavior of RAPD markers as described previously. Another potential limitation seen by some is the lack of prior genetic information on the genetic basis of a RAPD band. In other words a RAPD band can only be regarded as a genetic marker after its segregation has been confirmed from parents to progeny. This prerequisite, however, although is not exclusive to RAPD polymorphisms,

it seems to be more relevant in view of the large number of bands observed on a RAPD band profile.

The high levels of sequence polymorphisms detected with RAPD markers is generally regarded as a definite advantage. However it may also result in a limitation, which has to do with the ability to use the same genetic markers across divergent individuals within the species. Although the same set of primers can be used, this does not mean that the same fragments corresponding to the same loci will always be amplified. Technically speaking, a single primer samples only 20 base pairs of DNA sequence at the ends of what is defined as a marker locus. The sensitivity of the PCR is such that almost any single base pair difference, or change in the relative positions of the priming sites results in the inability of detecting the locus. So at the population level, new RAPD marker loci appear and some are not detected in different individuals, and this becomes more pronounced as the genetic distance between the individuals analyzed increases, as a reflection of the evolutionary process itself. In the RFLP this occurrence is rarer. Polymorphisms at an RFLP locus are typically lost or found across individuals, and different alleles can be detected. However, as the probes cover several hundred base pairs of DNA sequence and the hybridization process is less sensitive to single point mutations outside the restriction sites, RFLP loci tend to be more conserved, sometimes even across related species.

As with any other type of marker, RAPD's are not an exception in that some of the bands observed on the gel are easily and unambiguously interpreted, while others are less interpretable or repeatable. Ambiguity in the interpretation may result from: (1) low discriminatory power between distinct sites of a specific primer as a result of its nucleotide sequence; (2) competition between distinct amplification sites for substrate and enzyme, such that the occurrence of particular sites interfere or out compete the amplification of others, in a process equivalent to an epistatic interaction between marker loci; (3) problems related to the standardization of amplification conditions. Laboratory to laboratory variation, as a result of differences in thermal profiles of different PCR cyclers has been observed (Penner *et al.* 1993; Wang *et al.* 1993). It has also been suggested that RAPD markers should be used with caution for phylogenetic analysis, fingerprinting and paternity determination, because competition between different fragment may lead to ambiguous results (Heun and Helentjaris 1993; Thormann and Osborn 1992; Thormann *et al.* 1993; Riedy *et al.* 1992). Due to the sensitivity of the PCR protocol, the optimization and standardization of reaction components and

amplification conditions represent an important step in the application of the RAPD technology.

The RAPD procedure is intensive in gel electrophoresis. For some applications is could be more practical to prepare a single membrane and hybridize it with different probes than to prepare new RAPD reactions and gels every time. From the cost perspective, RAPD markers cost more than isozymes and less than RFLP's, however it is still an expensive technique. The reaction component responsible for 80% of the total cost of a data point is the *Taq* DNA polymerase. Its use is regulated by a patent that covers the Polymerase Chain reaction. Its production cost is about four orders of magnitude lower than the market price of the enzyme. The general expectation is that this price will fall substantially in the next years, or that special licensing agreements could be granted for its production for research purposes in academics. With developments in this direction, the cost of RAPD markers will become significantly lower than any other type of existing genetic marker.

9. SPECIFIC PCR-BASED GENETIC MARKERS

The recent development of large sequencing projects of complex eukaryotic genomes has resulted in a formidable technological progress in the ability to rapidly decode DNA sequence. The main advances have been: (1) automation of the traditional sequencing process; (2) the use of PCR as a more efficient method for cloning and production of large amounts of DNA necessary for sequencing; (3) the introduction of sequencing detection systems based of fluorescent labels. The accumulation of sequence information on specific genes (e.g. Taikawa *et al.* 1985) and complete organellar genomes in several crops (e.g. Hiratsuka *et al.* 1989) offers now the possibility to use specific PCR methodologies as a tool to generate large numbers of molecular markers.

The ease with which DNA sequence can now be determined, has led to the option of converting RFLP based markers into PCR based ones. This is done by sequencing the extremities of the RFLP probes (cDNA or genomic clones), and designing pairs of specific primers to these ends so as to recover potentially polymorphic fragments. Besides the conversion to PCR based markers, such an approach generates a large amount of sequence information about genes when the probes sequenced were derived from cDNA libraries (Adams *et al.* 1991). This procedure has been adopted

in some plant species where large number of cDNA probes had previously been mapped, like in maize (T. Helentjaris, personal communication) and rice (Williams *et al.* 1991). The conversion of RFLP markers into PCR based markers has also been used in green pepper for the determination of hybrid purity (Livneh *et al.* 1992). Markers obtained in this manner have been called Sequence Tagged Sites (STS) or Expressed Sequence Tags (EST) when the sequenced probes were cDNA's. The approach and names have been proposed initially by Olson *et al.* (1989) with the objective of unifying the language used among laboratories involved in physical mapping of the human genome. Today these markers constitute an important tool for fingerprinting YAC (Yeast Artificial Chromosomes) clones.

A variation of the STS markers called *Alu*-PCR has been used for the identification of molecular markers specifically in the human genome. Pairs of specific primers have been constructed based on conserved sequences of the 300 base pair *Alu* repeat element which is highly dispersed throughout the human genome. DNA amplification occurs between adjacent *Alu* repeats (Cole *et al.* 1991).

More recently, STS markers have been developed for mapping specific regions of plant genomes (Weining and Langridge 1991; Tragoonrung *et al.* 1992; Landry *et al.* 1992). Repeat elements similar to the *Alu* family in humans have also been identified in plants. For example in rye, specific primers were designed for the amplification of a highly polymorphic family of repeat elements (15,000 copies/genome) dispersed in the genome (Rogowsky *et al.* 1992).

Another variation of STS markers was proposed by Wu *et al.* (1989) with the development of allele specific PCR based markers (AS-PCR). The idea of assaying for specific alleles of known genes had previously been proposed using oligonucleotide probes (ASO - Allele Specific Oligonucleotide) for the detection of mutant alleles in the diagnosis of antitrypsin deficiency (Kidd *et al.* 1984). In plants, the development of AS-PCR based markers has not received much attention, probably because there are very few genes where such a strategy would be justified. However, a AS-PCR assay for the detection of specific alleles at a locus that encodes for an enzyme involved in anthocyanin biosynthesis has recently been developed in maize (B. Burr, personal communication).

A type of marker conversion that has received an increasing amount of attention lately has been that from RAPD markers to specific PCR based markers. This class of markers, although it is merely a variation of the STS theme, was named SCAR for

Sequence Characterized Amplified Region (Paran and Michelmore 1993). SCAR is a genomic DNA fragment from a genetically defined locus that is identified by PCR amplification using a specific pair of oligonucleotide primers. Although STS and SCAR markers follow the same principles, SCAR's can contain repetitive DNA sequences within the amplified region. Their uniqueness is determined by the sequence and spacing of the primer sequences (Paran and Michelmore 1993).

A common characteristic to STS and STS-like markers is the typically low level of genetic polymorphism encountered at the amplified loci. The amplified fragment length polymorphism detected is unacceptably low, considering the effort necessary to construct such markers. This represents a major limitation for a more general applicability in plant genetics and breeding. The source of sequence polymorphisms obtained in the RFLP assay is largely due to sequence variation outside the region to which the probe hybridizes. Thus, by designing primers to amplify sequences covered by the probe, the same polymorphisms are not recovered. The only source of useful polymorphism would be due to insertions or deletions within the size range supported by PCR. Furthermore, when STS are developed from cDNA probes, there are no intrinsic expectations of such types of genetic polymorphism in transcribed regions as they tend to be evolutionarily more conserved. Differential intron size and placement would be a potential source of polymorphism, however with the same restrictions in terms of size range that can be efficiently amplified by PCR. Genomic clone derived STS markers could potentially be more polymorphic, as non-transcribed regions could also be included.

The conversion of RAPD into SCAR's suffered from the same lack of polymorphism (Paran and Michelmore 1993). From the nine RAPD fragments that were cloned, sequenced and converted into specific PCR-based markers only two resulted in markers with detectable fragment length allelic variation. For one of them, polymorphism could be recovered when the PCR product was cut by a restriction enzyme. Five of the SCAR's continued to behave as dominant RAPD markers where the polymorphism was of the presence/absence type, and for one SCAR marker, the primer pair designed could not amplify the intended region. The success in converting a RAPD to a strictly PCR-based co dominant assay was therefore very limited, on the order of 20%.

10. SIMPLE SEQUENCE REPEAT (SSR) OR MICROSATELLITE BASED MARKERS

Since the mid 1980's it had been demonstrated that eukaryotic genomes are highly populated with different classes of repeated sequences, some more complex (see minisatellites) and some more simple (Hamada *et al.* 1982; Tautz and Rentz 1984). Simple Sequence Repeats (SSR) later also called microsatellites (Litt and Luty 1989) consist of mono to tetranucleotide sequence motifs that are tandemly repeated and display high levels of genetic polymorphism resulting from the variation in the number of repeat units. In eukaryotic genomes, SSR sites are more frequent and well dispersed in the genome than minisatellites sites. When these highly polymorphic sites are amplified by PCR they generate the most polymorphic class of genetic markers available today called Sequence Tagged Microsatellite Sites (STMS) (Beckmann and Soller 1990).

Microsatellites have been observed in several organisms including humans (Litt and Luty 1989), whales, *Drosophila* (Tautz 1989), mice (Love *et al.* 1990), rat (Serikawa *et al.* 1992), cattle and sheep (Moore *et al.* 1991). In plants their existence had previously been suggested with the observation oligonucleotides containing simple repeats such as TG and GATA/GACA detected high levels of polymorphism when used as RFLP probes (Weising *et al.* 1989; Beyermann *et al.* 1992). The most ubiquitous repeat units in mammals are extensions of CA and TG dinucleotides (Hamada *et al.* 1982). In plants, a database search of published sequences revealed that microsatellite sites are widely distributed with an estimated average frequency of one every 50,000 base pairs. Their presence has been detected in 34 plant species, and the most common repeat unit was the dinucleotide AT (Morgante and Olivieri 1993). The first report clearly documenting the presence of SSR in plants was that by Condit and Hubbell (1991) which included five tropical tree species besides maize. Microsatellite based markers are currently being developed for mapping applications for several major crops including soybean and rice (Akkaya *et al.* 1992; Zhao and Kochert 1993; M. Morgante, personal communication).

10.1 Genetic basis of microsatellite bases markers

Regions containing Simple Sequence Repeats are amplified individually by PCR, using a pair of specific primers (20 to 30 bases) designed based on unique flanking sequences to the microsatellite site. Fragments amplified from such sites almost invariably display extensive length polymorphism resulting from the presence of different numbers of simple repeat elements. Thus, each microsatellite island, independently of

the repeat element (CA, TG, AT, ATG etc.), constitutes a highly polymorphic multialleleic genetic locus with high Polymorphism Information Content. Each amplified fragment (typically from several tens to hundreds of base pairs) represents a different allele at the microsatellite locus defined by the specific primer pair. The average heterozygosity at microsatellite loci is generally above 0.7. In humans, for example, 72% of the SSR markers developed for genetic mapping showed three or more alleles (Weissenbach *et al.* 1992).

The detection of the PCR amplified fragments is done by electrophoresis in polyacrylamide or high resolution agarose gels. Depending on the length of the microsatellite sequence motif, resolution of a single base pair may be necessary. Band visualization on the gel can be done directly by ethidium bromide staining or by autoradiography when radioactive labeled nucleotides are used as substrate for the PCR reaction.

Each microsatellite site is analyzed individually by using the pair of primers specifically designed for its amplification. Occasionally the same pair of primers amplifies more than one locus. More than one locus can be analyzed simultaneously when the expected fragments sizes for each locus are sufficiently different so as to migrate to different zones of the gel. In this procedure, called multiplexing, more than one pair of specific primers is used simultaneously in the same PCR reaction. SSR loci seem to be sufficiently somatically stable to allow tracking them in subsequent generations. They behave as co dominant markers and are highly multiallelic at the population level where potentially all the alleles at each locus can be detected and discriminated.

10.2 Advantages of microsatellite based markers

Due to their co dominant expression, their intrinsic nature of polymorphism and detectable multiallelism, microsatellite based markers are the class of markers displaying the highest Polymorphism Information Content (PIC). Essentially any segregating population can be used as reference population for linkage studies. The choice of mapping population does not have to be made as to maximize genetic distance, i.e. sequence polymorphism. Rather, this choice can be made from the informativeness of the cross in terms of the biological or economical traits of interest.

STMS are highly frequent both in expressed and repetitive genomic regions allowing the most complete coverage of any eukaryotic genome. At least in mammals,

the observation that STMS show sequence conservation across diverse species, might suggest the use of heterologous primers (Moore *et al.* 1991). Besides these attractive genetic properties, the polymorphic fragments are sufficiently short to be detected by PCR. The practical advantages of PCR-based markers discussed previously, are therefore shared by STMS. All these attributes make STMS ideal markers for physical and genetic mapping, genotype identification and population genetic studies.

The overall relevance, resource level and number of laboratories involved in the human genome project has led to the development of several thousands STMS markers in recent years and the construction of the first STMS based linkage maps (Weissenbach *et al.* 1992; NIH/CEPH Collaborative Mapping Group 1992). Microsatellite markers represented a very important step in the human genome project because they resolved a major problem associated with the low levels of genetic polymorphism of the RFLP markers used previously (Goodfellow 1992). Similar maps, however less extensive have recently been reported for mice (Dietrich *et al.* 1992), rat (Serikawa *et al.* 1992), and are in development for other domestic animals including cattle, swine and sheep (Steffen *et al.* 1993; Coppieters *et al.* 1993; Moore *et al.* 1991). In soybean several groups are developing marker maps based on STMS. This interest stems both from the high value of this crop and the very low level of genetic polymorphism detected at the DNA level. The low polymorphism reflects the narrow genetic basis of the cultivated germplasm and has made the construction of genetic maps a tedious endeavor with RFLP's (Keim *et al.* 1989) and even RAPD (S.V. Tingey, personal communication).

10.3 Limitations of microsatellite based markers

The major limitation of this technology today is certainly the large amount of work necessary to develop these markers. Briefly, the procedure to obtain STMS markers involves the following steps. A genomic library of small fragments derived from digestion with frequent cutting restriction enzymes is constructed for the organism of interest. The genomic clones are screened for the presence of microsatellite sites using synthetic oligonucleotide probes complementary to the target repeat elements. For example poly-GT for mammals where CA repeats are more common or poly-AT for plants where AT repeats are typically ubiquitous. The positive clones are sequenced and specific primer pairs are constructed for unique sequences carefully selected flanking the microsatellite site (specific computer software exists to facilitate this step). During the

development of STMS markers for the human genome, from a total of 12,014 sequenced clones, only 2995 were selected for the construction of STMS (Weissenbach *et al.* 1992). The majority of the discarded clones generally contain too few repeat units (only clones with more than 10 repeats are typically useful) or the position of the microsatellite is at the ends of the clone thus impeding the design of flanking primers. With a selection factor of one for every six clones, a large amount of cloning and sequencing work is necessary for the development of each marker. To reduce the amount of work, short cuts can be used. For example, instead of sequencing all four bases, only one lane ocorresponding to one of the bases contained in the repeat is sequenced. This partial sequence information is sufficient to discard clones based on the number of repeat units and their position (M. Morgante, personal communication).

The large amount of work involving specialized personnel, sophisticated sequencing facilities and the costs associated with such an enterprise represent the major barriers that exist today for a more widespread application of this technology in plant genetics and breeding. Furthermore, with the exception of species with reduced levels of diversity such as soybean, for the great majority of plants, especially those of allogamous habit, the high levels of genetic polymorphism detectable with more accessible technologies does not justify the magnitude of investment necessary for the development of STMS based markers. However, as the techniques necessary in this process become simpler, automatable and economically more accessible, this technology will certainly be used for a growing number of species, until better technologies emerge.

11. OTHER TECHNOLOGIES FOR THE DETECTION OF DNA SEQUENCE POLYMORPHISMS

Besides the established technologies discussed so far, it is appropriate to briefly describe some existing and emerging technologies that are extremely sensitive in the detection of DNA polymorphisms. Although they have been developed with objectives other than the generation of large numbers of molecular markers, they represent new concepts in the detection of variation at the DNA sequence level. For example the first two methods described (Ligase Chain Reaction and Single Stranded Conformation Polymorphism) were originally developed for the detection of single base pair mutations in known genes. Such a diagnostic approach could become extremely useful as useful genes of breeding interest such as disease resistance genes become available and

acquisition of sequence information becomes faster. The other two methods discussed join the power of restriction enzymes and PCR to detect sequence polymorphisms. All the concepts presented will likely lead to the development of entirely different classes of molecular markers in the decades to come.

11.1 Ligase mediated reaction

The ligase-mediated method was originally proposed by Landegren *et al.* (1988) as a way to detect single base pair mutations in known genes. This method exploits the fact that the ends of two single strands of DNA must be exactly aligned for a DNA ligase to join them. If the terminal nucleotides of either end are not properly base-paired to the complementary strand, then the ligase cannot join them. In the original experiment, this method was used to detect a known single base pair mutation in the beta-globin gene that causes sickle-cell anemia. Oligonucleotides complementary to the normal sequence and mutant sequence, 5' to and including the mutation site were synthesized and end labeled with biotin. A third oligonucleotide complementary to the rest of the sequence, common both to the normal and mutant genes was also synthesized and end labeled with radioactivity. The oligonucleotides were hybridized to denatured strands of normal and sickle-cell DNA. Only when hybridized to the normal gene, the 5' and 3' oligonucleotides formed a flush junction that could be joined by the added DNA ligase. The single base pair mismatch between the normal 5' oligonucleotide and the mutation site was sufficient to prevent the ligase from joining the oligonucleotides. After denaturation, the reaction mixtures were run on a steptavidin column; radioactivity was only retained on the column if the 5' biotinylated oligonucleotide had been joined to the 3' radioactively labeled common oligonucleotide.

11.2 Single Stranded Conformational Polymorphism (SSCP)

Opposite to the ligase mediated method that relies on prior sequence information for the detection of point mutations, the SSCP technique was conceived with the objective of detecting previously unknown mutations (Orita *et al.* 1989; Hayashi 1991). In this context, this method is particularly interesting and has been applied to the generation of anonymous polymorphic molecular markers in specific genomic regions,

particularly to enhance the level of detectable polymorphisms in previously non polymorphic segments.

In the SSCP technique, DNA fragments that differ in sequence are denatured into single strands. When subject to rapid cooling, some single strands base pair with the original complementary strands, while the majority self anneals by hydrogen bonds, resulting in particular molecular conformations that are specific depending on the sequence composition. This single stranded DNA is then size fractionated by polyacrylamide gel electrophoresis. The rate of migration in the gel depends on the specific conformation of the DNA molecule, which in turn depends on the DNA sequence. The method is very sensitive and can easily detect point mutations by the difference in rate of migration of different fragments. This method was originally developed with fragments produced by restriction enzymes, but recently has been optimized with PCR amplified fragments. The main advantage of this method is the high sensitivity to DNA polymorphisms. In rice, for example, SSCP's have been detected in RFLP derived STS markers. Amplified STS that did not display polymorphisms, showed SSC Polymorphisms due to a single base pair difference that both RFLP and PCR techniques could not detect (Miyao *et al.* 1993).

11.3 Representational Difference Analysis (RDA)

RDA has been recently described by Lisitsyn *et al.* (1993) as a technique that combines the power of restriction enzymes, PCR and subtractive hybridization to find small differences between the sequences of two DNA populations. In the context of molecular markers, the use of this technique has been suggested for the directed isolation of RFLP probes or PCR based markers that are polymorphic between two lines or individuals to be used in a mapping experiment. However, the most exciting application envisaged is the actual cloning of genes involved in monogenetic traits, such as disease resistance, dwarfism etc., or somatic mutations. This should be possible when two isogenic lines differ only by the presence of the gene that confers the phenotypic difference, or somatic normal and mutant sectors are available. The latter application could be potentially interesting in fruit species where somatic mutations have been a rich source of useful traits.

The method consists in the selective enrichment of target DNA through successive cycles of hybridization (subtractive enrichment) and amplification (kinetic

enrichment). The genome that contains the target DNA is called tester, while the genome that does not contain the target is called driver. Initially, the complexity of the two genomes is reduced with the objective of increasing the relative concentration of target DNA. To achieve this, representative portions of each genome ("representations") are prepared by cleaving the DNA with relatively infrequent cutting restriction enzymes. To the fragments generated, oligonucleotide adapters containing priming sites were ligated, and PCR amplification was performed. Only around 10% of the genome is effectively amplified and represented. This procedure is repeated with different restriction enzymes, such that different fragment representations (also called amplicons) are generated covering the whole genome. Once the amplicons of the tester and driver genomes are produced the adapters are removed and new adapters are ligated only to the tester DNA.

The representations are then used in a subsequent denaturation, subtractive hybridization and PCR amplification step. In this step, the amplified fragments from the driver genome are hybridized in large molar excess to the tester genome containing the target. The hybridization is carried out until a point when a small percentage of the target DNA suffers reassociation, however a large percentage of the DNA in common between the two genomes also reassociates as a result of the molar excess of the driver genome. PCR is then carried out with primers that are specific to sequences on the adapters present only on the tester genome. Exponential amplification will only take place in the reassociated fragments where both DNA strands are derived from the tester DNA. All the other fragments will be amplified in a linear fashion or will not be amplified. Thus, the driver genome functions as a competitive inhibitor for the reassociation of the fraction of the genome common to tester and driver. As a result, the amount of target DNA is enriched in relation to the rest of the tester genome. This cycle can be repeated a few times for further enrichment. If a 50 times molar excess of driver in relation to tester is used in the subtractive hybridization, after two cycles of RDA, the enrichment is on the order of 10^5 and at the end of three cycles, more than 10^{10} (Lisitsyn *et al.* 1993). The population of discrete fragments obtained at the end of this process represents the differences between the two genomes. These PCR fragments can be readily isolated and cloned from the agarose gel. These clones can be used as polymorphic probes between the two individuals or will contain the gene of interest. The combined used of isogenic lines and RDA should offer a rapid way of cloning genes involved in disease and pest resistance. This approach has been initiated by several

groups working with crops where such lines are available (R. Sederoff, personal communication).

11.4 Amplified Fragment Length Polymorphism (AFLP)

This technology represents the more recently developed procedure for obtaining large numbers of molecular markers. It combines the specificity and power of restriction enzyme digestion with the speed of detection offered by PCR amplification. This approach is remarkably similar to the first step of the Representational Difference Analysis (RDA) when amplicons are prepared. However, its development was carried out independently and began prior to the publication of the RDA method (R. Sederoff, personal communication).

The procedure is basically analogous to the amplicon preparation in RDA, in that total genomic DNA is cleaved to completion with relatively infrequent cutting restriction endonucleases. Specific oligonucleotide adapters are then ligated with DNA ligase, and PCR carried out toward the inside of the restriction fragments using primers specific to sequences on the oligonucleotide adapters. As pointed out in the work by Lisitsyn et al. (1993) when developing RDA, this step works like a size fractionation since after 20 rounds of PCR, only low molecular size fragments, below 1000 base pairs are effectively amplified. Although this step substantially reduces the total number of fragments that are amplified, there would still be too many fragments to allow resolution by gel separation. To resolve this problem, a creative stratagem is used which uniquely characterizes the AFLP technique. The oligonucleotide primers used to direct PCR are designed to be two or three bases longer at their 3' end than the oligonucleotide adapters, so that they actually have to base pair with two or three additional bases on the restriction fragments. These extra bases provide an additional selective force on the potentially amplifiable fragments so that only those fragments whose terminal bases are complementary to the extra bases on the primers are eventually amplified. This final population of amplified fragments is run on a high resolution polyacrylamide gel, and the fragments detected by autoradiography as one of the nucleotides used as PCR substrate contained radioactive label. The result is a large number (> 100) of discrete fragments that may differ in even a few base pairs, and are also visually reminiscent of a bar code. Different markers can be obtained using different combinations of extra

bases at the 3' end of the primers, or alternatively different restriction enzymes or combinations of them.

The source of polymorphisms in this technique derives from restriction site differences between individuals that result in the production of fragments of different lengths which in turn are selectively amplified based on their size and terminal sequence composition. The name Amplified Fragment Length Polymorphism originated from this combined approach. AFLP markers offer a relatively rapid way of generating and analyzing large numbers of genetic marker loci on a single gel. They are relatively labor intensive in that several steps are required for the preparation of the amplicons. However this is compensated by the large number of markers that are obtained, substantially diluting the labor and cost per data point. AFLP markers behave as RAPD markers in terms of genetic interpretation, i.e. are dominant. Allelic fragments derived form the same genetic locus cannot be distinguished, and therefore each fragment on a gel is scored as a separate entity.

The main advantage of this technology is certainly the large number of polymorphic markers that can potentially be obtained and scored on a single gel. It has been suggested that AFLP markers are very efficient when a few samples are to be analyzed with a large number of markers, such as in Bulk Segregant Analysis for disease resistance (see chapter I), but its efficiency is low when genome scans with a few markers for several hundred individuals is the objective, as usually the case in the first step of QTL mapping experiments. A combined genome scan using RAPD markers and a subsequent focusing on particular regions using AFLP markers could be a potential strategy for QTL mapping. Alternatively, AFLP's could be used to quickly generate linkage maps, and determining regions of QTL interest. In a subsequent step, Bulk Segregant Analysis on AFLP genotypes could be used to screen for RAPD markers in the region, that could then be more easily assayed in large populations. The AFLP technology has not yet been published. It has been disclosed at meetings and it is subject of a patent filed by the Dutch company Keygene (M. Zabeau, Plant Genome II, 1994).

12. CONCLUSION AND NEW PERSPECTIVES

Three important events marked the evolution of the technology of molecular markers for genetics and breeding studies in the last 30 years: (1) the introduction of

isozyme electrophoresis in the 60's; (2) the isolation of restriction enzymes and the analysis of RFLP's in the 70's and (3) the discovery of the Polymerase Chain Reaction (PCR) in the 80's. In recent years, revolutionary methods of cloning and genomic analysis (Burke *et al.* 1987; Eckert 1990) have led to the construction of physical maps of chromosomes such as the Y chromosome in humans (Foote *et al.* 1992) and of the six smaller chromosomes of yeast (Riles *et al.* 1993). Parallel to this burst in creativity on the part of molecular biologists, major technological progress in engineering of molecular detection and automation of DNA sequence analysis allows today the processing of large numbers of samples and the rapid acquisition of DNA sequence information.

This decade has been and will be increasingly more characterized by formidable advances in the ability to detect genetic variability at the DNA sequence level. From the standpoint of information on genetic variability, the ultimate molecular diagnostic will not be based on any type of molecular marker, but rather on the complete DNA sequence at any locus of interest. With this concept in mind, a great part of the development observed in this area is derived from the progress in the Human Genome Project. The use of molecular analysis in plant genetics and breeding has certainly benefited from this progress and will continue to do so. However it is important to realize that the level of funding typically available in any genetic mapping effort with plant species is several orders of magnitude smaller than that devoted to the human genome project. An economically more accessible and technologically more "breeder friendly" genetic analysis is critical for the effective integration of this technology in plant breeding. Furthermore, the specific needs of a diagnostic system based on DNA polymorphisms substantially varies with the level of genetic diversity inherent to each plant species, its reproductive habits and the strategies used in its breeding and recombination. The choice for one or other techniques has to be guided by the question or challenge that is posed, on a case by case basis, taking into account the advantages and limitations of each class of molecular marker available without any regard to the fact that this or that technique is more elegant or fashionable at the time.

Table 1. Comparative analysis of the main types of molecular markers used in plant genetics and breeding.

Characteristic	ISOZYMES	RFLP	RAPD	MICROSATELLITES
Molecular basis of the observed polymorphism	Expressed proteins	Restriction fragment length polymorphism	Arbitrarily amplified fragment	Variable number of sequence repeats
Genetic expression	Co dominant	Co dominant	Dominant	Co dominant
Number of alleles/locus	Generally 2 alleles	Multiallelic	2 alleles	Highly multiallelelic
Availability of marker loci	Generally 20 to 50	Unlimited	Unlimited	Unlimited
Distribution in the genome	Single copy regions	Single copy regions	More or less random	More or less random
Transfer and presence of markers across individuals	General occurrence	Intra specific; fair across related species	Intra populational; low to fair at the intra specifc level	Intra specific; low to fair across related species
Steps for marker detection	Protein extraction Electrophoresis (starch) Histochemical staining	DNA extraction $(10^{-6}g)$ Rest. enz. digestion Electrophoresis Southern blotting Probe cloning Insert purification Probe labeling Probe hybridization Stringency washes Autoradiography	DNA extraction $(10^{-9}g)$ PCR amplification Electrophoresis(agarose) Ethidium bromide staining	DNA extraction $(10^{-9}g)$ PCR amplification Electrophoresis (high Et-Br staining or autogradiography
Prior information and/or development necessary	Optimization of buffer system	Specific library constr. (cDNA/genomic) Screening of single copy probes	Screening of primers available on the market	Genomic library Selection of clones DNA sequencing Primer engineering Primer synthesis

Table 1. Continued

Characteristic	ISOZYMES	RFLP	RAPD	MICROSATELLITES
Efficiency in genotypic data point acquisition	Very high in routine	Low during probe development; Medium in routine	Very high in routine	Low during marker development; High in routine
Technological accessibility in the plant breeding context	Very high	Average	Very high	Very low
Costs of implementation/ routine operation (per data point, including labor costs)	Low/low	High/average	Low/low	Very high/low
Adequacy for different applications in plant breeding				
Genotype identification	Low	High	Very high	Very high
Germplasm evaluation	Average	High	Average	High
Genetic mapping	Low	High	High	Very high
Directed mapping to specific genomic regions	Low	Average	Very high	Average
Comparative mapping	Low	Very high	Low	Average
Genetics of autogamous crops	Low	Average	High	Very high
Genetics of allogamous crops	Average	Very high	Very high	Very high

13. LITERATURE CITED

Adams, W.T 1983 Application of isozymes in tree breeding. In: Isozymes in plant genetics and breeding. Part A. Tanksley, S.D. and Orton, T.J. (eds.) Elsevier Science Publishers, Amsterdam pp. 60-64.

Adams, M.D.; Kelley, J.M.; Gocayne, J.D.; Dubnick, M.; Polymeropoulos, M.H.; Xiao, H.; Merril, C.R.; Wu, A.; Olde, B.; Moreno, R.F.; Kerlavange, A.R.; McCombie, W.R. and Venter, J.C. 1991 Complementary DNA sequencing: expressed sequence tags and human genome project. Science 252:1651-1656.

Akkaya, M.S.; Bhagwat, A.A. and Cregan, P.B. 1992 Length polymorphisms of Simple Sequence Repeat DNA in soybean. Genetics 132:1131-1139.

Alfenas, A.C.; Peters, I.; Brune, W. and Passador, G.C. 1991 Eletroforese de proteinas e isoenzimas de fungos e essências florestais. Universidade Federal de Viçosa, MG.

Arnheim, N.; White, T. and Rainey, W.E. 1990 Application of PCR: organismal and population biology. BioScience 40(3):174-182.

Beckmann, J.S. 1988 Oligonucleotide polymorphisms: a new tool for genomic genetics. Bio/Technology 6:1061-1064.

Beckmann, J.S. 1991 Genomic genetics and plant genetic improvement. In: Gene-mapping techniques and applications. Schook, L.B.; Lewin, H.A. and McLaren D.G. (eds.) Marcel Dekker Inc., New York pp. 201-230.

Beckmann, J.S. and Soller, M. 1990 Toward a unified approach to genetic mapping of eukaryotes based on sequence tagged microsatellite sites. Biotechnology 8:930-932.

Beckmann, J.S. and Soller, M. 1983 Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. Theor. Appl. Genet. 67:35-43.

Beckmann, J.S. and Soller, M. 1986 Restriction fragment length polymorphisms and genetic improvement of agricultural species. Euphytica 35:111-124.

Besse, P.; Lebrun, P.; Seguin, M. and Lanaud, C. 1993 DNA fingerprints in Hevea brasiliensis (Rubber tree) using human minisatellite probes. Heredity 70:237-244.

Beyermann, B.; Nurnberg, P.; Weihe, A.; Meixner, M. and Epplen, J.T. 1992 Fingerprinting plant genomes with oligonicleotide probes specific for simple repetitive DNA sequences. Theor. Appl. Genet. 83:691-694.

Botstein, D.; White, R.L.; Skolnick, M. and Davis, R.W. 1980 Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32:314-331.

Broun, P.; Ganal, M.W. and Tanksley, S.D. 1992 Telomeric arrays display high levels of heritable polymorphism among closely related plant varieties. Proc. Natl. Acad. Sci. USA 89:1354-1357.

Burke, D.T.; Carle, G.F. and Olson, M.V. 1987 Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science 236:806-812.

Burr, B.; Evola, S.V.; Burr, F.A. and Beckmann, J.S. 1983 The application of restriction fragment length polymorphism to plant breeding. In: Genetic engineering: principles and methods. Setlow, J.K. and Hollaender, A. (eds.). Plenum Press, Vol. 5, pp. 45-59.

Caetano-Annóles, G.; Bassam, B.J. and Gresshoff P.M. 1991b DNA Amplification Fingerprinting: a strategy for genome analysis. Plant Mol. Biol. Rep. 4:294-307.

Caetano-Anólles, G.; Bassam, B.J. and Gresshoff, P.M. 1991a High resolution DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. Biotechnology 9:553-557.

Cheliak, W.M.; Yeh, F.C.H. and Pitel, J.A. 1987 use of electrophoresis in tree improvement programs. For. Chron. 63:89-96.

Cole, C.G.; Goodfellow, P.N.; Bobrow, M.; Bentley, D.R. 1991 Generation of novel sequence tagged sites (STSs from discrete chromosomal region) using Alu-PCR. Genomics 10:816-826.

Condit R. and Hubbell, S.P. 1991 Abundance and DNA sequence of two-base repeat regions in tropical tree genomes. Genome 34:66-71.

Coppieters, W.; Vandeweghe, A.; Peelman, L.; Depicker, A.; Vanzeveren, A. and Bouquet Y. 1993 Characterization of Porcine Polymorphic Microsatellite Loci. Animal Genetics 24:163-170.

Dallas, J.F. 1988 Detection of DNA "fingerprints" of cultivated rice by hybridization with a human minisatellite DNA probe. Proc. Nat. Acad. Sci. (USA) 85:6831-6835.

Dietrich, W. ; Katz, H.; Lincoln, S.E.; Shin, H.S.; Friedman, J.; Dracopoli, N.C. and Lander, E.S. 1992 A genetic map of the mouse suitable for typing intraspecific crosses. Genetics 131:423-447.

Eckert, J.R. 1990 PFGE and YAC analysis of the Arabidopsis genome. Methods 1:186-194.

Erlich, H.A.; Gelfand, D. and Sninsky, J.J. 1991 Recent advances in the polymerase chain reaction. Science 252:1643-1651.

Foolad, M.R.; Jones, R.A. and Rodriguez, R.L. 1993 RAPD markers for constructing intraspecific tomato genetic maps. Plant Cell Reports 12:293-297.

Foote, S.; Vollrath, D.; Hilton, A. and Page, D.C. 1992 The human chromosome: overlapping DNA clones spanning the euchromatic region. Science 258:60:66.

Gepts, P.; Stockton, T. and Sonnante, G. 1992 Use of hypervariable markers in genetics diversity studies. Proceedings of the Symposium on Applications of RAPD Technology to Plant Breeding. CSSA/ASHS/AGA. pp. 41-45.

Goodfellow, P.N. 1992 Variation is now the theme. Nature 359:777-778.

Gottlieb, 1982 Conservation and duplication of isozymes in plants. Science 216:373:380.

Grattapaglia, D.; Wilcox, P.; Chaparro, J.X.; O'Malley, D.; McCord, S.; Whetten, R.; McIntyre, L. and Sederoff, R. 1991 A RAPD map of loblolly pine in 60 days. Third International Congress of Intl. Soc. for Plant Mol. Biol. abs. 2224.

Grodzicker, T.; Williams, J.; Sharp, P. and Sambrook, J. 1974 Physical mapping of temperature-sensitive mutations of adenoviruses. Cold Spring Harbor Symp. Quant. Biol. 1974 39:439-446.

Hamada, H.; Petrino, M.G. and Kakunaga, T. 1982 A novel repeated element with Z-DNA forming potential is widely found in evolutionarily diverse eukayotic genomes. Proc. Natl. Acad. Sci. USA 79:6465-6469.

Hayashi, K. 1991 PCR-SSCP: a simple and sensitive method for detection of mutations in the genomic DNA. PCR Meth. Appl. 1:34-38.

Hedrick, P. 1992 Shooting the RAPD'S. Nature 355:679-680.

Helentjaris, T.; King, G.; Slocum, M.; Siedenstrang, C. and Wegman, S. 1985 Restriction fragment length polymorphisms as probes for plant diversity and their development as tools for applied plant breeding. Plant Mol. Biol. 5:109-118.

Heun, M. and Helentjaris, T. 1993 Inheritance of RAPDs in F1 hybrids of corn. Theor. Appl. Genet. 85:961-968.

Hillel, J.; Schaap, T.; haberfeld, A.; Jeffreys, A.J.; Plotzky, Y.; Cahaner, A. and lavi, U. 1990 DNA fingerprints applied to gene introgression in breeding programs. genetics 124:783-789.

Hiratsuka, J.; Shimada, H.; Whittier, R.; Ishibashi, T.; Sakamoto, M.; Mori, M.; Kondo, C.; Honji, Y.; Sun, C.R.; Meng, B.Y.; Li, Y.Q.; Kanno, A.; Nishizawa, Y.; Hirai, A.; Shinozaki, K. and Sugiura, M. 1989 The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of cereals. Mol. Gen. Genet. 217:185-194.

Hunter, R.L. and Markert, C.L. 1957 Histochemical demonstration of enzymes separated by zone electrophoresis in strach gels. Science 125:1294-1295.

Jeffreys, A.J.; Wilson, V. and Thein, S.L. 1985a Hypervariable 'minisatellite' regions in human DNA. Nature 314:67-73.

Jeffreys, A.J.; Wilson, V. and Thein, S.L. 1985b Individual-specific "fingerprints" of human DNA. Nature 316:76-79.

Keim, P.; Shoemaker, R.C. and palmer, R.G 1989 Restriction fragment length polymorphism diversity in soybean. Theor. Appl. Genet. 77:786-792.

Kidd, V.J.; Golbus, M.S.; Wallace, R.B.; Itakura, K. and Woo, S.L.C. 1984 Prenatal diagnosis of alfa-antitripsin deficiency by direct analysis of the mutation site in the gene. N.E.J. Med. 310:639-642.

Landegren, U.; Kaiser, R.; Sanders, J. and Hood, L. 1988 A ligase-mediated gene detection technique. Science 241:1077-1080.

Landry, B.S.; Hubert, N.; Crete, R.; Chang, M.S. and Lincoln, S.E. 1992 A genetic map for *Brassica oleracea* based on RFLP markers detected with expressed DNA sequences and mapping of resistance genes to race 2 of *Plasmodiophora brassicae* (Woronin). Genome 35:409-420.

Lewin, B. 1990 Genes IV. Oxford University Press, Oxford.

Lewontin, R.C. and Hubby, J. 1966 A molecular approach to the study of genetic heterozygosity in natural populations. II. Amounts of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. genetics 54:595-609.

Lisitsyn, N.; Lisitsyn, N. and Wigler, M. 1993 Cloning the differences between two complex genomes. Science 259:946-951.

Litt, M. and Luty, J.A. 1989 A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am. J. Hum. Genet. 44:398:401.

Livneh, O.; Vardi, E.; Stram, Y.; Edelbaum, O and Sela, I. 1992 The conversion of a RFLP assay into PCR for the determination of purity in a hybrid pepper cultivar. Euphytica 62:97-102.

Love, J.M.; Knight, A.M.; McAleer, M.A. and Todd, J.A. 1990 Toward construction of a high resolution map of the mouse genome using PCR-analysed microsatellites. Nucl. Acid Res. 18:4123-4130.

Miyao, A.; Fukuoka, S.; Inoue, T.; Monna, L.; Zhong, H.S. and Sasaki, T. 1993 SSCP anlysis of rice genome DNA. Rice Genome 2:4.

Moore, S.S.; Sargeant, L.L.; King, T.J.; Mattick, J.S.; Georges, M. and Hetzel, D.J.S. 1991 The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. Genomics 10:654-660.

Morgante, M. and Olivieri, A.M. 1993 PCR-amplified microsatellites as markers in plant genetics. The Plant J. 3:175-182.

Moss, D.W. 1982 Isoenzymes. Capman and Hall, London and New York.

Mullis, K. and Faloona, F. 1987 Specific synthesis of DNA in vitro via a polymerase catalysed chain reaction. Methods Enzymol. 55:335-350.

Murphy, R.W.; Sites, J.W.Jr.; Buth, D.G. and Haufler, C.H. 1990 Proteins I: Isozyme electrophoresis. In: Molecular Systematics. Hillis, D.M. and Moritz, C. (eds.). Sinauer Associates, Sunderland MA. pp. 45-126.

Nakamura, Y.; Leppert, M.; O'Connell, P.; Wolff, R.; Holm, T.; Culver, M.; Martin, C.; Fujimoto, E.; Hoff, M.; Kumlin, E. and White, R. 1987 Variable Number of Tandem Repeat (VNTR) markers for human gene mapping. Science 235:1616-1622.

Neale, D.B. and Williams, C.G. 1991 Restriction fragment length polymorphism mapping in conifers and applications to forest genetics and tree improvement. Can. J. For. Res. 21:545-554.

NIH/CEPH Collaborative Mapping Group 1992 A comprehensive genetic linkage map of the human genome. Science 258:67-102.

Nybom, H. and Hall, H.K. 1991 Minisatellite DNA 'fingerprints' can distinguish Rubus and estimate their degree of relatedness. Euphytica 53:107-114.

Olson, M.; Hood, L.; Cantor, C. and Doststein, D. 1989 A common language for physical mapping of the human genome. Science 254:1434-1435.

Orita, M.; Suzuki, Y.; Sekiya, T. and Hayashi, K. 1989 Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. genomics 5:874-879.

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

Paran, I.; Kesseli, R. and Michelmore, R. 1991 Identification of restriction fragment length polymorphism and random amplified polymorphic DNA markers linked to downy mildew resistance genes in lettuce, using near-isogenic lines. Genome 34:1021-1027.

Penner, G.A.; Bush, A.; Wise, R.; Kim, W.; Domier, L.; Kasha, K.; Laroche, A.; Scoles, G.; Molnar, S.J. and Fedak, G. 1993 Reproducibility of random amplified polymorphic DNA (RAPD) analysis among laboratories. In: PCR methods and applications Cold Spring Harbor Laboratory Press, pp. 341-345.

Rafalski, J.A.; Tingey, S.V. and Williams, J.G.K. 1991 RAPD markers - a new technology for genetic mapping and plant breeding. AgBiotech News and Information 3:645-648.

Reiter, R.S.; Williams, J.G.K.; Feldmann, K.A.; Rafalski, J.A.; Tingey, S. V. and Scolnik, P.A. 1992 Global and local genome mapping in Arabidopsis thaliana by using recombinant inbred lines and random amplified polymorphic DNAs. Proc. Natl. Acad. Sci. USA 89:1477-1481.

Riedy, M.F.; Hamilton, W.J.III; Aquadro, C.F. 1992 Excess of non-parental bands in offspring from known primate pedigrees assayed using RAPD-PCR. Nuc. Acid Res. 20:918.

Riles, L.; Dutchik, J.E.; Baktha, A.; McCauley, B.K.; Thayer, E.C.; Leckie, M.P.; Braden, V.V.; Depke, J.E. and Olson, M.V. 1993 Physical maps of the six smallest chromosomes of Saccharomyces cerevisiae at a resolution of 2.6 kilobase pairs. Genetics 134:81-150.

Rogowsky, P.M.; Shepherd, K.W. and Langridge, P. 1992 Polymerase chain reaction based mapping of rye involving repeated DNA sequences. Genome 35:621-626.

Rogstad, S.H.; Patton, J.C.II and Schaal, B.A. 1988 M13 repeat probe detects DNA minisatellite-like sequences in gymnosperms and angiosperms. Proc. Nat. Acad. Sci. USA 85:9176-9178. .

Royle, N.J.; Clarkson, R.E. ;Wong, Z. and Jeffreys, A.J. 1988 Clustering of hypervariable minisatellites in proterminal regions of human autossomes. Genomics 3: 352-360.

Saiki, R.K.; Gelfand, D.H.; Stoeffel, S.; Shcarf, S.J.; Higuchi, G.; Horn, G.T.; Mullis, K.B. and Erlich, H.A. 1988 Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487-491.

Saiki, R.K.; Scharf, S.J.; Faloona, F., Mullis, K.B.; Horn, G.T.; Erlich, H.A. and Arnheim, N. 1985 Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230:1350-1354.

Schmidt, T. 1991 Distribution and evolution of satellite DNA families in the genus *Beta*. Theor. Appl. Genet. 82:793-799.

Serikawa, T.; Kuramoto, T.; Hilbert, P.; Mori, M.; Yamada, J.; Dubay, C.J.; Lindpainter, K.; Ganten, D.; Guenet, J.; Lathrop, G.M. and Beckmann, J.S. 1992 Rat gene mapping using PCR-analyzed microsatellites. Genetics 131:701-721.

Smithies, O. 1955 Zone electrophoresis in strach gels: group variation in the serum proteins of normal human adults. Biochemical J. 61:629-641.

Soller, M. and Beckmann, J.S. 1983 genetic polymorphism in varietal identification and genetic improvement. Theor. Appl. Genet. 67:25-33.

Southern, E.M. 1975 Detection of specific sequences among DNA fragments separated by electrophoresis. J. Mol. Biol. 98:503-517.

Steffen, P.; Eggen, A.; Dietz, A.B.; Womack, J.E.; Stranzinger, G. and Fries, R. 1993 Isolation and Mapping of Polymorphic Microsatellites in Cattle Animal Genetics 24: 121-124.

Stuber, C.W.1992 Biochemical and molecular markers in plant breeding. In: Plant breeding reviews Vol. 9. Dudley, J.W.; Hallauer, A.R. and Ryder, M. (eds) John Wiley and Sons, Inc.

Takaiwa, F.; Oono, K.; Lida, Y. and Sugiura, M. 1985 The complete nucleotide sequence of a rice 25S rRNA gene. Gene 37:255-259.

Tanksey, S.D. and Orton, T.J. (eds.) 1983 Isozymes in plant genetics and breeding (Parts A and B). Elsevier Science Publishers, Amsterdam.

Tanksley, S.D. 1983a Gene mapping. In: Isozymes in plant genetics and breeding, part A. Tanksley, S.D. and Orton, T.J. (eds). Elsevier Science Publishers B.V., Amsterdam.

Tanksley, S.D. 1983b Molecular markers in plant breeding. Plant Mol. Biol. Report 1(1):3-8.

Tanksley, S.D.; Young, N.D.; Paterson, A.H. and Bonierble, M.W. 1989 RFLP mapping in plant breeding: new tools for an old science. Bio/Technology 7:257-264.

Tanksley, S.D.; Medina-Filho, H. and Rick, C.M. 1982 Use of naturally occurring enzyme variation to detect and map genes controlling quantitative traits in an interspecific backcross of tomato. Heredity 49:11-25.

Tautz, D. 1989 Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucl. Acid Res. 17:6463-6471.

Tautz, D. and Renz, M. 1984 Simple sequences are ubiquitous repetitive componenets of eukaryotic genomes. Nucleic Acids Res. 12:4127-4138.

Thomas, M.R.; Matsumoto, S.; Cain, P. and Scott, N.S. 1993 Repetitive DNA of grapevine: classes present and sequences suitable for cultivar identification. Theor. Appl. Genet. 86:173-180.

Thormann, C.E. and Osborn, T.C. 1992 Use of RAPD and RFLP markers for germplasm evaluation. Proceedings of the Symposium on Applications of RAPD Technology to Plant Breeding. CSSA/ASHS/AGA. pp. 9-11.

Thormann, C.E.; Ferreira, M.E.; Camargo, L.E.A.; Tivang, J.G. and Osborn, T.C. 1993 Comparison of genetic relationship estimates among cruciferous species based on RFLP and RAPD markers. Theor. Appl. Genet. (no prelo).

Tingey, S.V. and DelTufo J.P. 1992 Genetic analysis with Random Amplified Polymorphic DNA markers. Plant Physiol. 101:349-352.

Tragoonrung, S.; Kanazin, V.; hayes, P.M. and Blake, T.K. 1992 Sequence-tagged-site-facilitated PCR for barley genome mapping. Theor. Appl. Genet. 84:1002-1008.

Vassart, G.; Georges, M.; Monsieur, R.; Brocas, H.; Lequarre, A.S. and Christophe, D. 1987 A sequence in M13 phage detects hypervariable minisatellites in human and animal DNA. Science 235:683-684.

Wang, D.; Waye, M.M.Y.; Taricani, M.; Buckingham, K. and Sandham, H.J. 1993 Artifactual variation in randomly amplified polymorphic DNA banding patterns. Biotechniques 14(2):214-218.

Weining, S. and Langridge, P. 1991 Identification and mapping of polymorphisms in cereals based on the polymerase chain reaction. Theor. Appl. Genet. 82:209-216.

Weir, B. 1990 Genetic data analysis. Sinauer Associates, Sunderland, MA.

Weising, K.; Weigand, F.; Driesel, A.J.; Kahl, G.; Zischler, H. and Epplen, J.T. 1989 Polymorphic simple GATA/GACA repeats in plant genomes. Nucl. Acid Res. 17:120-128.

Weissenbach, J.; Gypay, G.; Dib, C.; Vignal, A.; Morissette, J.; Millasseau, P.; Vaysseix, G. and Lathrop, M. 1992 A second-generation leinkage map of the human genome. Nature 359:794-801.

Welsh, J. and McClelland, M. 1990 Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res. 18:7213-7218.

Welsh, J. and McClelland, M. 1991 Polymorphisms generated by arbitrarily primed PCR in the mouse: application to strain identification and genetic mapping. Nucleic Acids Res. 19:303-306.

White, T.J.; Arnheim, N. and Erlich, H.A. 1989 The polymerase chain reaction. Trends in Genetics 5(6):185-189.

Williams, J.G.; Kubelik, A.R.; Livak, K.J.; Rafalski, L.A. and Tingey, S.V. 1990 DNA polymorphism amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 18:6531-6535.

Williams, J.G.K.; Rafalski, J.A. and Tingey, S.V. 1992 genetic analysis using RAPD markers. Methods Enzymol. (in press).

Williams, M.N.V.; Pande, N.; Nair, S.; Mohan, M. and Bennett, J. 1991 Restriction fragment length polymorphism analysis of polymerase chain reaction products amplified from mapped *loci* of rice (*Oryza sativa* L.) genomic DNA. Theor. Appl. Genet. 82:489-498.

Wu, D.Y.; Ugozzoli, L.; Pal, B.K. and Wallace, R.B. 1989 Allele-specific enzymatic amplification of B-globin genomic DNA for diagnosis of sickle cell anemia. Proc. Natl. Acad. Sci. USA 86:2757-2760.

Wu, T.Y. and Wu, R. 1987 A new rice repetitive DNA shows sequences homology to both 5S RNA and tRNA. Nucleic Acids Res. 15:5913-5923.

Wyman, A.R. and White, R. 1980 A highly polymorphic locus in human DNA. Proc. Natl. Acad. Sci. USA 77:6754-6758.

Zhao, X.P and Kochert, G. 1993 Phylogenetic distribution and genetic mapping of a (GGC)(n) microsatellite from rice (Oryza sativa L.). Plant Mol. Biol. 21:607-614.

Zhao, X.; Wu, T.; Xie, Y. and Wu, R. 1989 Genome specific repetitive sequences in the genus *Oryza*. Theor. Appl. Genet. 67:835-840.

ABSTRACT

This paper reports the nuclear DNA content estimates obtained by flow cytometry for a group of twelve *Eucalyptus* species and five fast growing hybrids that includes those most widely planted throughout the world. Estimates of nuclear (2C) DNA content for the species surveyed ranged from 0.77 pg/2C for *Eucalyptus citriodora* (Subgenus *Corymbia*) to 1.47 pg/2C for *E. saligna* (Subgenus *Symphyomyrtus*). This range corresponds to a haploid genome size range of 3.7 to 7 x 10^8 bp. The average physical equivalent of a 1 cM distance could be as low as 200 kilobase pairs in *Eucalyptus*, an attractive feature for positional cloning efforts in woody plants. The closer the species were in phylogenetic relationship the more similar their nuclear DNA content values. All the interspecific hybrids surveyed displayed a nuclear DNA content in the expected intermediate range between the respective parental species, with the exception of one originating from Rio Claro, Brazil, whose exact parentage is unknown. No evidence of polyploidy was observed in any of the hybrids. The flow cytometry procedure employed in this study is an efficient method for investigating ploidy levels of high yielding hybrids of *Eucalyptus*.

INTRODUCTION

Eucalyptus (Myrtaceae) is a large genus of evergreen hardwood trees and shrubs that includes about 700 species. It is almost exclusively native to Australia, with a few species indigenous to adjacent islands. Only two species, *E. deglupta* and *E. urophylla* , are endemic outside Australia (Boland *et al.* 1984; Brooker and Kleinig 1990). The genus *Eucalyptus* includes the most widely used tree species for plantation establishment in tropical and subtropical regions of the world. Among the hundreds of species, around 30 have shown potential for high yielding plantations. Eucalypt plantations were reported to occupy 4 million hectares worldwide, and almost all of the area planted with just 10 species: *E. camaldulensis, E. globulus, E. grandis, E. maculata, E. paniculata, E. robusta, E. saligna, E. tereticornis, E. urophylla* and *E. viminalis* (FAO 1981). This area has steadily increased in recent years and today eucalypts constitute the majority of the world's exotic hardwood forest . Fast growth rates and a wide range in adaptability have contributed to the great interest that eucalypt species enjoy in many countries outside their native range. Besides the fast growth that allows for shorter

rotations, many species display wood properties that make them very suitable for fuel and charcoal production or pulp and paper manufacture.

In spite of its commercial importance and a world-wide effort in breeding and propagation research, very little time has been devoted to the investigation and development of molecular genetic information about species of this genus. The few studies include the analysis of RFLP patterns in chloroplast DNA (Steane *et al.* 1991), the development and screening of RFLP probe libraries for *E. grandis* (Wolff *et al.*, 1993); , phylogenetic and fingerprinting analyses of individuals and populations with RAPD markers (Grattapaglia *et al.* 1992), construction of linkage maps for *E. grandis* and *E. urophylla* (Grattapaglia and Sederoff 1992); linkage mapping in *E. nitens* and *E. globulus* (Moran *et al.* 1992; Song and Cullis 1992) and the isolation of genes involved in the biosynthesis of lignin in *E. gunnii* (Boudet *et al.* 1992). In this context estimates of nuclear DNA content constitute basic and important information for genome structure analysis, genetic mapping of qualitative and quantitative trait loci, and especially for devising strategies to isolate and clone genes of interest, particularly through map-based cloning.

A list of nuclear DNA content estimated by Feulgen microdensitometry was assembled for 753 angiosperm species by Bennett and Smith (1976). A supplementary list of 240 species was later reported by Bennett *et al.* (1982). These compilations did not include any tree species. In recent years flow cytometry has become the technique of choice for nuclear DNA content estimation. It allows rapid fluorescence measurement of large numbers of stained nuclei and yields estimates of improved accuracy over the traditional Feulgen microdensitometry (Leutweiler *et al.* 1984). Using this methodology, Arumuganathan and Earle (1991) reported estimates of the nuclear DNA content of over 100 important plant species. Only a few fruit tree species and no forest tree species were reported in their article. This paper is intended to report the nuclear DNA content estimates obtained by flow cytometry for a group of *Eucalyptus* species that includes those most widely planted throughout the world. In view of the importance of interspecific hybridization in the genus (reviewed by Griffin *et al.* 1988) we also surveyed some of the most common hybrids used in plantation forestry. DNA content determination was employed as an approach to investigate the level of ploidy for such hybrids that were suspected polyploids in view of their high yielding ability.

MATERIAL AND METHODS

Genetic material. The species and hybrids surveyed in this study with their respective provenance or hybrid origin are listed in Table I. Seeds were originally collected from native stands in Australia and germinated in a greenhouse. Hybrid individuals, maintained as rooted cuttings or micropropagules, were either produced through controlled pollination or as a result of natural hybridization events in exotic conditions as evidenced by their intermediate morphology. These hybrids were provided by the following Brazilian forestry companies: Aracruz Florestal, Acesita Florestal and Riocell . A leaf sample for measurement of DNA content consisted of a bulk of healthy fully expanded leaves from a minimum of 5 individuals per species, or 5 ramets per hybrid clone.

Preparation of nuclei and flow cytometry. Leaves, either lyophilized or frozen in liquid nitrogen, were pulverized in a small coffee grinder, suspended in nuclei buffer (10mM Tris-HCl pH 8/1mM EDTA/0.3M sucrose), and homogenized with a Brinkmann Polytron 10-TS tip for 1min at a power setting of seven to dissociate nuclei from cell wall debris. Crude nuclei were sedimented by centrifugation (5 min. 1000xg). The nuclear homogenate was filtered through two layers of cheesecloth and a layer of Miracloth (Calbiochem). The pellet from 10g of leaves was resuspended in 1ml of nuclei buffer, diluted into 0.15M NaCl containing 10ug/ml DAPI (Sigma), and the clumped nuclei dispersed with 10 passes through a 26 gauge syringe needle. Flow cytometry was performed on an ICP-22 (Ortho Diagnostic Systems, Westwood, MA) with chicken erythrocytes (2.34pg/ diploid nucleus; Mirsky and Ris 1949) used as an internal standard. Several hundred nuclei were assayed for each measurement of DNA content. Two measurements were performed per species/hybrid.

Estimation of nuclear DNA content. Estimation was based on the fluorescence histogram by comparing the mean position of the peak of the plant nuclei with the mean peak position of the internal standard (CE, chicken erythrocytes 2.34pg/2C).

RESULTS AND DISCUSSION

The flow cytometry procedure used for estimation of nuclear DNA content was simple and reliable. Although fairly tough and high in phenolics, the *Eucalyptus* leaf tissue was readily amenable to the analysis. Both fresh and lyophilized leaf tissues yielded equivalent results. Younger leaves gave higher nuclei yields than older leaves. Leaf tissue from micropropagated plantlets gave the highest yields of nuclei.

Figure 1 shows a frequency distribution of number of nuclei as a function of relative fluorescence intensity (DNA content) for *Eucalyptus grandis*. Indicated in the distribution are the sample peak of plant nuclei and the chicken erythrocytes peak (standard peak). The coefficient of variation (CV) of the G1 peak for the samples analyzed ranged from 6.3 to 12.8%, while the CV for the standard peak varied between 2.5 and 3.7%. The same range in CV was observed both for the species samples (minimum of five genotypes) and for the hybrid samples (clonal material). Therefore the greater variation in the eucalypt samples as compared to the chicken erythrocytes is more likely due to differences in purity of the nuclei preparation than a real variation between genotypes within the sample. Nuclear DNA content was calculated by comparing the mean peak position of the plant nuclei to the mean peak position of the standard as described in materials and methods. No significant variation was observed between the two measurements performed on each species/hybrid sample.

Nuclear DNA content and genome size in millions of base pairs (Mbp) for the 12 species and 5 hybrids surveyed were estimated (Table I). DNA contents ranged from as low as 2C= 0.77 pg for *Eucalyptus citriodora* to 1.47 pg for *E. saligna*. One picogram (pg) of DNA is approximately equivalent to 965 million base pairs (Bennett and Smith, 1976), therefore these values correspond to a haploid genome size range from 3.7 to 7 x 10^8 bp. In view of the estimation methodology used, the precision of the estimates is given to three significant figures as for the CE standard used (2.34 pg/2C). This precision corresponds to the nearest 10^7 bp in terms of haploid genome size.

In this study, the closer the species were in phylogenetic relationship the more similar their C values. Species within the same section tended to show very similar DNA contents, and species from different subgenera showed the most pronounced differences. Species of subgenus *Symphyomyrtus* have DNA contents varying from 1.09 pg/2C for *E.globulus* to 1.47pg/2C for *E. saligna*, the highest value encountered. The variation in genome sizes of species of this subgenus is similar to species of *Brassica*,

(average size of 1.3 pg/2C), corresponding approximately to a haploid genome size of 6×10^8 bp (Arumuganathan and Earle 1991). Species of subgenus *Corymbia*, namely *E. citriodora* and *E. torelliana* have a much smaller nuclear DNA content of around 0.8 pg/2C (3.7×10^8 bp), slightly over half the size of species of *Symphyomyrtus*. This low value is similar to another tree species, *Citrus sinensis* 0.76-0.82 pg/2C (Arumuganathan and Earle 1991).

The nuclear DNA content of the hybrids surveyed was found to be in the intermediate range between the respective parental species with the exception of the spontaneous hybrid *Eucalyptus urophylla* x *E. alba*. Its DNA content was estimated as 0.94 pg/2C, much lower than that of *E. urophylla* (1.34 pg/2C). The DNA content of *Eucalyptus alba* (Section *Exsertaria*) was not measured in this study. However based on its close phylogenetic relationship to *E. camaldulensis*, it is expected that it would fall between 1.2 and 1.3 pg/2C. The significantly lower DNA content of the hybrid could be explained by the fact that its correct parentage is unknown. Although morphologically classified as *E. urophylla* x *E. alba* , the estimate of DNA content obtained in this study and records about its origin jointly suggest that other species of lower DNA content were involved in its composition. This hybrid was selected from plantations originated from seeds collected in the Rio Claro arboretum in Sao Paulo, Brazil, where a large number of species were planted side by side in an exotic environment allowing for free hybridization (Zobel and Talbert, 1984).

Anecdotal speculations have been raised among eucalypt breeders that the increased productivity of fast growing intra and interspecific hybrids of *Eucalyptus* could be a result of an increased level of ploidy, similar to some interspecific hybrid *Populus* clones (Muntzing 1936; Bradshaw and Stettler 1993). The spectacular growth potential of eucalypt hybrids has been documented throughout the world for several hybrid combinations (Chaperon 1984; Sinoir 1984; Brandao *et al.* 1984; Kapoor and Sharma 1984; Meskimen and Franklin 1984). All the interspecific hybrids surveyed in this study are very fast growing genotypes used in vegetative propagation programs. They all displayed a nuclear DNA content in the expected diploid range. No multiple peaks corresponding to higher levels of ploidy were observed on any occasion in the fluorescence histograms. Results of this work clearly demonstrate that the five hybrid genotypes surveyed are not polyploids. These results however do not preclude this possibility for other hybrids. The flow cytometry procedure employed in this study

establishes a very straightforward and efficient method for determining ploidy levels of high yielding hybrids of *Eucalyptus*.

Subgenus *Symphyomyrtus* includes the majority of the most widely planted and bred species of *Eucalyptus*. Due to their importance, species of this genus have been and will continue to be the most likely candidates for genetic mapping, gene isolation and cloning efforts. In this study we have estimated an average haploid genome size of 650 megabase pairs for members of *Symphyomyrtus*. Total genomic map distances were estimated between 1200 and 1500 centiMorgans for *E. grandis* and *E. urophylla* through linkage mapping using RAPD markers (Grattapaglia and Sederoff 1992). These sizes provide an estimate of 400 to 600 kilobase pairs as the average physical equivalent of a 1 cM distance for these species. Total map distance per chromosome is relatively constant even across widely divergent species (Rees and Durant 1986). For *E. torelliana* and *E. citriodora* both with a genome size around 380 megabase pairs, assuming a similar range of total map distance as for *E. grandis* and *E. urophylla* , 1 cM would correspond to 200 to 300 kilobase pairs.

A small nuclear genome, with a low proportion of interspersed highly repetitive DNA is an important characteristic for cloning genes through map-based, or positional cloning strategies. Positional cloning requires the detection of DNA markers (e.g. RFLPs and RAPDs) flanking the gene of interest. These markers are then used to identify genomic clones which can be used as starting points for chromosome walking towards the gene. As total map distance per chromosome is relatively constant, the smaller the genome, the shorter the physical distance between any two markers thus facilitating the walk to the gene. *Arabidopsis thaliana* with a genome size of 70 megabase pairs has been the model plant of choice to clone genes using this strategy (Arondel *et al.* 1992). *Eucalyptus torelliana* and *E. citriodora* with a genome size of 380 megabase pairs would be a very suitable woody plant species for such a gene cloning approach. Besides the relatively small genome size, some commercially important species of *Eucalyptus* have other characteristics that would make it an attractive candidate model system for woody plant molecular genetics. These include: *i)* ability to induce abundant flowering in less than a year using paclobutrazol (Cauvin 1991); *ii)* ease of rejuvenation and vegetative propagation ; *iii)* ease of "in vitro" cell and tissue culture manipulations; *iv)* availability of methods for transient gene expression (Manders *et al.* 1992) and stable transformation (J. Purse, Shell Research, personal communication).

ACKNOWLEDGMENTS

LITERATURE CITED

Arondel, V., Lemieux, B., Hwang, I., Gibson, S., Goodman, H.M. and Somerville, C.R. 1992. Map-based cloning of a gene controlling Omega-3 fatty acid desaturation in *Arabidopsis*. Science **258**:1353-1354.

Arumuganathan, K. and Earle, E.D. 1991. Nuclear DNA content of some important plant species. Plant Mol. Biol. Reporter **9**:208-218

Bennett, M.D. and Smith, J.B. 1976. Nuclear DNA amounts in angiosperms. Phil. Trans. R. Soc. Lond. **B274**:227-274.

Bennett, M.D., Smith, J.B. and Heslop-Harrison, J.S. 1982. Nuclear DNA amounts in angiosperms. Proc. R. Soc. Lond. **B216**:179-199.

Boland, D.J., Brooker, M.I.H., Chippendale, G.M., Hall, N., Hyland, B.P.M., Johnston, R.D., Kleinig, D.A. and Turner, J.D. 1984. Forest trees of Australia. 4th ed. Nelson-CSIRO, Melbourne.

Boudet A.M., Grima-Pettenati, J., Feuillet, C., Goffner, D., Joffroy, I. and Hawkins, S. 1992. Cinnamyl alcohol dehydrogenase in *Eucalyptus gunnii*: polymorphism, molecular cloning and expression. 5th IUFRO Workshop Molecular Biology of Forest Trees S2.04.06. Carcans-Maubuisson, France, abstract # 1.1.

Bradshaw, H.D. and Stettler, R.F. 1993. Molecular Genetics of growth and development in *Populus*. I. Triploidy in hybrid poplars. Theor. Appl. Genet. **86**:301-307.

Brandao, L.G., Campinhos, E. and Ikemori, Y.K. 1984. Brazil's new forest soars to success. Pulp & Paper Int. **26**:38-40.

Brooker, M.I.H. and Kleinig, D.A. 1990. Field guide to *Eucalyptus*. Vol. 2. Inkata Press, Melbourne and Sydney.

Cauvin, B. 1991. Effet du Paclobutrazole sur la mise a fleur, la croissance de clones juveniles d'*Eucalyptus*. Annales de Recherches Sylvicoles - AFOCEL pp. 71-88.

Chaperon, H. 1984. Influence of propagation by cuttings on the breeding strategy of forest trees. *In*: Provenance and Genetic Improvement Strategies in Tropical Forest Trees. *Edited by* R.D. Barnes and G.L. Gibson. Dep. of Forestry, University of Oxford and Zimbabwe Forestry Commission. pp. 135-148.

F.A.O. 1981. Eucalypts for planting. Forestry Series 11, Rome: Food and Agriculture Organization of the United Nations.

Grattapaglia, D. and Sederoff, R.R. 1992. Pseudo-testcross mapping strategy in forest trees: single tree RAPD maps of *Eucalyptus grandis* and *E. urophylla*. Plant Genome I, San Diego, CA poster # 47.

Grattapaglia, D., O'Malley, D.M., Sederoff, R.R. 1992. Multiple applications of RAPD markers to genetic analysis of *Eucalyptus* sp. Proceedings of IUFRO International Conference "Breeding tropical trees" Section 2.02-08 Cali, Colombia (in press).

Griffin, A.R., Burgess, I.P. and Wolf, L. 1988. Patterns of natural and manipulated hybridisation in the *Eucalyptus* L'Hérit. - a review. Aust. J. Botany **36**:41-66

Kapoor, M.L. and Sharma, V.K. 1984. Hybrids between *Eucalyptus citriodora* Hook. and *E. torelliana* F.V. Muell. in India. Sil.v. Genet. **33**:42:46.

Leutwiler, L.S., Hough-Evans, B.R. and Meyerowitz, E.M. 1984. The DNA of *Arabidopsis thaliana*. Mol. Gen. Genet. **194**:15-23.

Manders, G., Dossantos, A.V.P., Vaz, F.B.D., Davey, M.R. and Power, J.B. 1992. Transient gene expression in electroporated protoplasts of *Eucalyptus citriodora* Hook. Plant Cell Tissue and Organ Culture **30**:69-75.

Meskimen, G. and Franklin, E.C. 1984. Hybridity in the *Eucalyptus grandis* breeding population in Florida. USDA For. Serv. Res. Paper SE-242

Mirsky A.E., Ris H.1949. Variable and constant components of chromosomes. Nature **163**:666-669.

Moran , G.F., Bell, J.C., Byrne, M., Devey, M.E. and Smith, D.1992. Molecular marker maps in *Pinus radiata* and *Eucalyptus nitens* for application in tree breeding programs. Plant Genome I, San Diego, CA poster # 93.

Muntzing, A. 1936. The chromosomes of a giant *Populus tremula*. Hereditas **21**:383-393.

Pryor, , L.D. and Johnson, L.A.S. 1971. A classification of the Eucalypts. Aust. Nat. Univ. Press, Canberra.

Rees, H. and Durant, A. 1986. Recombination and genetic size. Theor. Appl. Genet. **73**:72-76.

Sinoir, M. 1984. Les hybrides d'*Eucalyptus* au Congo, situation en Octobre 1984. Centre Technique For. Trop. Congo.

Song, Y. and Cullis, C.A. 1992. RFLP and RAPD mapping of *Eucalyptus globulus*. Plant Genome I, San Diego, CA poster # 133.

Steane, D.A., West, A.K., Potts, B.M., Ovenden, J.R. and Reid, J.B. 1991. Restriction fragment length polymorphism in chloroplast DNA from six species of *Eucalyptus*. Aust. J. Bot. **39**:399-414.

Wolff J.L.C., McDowell, S.W. and Machado, M.A. 1993. Analysis of the restriction fragment length polymorphism of *E. urophylla*, *E. grandis* and *E. citriodora*. Rev. Bras. Genet. **16**(2): 431-439

Zobel, B. and Talbert, J. 1984. Applied forest tree improvement. John Wiley & Sons Inc., New York.

Table 1. Species and hybrids of *Eucalyptus* and their respective nuclear DNA content and genome size as estimated by flow cytometry. Species were grouped by subgenus and section according to Pryor and Johnson (1971). Provenances (Australia) and origin of the hybrids (Brazil) are listed when available.

SPECIES	ORIGIN	NUCLEAR DNA CONTENT	
		pg/2C	~Mbp/1C
Subgenus *Symphyomyrtus*			
Section *Exsertaria*			
E. camaldulensis Dehnh.	Cooktown	1.23	590
E. tereticornis Smith	Mackay Dist.	1.20	580
Section *Transversaria*			
E. grandis Hill ex Maiden	Atherton	1.33	640
E. pellita F. Muell.	Helenvalle	1.37	660
E. resinifera Smith	S8285 (CSIRO)	1.28	620
E. robusta Smith	n.a.[1]	1.35	650
E. saligna Smith	Coffs Harbour	1.47	710
E. urophylla S.T. Blake	Timor	1.34	650
Section *Maidenaria*			
E. dunnii Maiden	n.a.	1.10	530
E. globulus Labill.	n.a.	1.09	530
Subgenus *Corymbia*			
Section *Ochraria*			
E. citriodora Hook.	n.a.	0.77	370
E. torelliana F. Muell.	Atherton	0.80	390
Hybrids			
E. grandis x E. urophylla	C.P.[2] (Aracruz - BR)	1.33	640
E. grandis x E. camaldulensis			
Clone 016	S.H.[3] (Acesita - BR)	1.36	660
Clone 252	S.H.[3] (Acesita - BR)	1.21	580
E. tereticornis x E. grandis	S.H. (Riocell - BR)	1.38	670
E. citriodora x E. torelliana	C.P. (Acesita - BR)	0.77	370
E. urophylla x E. alba	S.H. (Rio Claro -BR)	0.94	450

[1]n.a. not available; [2]C.P. Controlled Pollination; [3]S.H. Spontaneous Hybrid

ABSTRACT

In this paper we present results of a series of genetic analyses carried out on *Eucalyptus* species using RAPD (Random Amplified Polymorphic DNA) genetic markers. These results describe the power, reliability, ease of use and multiplicity of applications of this new class of molecular markers for genetic studies in the genus *Eucalyptus* as well as in any forest tree species. A "Population Bulking" technique was used to establish genetic relationships among nine species and a known hybrid. The most parsimonious phylogenetic tree constructed agreed closely with the proposed taxonomy. Genetic fingerprints were easily obtained with two highly polymorphic RAPD primers that discriminated a group of ten commercially propagated hybrid clones. Preliminary results are presented on the estimation of mating parameters for a seed orchard of *E. urophylla*. The multilocus outcrossing rate estimate based on 26 dominant RAPD markers was t_m = 0.885 ± 0.039. It reveals that outcrossing is prevalent in the seed orchard but that a significantly low level of inbreeding may be occurring. This estimate agrees very closely with results from previous studies that used codominant isozymes markers. Genetic inheritance, segregation and linkage of RAPD markers in *Eucalyptus* were established for 58 RAPD markers using a controlled F1 cross between *E. grandis* and *E. urophylla*. A pseudo-testcross mapping strategy is briefly described. Such strategy in combination with the DNA polymorphism uncovering power of RAPD allows very quick generation of RAPD genetic linkage maps for elite individuals. We conclude that RAPD is a DNA marker technology that is very accessible to tree geneticists and breeders both technically and economically. Large number of neutral markers are available which allow extensive sampling of the target genome. RAPD markers are well suited for the analysis of relatedness and genotype discrimination at the individual and group levels. For studies of mating parameters the large number of marker loci obtained with RAPD compensate for the small amount of information given by each locus due to dominance. The universality of the RAPD assay makes it particularly attractive for studies of mating systems in tropical tree populations. Genetic linkage maps constructed by the pseudo-testcross strategy will be a powerful tool to study the architecture of important quantitative traits in forest trees, and potentially to carry out marker aided breeding.

INTRODUCTION

Genetic markers have been a useful tool in forest tree genetics and breeding in the last two decades. Markers detected as isozyme polymorphisms have been widely used to label individuals or populations and to estimate levels and distributions of genetic variation in populations, (Mueller-Starck 1982; Shaw and Allard 1982; Adams 1983; Brown *et al.* 1985). High levels of isozyme variability have been found in most forest trees to date. In the genus *Eucalyptus*, isozymes have been used for the study of mating systems in natural (Moran and Bell 1983; Fripp *et al.* 1987; Sampson *et al.* 1990) and exotic populations (Yeh *et al.* 1983) and for the determination of phylogenetic relationships (Burgess and Bell 1983; Prober *et al.* 1990). Isozymes are codominant, relatively inexpensive and easy to use, and as such they will continue to be an important class of markers. However, for applications that require a broader genome coverage, isozymes are severely limited by the number of available loci.

With the development of molecular techniques, large numbers of polymorphic loci can be obtained at the DNA level by codominant restriction fragment length polymorphisms. RFLP's have been used to construct linkage maps for identifying QTL's in crop species (Lander and Botstein 1989; Paterson *et al.* 1988) and trees (Devey *et al.* 1990; Neale and Williams 1991; Bradshaw and Foster 1992). In spite of the power of RFLP technology, it is still expensive, laborious and requires more complex molecular expertise that make this assay less desirable for tree genetics and breeding projects that require the analysis of large number of samples in modest laboratory settings.

The recently published PCR based technique called RAPD for Randomly Amplified Polymorphic DNA (Williams *et al.* 1990; Welsh and McClelland 1990) provides a new way of detecting polymorphism at the DNA level. This technique is based on the amplification of random DNA segments with single primers of arbitrary nucleotide sequence. RAPD fragments are produced by PCR amplification of DNA, based on the frequency and location of short inverted DNA repeats. The fragment is identified by the primer sequence used to amplify it and its size in base pairs. RAPD markers are highly polymorphic, at least equivalent to RFLP markers and far more than isozymes. The products of RAPD amplification reactions can be viewed directly by agarose gel electrophoresis much like isozymes without the need for Southern blots. RAPD does not require specific probe library construction and the amount of DNA required for the RAPD assay is about one thousand fold less than for RFLP. The RAPD methodology is

much more accessible to the nonmolecular biologist than RFLP methods while providing the same virtually unlimited supply of neutral markers to cover the entire genome. For some applications, RAPD markers might be limited by the fact that in contrast to RFLP and isozymes, the allelic forms detected are in the great majority dominant rather than codominant.

RAPD's have been used as molecular markers for several purposes: linkage map construction in conifers (Grattapaglia *et al.* 1991; Tulsieram *et al.* 1992), mapping of disease resistance genes through Bulk Segregant Analysis (Michelmore *et al.* 1991), cultivar and clone fingerprinting (Hu and Quiros 1991, Wilde *et al.* 1992), phylogenetic studies (van Heusden and Bachmann 1992) and the study of introgression in natural populations (Arnold *et al.* 1991). In this paper we present results of a series of genetic analyses of individuals and species of *Eucalyptus* carried out using RAPD markers. The objective of this paper is to describe the power, reliability, ease of use and multiplicity of applications of this new class of molecular markers in forest trees genetics and breeding with special emphasis to *Eucalyptus*. Phylogenetic relationships were established among species using a DNA bulking technique. DNA fingerprints were obtained for a group of commercially propagated hybrid clones. Genetic inheritance, segregation and linkage of RAPD markers were established, and a pseudo-testcross mapping strategy is described. Finally, preliminary results are presented where RAPD dominant markers were used for the estimation of mating parameters in a seed orchard.

MATERIAL AND METHODS

Plant Material. Seeds from *E. grandis, E. urophylla, E. grandis x E. urophylla, E. pellita, E. resinifera, E. tereticornis* were obtained from Aracruz Florestal S.A. (ES, Brazil) and seeds from *E. paniculata, E. dunnii, E. torelliana and E. citriodora* were obtained from Acesita Florestal S.A. (MG, Brazil). Open pollinated seeds of *E. urophylla* collected by mother tree from a three-year old seed orchard were obtained from IPEF (Instituto de Pesquisa Florestal), Brazil. Seeds from a controlled cross between *E. grandis* (clone 44) and *E. urophylla* (clone 28) were also obtained from Aracruz. Seeds were surface sterilized and germinated "in vitro" on agar solidified 1/2 strenght MS medium and grown under 12 hours light. Adult expanded leaves from 10 commercially planted hybrid clones and the parents of the controlled cross were collected from field plantation at Aracruz Florestal S.A. These leaves were freeze dried and stored at - 20ºC for later use.

DNA isolation and bulking. Total genomic DNA was isolated from freeze dried adult leaf tissue of the parents and from fully expanded leaves of *in vitro* plantlet progeny. A quick DNA miniprep procedure modified from Doyle and Doyle (1987) was used. Approximately 300 mg of fresh or 50 mg of freeze dried tissue was ground to a fine powder in liquid nitrogen directly in 1.5 ml microtubes using a plastic pestle. Eight hundred µl of extraction buffer (2% CTAB, 1% PVP, 1.4 M NaCl, 100mM Tris-HCL pH 8.0, 20 mM EDTA pH 8.0 and 1% 2-mercaptoethanol) was added. Tubes were incubated at 65°C for 30 to 60 min and inverted every 15 min. The tubes were brought to room temperature, then 700 µl of chloroform:isoamyl alcohol 24:1 was added and the tubes repeatedly inverted until a good emulsion was obtained. Tubes were centrifuged for phase separation (12,000 xg, 5 min). The upper aqueous phase was transferred to a new tube containing 700 µl of ice-cold isopropanol. Tubes were gently inverted a few times until a precipitate could be seen. When precipitate was not easily visible the tubes were chilled at -20°C for 30 min. DNA was pelleted by centrifugation (12,000 xg, 10 min). Pellets were washed once in 1.5 ml of 70% ethanol and once in 500 µl of 95% ethanol. Pellets were either dried in a speed-vac (Savant) or air-dried over night and then dissolved in 50 µl of TE buffer (10mM Tris-HCl pH 8.0 and 1mM EDTA pH 8.0) containing 50 µg/ml RNAse. DNA quality was checked on an agarose gel and the concentration estimated by comparing the fluorescence intensities of the ethidium bromide stained samples to those of lambda DNA standards. Samples were diluted to between 2.0 and 4.0 ng/µl DNA in sterile water for RAPD analysis. Bulked DNA samples for each species or provenance were prepared by pooling 200 ng of genomic DNA from each one of 20 randomly chosen individuals.

RAPD analysis. Random ten base primers were obtained from Operon Technologies (Alameda, CA). Amplification reactions were performed in volumes of 13 µl containing 1.3 µl of the 10X buffer supplied by the *Taq* polymerase manufacturer, 100µM each of dATP, dCTP, dGTP, dTTP (Promega), 0.2 mM primer, 5 to 10 ng of genomic DNA and 0.8 units of Taq DNA polymerase. Amplification was performed in 96-well plates using a MJ Research PT-100 thermal controller programmed for 41 cycles of 1 min. at 92°C, 1 min. at 35°C, 2 min. at 72°C. Amplification products were analyzed by horizontal gel electrophoresis in 1.5% agarose TBE gels and detected by ethidium bromide staining. Gels were photographed under U.V. light with Polaroid film 667. RAPD

fragments visualized as bands on a gel were scored for presence or absence across the individuals or bulks analyzed. Fragment sizes were calculated using the software SEQAID II (Rhoads and Roufa 1990).

Data analysis. RAPD markers that were reproducible in multiple runs were used in the studies. For the phylogenetic study, polymorphic RAPD bands were listed as discrete characters (present/absent) per species bulks. Most parsimonious trees were determined by branch and bound search and bootstrap resampling using the PAUP (Phylogenetic Analysis Using Parsimony) software (Swofford 1990). In the segregation study, $\chi2$ tests of goodness-of-fit to the 1:1 segregation ratio and of independence of segregation for RAPD marker pairs were performed. For the estimation of mating system parameters, a modification of the MLT program, MLDT (Multilocus estimation of outcrossing with dominant markers) (Ritland 1990) was used.

RESULTS

Optimization of the RAPD assay for Eucalyptus. Among the several DNA isolation procedures tested, the CTAB procedure used proved to be the most efficient to obtain genomic DNA that would consistently amplify by PCR. Starting tissue, however, is a critical variable. Young, tender tissue from expanding leaflets was the best starting material to obtain high yields of clean DNA. In order to obtain such tissue, plants were either grown in the greenhouse under intensive care or maintained 'in vitro'. For adult leaf tissue, the DNA procedure was essentially the same. However, oxidized phenolics and polyssacharides occasionally co-precipitated with the DNA and tended to inhibit the PCR reaction. This occurrence was genotype-dependent. To overcome this problem the strategy we used was not to try cleaning the DNA but rather to either establish an 'in vitro' stock of the genotype of interest or to induce epicormic shoots from branches under mist irrigation to obtain cleaner starting tissue.

The optimal RAPD reaction conditions for *Eucalyptus* were established after testing variable concentrations of template genomic DNA, magnesium and Taq polymerase (data not shown). Concentrations between 5 and 10 ng of genomic DNA are optimal. Magnesium concentration between 1.5 and 2.0 mM can be used, however 1.5mM was the concentration at which the RAPD reaction shows less primer dependency. Taq polymerase concentrations down to 0.5 units (in 13 µl) can be used,

however with 0.8 units consistency is better. Over 200 different random ten-base oligonucleotides were tested in different studies. The originally described primer concentration of 0.2 mM was optimal. However variable success was obtained in the assay depending on the primer used. Occasionally, raising the primer concentration to 0.4 mM significantly improved the RAPD band profiles. In general about 50 to 70% of the random primers screened yielded informative DNA polymorphisms.

Population Bulking and phylogenetic analysis. Figure 1 shows the effect of bulking DNA on the RAPD banding pattern. On the left side of the gel, the DNA of 15 randomly chosen individuals of *Eucalyptus grandis* (Atherton provenance) were amplified. On the right side of the gel 15 bulks of the same DNA samples were amplified with the same primer. These bulks referred as "bootstrap" bulks were constructed by leaving out one at a time one of the individual's DNA. So bulk # 1 has all but individual # 1, bulk # 2 has all but # 2 and so on. This experiment was carried out with 10 different random primers yielding the same result. The RAPD fragment variation among individuals is lost by bulking and an average RAPD banding profile is obtained for the group of bulked individuals. For composing a representative bulk of the Eucalyptus grandis population studied, bulks of a minimum of 15 individuals were necessary and only fragments with frequencies above 0.2 were amplified in the bulk due to competition among amplification sites (data not shown). We refer to this technique as "population bulking". It follows the same basic principles of the Bulk Segregant Analysis described by Michelmore *et al.* (1991) for targeting specific genomic regions. This population bulking strategy was used to study the genetic distance between 9 *Eucalyptus* species and a hybrid. Bulks of 20 individuals were used for the analysis. A total of 20 primers amplified 39 RAPD fragments that were selected as being informative for the study. Figure 2 shows the RAPD banding profiles for six of those primers. Within each RAPD primer, the first six lanes show quite similar banding profiles sharing most of the RAPD fragments, while the other four, and particularly the last two show different size bands. The similarity of the banding profiles observed were expected on the basis of the proposed taxonomic relationships among the species bulks analyzed.

The branch and bound search yielded two very similar phylogenetic trees. After bootstrap resampling the consensus tree obtained is depicted in Figure 7. The phylogenetic distances obtained closely agree with the taxonomic subdivisions described by Pryor and Johnson (1981). Within the subgenus *Symphyomyrtus, E. grandis,*

E. urophylla, E. pellita and *E. resinifera* which belong to the section *Transversaria*, group together. *E. tereticornis*, the next closest species, belongs to a different section (*Exsertaria*), for which several hybrids with members of *Transversaria* have been described (Griffin *et al.* 1988). *E. dunnii* belongs to a third section (*Maidenaria*) for which hybrids both with *Transversaria* and *Exsertaria* have also been reported (Griffin *et al.* 1988). *E. paniculata*, which outgroups within *Symphyomyrtus* belongs to a fourth section, *Adnataria* which does not cross readily with species in other sections (Pryor and Willing 1974). Finally *E. citriodora* and *E. torelliana* which belong to subgenus *Corymbia* group togheter in a separate branch.

Genetic fingerprinting of Eucalyptus clones. Discriminating RAPD fingerprints were obtained for all 10 clones analyzed. A total of 20 random primers were screened for discriminatory polymorphisms. Figure 4 shows an example of RAPD genetic fingerprints. After selecting the highest polymorphisms yielding primers, only two primers were necessary to unambiguously discriminate every clone in the group. Table 1 shows the fingerprinting genotypes for the 10 clones. Several RAPD markers are shared among subsets of clones, however each clone has a specific combination of RAPD markers (identified by the primer number and fragment size) when both primers are considered together.

Genetic inheritance, segregation and linkage of RAPD markers. To assess the inheritance and segregation of RAPD markers, forty random primers were used to carry out RAPD analysis on a sample of 6 F1 individuals plus the two parents of a controlled cross. This screening step consisted in looking for the presence of RAPD fragments that were transmitted from parents to their F1 and that segregated for presence and absence among the progenies. Figure 5 shows the result of this screening step with six different primers. Several RAPD fragments were found to be polymorphic between the two parents and to segregate in the progeny. From a total of 40 primers screened, 28 were found to yield at least one polymorphic segregating RAPD marker. For a total of 58 markers, segregation was analyzed on 30 F1 individuals. Figure 6 shows the genetic inheritance and segregation of several RAPD markers amplified with a single primer. Results of the $\chi 2$ analysis of goodness-of-fit are given in Table 2. The observed 1:1 ratio follows the expected Mendelian segregation of alleles in a test cross, where one of the parents is heterozygous (genotype +/-) and the other parent is homozygous null (-/-) for

the locus under consideration. This configuration is here referred as pseudo-testcross since the marker configuration of the parents is inferred from the progenies "a posteriori", i.e. without prior planning as in a true testcross. For only four RAPD loci a significant χ^2 statistic was obtained, which indicates distortion from a 1:1 ratio. To minimize type I errors, a cut-off χ^2 of 15.00 was used as criterion to determine linkage between pairs of markers. The observed number of linkages agrees with the expectation based on the prior probability of linkage with the available number of markers as described earlier (Grattapaglia *et al.* 1992). With such stringency, only close linkages (recombination fraction $\theta < 0.15$) were estimated. In order to obtain more precise estimates of recombination fraction for linkage map construction, a larger number of progeny needs to be genotyped.

Estimation of mating parameters. A total of 26 dominant RAPD marker loci were selected based on their amplification intensity and occurrence across families. As an example, Figure 3 shows the presence of two RAPD loci in two different families. A total of 128 individuals, 32 for each of four families were scored for the presence or absence of the selected RAPD markers. These markers (identified by primer number and fragment size) together with the maximum likelihood estimates of their gene frequencies (presence of the fragment), standard error on these estimates and singlelocus χ^2 goodness of fit of t_m (population multilocus outcrossing rate) to the data are shown in Table 3. RAPD marker frequencies estimates varied in the range of 0.05 to 0.65. These frequencies are very appropriate for studies of outcrossing rates. Standard error on these estimates were obtained by 100 bootstrap resampling within families. Significant departures from the mixed mating model were detected for 6 of the 26 RAPD loci scored. The multilocus outcrossing estimate based on all 26 RAPD loci was $t_m = 0.885 \pm 0.039$, which did not differ from t_s (average singlelocus outcrossing rate) (Table 3). It reveals that outcrossing is prevalent in the *E. urophylla* seed orchard studied but that a significantly low level of inbreeding may be occurring. This estimate agrees very closely with the only estimate available in exotic conditions (0.853) obtained with isozyme loci for a different species, *E. citriodora*, in Brazil. Estimates for a closely related species, *E. grandis*, in natural stands was 0.84 (Moran 1983).

To check the robustness of the outcrossing rate estimates, several analyses were performed on different subsets of 10 or 15 RAPD loci, as well as by leaving out the loci that departed from the mixed mating model. The maximum likelihood estimates for t_m

and t_s always converged to values between 0.85 and 0.9 independently of which set of loci was used or from which starting value (higher or lower) of t_m was used for the iterative process. In spite of the fact that only four families were used in this preliminary study the estimate of outcrossing rate seems quite robust. The large number of loci used is probably responsible for this, since for t_m near one as in *Eucalyptus*, it has been shown that it is better to assay more loci instead of more plants (Shaw and Brown 1982). Since a very small number of mother trees were analyzed, the estimate of F Wright's fixation index of maternal parents has to be viewed with caution. The negative estimate denotes a slight excess of heterozygotes, which agrees with earlier estimates (Yeh *et al.* 1983) and is in concordance with the hypothesis that selection operates against homozygotes through the life cycle in *Eucalyptus*. Such a selection mechanism has been reported operating in *E. regnans* (Griffin and Cotterill 1988).

DISCUSSION

RAPD markers constitute a very powerful tool for genetic analysis in species of the genus *Eucalyptus* as well as in any highly heterozygous forest tree species. The "population bulking" technique here described allows for a very simple way of finding DNA markers that are specific to the groups that constitute the bulks and to find distances/similarities among such groups. An important aspect of this procedure is that the bulking mechanism removes detectable variability among individuals within a bulk and yields an "average genome" RAPD profile that characterizes the bulked group. This short cut is useful when seeking informative sequence differences among groups of individuals, e.g. populations, provenances, species. The combined use of bulks and RAPD markers significantly reduces the number of individual samples to be analyzed to uncover those differences. In this process, the larger the portion of the genome that is sampled, the larger the number of potentially informative markers obtained and therefore the more reliable will be the estimates of differences or similarities among the groups. In the phylogenetic study described herein, a total of 39 markers were used and those were obtained with only 12 primers.The phylogenetic grouping of the species agreed closely with the proposed taxonomy. Using a similar approach we found 129 informative markers to study phylogenetic relationships among closed cone pines (Grattapaglia *et al.* 1992). Besides studying phylogenetic relationships, this bulking strategy potentially allows finding species/provenance specific markers that could be

used to determine the species composition of natural and spontaneous hybrids of *Eucalyptus* used in clonal deployment programs throughout the world.

Along the same line stands fingerprinting analysis. We have shown here with *Eucalyptus* that primers can be selected to simultaneously amplify a large number of polymorphic fragments that can be easily used to discriminate clones. In this study with only two primers it was possible to establish clone specific RAPD banding genotypes within a group of 10 clones. Evidently, as the number of clones to be discriminated increases, more RAPD primers/markers will be necessary. Also, as the clones to be doscriminated are more closely related by species or provenance origin, more RAPD markers will have to be surveyed for informative differences. RAPD genetic fingerprinting is a very simple and inexpensive tool for quality control of clonal deploymnet programs, and for protection of proprietary rights on elite clonal material.

Our preliminary results show that RAPD markers although dominant can be reliably used to determine population outcrossing rates. The large number of markers that can be found segregating in different families compensates for the small amount of information given by each locus. Ritland and Jain (1981) when describing the mixed mating model have shown that relatively more loci are needed with dominance to obtain satisfactory estimates of gene frequencies and outcrossing rates. Isozyme markers are quite efficient and have traditionally been used for such studies. However a potential advantage of RAPD markers for outcrossing rate studies is the universality of the assay protocols which would allow analyzing a broad range of tree species without prior need for optimizing isozyme detection systems. This feature is particularly relevant to research on the genetic diversity and conservation of tropical tree species.

We have reported the genetic inheritance, segregation and linkage of RAPD markers in *Eucalyptus*. We have observed that due to the high level of heterozygosity in the individuals it was possible to find a considerable large number of markers in a pseudo-testcross configuration between the parents which yield a segregation ratio of 1:1 in the F1. When analyzed for co-segregation, RAPD markers showed linkages. As pointed out earlier (Ritter *et al.* 1990, Carlson *et al.* 1991; Grattapaglia *et al.* 1992), the pseudo-testcross mapping strategy in combination with the polymorphism uncovering power of RAPD is a very efficient tool for constructing linkage maps for individual diploid trees such as *Eucalyptus* In a very similar way as the haploid biology allowed such mapping in conifers (Grattapaglia *et al.* 1991; Tulsieram *et al.* 1992). The concept of screening primers in search of the informative markers as opposed to relying on known

loci is a new one that has become possible with the technological advance brought about by RAPD.

In conclusion we have shown that RAPD is a powerful DNA marker technology that is also very accessible to tree geneticists and breeders both technically and economically. It is particularly interesting in developing countries since it does not depend on a large number of reagents, uses few and simple equipments and does not require constant supply of radioisotopes. Large numbers of neutral markers are available which allow detailed study of the target genome. RAPD makes pre-selection of informative markers easy allowing very fast generation of marker data for multiple applications. RAPD markers are well suited for the analysis of relatedness and genotype discrimination at the individual level. The described "population bulking" strategy allows a similar approach at the group level, i.e. species or provenances. The pseudo-testcross mapping strategy will lead to very quick generation of individual genetic linkage maps. Such maps can be used to study the architecture of important quantitative traits in forest trees, and potentially carry on marker aided breeding strategies in structured elite populations. In *Eucalyptus* breeding, linkage disequilibrium generated by hybridization coupled with the possibility of clonal propagation are conditions that could greatly favour the use of such strategies.

ACKNOWLEDGMENTS

LITERATURE CITED

Adams, W.T. 1983. Applications of isozymes in tree breeding. In: Tanskley, SD, Orton, TJ (eds) Isozymes in Plant Genetics and Breeding, part A. Elsevier, Amsterdam, pp. 381-400.

Arnold, M.L.; Buckner, C.M. and Robinson, J.J. 1991. Pollen mediated introgression and hybrid speciation in Louisiana irises. Proc. Natl. Acad. Sci. USA 88:1398-1402.

Bradshaw, H.D. Jr. and Foster, G.S. 1992. Marker aided selection and propagation

systems in trees: advantages of cloning for studying quantitative inheritance. Can. J. For. Res. 22:1044-1049.

Brown, A.H.D.; Barrett, S.C.H. and Moran, G.F. 1985. Mating systems estimation in forest trees: models, methods, meanings. In: Gregorius, HR (ed) Population Genetics in Forestry. Lecture Notes in Biomathematics 60, Springer Verlag NY pp 32-49

Burgess, I.P. and Bell, J.C. 1983. Comparative morphology and allozyme frequencies of *Eucalyptus* grandis Hill ex Maiden and E. saligna Sm. Aust. For. Res. 13(2):133-149.

Carlson, J.E.; Tulsieram, L.K.; Glaubitz, J.C.; Luk, V.W.K.; Kauffeldt, C. and Rutledge, R. 1991. Segregation of random amplified DNA markers in F1 progeny of conifers. Theor. Appl. Genet. 83:194-200.

Devey, M.E.; Jermstad, K.D.; Tauer, C.G. and Neale, D.B. 1991. Inheritance of RFLP loci in a loblolly pine three-generation pedigree. Theor. Appl. Genet. 83:238-242.

Fripp, Y.J.; Griffin, A.R. and Moran, G.F. 1987. Variation in allele frequencies in the outcross pollen pool of *Eucalyptus* regnans F Muell. throughout a flowering season. Heredity 59:161-171.

Grattapaglia, D.; Wilcox, P.; Chaparro, J.X.; O'Malley, D.M.; McCord, S; Whetten, R.; McIntyre, L. and Sederoff, R.R. 1991. A RAPD map of loblolly pine in 60 days. International Society for Plant Molecular Biology Third International Congress, Tucson AZ, Abstr. # 2224

Grattapaglia, D.; O'Malley, D.M. and Dvorak, W.S. 1992. Phylogenetic analysis of Central American and Mexican pines using RAPD markers on bulked DNA samples. Proceedings of IUFRO Conference Section 2.02-08, Breeding tropical trees. Cali, Colombia , pp. 132-147.

Grattapaglia, D.; O'Malley, D.M. and Sederoff, R.R. 1992. Linkage mapping of *Eucalyptus* using an F1 cross between heterozygous parents. Abstracts of the Sixth Meeting of the International Conifer Biotechnology Working Group. Research Triangle Park, NC.

Griffin, A.R.; Burgess, I.P. and Wolf, L. 1988. Patterns of natural and manipulated hybridisation in the genus *Eucalyptus* L'Herit. - a review. Aust. J. Bot. 36:41-66.

Griffin, A.R. and Cotterill, P.P. 1988. Genetic variation in growth of outcrossed, selfed and open pollinated progenies of *Eucalyptus* regnans and some implications for breeding strategy. Silvae Genet. 37 (3-4):124-131.

Hu, J. and Quiros, C.F. 1991. Identification of broccoli and cauliflower cultivars with RAPD markers. Plant Cell Reports 10:505-511.

Lander, E.S. and Botstein, D. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185-199.

Michelmore, R.W.; Paran, I. and Kesseli, R.V. 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.

Moran, G.F. and Bell, J.C. 1983. *Eucalyptus*. In: Tanskley, SD, Orton, TJ (eds) Isozymes in Plant Genetics and Breeding, part B. Elsevier, Amsterdam, pp 423-441.

Muller-Starck, G. 1982. Reproductive systems in conifer seed orchards. I. Mating probabilities in a seed orchard of *Pinus sylvestris* L. Sivae Genet. 31(5-6):188-197.

Neale, D.B. and Williams, C.G. 1991. Restriction fragment length polymorphism mapping in conifers and applications to forest genetics and tree improvement. Can. J. For. Res. 21(5):545-554.

Paterson, A.H.; Lander, E.S.; Hewitt, J.D.; Peterson, S.; Lincoln, S.E. and Tanksley, S.D. 1988. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721-726.

Prober, S.; Bell, J.C. and Moran, G. 1990. A phylogenetic and allozyme approach to understanding rarity in three "green ash" eucalypts (*Myrtaceae*). Plant Syst. Evol. 172:99-118.

Pryor, L.D. and Johnson, L.A.S. 1981. *Eucalyptus*, the universal Australian. In: Ecological Biogeography of Australia. (ed. A Keast.) Junk: The Hague. pp 499-536.

Pryor, L.D. and Willing, R.R. 1974. The production of interspecific hybrids between incompatible pairs of eucalypts. Int. Plant Prop. Soc. 24:265-268.

Rhoads, D.D., and Roufa, D.J. 1990. Seqaid II 3.80. Kansas State University. Molecular Genetics Laboratory.

Ritland, K. 1990. A series of FORTRAN computer programs for estimating plant mating systems. J. Heredity 81:235-237.

Ritland, K. and Jain, S. 1981. A model for the estimation of outcrossing rate and gene frequencies using n independent loci. Heredity 47(1):35-52.

Ritter, E.; Gebhardt, C. and Salamini, F. 1990. EStimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents. Genetics 125:645-654.

Sampson, J.F.; Hopper, S.D. and James, S.H. 1990. Temporal variation in allele frequencies in the pollen pool of *Eucalyptus rhodantha*. Heredity 65:189-199.

Shaw, D.V. and Allard, R.W. 1982. Estimation of outcrossing rates in Douglas-fir using isozyme markers. Theor. Appl. Genet. 62:113-120.

Shaw, D.V. and Brown, A.H.D. 1982. Optimum number of marker loci for estimating outcrossing in plant populations. Theor. Appl. Genet. 61:321-325.

Swofford, D.L. 1990. PAUP: Phylogenetic Analysis Using Parsimony. Version 3.0. Illinois Natl. Hist. Surv., Champaign, IL.

Tulsieram, L.K.; Glaubitz, J.C.; Kiss, G. and Carlson, J.E. 1992. Single tree genetic linkage mapping in conifers using haploid DNA from megagametophytes. Bio/Technology 10:686-690.

Van Heusden, A.W. and Bachmann, K. 1992. Genotype relationships in *Microseris elegans* (*Asteraceae, Lactuceae*) revealed by DNA amplification from arbitrary primers (RAPD's). Plant Syst. Evol. (in press).

Welsh, J. and McClelland, M. 1990. Fingerprinting genomes using PCR with arbitrary primers. NAR 18:7213-7218.

Wilde, J.; Waugh, R. and Powell, W. 1992. Genetic fingerprinting of *Theobroma* clones using randomly amplified polymorphic DNA markers. Theor. Appl. Genet. 83:871-877.

Williams, J.G.K.; Kubelik, A.R.; Livak, K.J.; Rafalski, J.A. and Tingey, S.V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acid Research 18:6531-6535.

Yeh, F.C.; Brune, A.; Cheliak, W.M. and Chipman, D.C. 1983. Mating system of *Eucalyptus* citriodora in a seed production area. Can. J. For. Res. 13:1051-1055.

Table 1. RAPD fingerprinting genotypes of 10 *Eucalyptus* clones obtained with two random primers that amplified 14 polymorphic discriminatory markers identified by the primer code and fragment sizein bp; (++ = presence of marker).

	F4_344	F4_370	F4_490	F4_530	F4_540	F4_550	F4_600	F4_1650	F4_2700	B8_360	B8_520	B8_700	B8_750	B8_900
CLONE 34		++					++		++	++			++	++
CLONE 351		++					++		++		++		++	++
CLONE 567	++		++					++	++					
CLONE 1205	++		++	++	++		++	++		++	++	++		
CLONE 1225			++	++		++	++		++		++			
CLONE 1248		++	++				++						++	
CLONE 1501		++	++			++		++		++	++			
CLONE 2225	++		++					++	++				++	
CLONE 2277	++		++					++	++					
CLONE 4619			++	++	++				++		++		++	

Table 2. Genetic segregation and linkage of RAPD marker loci in *Eucalyptus*. Columns list from left to right: (1) and (2) pair of linked RAPD markers (arbitrary primer code_fragment size in bp); (3) to (6) counts of the four haplotypic marker classes ; (7) recombination fraction (θ) between linked loci; (8) and (9) χ2 statistics of goodness-of-fit test to 1:1 segregation ratio for each locus (*indicates significantly distorted locus at α=0.05); (10) χ2 statistics for linkage (significance threshold χ2 = 15.00).

LOCUS 1	LOCUS 2	(+/+)	(+/-)	(-/+)	(-/-)	θ	χ2 LOCUS 1	χ2 LOCUS 2	χ2 LINKAGE
G6_1140	H3_1323	16	2	1	9	0.11	2.29	1.29	17.29
A10_635	G12_1231	12	0	1	16	0.03	0.86	0.31	25.14
A10_635	G5_1898	11	1	3	15	0.13	1.20	0.13	16.13
A10_562	G12_1231	1	16	12	0	0.03	0.86	0.31	25.14
A10_562	G5_1898	3	15	11	1	0.13	1.20	0.13	16.13
A11_2374	B6_515	14	1	0	15	0.03	0.00	0.13	26.13
A11_1439	A20_2127	18	0	2	9	0.07	1.69	4.17*	21.55
A11_1439	A19_2448	13	1	0	10	0.04	0.67	0.17	20.17
A11_635	G12_1459	10	1	1	17	0.07	1.69	1.69	21.55
A11_635	A7_865	1	10	16	1	0.07	1.29	1.29	20.57
A11_635	A2_1744	2	6	17	1	0.12	3.85*	5.54*	15.38
D2_1095	G12_1459	9	2	0	16	0.07	0.93	3.00	19.59
D2_1095	A7_865	2	9	15	0	0.08	0.62	2.46	18.62
D2_1095	A19_398	0	9	13	1	0.04	1.09	0.39	19.17
D2_848	G2_1444	0	13	13	1	0.04	0.04	0.04	23.15
D2_848	G5_304	12	1	0	15	0.04	0.14	0.57	24.14
D3_508	G12_1459	9	1	1	17	0.07	2.29	2.29	20.57
D3_508	A7_865	1	9	16	1	0.07	1.81	1.81	19.59
D3_508	A2_1744	1	6	17	1	0.08	4.84*	4.84*	17.64
B6_810	A20_2127	18	0	2	9	0.07	1.69	4.17*	21.55
B6_810	G2_584	2	16	11	1	0.10	1.20	0.53	19.20
B6_810	G2_519	16	2	1	11	0.10	1.20	0.53	19.20
B6_810	A19_2448	13	1	0	10	0.04	0.67	0.17	20.17
B6_780	A20_1068	17	1	0	11	0.03	1.69	0.86	25.14
B6_780	A2_1691	17	0	2	7	0.08	2.46	5.54*	18.62
B7_1732	A18_509	13	1	1	11	0.08	0.15	0.15	18.62
B7_1549	A18_509	1	11	13	1	0.08	0.15	0.15	18.62
G12_1459	A19_398	0	9	12	2	0.09	1.09	0.04	15.70
G12_1231	G5_1898	12	1	1	15	0.07	0.31	0.31	21.55
A7_865	A19_398	12	2	0	9	0.09	1.09	0.04	15.70
A20_1068	G5_1898	3	14	11	1	0.14	0.86	0.03	15.21
G2_1444	G5_304	0	14	13	2	0.07	0.03	0.31	21.55

Table 3. List of RAPD markers used in the estimation of outcrossing rate in the *Eucalyptus urophylla* seed orchard. Gene frequencies for the (+) RAPD allele were estimated by maximum likelihood under a mixed mating model using MLTD. Standard errors (S.E.)were obtained by bootstrapping. RAPD marker loci displaying significant departure (α=0.05) from the mixed mating model ($\chi2$ goodness-of-fit of t_m (outcrossing rate) to the data) are indicated by an asterisk.

RAPD MRK.	GENE FREQ.	S.E.	χ^2
D8_1200	0.246	0.074	9.08*
D8_1000	0.155	0.035	0.62
D8_800	0.355	0.067	4.97*
G2_1650	0.382	0.055	2.51
G2_1400	0.495	0.066	0.12
G2_1200	0.049	0.031	4.66*
A4_900	0.178	0.055	2.73
A4_750	0.250	0.068	2.46
A4_360	0.373	0.067	2.38
A4_250	0.204	0.042	0.84
F4_1500	0.058	0.024	0.15
F4_1400	0.117	0.036	1.03
F4_550	0.159	0.038	0.34
F4_520	0.298	0.051	1.03
F4_490	0.128	0.032	2.32
F4_344	0.111	0.041	14.81*
A18_1400	0.149	0.044	0.31
A18_850	0.395	0.056	0.59
A18_450	0.135	0.053	0.11
H7_1500	0.510	0.068	0.79
H7_1100	0.477	0.070	0.21
H7_700	0.232	0.072	0.24
H7_550	0.632	0.067	0.12
H7_520	0.135	0.063	3.51
I4_1650	0.055	0.020	5.63*
I4_1200	0.252	0.068	5.19*
MULTILOCUS OUTCROSSING RATE		0.885 (0.036)	
SINGLELOCUS OUTCROSSING RATE		0.885 (0.039)	
WRIGHTS FIXATION INDEX F		- 0.059	

15 INDIVIDUAL   15 "BOOTSTRAP"
DNA SAMPLES ⟷ DNA BULKS
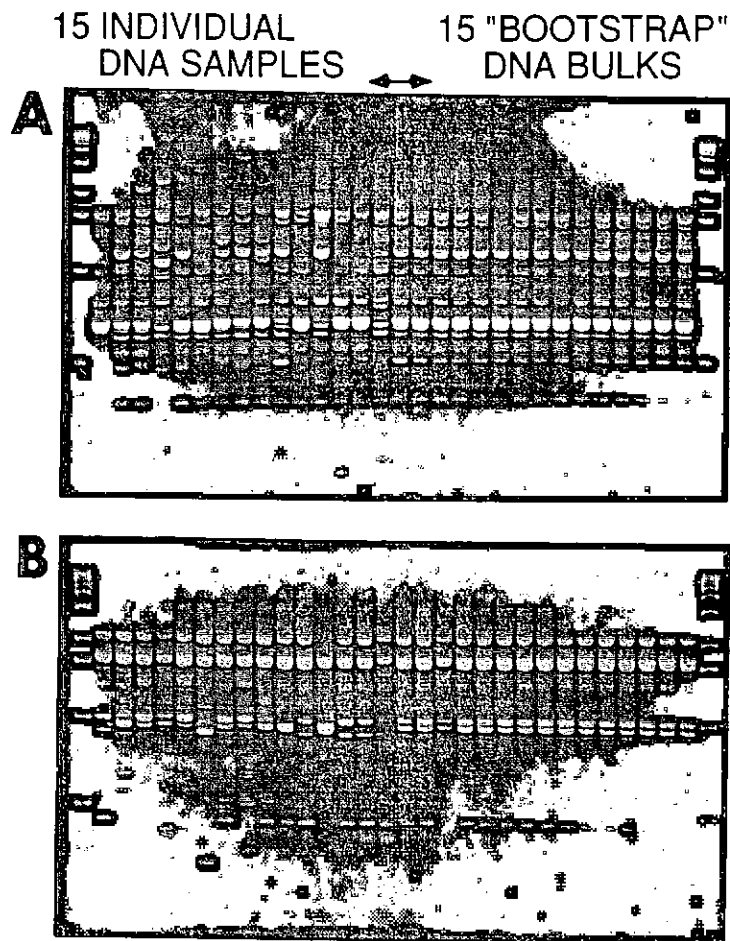
Figure 1. "Population Bulking" of genomic DNA. Genomic DNA of 15 randomly chosen individuals (left side of panels) of *Eucalyptus grandis* were assayed using RAPD. In a procedure similar to a "bootstrap", (resampling however was not random) 14 out of the same 15 DNA samples were pooled, leaving one of the samples out in each bulk (right side of panels). By bulking the DNA samples variation among individuals is canceled and a RAPD banding profile for a representative "average genome" for the population is obtained. Panel A: primer H7 (Operon); Panel B: primer B8.
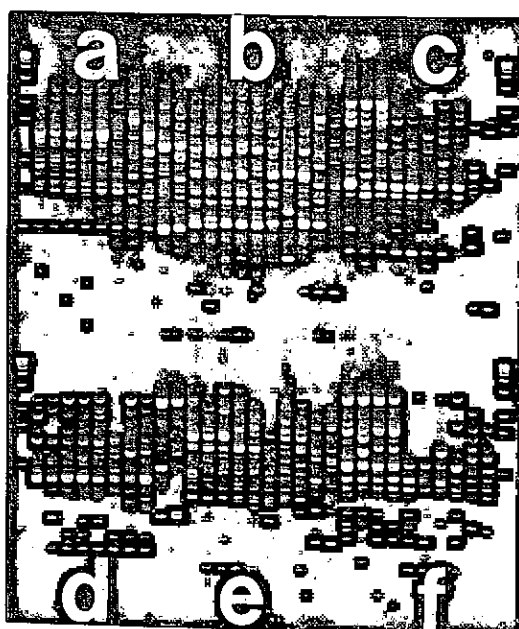
Figure 2. RAPD assays on "Population bulks" of nine species of *Eucalyptus* and a known interspecific hybrid. Letters (a) through (f) correspond to primers (Operon): A10, A11, B8, D2, F4, H7. For each primer set, 10 lanes correspond to the following genetic entries from left to right: *Eucalyptus grandis, E. grandis x E. urophylla* hybrid; *E. pellita; E. resinifera; E. tereticornis, E. paniculata; E. dunnii; E. torelliana; E. citriodora.*
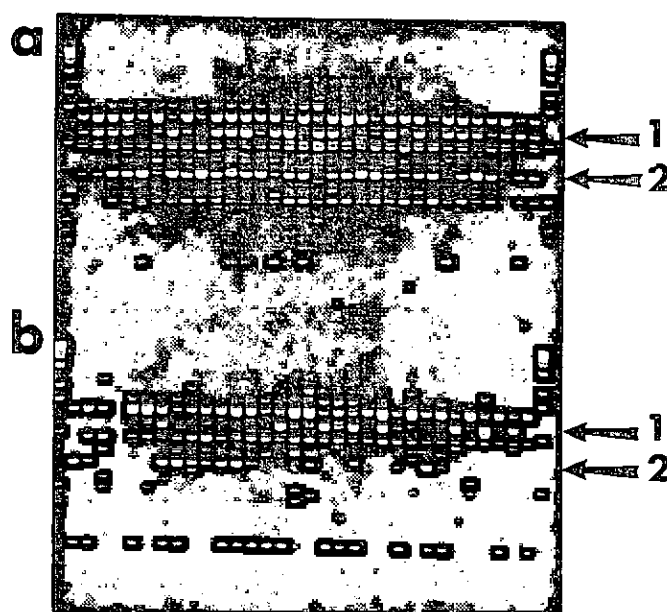
Figure 3. RAPD analysis of open pollinated progeny arrays of *Eucalyptus urophylla*. Panel (a): family 1; Panel (b): family 27. Arrows indicate two of the 26 RAPD markers used to estimate outcrossing rate (t_m). The frequency of the markers in the two families suggest that mother tree # 1 is heterozygous for markers 1 and 2, while tree # 27 is homozygous null for both markers, i.e. all progeny with the marker indicate an outcrossing event.
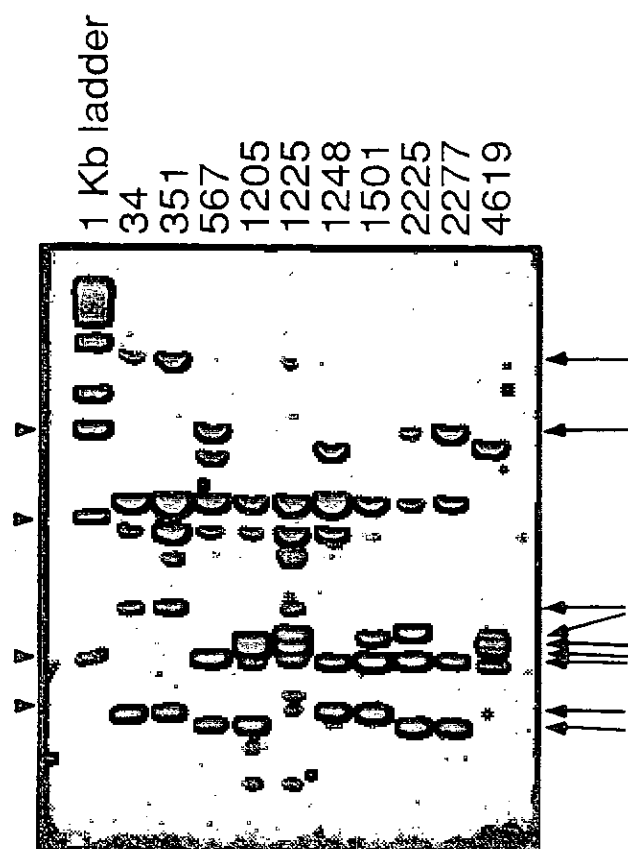
Figure 4. Genetic fingerprinting of clones using RAPD markers. Ten commercially propagated "plus" clones of *Eucalyptus* (identified by numbers on top) were assayed using RAPD primer F4 to obtain discriminatory genetic fingerprints. Arrows indicate the informative markers used in the analysis (see also Table 1 for gel interpretation).
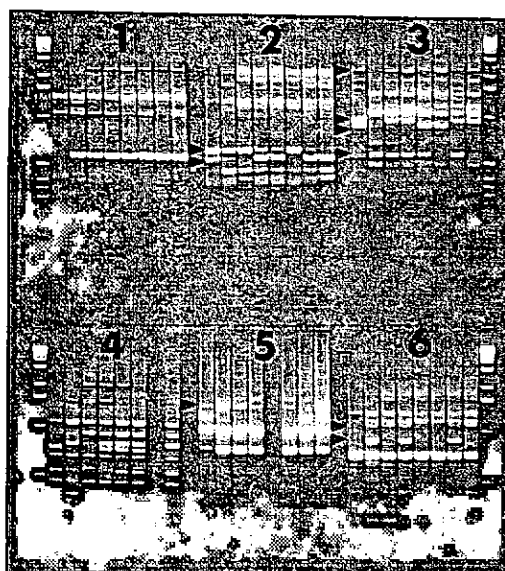
Figure 5. Screening of arbitrary primers for segregating RAPD polymorphisms. genomic DNA of the parents (first 2 lanes of each set) and 6 F1 individuals were assayed with 6 different primers (Operon) (1) through (6): A9, A10, A11, B6, B7 and D2. Selected informative markers in the pseudo-testcross configuration are indicated by arrows.
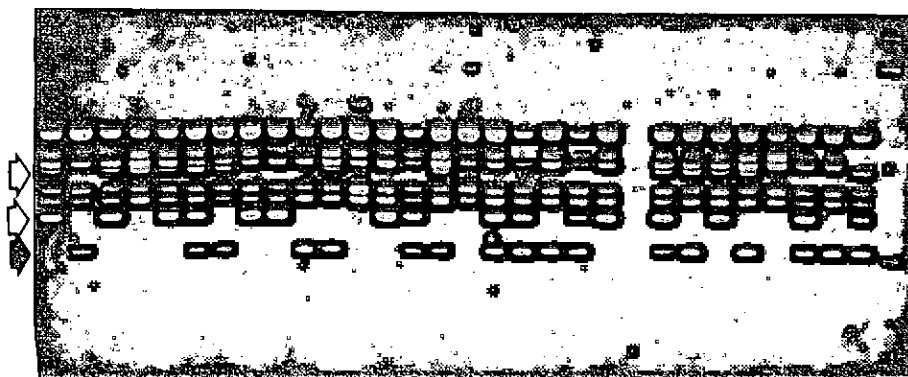
Figure 6. Genetic inheritance and segregation of RAPD markers in *Eucalyptus*. First and last lanes are 1 Kb ladder fragment size standards. Lane 2 and 3 are the parents, *E. grandis* and *E. urophylla*. The following 28 lanes are F1 progeny individuals. Blank arrows indicate markers inherited from *E. grandis* and filled arrow from *E. urophylla*.

ABSTRACT

We have used a "two-way pseudo-testcross" mapping strategy in combination with the RAPD assay to construct two moderate density genetic linkage maps for species of *Eucalyptus*. In the cross between two heterozygous individuals many single-dose RAPD markers will be heterozygous in one parent, null in the other and therefore segregate 1:1 in their F1 progeny following a test cross configuration. Meiosis and gametic segregation in each individual can be directly and efficiently analyzed using RAPD markers. We screened 305 primers of arbitrary sequence, and selected 151 to amplify a total of 558 markers. These markers were grouped at LOD 5.0, $\theta = 0.25$ resulting in the maternal *E. grandis* map having a total of 240 markers into 14 linkage groups (1552 cM) and the paternal *E. urophylla* map with 251 markers in 11 linkage groups (1101 cM) (n=11 in *Eucalyptus*). Framework maps ordered with a likelihood support $\geq 1000:1$ were assembled covering 95% of the estimated genome size in both individuals. Characterization of genome complexity of a sample of 48 mapped RAPD markers indicate that 53% amplify from low copy regions. These are the first reported high coverage linkage maps for any species of *Eucalyptus* and among the first for any hardwood tree species. We propose the combined use of RAPD markers and the pseudo-testcross configuration as a general strategy for the construction of single individual genetic linkage maps in outbred forest trees as well as in any highly heterozygous sexually reproducing living organism. A survey of the occurrence of RAPD markers in different individuals suggests that the pseudo-testcross/RAPD mapping strategy should also be efficient at the intraspecific level and increasingly so with crosses of genetically divergent individuals. The ability to quickly construct single-tree genetic linkage maps in any forest species opens the way for a shift from the paradigm of a species index map to the heterodox proposal of constructing several maps for individual trees of a population, therefore mitigating the problem of linkage equilibrium between marker and trait loci for the application of marker assisted strategies in tree breeding.

INTRODUCTION

The genus *Eucalyptus*, family *Myrtaceae*, is native to Australia and adjacent islands. It includes over 700 species of hardwood trees and shrubs (BOLAND *et al.* 1984) of which approximately thirty are grown throughout the world for fiber and energy

production. Due to their fast growth and wide range of adaptability, eucalypts are the most widely used species for plantation establishment in tropical and subtropical regions of the world. Today, they constitute the majority of the world's exotic hardwood forest and one of the world's main sources of biomass (ELDRIDGE *et al.* 1993).

In spite of its commercial importance and a world-wide effort in breeding and propagation research, very little effort has been devoted to the development and use of molecular genetic markers for species of this genus. Isozyme markers have been used for the study of mating systems in natural and exotic populations of *Eucalyptus* (MORAN and BELL 1983; FRIPP, GRIFFIN and MORAN 1987; SAMPSON, HOPPER and JAMES 1990; YEH *et al.* 1983). Isozymes are inexpensive and technically accessible markers, however they are limited when a broader genome coverage is required.

In the last decade, detailed linkage maps based on RFLP markers have been developed for a number of species (see reviews by STUBER 1992 and GRATTAPAGLIA 1994). To date only two studies analyzed RFLP patterns in *Eucalyptus*: a phylogenetic analysis of chloroplast DNA (STEANE *et al.* 1991), and the development and screening of a genomic library for RFLP probes (WOLFF, MCDOWELL and MACHADO 1993). RFLP analysis requires the "up front" development of probe libraries and involves Southern blot hybridization, procedures that are time consuming and labor intensive. The advent of a PCR-based arbitrarily primed genetic assay called RAPD (Random Amplified Polymorphic DNA), AP-PCR (Arbitrarily Primed PCR) or DAF (DNA Amplification Fingerprinting) (WILLIAMS *et al.* 1990; WELSH and MCCLELLAND, 1990; CAETANO-ANOLLÉS, BASSAM and GRESSHOFF 1991), has provided an extremely efficient way to detect DNA polymorphisms and generate large numbers of molecular markers for genetic mapping and genomic fingerprinting applications (TINGEY and DELTUFO 1992). RAPD markers do not require prior sequence information and can be viewed directly by agarose gel electrophoresis without the need of specific probe libraries and radioisotope detection. RAPD markers are therefore much more accessible to the nonmolecular biologist than RFLP's.

In recent years, RAPD markers have allowed a significant advance in the ability to generate linkage maps quickly. Maps of RAPD markers were reported for loblolly pine (GRATTAPAGLIA *et al.* 1991), white spruce (TULSIERAM *et al.* 1992), *Arabidopsis* (REITER *et al.* 1992), apple (LAWSON, HEMMAT and WEEDEN 1992) sugar cane (AL-JANABI *et al.* 1993), faba bean (TORRES, WEEDEN and MARTIN 1993), slash pine (NELSON, NANCE and DOUDRICK 1993) and peach (CHAPARRO *et al.* 1994). In *Eucalyptus*, RAPD markers have been used in

genetic analyses of individuals and populations including clone fingerprinting, outcrossing rate estimation and phylogenetic relationship studies (GRATTAPAGLIA, O'MALLEY and SEDEROFF 1992). In *Eucalyptus*, no reports of genetic linkage of morphological or isozyme markers are available. However, in the last two years the construction of linkage maps of molecular markers has been undertaken for *E. grandis* and *E. urophylla* (GRATTAPAGLIA and SEDEROFF 1992), *E. globulus* (SONG and CULLIS 1992) and *E. nitens* (MORAN *et al.* 1992).

Most linkage maps in plants have been obtained from segregating populations derived from crosses between inbred lines. Such populations or even three-generation pedigrees are generally not available in trees and are difficult to obtain due to a significant genetic load and time constraints. In spite of this obstacle, existing three-generation outbred *Pinus taeda* and inbred *Populus* pedigrees (DEVEY *et al.* 1992; BRADSHAW and STETTLER 1993) have been used as mapping populations. Alternatively in conifers, haploid megagametophytes have allowed the direct analysis of linkage in gametes and the construction of genetic maps (CONKLE 1980; GRATTAPAGLIA *et al.* 1991; TULSIERAM *et al.* 1992; NELSON, NANCE and DOUDRICK 1993). Available pedigrees for the majority of angiosperm outbred tree species, including the eucalypts, generally involve only two parents and their full-sibs, or maternal half-sib families. To incorporate molecular marker assisted strategies into forest tree breeding it is imperative to explore alternative approaches for the construction of linkage maps that make use of pedigree structures already existing and commonly generated in tree breeding programs. Particularly in the case of trees, genetic linkage maps could prove a very powerful tool for accelerating breeding through marker assisted selection and recombination. Breeding programs of *Eucalyptus* species are based on simple or reciprocal recurrent selection, interspecific hybridization and, in some cases, clonal selection and deployment. Linkage disequilibrium generated by hybridization coupled with the possibility of capturing non additive genetic variance through clonal propagation are conditions that would greatly enhance the potential use of molecular marker assisted breeding strategies in these species.

In this paper we report the genetic inheritance, segregation and linkage of 558 single-dose RAPD markers for two closely related species of *Eucalyptus*, *E. grandis* Hill ex Maiden and *E. urophylla* S.T. Blake. We used a two-generation interspecific pedigree to construct single-tree genetic linkage maps using a two-way "pseudo-testcross" mapping strategy. We propose this mapping approach as a general strategy to

generate single individual linkage maps quickly for highly heterozygous organisms. These are the first reported genetic linkage maps for *Eucalyptus* and among the first for any hardwood tree species.

MATERIALS AND METHODS

Plant material. A single controlled cross between two highly heterozygous elite trees was selected for genetic mapping. *Eucalyptus grandis* (clone 44, Coffs Harbor provenance, Australia - selection from a Zimbabwe seed source), used as the female parent was crossed to *E. urophylla* (clone 28 selection from Rio Claro land race, Brazil), used as male, in 1989 at Aracruz Florestal S.A., Brazil. These two species belong to the same subgenus and section. They are easily crossed and their progeny are fully fertile. A second controlled cross used in the study was performed in the same year and location between a different female parent, *Eucalyptus grandis* (clone 816/2 Atherton provenance, Australia) and the same *E. urophylla* male parent. Seeds obtained from these crosses were surface sterilized and germinated on solid agar containing half-strength MS medium (MURASHIGE and SKOOG 1962) under a 14 hour photoperiod. The mapping population consisted of 62 F1 individuals. This population was immortalized by establishing clonal cultures of the individuals by vegetative propagation *in vitro* on half-strength MS medium supplemented with 0.005 mg/l IBA (indol-butyric acid) to stimulate rooting of microcuttings.

DNA extraction. Total genomic DNA was isolated from freeze dried adult leaf tissue of the parents and from fresh leaves of *in vitro* plantlet progeny using the protocol of DOYLE and DOYLE (1987), modified by the addition of 1% PVP and 1% 2-mercaptoethanol to the isolation buffer. DNA concentration was estimated by gel electrophoresis comparing the fluorescence intensities of the ethidium bromide stained samples to those of λ DNA standards.

RAPD assay. Random ten-base primers (kits OP-A through OP-Z) were obtained from Operon Technologies Inc. (Alameda CA). Amplification reactions (13 µl) were carried out according to WILLIAMS *et al.* (1990) with the following modifications: 0.4 µM ten-base primer, 10 µg/µl Non-acetylated Bovine Serum Albumin (New England Biolabs), 5 to 10 ng of genomic DNA and 1 U of *Taq* DNA Polymerase. With a genome size of 0.6

pg/1C (GRATTAPAGUA and BRADSHAW 1994) the amount of genomic DNA used in a RAPD reaction corresponded to between 8,000 and 16,000 haploid genome equivalents. Amplifications were performed in 96-well microtest plates using an MJ Research PT-100 thermal controller. RAPD products were analyzed by electrophoresis in 1.5% or 2.0% agarose gels containing 0.2 µg/ml ethidium bromide. On a custom made gel tray a full 96-well plate was run on a single gel. Gels were photographed under transmitted U.V. light using a MP4 Polaroid camera or an Eagleye^{TM} video imaging system (Stratagene) and printed on 20x15 cm thermal paper. Gel scoring was performed directly from the photographs or thermal prints.

Primer screening. A total of 305 ten-base random primers were screened against the two parents and a progeny sample of 6 individuals. RAPD fragments that are polymorphic between the two parents and segregate in the progeny can be detected. With 6 individual progeny the probability of missing a polymorphic marker segregating 1:1 is 0.094. Twelve primers were conveniently screened on one 96-well plate and one gel. A total of 151 primers were selected during this step based on the number of RAPD polymorphisms amplified, their size and amplification intensity. Selected primers were used on the mapping population.

Scoring of RAPD markers. Segregation of RAPD markers in the mapping population was recorded in two independent replications, each one with a different set of individuals. In the first replicate the two parents and 30 progeny were assayed for RAPD markers with all the selected primers. In a second replicate, totally independent DNA extractions, reaction mixture preparations, gel analysis and genotype scoring were performed for another set of 32 progeny individuals. This procedure provided an internal control and an estimate of repeatability for all the scored markers. Markers that did not amplify consistently or could not be scored reliably across the two replicates were dropped from further analysis.

Segregating RAPD markers were identified by the manufacturer primer code corresponding to a particular ten-base sequence, followed by a number indicating the fragment size in base pairs. Fragment sizes were estimated using the software SEQAID II (RHOADS and ROUFA 1990). Following the fragment size, separated by a slash, a subjective score was given from 1 to 3 denoting the fragment amplification intensity, 3 being the most intense. For example RAPD marker A11_980/3 corresponds to a RAPD

fragment amplified by Operon primer A11 (corresponding to the sequence 5'-CAATCGCCGT-3'), with size 980 base pairs, of high (score 3) amplification intensity. All the scored RAPD fragments were sampled from the agarose gel by gently stabbing the fluorescing band with a pipette tip and rinsing the tip into 20 µl of sterile TE buffer (10 mM Tris-HCl and 0.2 mM EDTA). All samples were stored at -20°C until required for reamplification.

Characterization of genomic sequence complexity of RAPD marker loci. Total genomic DNA was extracted from a bulked leaf sample of individuals of the mapping population. In the wells of a dot-blot apparatus, an appropriate amount of DNA was denatured and vacuum transferred and cross linked to a nylon membrane. Membranes were cut into strip blots with 4 contiguous dots containing 5 µg, 0.5µg and 0.05 µg of eucalypt DNA and 5 µg of herring sperm DNA as a negative control.

Dot blots to be used as controls for comparative analysis were prepared by performing a reconstruction experiment where known picogram amounts of pure RAPD fragment corresponding to 1, 10, 100 and 1000 copies were immobilized onto a membrane using 5.0 µg of herring sperm DNA as a carrier following the same contiguous dot blot format. Control blots were prepared for twelve different RAPD fragments varying in size from 300 to 1500 base pairs. A total of 48 RAPD fragments corresponding to mapped marker loci were labeled and used as probes on the dot blots to characterize the copy number of their internal sequences. Control blots were simultaneously probed with the corresponding fragment used in the reconstruction experiment. To further confirm the results, two RAPD fragments from each copy number class were then used as probes on genomic Southern blots (SOUTHERN 1975).

RAPD fragment hybridizations. RAPD fragments to be used as hybridization probes were reamplified using the RAPD assay conditions described. Template DNA for reamplification consisted of a 3 µl volume of the 20 µl RAPD band sample. Non-radioactive labelling of the probes was performed with dUTP-digoxigenin according to manufacturer's recommendations (Boheringer-Mannheim). Reamplified probes were checked for single band purity on a minigel. Probe hybridization, washes and detection procedures were carried out following manufacturer recommendations using the chemiluminescent substrate solution CSPD (Tropix Inc.).

Confirmation of RAPD marker inheritance, codominance and presence in different individuals. Putative codominance (size-variation) of RAPD markers was investigated by DNA hybridization of gel blots of RAPD products with the putative codominant fragment used as probe. A subset of mapped RAPD markers were surveyed for their presence and segregation in a different individual of *E. grandis* by analyzing the second F1 cross described previously. RAPD assay was carried out on the two parents and 10 progeny for each one of the two crosses. Confirmation of homology for RAPD markers of same size was carried out through DNA hybridization, using the fragment of interest as a probe. RAPD gels were blotted onto nylon membranes and hybridized as described.

Segregation and linkage analysis of RAPD markers. Segregating markers were scored for presence (1) or absence (2) of the amplified RAPD band. The parental origin of the marker was also recorded. Two separate data sets were obtained, one for each parent. In the pseudo-testcross configuration markers are present in one parent and absent in the other or vice versa, and are expected to segregate 1:1 in the F1 generation. A $\chi 2$ test ($\alpha=0.05$) was performed to test the null hypothesis of 1:1 segregation on all scored markers. Preliminary grouping was done using a $\chi 2$ test for independence of segregation at a threshold of 15.00 (~ LOD= 5.0, R. DOERGE, personal communication). The linkage analysis was done using MAPMAKER (LANDER *et al.* 1987). The software program GMENDEL (LIU and KNAPP 1990) was also used during linkage analysis particularly for ordering linkage groups. To allow the detection of linkage of RAPD markers in repulsion phase the data set was duplicated and recoded. A LOD score of 5.0 and maximum $\theta= 0.25$ were set as linkage thresholds for grouping markers. Map distances in centimorgans were calculated using Kosambi's mapping function. Preliminary orders of marker loci in each linkage group were established using a matrix correlation method implemented by MAPMAKER. From this initial order, a subset of evenly spaced loci that could be ordered with a likelihood ratio support $\geq$ 1000:1 established a framework map. Error detection functions of MAPMAKER were employed to check potential genotyping errors in the framework markers. The final framework order obtained was then compared to the order outputted for the same subset of marker loci by GMENDEL that employs a simulated annealing algorithm. Markers that could not be ordered with equal confidence were indicated as accessory markers at an already specified locus on the map. Genome map sizes were estimated according

to HULBERT *et al.* 1987 taking into consideration only pairwise comparisions between markers placed on the framework map.

RESULTS

Primer screening. In our standard conditions, RAPD reactions amplified an average of 10.7 visible bands on an ethidium bromide stained agarose gel. Primer screening was efficiently carried out using both parents and a sample of F1 individuals (Figure 1). With this format, parental origin of the polymorphic loci as well as their allelic state (homozygous or heterozygous) was directly inferred from the presence of the fragment in one parent, the absence in the other and at least one presence/absence in the F1 progeny sample. Of the 305 arbitrary primers screened, 57 (18.7%) did not yield any amplified product, 50 (16.4%) did not detect any visible sequence polymorphism in our particular cross and 198 (64.9%) uncovered at least one RAPD fragment polymorphism. From these 198 primers, a total of 151 were selected, aiming at maximizing the number of scorable markers per primer in the following mapping step. A total of 558 RAPD markers were scored on the mapping population, yielding an average of 3.69 markers/selected primer or 1.82 markers/any arbitrary primer . Similar ratios were recently reported for single-dose RAPD polymorphisms in sugar cane (1.88) (SOBRAL and HONEYCUTT 1993) and *Stylosanthes* (1.8) (KAZAN, MANNERS and CAMERON 1993). Although the majority (64.9%) of primers screened detected at least one polymorphism, the screening step essentially doubled the time efficiency and halved the cost of data gathering in the mapping phase.

Scoring of markers on the mapping population. Segregation of a total of 558 RAPD markers was scored on the mapping population (Figure 2). The numbers of markers inherited from each parent were very similar: 272 from *E. grandis* and 286 from *E. urophylla.* RAPD fragments sizes ranged from 3335 bp to 207 bp, with an average of 979 ± 570 bp for *E. grandis* and 910 ± 521 bp for *E. urophylla.* The replicated design used (see Material and Methods) throughout the mapping phase yielded a repeatability estimate of 92.4%. From the 558 RAPD markers, 516 were fully repeatable in both replicates while 42 were not, 20 from *E. grandis* and 22 from *E. urophylla.* These markers either could be scored in the first set of progeny and not in the second or vice-versa. With no exception, all of them were originally classified as class 1 markers denoting low

amplification intensity and/or difficulty in scoring due to co-migrating fragments. Such fragments were not considered in further analysis.

Linkage analysis. Segregation ratios that departed from the Mendelian expectation of 1:1 at α= 0.05 were detected at 10 marker loci in *E. grandis* and 10 marker loci in *E. urophylla.* No departure was detected at α= 0.01. These apparently distorted markers (denoted by an asterisk following the marker identification) are clustered on only 2 linkage groups in *E. grandis* (groups 6 and 7) but are scattered in 6 linkage groups in *E. urophylla* (Figures 3 and 4). Note that at α= 0.05, considering a total of 500 marker loci, around 25 of these are expected to display this behavior due to chance alone. Therefore, at this point we have no reason to suggest that such distortions have a biological basis. Such an indication may be drawn from the examination of a larger sample of meiosis. Furthemore, the parents of the mapped trees would be necessary to study the specific origin of the observed distortion, i.e. excess of the allele of one grandparent versus deficiency of the allele from the other. Only markers that passed the single-locus segregation test were initially used in the linkage grouping analysis. Distorted markers were later placed on the map by determining their most probable location in an already established gene order.

Linkage relationships of the segregating markers were established using both a $\chi 2$ test for independence of segregation at a threshold value of 15.00 and by two-point mapping (LOD 5.0 and max. θ= 0.25). Both linkage analyses agreed very closely. Overall, linkages were robust at a LOD score range from 4.0 to 6.0. In view of the large number of markers, at lower LOD scores, especially below 4.0, occasional spurious linkages resulted in the agglomeration of some linkage groups, while increasing LOD scores beyond 6.00 would result in fragmentation of linkage groups. At LOD 5.0 the maternal *E. grandis* map has a total of 240 markers into 14 linkage groups and the paternal *E. urophylla* map 251 markers in 11 linkage groups (Figures 3 and 4). Twelve markers for *E. grandis* and 13 for *E. urophylla* remained unlinked at LOD 5.0. Although they were linked at a lower LOD (3.0), and higher θ (0.35), they did not contribute any significant additional information in terms of genome coverage and therefore they were left out of the final map versions. The proportion of unlinked markers found (4.5%) is smaller than those reported for other single-dose marker linkage mapping studies (e.g. 10% in potato (GEBHARDT *et al.* 1988); 15.4% and 12.9% in sugar cane (AL-JANABI *et al.* 1993; DA SILVA *et al.* 1993)).

Locus ordering and map construction. Both *Eucalyptus* species have n=11 chromosomes, and therefore 11 linkage groups were expected in each map. This expectation was met for the *E. urophylla* map but not for the *E. grandis* map. However, for *E. grandis*, lowering the threshold LOD score to 4.0 and increasing θ to 0.35 would result in the merging of 3 pairs of linkage groups leading therefore to a correspondence between number of linkage groups and number of chromosomes. The following mergers with the respective highest LOD score between markers on separate groups would take place: groups 8 and 12 (LOD score 4.3 between markers N6_634/1 and K9_534/2); groups 11 and 13 (LOD score 4.4 between markers B6_759/1 and X12_530/2); groups 1 and 6 (LOD score 4.05 between V7_450/2 and L16_389/2) (Figure 3). In spite of the possibility of merging some linkage groups to attain the expected number of chromosomes, it was found more appropriate to assemble both maps with the same statistical stringency allowing for more meaningful comparisions between maps. An excess of linkage groups in relation to the haploid chromosome number has been reported for other species (e.g. bean (NODARI *et al.* 1993); lettuce (KESSELI, PARAN and MICHELMORE 1990)).

The linkage groups were constructed using markers in both linkage phases. Markers on one linkage phase are indicated with a "+" sign following the marker identification code, while markers on the alternative phase are indicated with a "-" sign (Figures 3 and 4). Matrix correlation was used to get a preliminary locus ordering. Based on this first approximate order, a subset of candidate framework marker loci was selected spanning the whole linkage group at distances varying between 5 and 20 cM. This selection was based on a sequence of criteria, that by order of priority were as follows: i) fragment intensity of amplification score; ii) ease of marker scoring in view of co-migration of other fragments or smearing that could lead to errors in genotyping (gel photo was reviewed); iii) number of missing data; iv) size of amplified fragment (below 2000 and above 300 basepairs). Candidate framework markers selected were again ordered using matrix correlation. This linear order was then tested by permutating all possible sets of 3 adjacent markers. A final framework order was accepted when the log-likelihood difference between the initial order and all the alternative local permutations was at least -3.0. Therefore, the framework map orders presented are approximately 1000 times more likely than the next best orders. When local differences of less than -3.0 were observed additional markers were excluded from the framework at the regions of ambiguous ordering, and the analyses performed again. Generally

after one or two such iterations, a final order was attained. This final framework order was then compared to the order obtained by simulated annealing implemented by GMENDEL. With the exception of a few two-marker order permutations the framework orders obtained were the same. Segments of the framework where order ambiguities still persisted, were analyzed using error detection functions of MAPMAKER 3.0 that display potential genotyping errors. Gel photos were re-checked for potential scoring errors. In the few instances where the data point was dubious, it was treated as missing data and the ordering analysis performed again.

Markers that could not be placed on the map with a 1000:1 odds were designated as accessory markers and were positioned on the map in relation to the closest framework marker. Their most likely position was obtained by looking for the framework locus that displayed the highest LOD score and lowest two-point θ, or alternatively by looking for the interval with log-likelihood closest to zero. Accessory markers are listed on the right of linkage groups (Figures 3 and 4).

Clustering of markers seems to occur throughout both linkage maps, particularly in *E. urophylla* that displays large clusters on groups 2, 5 and 6. However no formal test for clustering was carried out. Clustering is a common occurrence and has been reported in essentially all relatively dense linkage maps constructed to date irrespective of the organism or technique used to assay DNA polymorphisms: RFLP in tomato (TANKSLEY *et al.* 1992) and common bean (VALLEJOS, SAKIYAMA and CHASE 1992); RAPD in *Arabidopsis* (REITER *et al.* 1992) or microsatellites in humans (WEISSENBACH *et al.* 1992). Clustering of markers could be an artifact resulting from the limited resolution of our mapping population. For example, while the *Arabidopsis* map by REITER *et al.* (1992) displayed clustering, the integrated map based on larger populations reported by HAUGE *et al.* (1993) did not. Alternatively, clustering might have a biological basis reflecting supressed genetic recombination in heterochromatin around the centromeres and/or in telomeric regions as discussed by TANKSLEY *et al.* (1992) following studies on the correlation of genetic and physical structure in the tomato genome (SEGAL *et al.* 1992). As *Arabidopsis* has a comparatively much smaller proportion of repetitive DNA than *Eucalyptus*, tomato and humans, less clustering may be expected *a priori* if clustering is due to restricted recombination in regions rich in repetitive DNA.

The great majority of the accessory markers are within 5 cM of the nearest framework marker. It is likely that their ordering ambiguity results largely from the relatively small recombinational distance estimated from a limited number of meioses

analyzed. With 62 meioses, the standard error on a recombination fraction of 0.05 is approximately 0.03. However we also observed that 25% of the accessory markers (22 in 88 for *E. grandis*) and 18.5% (25 of 135 in *E. urophylla*) were at distances greater than 6 cM (θ ~ 0.05) and could still not be placed in the framework. The ambiguity in the placement of these markers might be the result of missing data or to errors in genotyping. The overall fraction of missing data points including both data sets was 8.3%. For the framework markers only, this fraction was 4.5%. In our experimental conditions, a genotyping error rate ≤3% was estimated, varying with the RAPD fragment amplification intensity.

Linkage maps statistics and estimates of genome size and coverage.
Approximately 59% of the markers for *E. grandis* could be placed on the framework defining a total of 142 loci or loci clusters and 1551 cM of total map distance. For *E. urophylla* 47% of the markers could be placed on a framework of 119 loci or loci clusters covering 1101 cM (Figures 3 and 4). Linkage groups were numbered sequentially from the longest to the shortest. For *E. grandis* the average size of linkage groups was 110 ± 35 cM and the range from 41.6 cM to 156.9 cM. For *E. urophylla* the average size was 99 ± 32 cM and the range from 46.7 to 141 cM. The total number of markers per linkage group (framework and accessory) varied from 6 (group 14) to 30 (group 5) for *E. grandis*, and from 6 (group 11) to 39 (group 6) for *E. urophylla*. The average distance between two framework markers was 12.2 ± 6.3 cM for *E. grandis* and 10.2 ± 6.6 cM for *E. urophylla*. Both maps have a density of 27 cM, which corresponds approximately to a recombination fraction of 0.25 i.e. no interval between two markers is greater than 27 cM.

Total genome size was estimated for both parents using the method of HULBERT *et al.* (1987), as described by VALLEJOS, SAKIYAMA and CHASE (1992). Only framework markers were used in this procedure to avoid an overestimate of genome coverage. For *E. grandis*, the maternal parent, a total map distance of 1620 cM was estimated, of which 1552, i.e. 95.8% were covered by the framework map. For *E. urophylla*, the paternal parent, the total map distance estimated was 1156 cM, of which 1101 cM, i.e. 95.2% were covered. A reasonably equivalent genome coverage in both species and sexes was therefore achieved with the pseudo-testcross mapping strategy. Given the estimated total map distances and genome sizes of 641 and 646 Mbp/1C (GRATTAPAGLIA

and BRADSHAW 1994), the average physical equivalent of 1 cM would correspond to 395 and 559 kilobase pairs, respectively for *E. grandis* and *E. urophylla*.

Confirmation of inheritance and segregation of RAPD markers. The inheritance of segregating RAPD markers from both parents following the pseudo-testcross configuration was confirmed by DNA hybridization experiments (Figures 5 and 6). A case of allelism between RAPD fragments in the two parents was tested and confirmed using DNA hybridization. In *E. grandis*, the RAPD marker G14_927/3 is present in one allelic form with a fragment size of 927 bp. In *E. urophylla*, the same marker is present in two allelic forms: a 917 bp and a 960 bp fragment. In the F1 a 1:1 segregation is observed for the heterozygote 917/927 (lanes 4,5 and 9) bp versus the heterozygote 960/927 bp (lanes 6-8 and 10-13). The hypothesized allelism between the three bands in the two parents, was confirmed by probing the RAPD gel blot with the 927 bp RAPD fragment from *E. grandis*, and detecting signal in all three bands (Figure 5, panel A). The same principle was used to confirm the homology of markers that segregated 3:1. Both parents showed bands of equivalent size, and the progeny showed a segregation ratio that fit a 3:1. DNA hybridization of the RAPD gel blot with a fragment from one of the parents confirmed the hypothesis (Figure 5, panel B). A total of 11 markers were found in this configuration, i.e. present in both parents in a heterozygous state therefore segregating 3:1 in the progeny. In principle, such markers could be helpful to define homologies between linkage groups in the two maps. We attempted to do this using GMENDEL that analyzes mixtures of segregation ratios (1:1 and 3:1). Although a few linkages were found at relaxed thresholds, no reliable map position could be established with our sample size. We estimated that the mean amount of information (ALLARD 1956) supplied by a single individual for such mixture of mating configurations is only 1/4 (at θ= 0.05) and 1/3 (θ= 0.25) of the information in the backcross. Therefore to achieve adequate power to estimate linkage in this case, a larger sample size is necessary. Fully informative codominant markers such as microsatellites, RFLP's or isozymes would be highly desirable for this task.

Codominant (size-variant) RAPD markers. Although rare, codominant (size-variant) RAPD markers were found on both maps. Codominant RAPD markers can result from small insertions or deletions between priming sites (WILLIAMS, RAFALSKI and TINGEY 1992). Codominant RAPD's were initially hypothesized from the following observation: (1)

both allelic fragments are present in the same parent and are amplified with the same primer; (2) F1 individuals receive either one or the other allele, i.e. the two RAPD fragments are in repulsion and (3) no recombinant genotypes are observed in the F1, i.e. no individuals with both fragments or null for both fragments. In *E. grandis* four marker pairs fit these observations: A10_635/562 (group 5); A11_980/920 (group 7); Y17_525/515 (group 8); Y15_760/740 (group 11). In *E. urophylla* also four pairs were found : U7 1100/850 (group 3); Y20_400/390 (group 5) U13_350/320 (group 5); Z11_550/480 (group 6). DNA hybridization experiments confirmed the codominance of such sets of markers (Figure 6, panel C). Several pairs of RAPD markers were observed that satisfied all the observations outlined above except that the two fragments were amplified with different primers (e.g. in *E. grandis* the pair K9_884/3 and K19_448/3 on group 4). For this category of tightly linked markers in repulsion no DNA hybridization experiments were carried out. Functionally, as pointed out originally by WILLIAMS *et al.* (1990) such pairs of markers could also be used as a single codominant marker. Overall, however, the frequency of truly codominant and functionally codominant RAPD markers for the *Eucalyptus* species surveyed remains below 3%.

Survey of the presence and allelic state of RAPD markers in different individuals.
A subset of markers from the *E. grandis* clone 44 map were surveyed for their presence and allelic state in a second individual tree of the same species (*E. grandis* 816/2) by analyzing their segregation in a second F1 progeny set involving 816/2 as a parent. From a total of 112 RAPD markers surveyed, 37 (33%) were found to amplify in the second tree. Of these, 20 were also in a heterozygous state and segregated 1:1 in the F1, while 17 were homozygous, i.e. did not segregate. DNA hybridization experiments were carried out for 5 markers that were not shared (Figure 6, panel A), and 10 that were shared confirming the homology (Figure 6, panels B and C). Only one marker of the ones tested was found to be misinterpreted, i.e. the RAPD bands were scored as being the same but in fact were not homologous. This was a relatively large (1500 bp) fragment. Misinterpretations of this kind are more likely for larger fragments that are not as efficiently size fractionated as smaller fragments. In carrying out this kind of RAPD marker survey we found imperative to run the gels for at least 15 cm, to minimize errors resulting from co-migrating fragments.

For *E. urophylla*, after surveying 34 randomly chosen RAPD markers, it was found that all of them were present in what was thought to be a different individual of

the same species. Moreover, all the markers surveyed also segregated in the second F1 progeny. The possibility that the same *E. urophylla* clone 28 had actually been used as the male parent in the second cross was tested and confirmed. A subset of five markers that were not recombinationally separated in a locus cluster on group 5 (defined by marker U13_350/2) and a second subset of four markers in a locus cluster on group 8 (defined by marker M4_477/2) were surveyed for linkage on a set of 16 progenies. All the markers were found to be present and no recombinants were found. Further DNA hybridization experiments also confirmed that in fact the same *E. urophylla* (clone 28) had been used as the male parent in both crosses (data not shown). This result did not allow us to explore the extent of conservation of RAPD markers in different individuals of *E. urophylla*, however it was useful to confirm the stable behavior of RAPD markers in terms of segregation and linkage relationships in a second cross involving the same individual tree.

Characterization of genomic sequence complexity of RAPD marker loci. Over 50% of the 48 RAPD fragments surveyed were found to amplify from low copy genomic regions (1 to 10 copies) and less than 10% originated from very highly repeated regions ($\leq$ 1000 copies). Approximately equal frequencies (~20%) were found for fragments amplified from moderately repeated (10 to 100) and highly repeated regions (100 to 1000) (Figure 7, panels A and B). Similar estimates of genomic sequence complexity of RAPD marker loci was observed in soybean (WILLIAMS *et al.* 1990) and *Arabidopsis* (REITER *et al.* 1992). Based on 48 data points, a simple correlation analysis was carried out between the following variables: RAPD fragment size in basepairs, amplification intensity score and copy number class. The results were as follows: fragment size x intensity score r = 0.25; fragment size x copy number r = 0.04; intensity score x copy number r = 0.18. In conclusion, no significant correlation (α= 0.05) was found for any of the three pairwise analyses, suggesting no particular dependency of the fragment size or amplification efficiency of RAPD marker loci on the complexity of the genomic region sampled.

It is important to point out that the sample of RAPD fragments surveyed in this experiment are not randomly chosen RAPD fragments, rather they correspond to a subset of fragments that behave as genetic markers, map to a single location and therefore originate from a unique site in the genome. WOLFF, McDOWELL and MACHADO (1993) found that 48% of random *Pst*I genomic clones of *E. grandis* and *E. urophylla* were useful as RFLP probes. In our study we found that 53% of the mapped RAPD

fragments amplified from low copy regions and could potentially be used as RFLP probes. These results indicate that the genomic library of mapped RAPD fragments, obtained in *Eucalyptus* as a byproduct of this mapping experiment, closely resembles a genomic library of RFLP probes constructed by the traditional approach.

DISCUSSION

Pseudo-testcross mapping strategy using RAPD markers. We have used a "pseudo-testcross" mapping strategy in combination with the RAPD assay to construct the first reported linkage maps for species of *Eucalyptus*. In a cross between heterozygous parents, many single-dose polymorphic markers will be heterozygous in one parent, null in the other and therefore segregate 1:1 in their progeny as in a testcross. We use the name "pseudo-testcross" for this strategy because the testcross mating configuration of the markers is not known *a priori* as in a conventional testcross where the tester is homozygous recessive for the locus of interest. Rather, the configuration is inferred *a posteriori* after analyzing the parental origin and genetic segregation of the marker in the progeny of a cross between highly heterozygous parents with no prior genetic information. When this inference is done for both parents involved in the cross, the term "two-way pseudo-testcross" is more appropriately used.

RITTER, GEBHARDT and SALAMINI (1990) described the theoretical background for linkage analysis of markers segregating in crosses between heterozygous parents. As mentioned in that work, map construction in allogamous plant species for which only heterozygous individuals are available can make use of single-dose polymorphic markers behaving as dominant markers in an F1, segregating 1:1 for the presence or absence of the fragment. These markers were used for genetic mapping in potato (BONIERBALE, PLAISTED and TANKSLEY 1988; GEBHARDT *et al.* 1989), and recently allowed genetic mapping in polyploid sugar cane (WU *et al.* 1992; DA SILVA *et al.* 1993; SOBRAL and HONEYCUTT 1993; AL-JANABI *et al.* 1993). We and others observed this same mating configuration when analyzing genetic segregation of RAPD markers in F1 crosses of forest and fruit trees, and suggested its wide applicability for genetic mapping in this group of highly heterozygous largely undomesticated species (CARLSON *et al.* 1991; GRATTAPAGLIA, O'MALLEY and SEDEROFF 1992; ROY *et al.* 1992; LAWSON, HEMMAT and WEEDEN 1992).

The pseudo-testcross mapping strategy is conceptually simple to implement and can be applied with any type of molecular marker. However its potential can be better explored with the efficiency of the RAPD assay in pre-screening marker polymophisms in search of the informative test cross configurations. The fact that the RAPD assay is sensitive to single base changes, contributes to a higher efficiency in scanning the genomes for polymorphisms. Moreover, the fact that RAPD detects only one allele at a locus facilitates the occurrence of pseudo tescross configurations, because the necessary null genotype of one of the parents actually corresponds to undetected alleles. In addition to that, the RAPD assay is technically simple and fast to perform facilitating the initial screening step. Following our screening procedure in *Eucalyptus*, 36 arbitrary primers could be easily screened in a single working day, yielding an estimated 1.82 markers/primer, i.e. 65 markers from both parents taken together. Finally, the segregation ratio observed for a dominant RAPD marker in this configuration has the same information content as that of a codominant marker. Evidently, a highly polymorphic, multiallelic marker that detected all four allelic variants of the mating configuration, (e.g sequence tagged microsatellite site), would contain more genetic information (MORGANTE and OLIVIERI 1993).

The use of RAPD markers in a pseudo-testcross configuration is a general strategy for the construction of genetic linkage maps in outbred forest trees as well as in any highly heterozygous living organisms. It can be immediately applied to any species without any prior genetic information. The only requirements are sexual reproduction between two individuals that results in the generation of a progeny large enough to allow the estimation of recombination frequencies between segregating markers. Its efficiency will be directly proportional to the level of genetic heterozygosity of the species under study, which is a function of the mating system, and the genetic divergence between the individuals crossed. In our study we employed an interspecific cross between highly heterozygous individuals from two closely related outcrossing species, thus increasing the probability of finding pseudo-testcross marker configurations. We only found 11 markers heterozygous in both parents thus segregating 3:1 compared to 558 markers in a test cross configuration, segregating 1:1.

The pseudo-testcross strategy should also be efficient at the intraspecific level and increasingly so with crosses of genetically divergent individuals from geographically distinct origins. In a survey of 112 mapped markers, we found that only 33% were shared between two individuals of the same species and different provenances. We suggest

that at the intraspecific level, the mapping efficiency of the pseudo-testcross strategy, measured by the number of informative markers/arbitrary primer should reach between 60 to 70% of the one reported in this study, that is ~2.4 instead of 3.69 markers/selected primer. With individuals from the same population, this number will tend to be lower, as more markers will be shared. In a group of 38 heterozygous clones of *Solanum tuberosum* the informativeness of RFLP probes for direct segregation analysis in F1 populations varied from 49% to 95% indicating that linkage mapping using F1 progeny should be feasible for most combinations (GEBHARDT *et al.* 1989). Test cross RAPD marker configurations were often observed at the intraspecific level in other highly heterozygous forest tree species, however no estimates of frequencies per arbitrary primer were given (CARLSON *et al.* 1991; ROY *et al.* 1992).

The pseudo-testcross strategy basically extends the haploid mapping approach used for conifers, to any other angiosperm tree species. The final result is essentially the same, i.e. linkage maps for individual trees, however it requires performing a controlled cross. On the other hand it is more time and cost efficient since gametic segregation from two individuals or twice the heterozygosity is surveyed simultaneously in the same PCR reaction, both in the primer screening and in the mapping phase. Therefore, even in conifers, the pseudo-testcross could potentially be the mapping strategy of choice for quickly generating single-tree linkage maps.

Genetic linkage maps of single individuals. The genetic linkage maps constructed in this study (Figures 3 and 4) are individual-specific. The pseudo-testcross strategy is specifically based on the selection of single-dose markers present in one parent and absent in the other. In our maps, no RAPD markers are in common and so it is not possible to determine homologies of linkage groups in the two maps or integrate the two maps into one. Overlap of RAPD marker occurrence and linkage relationships in genetic maps of different individuals will depend on the presence of the same RAPD marker loci and their allelic state. While at the interspecific level, the overlap will be very low, at the intraspecific level, it will be increasingly high as individuals from the same population are used. In this study we found that 33% of the mapped markers in *E. grandis* were present in a second individual of the same species but from a very distinct origin, and 54% of those were also in a heterozygous state. Indirect evidence for the occurrence of the same RAPD markers across different individuals of the same population come also from studies that employed RAPD markers to estimate outcrossing

rates in stands of *E. urophylla* (GRATTAPAGLIA, O'MALLEY and SEDEROFF 1992) and *Datisca glomerata* (FRITSCH and RIESBERG 1992).

To integrate linkage maps constructed by the pseudo-testcross strategy, multiallelic codominant markers with alleles segregating from both parents would be most efficient, providing a set of common loci which could be used as locus bridges. In this study, 11 dominant RAPD markers were shared by the parents and could in principle be used to assign linkage groups homologies between the two maps. However, due to the low information content of the mixed mating configuration, a larger sample size would be necessary to achieve adequate power to accomplish such assignment. More than 50% of the RAPD fragments mapped in this study are low-copy, and could potentially be used to detect codominant RFLP's. Although such markers could be helpful in connecting linkage groups, a complete map merging would still be fairly difficult to achieve, since correct locus ordering among the markers not in common between homologous linkage groups would not be known (B. H. LIU, personal communication). Statistical integration of mapping data from different populations of *Arabidopsis* was shown to be problematic, especially in regions containing low densities of common markers between maps (HAUGE *et al.* 1993). As mentioned earlier, multiallelic markers such as microsatellites would be very powerful and desirable for this purpose. In the context of the breeding applications envisaged for these linkage maps in forest tree breeding, map merging is not immediately necessary , but it will become important in subsequent generations of breeding and selection (see discussion below).

No morphological traits or other single gene traits that could potentially be placed on these maps, segregated in the population used for map construction. To our knowledge no simply inherited traits are presently known in *Eucalyptus* that could be placed on any genetic map, with the exception of isozyme loci. The same screening procedure with parents and a subset of progeny could be used for mapping isozymes loci. Informative configurations of isozymes genotypes in the parents would result in either 1:1 or 1:2:1 segregation ratios in the F1, and mapping of the isozyme locus could be achieved in only one or both maps respectively. The existing RAPD linkage maps provide a scaffold where even distant linkage relationships of isozymes could be determined. On the other hand, isozyme loci would provide anchor loci for single-tree map comparisons and merging.

We obtained equivalent genome coverages on the two maps in spite of different estimates of total map distances. This suggests that the difference in total map

distance observed between the two species and sexes are biologically significant. However no distinction is possible at this point between a species specific, sex specific or individual specific difference in genetic recombination. If applied within species, however, the pseudo-testcross strategy should provide a valuable tool to study specific differences in general recombination rate. Genome sizes estimated in our study are well within the range of several other species (listed by NODARI *et al.* 1993). However, genome coverages found in our study are slightly higher than those found in other maps. A good comparative example in this respect is common bean, that has genome characteristics similar to eucalypts, (n= 11 chromosomes; genome size around 600 Mbp). Approximately 80% of the genome could be covered with a framework map of 145 RFLP loci (VALLEJOS, SAKIYAMA and CHASE 1992). Besides intrinsic biological differences in levels of DNA polymorphism and rates of recombination, one of the possible reasons for the observed difference in genome coverage could be the result of a more efficient genome sampling by RAPD markers as compared to the RFLP technique, particularly for genomic regions rich in repetitive DNA.

Framework maps at a likelihood ratio support ≥1000:1 were constructed (Figures 3 and 4). This presentation of the data is convenient for selecting a subset of evenly spaced framework markers to initially scan the genome for QTL mapping. Recent simulation studies have shown that wide marker spacings of 20 or even 50 cM are optimal for this task (DARVASI *et al.* 1993). A more focused search for the exact position of QTL's can then be done with the available nearby accessory markers. If more markers are still needed in a region of interest, genetic walking based on genotype pooling techniques could be used (MICHELMORE, PARAN and KESSELI 1991; GIOVANNONI *et al.* 1991; REITER *et al.* 1992).

Toward marker assisted breeding strategies in forest trees. It has been long recognized that one of the problems facing marker assisted breeding in outbred species such as forest trees is the linkage equilibrium between marker loci and genetic loci of interest (SOLLER 1978; BECKMANN and SOLLER 1983; NEALE and WILLIAMS 1991; LANDE and THOMPSON 1990; STRAUSS, LANDE and NAMKOONG 1992). With linkage equilibrium, marker-trait associations established in one cross, would not hold in a second pedigree, since marker and QTL alleles would be randomly associated. As pointed out earlier, one solution would be to construct maps for each genotype in the breeding population (NEALE and WILLIAMS 1991; GRATTAPAGLIA *et al.* 1992). This was considered a task that

depended on significant advances in the ability to obtain marker data. The RAPD technology provided this advance by allowing the construction of a 200 marker linkage map in *Pinus taeda* in 6 person per month (GRATTAPAGLIA *et al.* 1991). In this study, with the pseudo-testcross strategy and RAPD markers we constructed two linkage maps simultaneously in approximately 5 person per month.

The ability to construct genetic linkage maps quickly in any forest tree opens the way to the heterodox proposal of constructing maps for individual trees in a breeding population. The paradigm of an index linkage map for a species is an attractive one for comparative mapping applications. However it does not seem adequate as the initial approach for establishing marker/traits associations for breeding in allogamous populations with a wide genetic base such as those of forest trees. Rather, the progressive accumulation of individual linkage maps with subsets of common markers among them will make obvious the relationships of linkage groups in different maps. This will eventually lead to a unified map where general regions associated with trait expression could be identified. Multiallelic co dominant markers such as microsatellites would then be highly desirable specifically bracketing such regions to facilitate their manipulation in breeding. Although such "population level" or general quantitative trait loci (QTL) should exist, their relative importance in the overall level of genetic variation in quantitative traits in forest trees is still unclear. The identification and manipulation of QTL's specific to individual trees might emerge as being more important for the advancement of quantitative traits by marker assisted breeding.

Based on the proposal of individual-specific linkage maps, the integration of mapping information into tree breeding programs would involve four basic steps, briefly: (1) construction of moderate density individual tree maps for elite genotypes, in a two-by-two fashion using the "two-way pseudo-testcross" strategy; (2) localization of favorable alleles at qualitative and quantitative trait loci of interest on these maps, by analyzing the performance of an extended set of the full-sib family used for map construction. This extended set of progeny would be initially genotyped only for a subset of evenly spaced framework markers, followed by a finer search with accessory markers in regions of interest. Alternatively half-sib families of the mapped individuals could also be used for this purpose. In this step, retrospective QTL analysis using existing full and half-sib families at harvest age would be highly preferrable in order to gather the necessary quantitative data in acceptable time. Power and precision in QTL mapping would be greatly enhanced by using large family block plantations (> 1000

individuals) and clonal replication of genotypes when possible; (3) validation of marker-trait associations by replication and prediction experiments; (4) marker assisted selection of progeny, or retrospective selection of parents for planned recombination in subsequent generations of breeding. Given that close linkages are established in the mapping phase, the decay of marker-trait associations with time, would not be of immediate concern in the context of the long generations of tree breeding.

ACKNOWLEDGMENTS

LITERATURE CITED

Al-Janabi, S. M., R. J. Honeycutt, M. McClelland and B. W. S. Sobral, 1993 A genetic linkage map of *Saccharum spontaneum* L. 'SES 208'. Genetics **134**: 1249-1260.

Allard, R., 1956 Formulas and tables to facilitate the calculation of recombination values in heredity. Hilgardia **24**: 235-278.

Beckmann, J. S. and M. Soller, 1983 Restriction fragment length polymorphisms in genetic improvement methodologies, mapping and costs. Theor. Appl. Genet. **67**: 35-43.

Boland, D. J., M. I. H. Brooker, G. M. Chippendale, N. Hall, B. P. M. Hyland, R. D. Johnston, D. A. Kleinig and J. D. Turner, 1984 *Forest trees of Australia*. 4th ed. Nelson-CSIRO, Melbourne.

Bonierbale, M. W., R. L. Plaisted and S. D. Tanksley, 1988 RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics **120**: 1095-1103.

Bradshaw, H. D. and R. F. Stettler, 1993. Molecular genetics of growth and development in *Populus*. I. Triploidy in hybrid poplars. Theor. Appl. Genet. **86:** 301-307.

Caetano-Anólles, G., B. J. Bassam and P. M. Gresshoff, 1991 High resolution DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. Bio/Technology **9:** 553-557.

Carlson, J. E., L. K. Tulsieram, J. C. Galubitz, V. W. K. Luk, C. Kauffeldt and R. Rutledge, 1991 Segregation of random amplified DNA markers in F1 progeny of conifers. Theor. Appl. Genet. **83:** 194-200.

Chaparro, J. X., D. J. Werner, D. M. O'Malley, and R. R. Sederoff, 1994 Targeted mapping and linkage analysis of morphological, isozyme and RAPD markers in peach. Theor. Appl. Genet. (in press).

Conkle, M. T. 1981 Isozyme variation and linkage in six conifer species. In: Proceedings of the Symposium on Isozymes in North American Forest Trees and Forest Insects. Pacific Southwest Exp. St. USDA Genet. tech. Rep. PSW-48. pp. 11-17.

Darvasi, A., A. Weinreb, V. Minke, J. I. Weller and M. Soller, 1993 Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics **134:** 943-951.

DaSilva, J. A. G., M. E. Sorrels, W. L. Bumquist and S. D. Tanksley, 1993 RFLP linkage map and genome analysis of *Saccharum spontaneum*. Genome **36:** 782-791.

Devey, M. E., K. D. Jermstad, C. G. Tauer and D. B. Neale, 1991 Inheritance of RFLP loci in a loblolly pine three-generation pedigree. Theor. Appl. Genet. **83:** 238-242.

Doyle, J. J. and J. L. Doyle, 1987 Isolation of plant DNA from fresh tissue. Focus **12:** 13-15.

Eldridge, K., J. Davidson, C. Harwood and G. Van Wyk, 1993 *Eucalypt domestication and breeding*. Oxford Science Publications. Oxford University Press, Oxford.

Fripp, Y. J., A. R. Griffin and G. F. Moran, 1987 Variation in allele frequencies in the outcross pollen pool of *Eucalyptus regnans* F Muell. throughout a flowering season. Heredity **59:** 161-171.

Fritsch, P. and L. H. Rieseberg, 1992 High outcrossing rates maintain male and hermaphrodite individuals in populations of the flowering plant *Datisca glomerata* Nature **359:** 633-636.

Gebhart, C., E. Ritter, T. Debener, U. Schachtschabel, B. Walkemeier, H. Uhrig and F. Salamini, 1989 RFLP analysis and linkage mapping in *Solanum tuberosum*. Theor. Appl. Genet. **78:** 65-75.

Giovannoni, J. J., R. A. Wing, M. W. Ganal and S. D. Tanksley, 1991 Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. Nucl. Acid. Res. **19:** 6553-6558.

Grattapaglia, D. , 1994 Applications of molecular markers in plant breeding. In *Genetic mapping of quantitatively inherited and economically important traits in Eucalyptus.* Ph.D. Dissertation, North Carolina State University, Raleigh.

Grattapaglia, D. and H. D. Bradshaw Jr. , 1993 Nuclear DNA content of commercially important *Eucalyptus* species and hybrids. Can. J. For. Res. (in press).

Grattapaglia, D., D. M. O'Malley and R. R. Sederoff, 1992 Multiple applications of RAPD markers to genetic analysis of *Eucalyptus* sp. pp. 436-450 in *Proceedings of IUFRO International Conference "Breeding tropical trees"* Section 2.02-08 Cali, Colombia.

Grattapaglia, D. and R. R. Sederoff, 1992 Pseudo-testcross mapping strategy in forest trees: single tree RAPD maps of *Eucalyptus grandis* and *E. urophylla.* Plant Genome I, San Diego, CA poster # 47.

Grattapaglia, D., J. X. Chaparro, P. Wilcox, S. McCord, D. Werner, H. Amerson, S. McKeand, F. Bridgwater, R. Whetten, D. O'Malley and R. R. Sederoff, 1992 Mapping in woody plants with RAPD markers: applications to breeding in forestry and horticulture pp. 37-40 In: *Proceedings of the Symposium on Applications of RAPD Technology to Plant Breeding.* CSSA/ASHS/AGA.

Grattapaglia, D., P. Wilcox, J. X. Chaparro, D. O'Malley, S. McCord, R. Whetten, L. McIntyre and R. R. Sederoff, 1991 A RAPD map of loblolly pine in 60 days. Third International Congress of Intl. Soc. for Plant Mol. Biol. abs. 2224.

Hauge, B. M., S. M. Hanley, S. Cartinhour, J. M. Cherry, H. M. Goodman, M. Koornneef, P. Stam, C. Chang, S. Kempin, L. Medrano and E. M. Meyerowitz, 1993 An integrated genetic/RFLP map of the *Arabidopsis thaliana* genome. The Plant Journal **3**: 745-754.

Hulbert, S. H., T. W. Ilot, E. J. Legg, S. E. Lincoln, E. S. Lander and R. W. Michelmore, 1987 Genetic analysis of the fungus, *Bremia lactucae,* using restriction fragment length polymorphisms. Genetics **120**: 947-858.

Kazan, K., J. M. Manners and D. F. Cameron, 1993 Inheritance of random amplified polymorphic DNA markers in an interspecific cross in the genus *Stylosanthes.* Genome **36**: 50-56.

Kesseli, R. V., I. Paran and R. W. Michelmore, 1990 Genetic linkage map of lettuce (*Lactuca sativa,* 2n=18), pp. 6100-6102 in *Genetic Maps,* edited by S. J. O'Brien, Cold Spring harbor Laboratory press. Cold Spring harbor, N.Y.

Lande, R. and R. Thompson, 1990 Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics **124**: 743-756.

Lander, E. S., P. Green, J. Abrahamson, A. Baarlow, M. J. Daly, S. E. Lincoln and L. Newburg, 1987 MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics **1**: 174-181.

Lawson, D., M. Hemmat and N. Weeden, 1992 Identification of genes influencing morphological and physiological traits in apple. Abstracts of the International Conference on the Plant Genome I, San Diego, CA poster # 148.

Liu, B. H. and S. J. Knapp, 1990 GMENDEL: a program for Mendelian segregation and linkage analysis of individual or multiple progeny using log-likelihood ratios. J. Hered. **81**: 407.

Michelmore, R. W., Paran, I. and Kesseli, R.V. 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.

Moran, G. F. and J. C. Bell, 1983 *Eucalyptus*. In: Tanskley, SD, Orton, TJ (eds) Isozymes in Plant Genetics and Breeding, part B. Elsevier, Amsterdam, pp. 423-441.

Moran, G. F., J. C. Bell, M. Byrne, M. E. Devey and D. Smith, 1992 Molecular marker maps in *Pinus radiata* and *Eucalyptus nitens* for application in tree breeding programs. Plant Genome I, San Diego, CA poster # 93.

Morgante, M. and A. M. Olivieri, 1993 PCR-amplified microsatellites as markers in plant genetics. The Plant Journal **3**: 175-182.

Murashige, T. and F. Skoog, 1962 A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol. Plant. **15**: 473-497.

Neale, D. B. and C. G. Williams, 1991 Restriction fragment length polymorphism mapping in conifers and applications to forest genetics and tree improvement. Can. J. For. Res. **21**: 545-554.

Nelson, C. D., W. L. Nance and R. L. Doudrick, 1993 A partial genetic linkage map of slash pine (*Pinus elliotti* Engelm var. *elliottii*) based on random amplified polymorphic DNA's. Theor. Appl. Genet. **87**: 145-151.

Nodari, R. O., S. M. Tsai, R. L. Gilbertson and P. Gepts, 1993 Towards an integrated linkage map of common bean. II. Development of an RFLP-based linkage map. Theor. Appl. Genet. **85**: 513-520.

Reiter, R. S., J. G. K. Williams, K. A. Feldmann, J. A. Rafalski, S. V. Tingey, and P. A. Scolnik, 1992 Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random amplified polymorphic DNAs. Proc. Natl. Acad. Sci. USA **89**: 1477-1481.

Rhoads, D. D. and D. J. Roufa, 1990 Seqaid II 3.80. Kansas State University. Molecular Genetics Laboratory.

Ritter, E., C. Gebhardt and F. Salamini, 1990 Estimation of recombination frequencies and construction of linkage maps from crosses between heterozygous parents. Genetics **125**: 645-654.

Roy, A., N. Frascaria, J. MacKay and J. Bousquet, 1992 Segregating random amplified polymorphic DNAs (RAPDs) in *Betula alleghaniensis*. Theor. Appl. Genet. **85**: 173-180.

Sampson, J. F., S. D. Hopper, and S. H. James, 1990 Temporal variation in allele frequencies in the pollen pool of *Eucalyptus rhodantha*. Heredity **65**: 189-199.

Segal, G., M. Sarfatti, M. A. Schaffer, N. Ovi, D. Zamir and R. Fluhr, 1992 Correlation of genetic and physical structure in the region surrounding the *I2 Fusarium oxysporum* resistance locus in tomato. Mol Gen. Genet. **231**: 173-185.

Sobral, B. W. S. and R. J. Honeycutt, 1993 High output genetic mapping of polyploids using PCR-generated markers. Theor. Appl. Genet. **86**: 105-112.

Soller, M., 1978 The use of loci associated with quantitative traits in dairy cattle improvement. Anim. Prod. **27**: 133-139.

Song, Y. and C. A. Cullis, 1992 RFLP and RAPD mapping of *Eucalyptus globulus*. Plant Genome 1, San Diego, CA poster # 133.

Southern, E. M. 1975 Detection of specific sequences among DNA fragments separated by electrophoresis. J. Mol. Biol. **98**: 503-517.

Steane, D. A., A. K. West, B. M. Potts., J. R. Ovenden and J. B. Reid, 1991 Restriction fragment length polymorphism in chloroplast DNA from six species of *Eucalyptus*. Aust. J. Bot. **39**: 399-414.

Strauss, S. H., R. Lande and G. Namkoong, 1992 Limitations of molecular-marker aided selection in forest tree breeding. Can. J. For. Res. **22**: 1050-1061.

Tanksley, S.D., M. W. Ganal, J. P. Prince, M. C. de Vicente, M. W. Bonierbale, P. Broun, T. M. Fulton, J. J. Giovannoni, S. Grandillo, G. B. Martin, R. Messeguer, J. C. Miller, L. Miller, A. H. Paterson, O. Pineda, M. S. Roder, R. A. Wing, W. Wu and N. D. Young, 1992 High density molecular linkage maps of the tomato and potato genomes. Genetics **132**:1141-1160.

Tingey, S. V. and J. P. DelTufo, 1992 Genetic analysis with Random Amplified Polymorphic DNA markers. Plant Physiol. **101**: 349-352.

Torres, A. M., N. F. Weeden and A. Martin, 1993 Linkage among isozyme, RFLP and RAPD markers in *Vicia faba*. Theor. Appl. Genet. **85**: 937-945.

Tulsieram, L. K., J. C. Glaubitz, G. Kiss and J. Carlson, 1992 Single tree genetic linkage mapping using haploid DNA from megagametophytes. Bio/Technology **10**: 686-690.

Vallejos, C. E., N. S. Sakiyama and C. D. Chase, 1992 A molecular marker-based linkage map of *Phaseolus vulgaris* L. Genetics **131**: 733-740.

Weissenbach, J., G. Gypay, C. Dib, A. Vignal, J. Morissette, P. Millasseau, G. Vaysseix and M. Lathrop, 1992 A second-generation linkage map of the human genome. Nature **359:** 794-801.

Welsh, J. and M. McClelland, 1990 Fingerprinting genomes using PCR with arbitrary primers Nucleic Acids Res. **18:** 7213-7218.

Williams, J. G. K., J. A. Rafalski and S. V. Tingey, 1992 Genetic analysis using RAPD markers. Methods Enzymol. (in press).

Williams, J. G. K., A. R. Kubelik, K. J. Livak, J A. Rafalski and S. V. Tingey, 1990 DNA polymorphism amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. **18:** 6531-6535.

Wolff J. L. C., S. W. McDowell and M. A. Machado, 1993 Analysis of the restriction fragment length polymorphism of *E. urophylla*, *E. grandis* and *E. citriodora*. Braz. J. Genet. **16:** 431-439.

Wu, K. K., W. Burnquist, M. E. Sorrells, T. L. Tew, P. H. Moore and S. D. Tanksley, 1992 The detection and estimation of linkage in polyploids using single-dose restriction fragments. Theor. Appl. Genet. **83:** 294-300.

Yeh, F. C., A. Brune, W. M. Cheliak and D. C. Chipman, 1983 Mating system of *Eucalyptus citriodora* in a seed production area. Can. J. For. Res. **13:** 1051-1055.
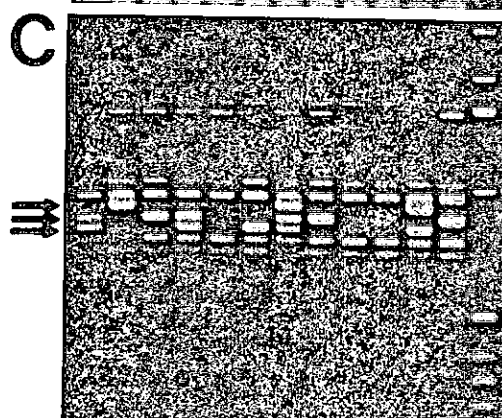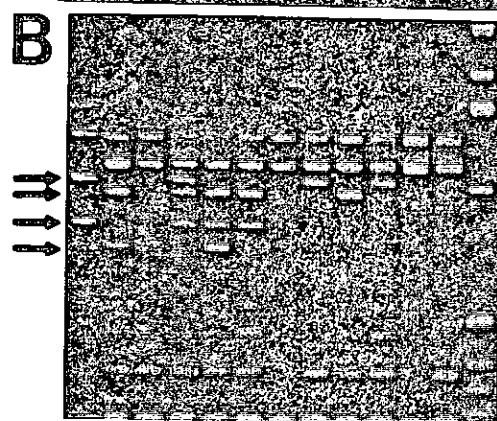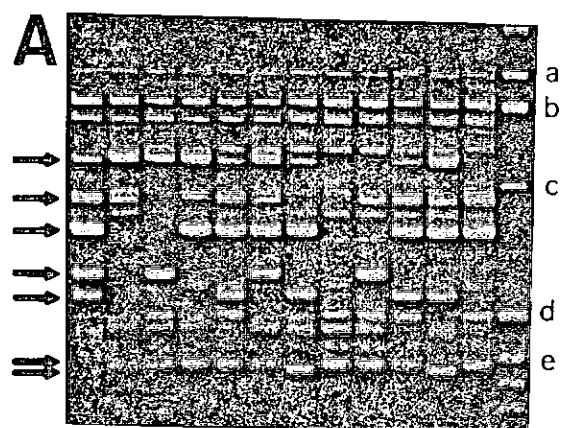
Figure 3. GENETIC LINKAGE MAP OF *Eucalyptus grandis* CLONE 44. Linkage relationships of 240 RAPD markers in 14 linkage groups were established at a threshold LOD score 5.0 and maximum $\theta = 0.25$. A framework map of 142 loci or loci clusters (indicated in bold letters along the linkage groups) that could be ordered with a likelihood support $\geq 1000:1$ was assembled covering 95.8 % of the estimated genome size. RAPD marker loci that could not be ordered with equal confidence were designated as accessory markers and are listed on the right of the linkage groups along with the approximate cM distance to the closest framework locus. RAPD marker loci are identified by the Operon primer code, fragment size in base pairs, fragment amplification intensity and linkage phase (+ or -) (see Materials and Methods). A RAPD marker locus showing significant distortion from 1:1 segregation ratio, is indicated by an asterisk.

GROUP 1

	P10_1250/2+
10.1	J6_510/2-
3.2	X16_560/3- — Z19_516/2-; 0.9
10.7	R7_490/1- ;1.8
	V7_400/3-
10.7	
	K10_1100/1-
9.8	
	Y20_520/3 +
8.1	
	J6_997/2 +
13.2	
	R6_670/1+
12.2	
	U16_630/3 + — R11_1250/2-; 1.8
	Y3_470/2+ ;0.0
19.0	G12_337/1+; 18.3
	X16_1050/1+
13.4	
	I4_1300/1+
5.9	Y18_350/2 + — Z3_830/3-; 3.3
12.7	
3.5	V7_450/2 -
5.3	R13_490/2 -
	K8_1000/2 +
13.2	
	Z17_1490/2 - — U7_1200/1-; 3.3
5.0	M4_533/2 -

GROUP 2

	U19_900/2 + — N7_739/1+; 3.6
10.5	
	U29_1000/3 +
11.5	
	L11_700/2 - — U19_450/1-; 18.7
7.0	L6_711/2 +
18.1	
	X3_2600/1-
18.6	
	R16_1650/2 -
15.0	
	P9_1350/2 + — G12_591/1+; 1.7
	K10_549/2+ ;1.7
8.2	X12_620/2 +; 1.7
	L17_533/3- ;1.7
	R3_580/3 + — J17_3335/1+ ;3.5
21.7	
	X3_1200/3+ — D3_1748/2 <7.1
19.0	K16_635/2-; 1.9
	K16_913/3 +
19.3	
	T1_690/1 -

GROUP 3

5.2	T9_900/3 - — P14_600/1+; 11.1
1.6	Z4_520/2 +
1.6	K4_230/3 -
	Y20_760/3 - — R8_1100/3- ;0.0
26.7	
	U16_550/1+
11.5	
1.8	K16_727/2-
4.9	N20_1850/2 + — R20_1650/1+ ;1.9
	M5_600/1- — Y15_1220/2+; 5.3
9.9	
	J7_580/2-
15.0	
	Y17_340/1 + — K12_2033/1-; 26.3
15.0	
	J1_363/1-
11.5	
	K19_304/2 -
8.7	J1_924/3+ — R15_210/1+; 1.7
	V7_770/2 < 1.7
8.7	J13_407/3-
10.5	
6.5	G13_1160/2-
3.3	X4_1850/2+ — X4_1780/2+ ;0.0
	P11_1800/2 -

GROUP 4

	N9_736/2-
9.8	R20_1090/3 -
9.1	U19_900/2 +
16.3	
	R20_2700/2 +
25.8	
	X1_720/2-
17.4	
	K1_1340/2+
9.3	K19_1495/2+
8.2	A20_1706/2+
6.5	F4_341/3-
6.5	K9_964/3 + — K19_449/3- ;0.0
12.4	F4_377/3+; 1.7
	I4_915/2-
13.8	
4.1	Y13_750/2 -
	N15_629/2-

GROUP 5

	Y3_860/2-
13.8	
	M12_759/2+ — R13_430/1+ ; 5.2
12.1	
	X1_1450/1+
14.0	
	R4_1300/2+ — W9_1830/1- ;0.9
	X7_580/2- ;0.0
13.5	
	M6_991/2+
16.9	
	Y18_1400/1- — Y17_1500/2- ;11.9
	H7_2350/2+ ;12.5
	A20_1068/2- ;7.9
11.6	U15_600/2-; 6.9
	X15_600/3- — K1_1000/2+ ;0.0
	A2_1691/2- ;3.6
	M15_786/2+; 1.8
11.5	T12_220/2+ ;1.7
	G6_1990/1+ — B6_780/1- ;3.4
	M10_1300/1+; 0.0
11.5	
4.9	K12_954/1+ — V7_1200/1+ ;3.5
1.9	K16_567/1+
7.6	G12_1231/2+ — P11_1450/3+; 6.0
	A10_635/3+ — A10_562/3- ;0.0
18.0	Y14_1400/2+ ;3.6
	Z5_550/3-

Figure 3. Continued

Figure 4. GENETIC LINKAGE MAP OF *Eucalyptus urophylla* CLONE **28.** Linkage relationships of 251 RAPD markers in 11 linkage groups were established at a threshold LOD score 5.0 and maximum θ=0.25. A framework map of 119 loci or loci clusters (indicated in bold letters alo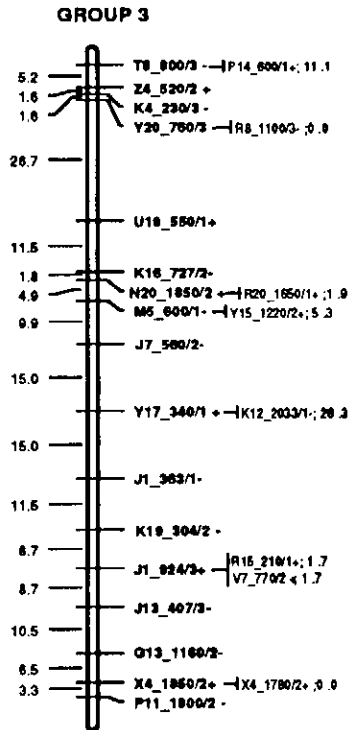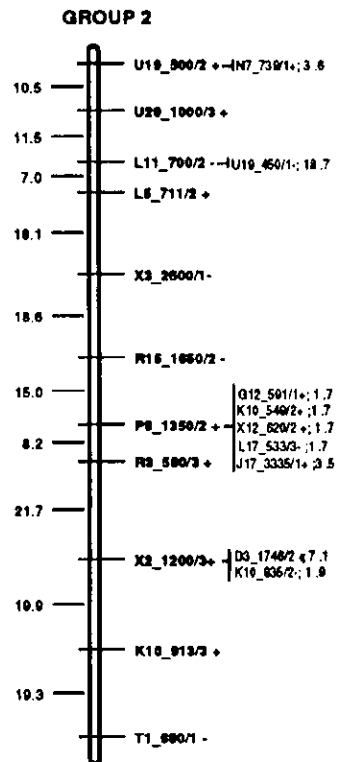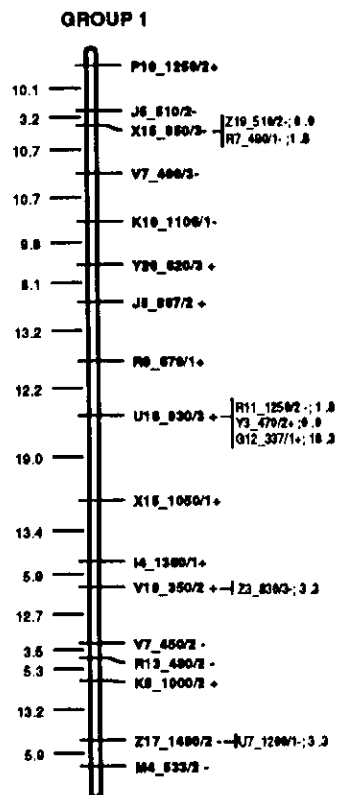ng the linkage groups) that could be ordered with a likelihood support $\geq$1000:1 was assembled covering 95.2 % of the estimated genome size. RAPD marker loci that could not be ordered with equal confidence were designated as accessory markers and are listed on the right of the linkage groups along with the approximate cM distance to the closest framework locus. RAPD marker loci are identified by the Operon primer code, fragment size in base pairs, fragment amplification intensity and linkage phase (+ or -) (see Materials and Methods). A RAPD marker locus showing significant distortion from 1:1 segregation ratio, is indicated by an asterisk.
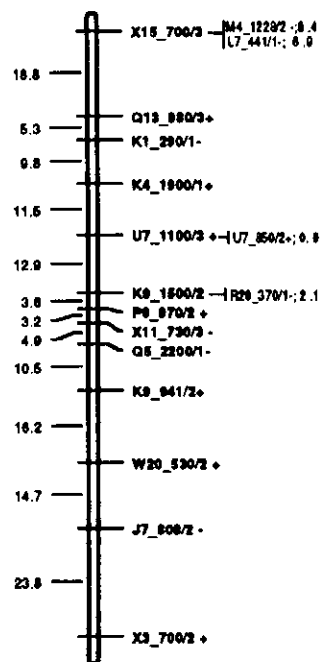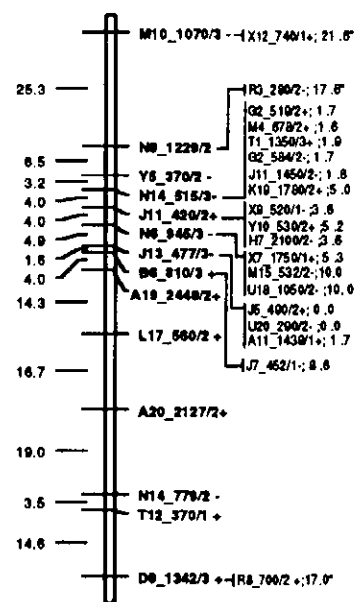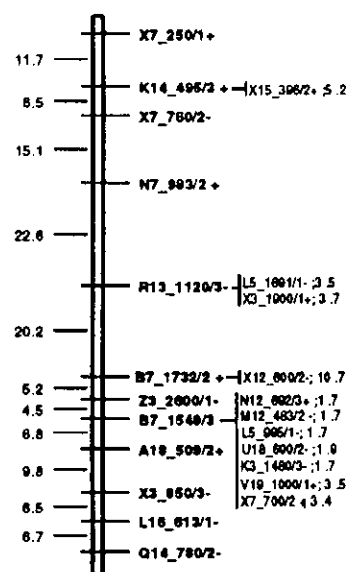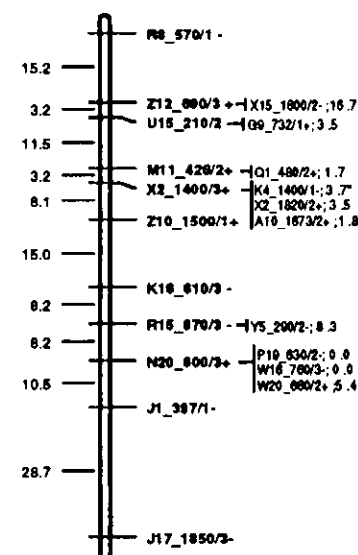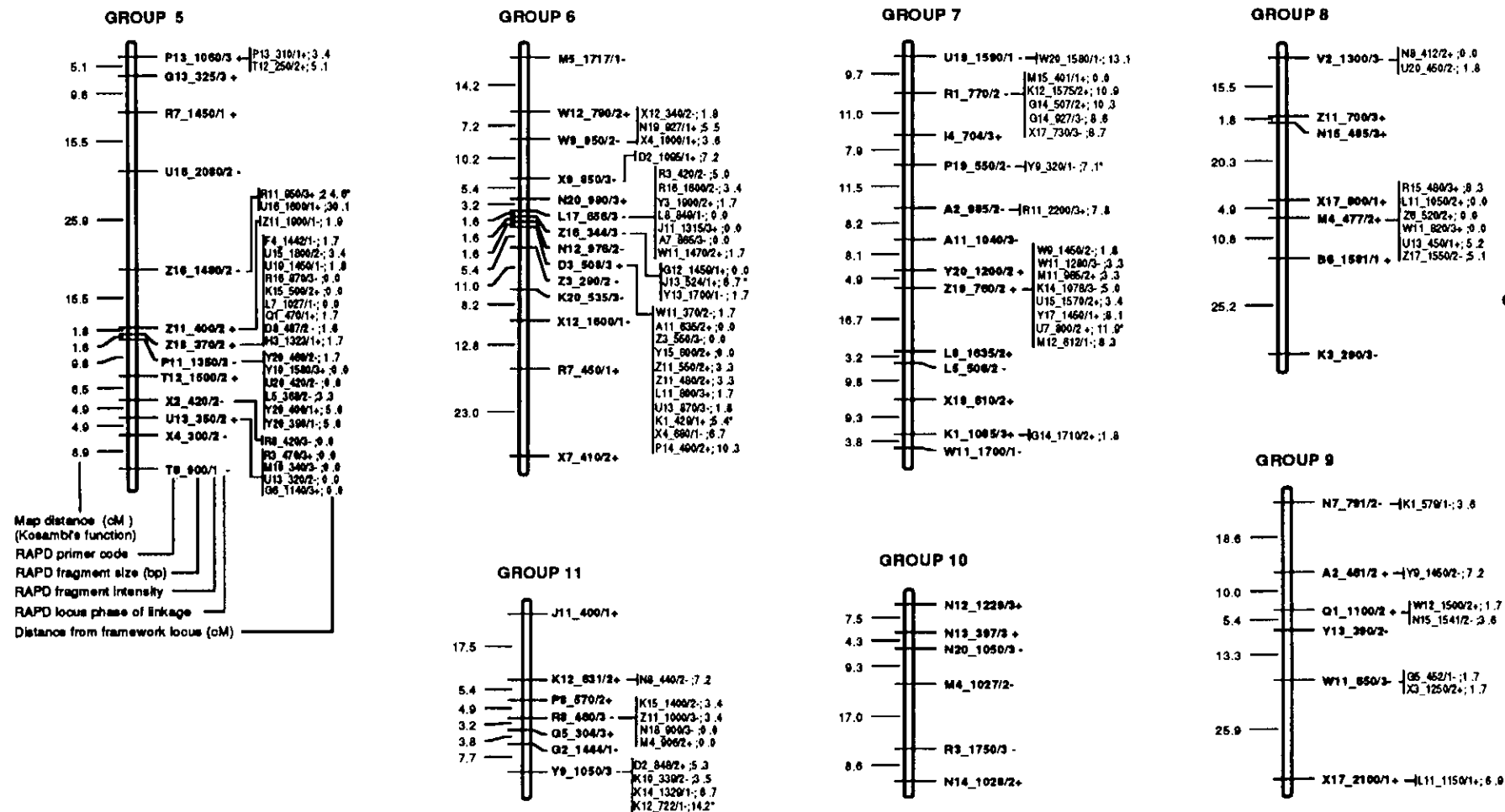
Figure 4. Continued

185

Figure 5. Sequence homology tests of RAPD markers. Panels show the RAPD gel profiles and the corresponding autoradiograms where the indicated RAPD fragments were used as hybridization probes under high stringency to confirm RAPD fragment homology within the parents/progeny set. RAPD assays were performed with genomic DNA of the parents *E. grandis* clone 44 and *E. urophylla* clone 28 (respectively second and third lanes of each panel) and a progeny sample of 10 F1 individuals. Segregating informative RAPD markers are indicated by arrows. First lane of gel profiles are 1 Kb DNA ladder size standards (BRL) (a= 1018 bp; b= 506 bp). Panel A: confirmation of allelism between RAPD fragments G14_917, G14_927 and G14_960 amplified in the two parents (see Results for details); Panel B: confirmation of homology of a RAPD marker locus (G2_720/3) heterozygous in both parents and therefore segregating 3:1 in the F1 progeny.
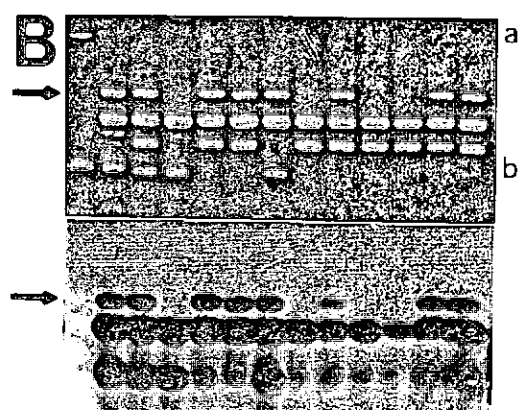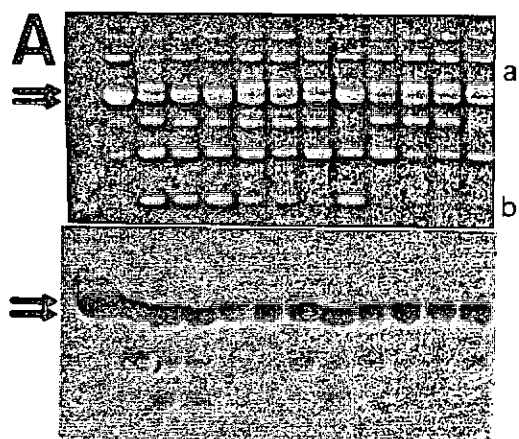
Figure 6. Survey of the occurrence of RAPD markers in different individuals of *E. grandis*.
Panels show the RAPD gel profiles and the corresponding autoradiograms where the indicated RAPD fragments were used as hybridization probes under high stringency to confirm RAPD fragment homology. RAPD assay was performed on genomic DNA of two sets of parents and 10 F1 progeny each. Lanes from left to right: *E. grandis* clone 44, *E. urophylla* clone 28 ; 10 F1 progeny; 1 Kb DNA ladder size standard (BRL) (letter a on the far left indicates the position of the 1018 bp standard), *E. grandis* clone 816/2, *E. urophylla* clone 28, 10 F1 progeny. Panel A: RAPD marker N15_1079 present in clone 44 and not in clone 816/2; Panel B: RAPD marker M5_961 present in both *E. grandis* clones in a heterozygous state thus segregating 1:1 in both crosses; Panel C: confirmation of codominance of marker pair A11_980/A11_920. The hybridization experiment also shows that this RAPD marker is also present in clone 816/2 in a heterozygous state with the 980 bp allele and a third size-variant allele, A11_970. This marker locus therefore segregates in both crosses in a codominant fashion, with one allele in common between the crosses.
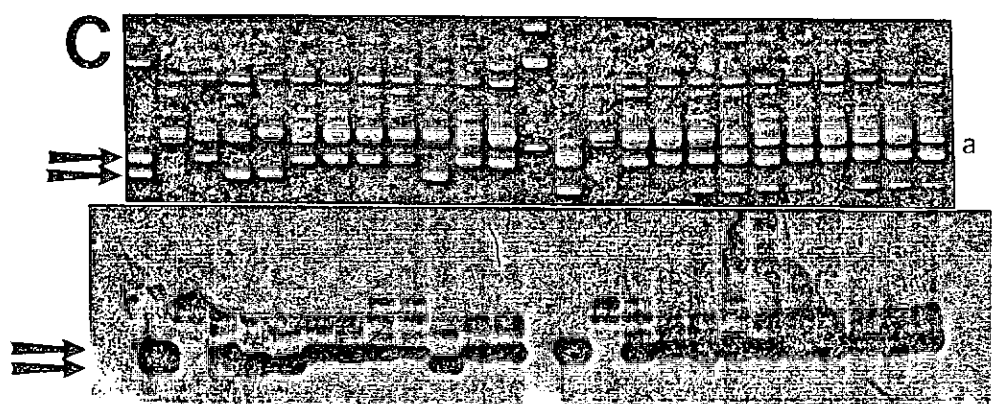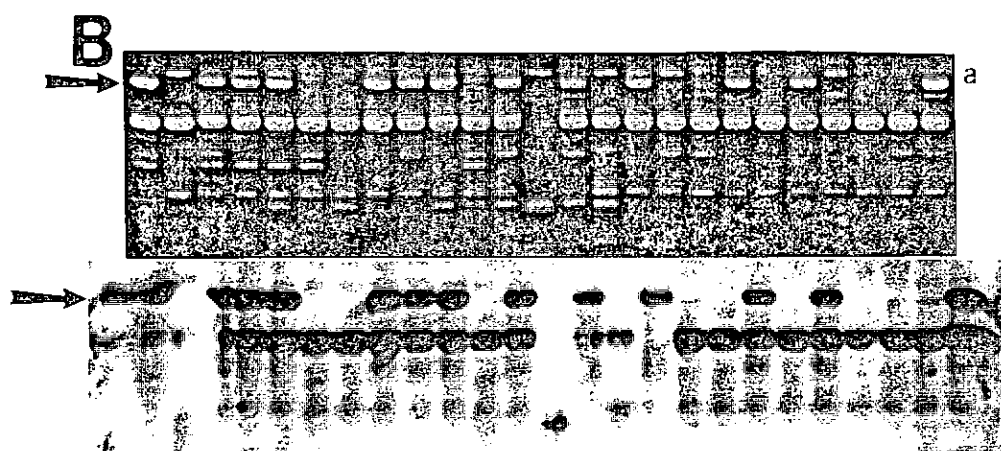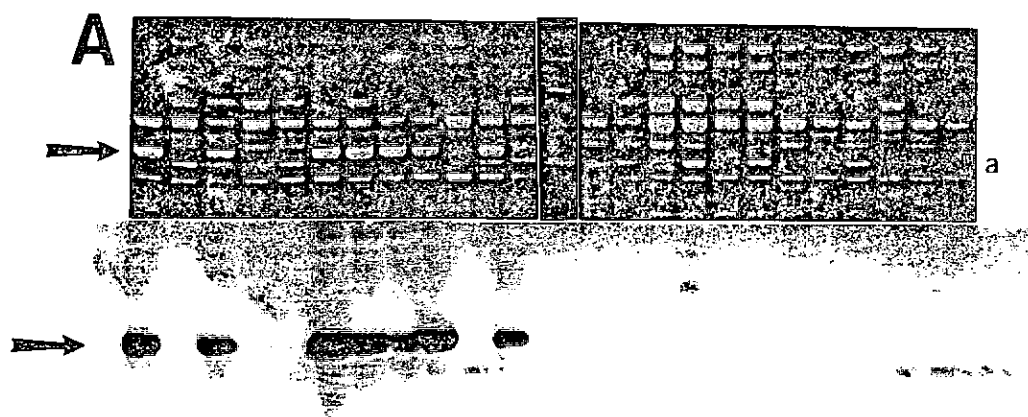
Figure 7. Characterization of genomic sequence complexity of mapped RAPD marker loci. Panel A: top horizontal dot blot corresponds to a control blot with a reconstruction experiment where signals indicate a sequence represented by (1) 1 copy; (2) 10 copies; (3) 100 copies; (4) 1000 copies (see Materials and Methods for details). Vertical dot blots and corresponding *EcoRI* Southern blots show RAPD fragments classified in the following copy number classes: (1) low copy (1-10 copies); (2) moderately repetitive (10-100 copies); (3) highly repetitive (100-1000 copies) and (4) very highly repetitive (>1000 copies). Dots contain (a) negative control, 5 µg herring sperm DNA; (b) 5 µg *Eucalyptus* genomic DNA; (c) 0.5 µg ; (d) 0.05 µg. Panel B: frequency histogram of genome complexity classes of mapped RAPD markers based on a sample of 48 RAPD markers. Numbers on X axis correspond to the same copy number classes described above.

A

1 2 3 4

a
b
c
d
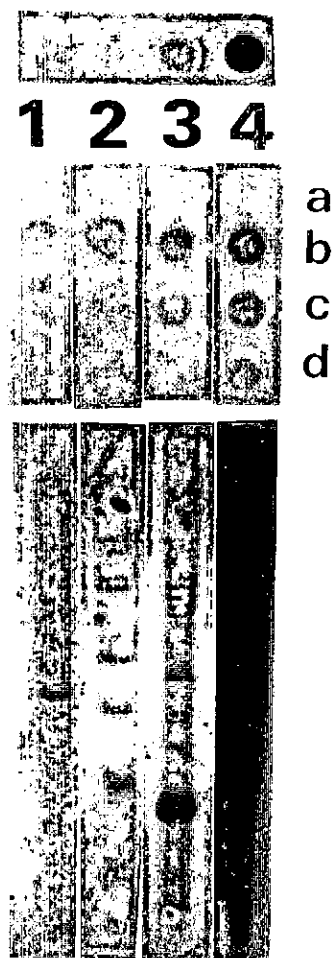
B
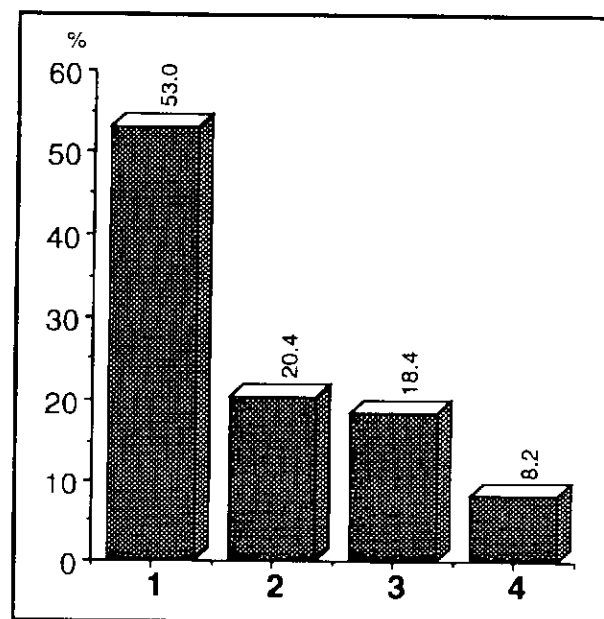
ABSTRACT

In this study we have extended the combined use of the "pseudo-testcross" mapping strategy and RAPD markers to map QTLs controlling traits related to the ability of vegetatively propagate trees in *Eucalyptus*. QTL analyses were performed using two different interval mapping methods. A total of 10 putative QTLs were detected for micropropagation response (measured as fresh weight of shoots, FWS), 6 for stump sprouting ability (measured as # stump sprout cuttings, #CUTT) and 4 for rooting ability (measured as % rooting of cuttings, %ROOT). With the exception of three QTLs, both interval mapping methods yielded similar results in terms of QTL detection. Discrepancies in the most likely QTL location were observed between the two methods, although in 75% of the cases the most likely position was either in the same or in an adjacent interval. Standardized gene substitution effects for the QTLs detected were typically between 0.46 and 2.1 phenotypic standard deviation (σp), while differences between the family mean and the favorable QTL genotype were between 0.25 and 1.07 σp. Multipoint estimates of the total phenotypic variation explained by the QTLs (52.5 for FWS, 28.2% for #CUTT, 32.6% for %ROOT) indicate that relatively large numbers of genes control the traits investigated, however major effect QTLs for rooting and micropropagation response were identified. The QTL mapping information indicates that *E. grandis* is responsible for most of the inherited variation in the ability to form shoots, while *E. urophylla* is for rooting. QTL mapping in the pseudo-testcross configuration relies on within-family linkage disequilibrium to establish marker/trait associations. With this approach one can contemplate QTL analysis in any available full-sib family generated from undomesticated and highly heterozygous organisms such as forest trees. QTL mapping on two-generation pedigrees opens the possibility of using already existing families in retrospective QTL analyses allowing one to gather at once the quantitative data necessary to move into marker-assisted tree breeding.

INTRODUCTION

Vegetative propagation is a powerful way to capture the genetic superiority of a selected individual. In clonal propagation, both additive and non-additive sources of genetic variation contribute to the gain, while in sexual propagation the gain is achieved exclusively on the basis of the interfamily component of the genetic variance.

Very little information is available on the genetic basis of such traits. Easy and hard-to-root species of *Eucalyptus* have been identified (Hartney 1980). However, most of the existing information is either anecdotal or proprietary in nature and therefore not published. No estimates of genetic parameters such as heritability or information on the genetic control and architecture of vegetative propagation traits are available for forest trees.

Genetic linkage maps of molecular markers offer a powerful tool to investigate the genetic architecture of polygenic traits and potentially assist in their manipulation through marker-assisted selection and breeding. A number of studies in recent years have used molecular markers to investigate the inheritance of quantitative traits. Results to date strongly support the existence of a few major genes controlling large proportions of the total variation in a wide range of quantitatively inherited traits (reviewed by Stuber 1992; Dudley 1993). These studies, however, have been limited to a few annual crop plants and have been performed using segregating populations derived from crosses between inbred lines. Such populations are not available in trees and are difficult to obtain due to a significant genetic load and time constraints. To circumvent this limitation we recently adopted a "two-way pseudo-testcross" approach and RAPD (Random Amplified Polymorphic DNA) markers to construct linkage maps for individual trees of *Eucalyptus* (Grattapaglia and Sederoff 1994). In this study we extend the use of this approach for QTL analysis. We report the identification of quantitative trait loci (QTLs) controlling significant proportions of the phenotypic variation in traits related to the ability to vegetatively propagate trees in *Eucalyptus grandis* and *E. urophylla*. The results presented provide the first insights into the possibility of marker-assisted breeding of these traits in forest tree improvement programs.

MATERIAL AND METHODS

Plant Material. The experimental material consisted of a single controlled cross between two highly heterozygous elite trees. *Eucalyptus grandis* (clone 44, Coffs Harbor provenance, Australia - selection from a Zimbabwe seed source), used as the female parent was crossed to *E. urophylla* (clone 28 selection from Rio Claro land race, Brazil), used as male, in 1989 at Aracruz Florestal S.A., Brazil. Sixty two F1 individuals of this population had been used for the construction of genetic linkage maps (Grattapaglia and Sederoff 1994). For this study, the mapping population was expanded to 122

Therefore the full benefit of broad sense heritability is realized rather than only some portion of the narrow-sense heritability.

In horticulture, vegetative propagation of desired plant phenotypes has been used successfully for centuries (Hartman and Kester 1983). In forestry, however, aside from a few genera like *Populus, Salix* and *Cryptomeria,* vegetative propagation of "plus" trees has not been used extensively in most operational forest planting programs (Zobel and Talbert 1984). This is largely because cuttings from physiologically mature trees of many species are difficult or impossible to root.

Species of the genus *Eucalyptus* constitute the majority of the world's planted hardwood forest and one of the world's main sources of cultivated biomass (Eldridge *et al.* 1993). In sprouting species such as the eucalypts, the stump sprouts are physiologically juvenile and can be rooted, therefore allowing in principle, clonal propagation of individuals. Propagation systems based on rooted cuttings have been optimized and implemented at the production level, resulting in outstanding gains in productivity and uniformity (Campinhos and Ikemori 1980; Delwaulle 1985). Currently, the largest operational clonal forestry programs are with species of *Eucalyptus.* In the tropics, such operations yield the highest productivity of woody biomass on earth (Brandao 1984).

As an alternative to rooted cuttings, methods of micropropagation by tissue culture have been developed for several species of *Eucalyptus* (DeFossard 1974; Gupta and Mascarenhas 1987). Some progress has been made in micropropagation of adult selected trees, trying to mitigate the problem of lack of propagation potential due to maturation. Although *in vitro* methods are not economically viable for large operations, they have been used as an efficient way to rapidly develop a "sprout nursery" to produce the number of cuttings needed for operational planting (Grattapaglia *et al.* 1990).

The ability to sprout, root and respond to tissue culture varies widely both within and particularly across species of *Eucalyptus* (Hartney 1980; Zobel 1993). Variation in rooting ability frequently dictates which trees will be available in a planting operation, severely limiting the use of clonal propagation when particular species are the most desired. For example *Eucalyptus globulus* which has some of the best wood properties for cellulose pulp production roots very poorly. The transfer of vegetative propagation traits by intra and interspecific hybridization is an increasingly important objective in many breeding programs, that aim at exploring the benefits of clonal propagation.

individuals. Seeds were originally germinated on solid agar containing half-strength MS medium (Murashige and Skoog 1962) under a 14 hour photoperiod. The population was immortalized by establishing clonal cultures of the individuals by vegetative propagation *in vitro* on maintenance medium (half-strength MS medium supplemented with 0.005 mg/l IBA, indol-butyric acid). Rooted plantlets produced *in vitro* were transplanted to containers containing a fertilized 1:1:1 mixture of vermiculite, soil and peat moss under mist irrigation for two weeks. After two months of growth, two similar looking plants per individual were transplanted to large (20 liters) pots and managed for fast growth and cutting production.

Experimental designs and traits measured. The following quantitative traits related to vegetative propagation response were evaluated: (1) micropropagation response (fresh weight of *in vitro* micropropagated shoot clumps, FWS); (2) sprouting (number of stump sprout cuttings, #CUTT); (3) adventitious rooting response (percentage of rooted cuttings, %ROOT). Two plants per individual were used for phenotype evaluations. Although limited, this clonal replication of the individuals provided some increased precision in trait measurement. Micropropagation response was analyzed by a randomized complete block design with two blocks and two explant plots. Explants were stem segments containing one axillary meristem derived from plants grown in maintenance medium. We had previously determined that the two basal axillary nodes are the most consistently responsive to induction of shoot growth in tissue culture (Grattapaglia, unpublished). Therefore only these explants were used to control for this source of inherent physiological variation. Fresh and dry (24 hours at 105°C) weight of shoot clumps were evaluated at 25 days of culture. Because dry and fresh weight were found to be highly correlated traits in our study (r=0.98), QTL analyses were performed only for fresh weight.

For coppice and rooting response, two potted plants per individual were grown for three months (~2 cm stem diameter) and then cut back to stimulate dormant buds to sprout. Sprouting was then evaluated as the number of operational quality (one node/two leaf) stump sprout cuttings that could be harvested after 60 days following cut back of the plant. Cuttings harvested in this evaluation were put to root to evaluate adventitious rooting response. The basal 1 cm of the cutting was dipped into a 5000 ppm IBA solution in talc and placed into a rooting medium consisting of a fertilized 3:1 mixture of vermiculite and pine bark. A variable number of cuttings was therefore used

to evaluate rooting response, and the measurement was recorded in terms of percent cuttings able to root. The experiment was a randomized complete block design with two blocks (corresponding to the two potted plants) and a variable number of cuttings per plot.

RAPD marker genotyping. DNA extractions, RAPD assay conditions, marker identification and scoring were performed as described elsewhere (Grattapaglia and Sederoff 1994). RAPD markers in the pseudo-testcross mating configuration are present in a heterozygous state in one parent and absent in the other or vice versa, and segregate 1:1 in the F1 generation. Two separate sets of linkage data are obtained, one for each parent. A total of 165 markers for *E. grandis* and 166 for *E. urophylla* that segregated accordingly ($\chi 2$ test at $\alpha=0.05$) were employed in this study. These included 100 and 83 framework markers respectively for *E. grandis* and *E. urophylla*. Markers assigned to the framework map were those that had been previously ordered with a likelihood support $\geq$ 1000:1 (Grattapaglia and Sederoff 1994).

Data analysis. Although all the RAPD markers used in this study have been mapped previously (Grattapaglia and Sederoff 1994), genetic maps were calculated from the genotypic data *de novo* and checked for consistency with the previously reported maps. Linkage relationships among markers were determined using MAPMAKER (Lander *et al.* 1987). To allow the detection of linkage of RAPD markers in repulsion phase the data set was duplicated and recoded. LOD 5.0, and maximum $\theta=0.30$ were used as linkage thresholds for grouping markers. The software program GMENDEL (Liu and Knapp 1990) with threshold P value= 0.0001 and $\theta=0.30$ was also used for the linkage analysis.

QTL mapping analysis was performed using interval mapping methods implemented by MAPMAKER-QTL (Lander and Botstein 1989) and QTLSTAT (Knapp *et al.* 1992). QTL analyses were performed on the mean trait value of the F1 individuals, computed as the average of plot means across the two blocks (stock plants). The genetic analysis was carried out under a backcross model. Separate analyses were performed on each parental linkage map. A LOD score threshold of 1.6 _or_ nominal significance level of p=0.01 were used to declare the presence of a linked QTL in the interval. With this stringency and given the number of markers per chromosome used, a per-chromosome false positive rate of 5% was ensured as estimated numerically by

Darvasi *et al.* 1993. For each LOD peak, the 1.0 LOD support intervals were determined. For all detected QTLs, the percentage of variance explained as estimated by MAPMAKER-QTL and QTLSTAT (as non-linear R^2), and the shifts in trait value in phenotypic standard deviations were also reported. When linked QTLs with no overlapping 1.0 LOD support intervals were detected, the locus with highest LOD score was fixed and the chromosome scanned again for the linked effect. Multipoint estimates of the total variation explained by the mapped QTLs were obtained by interval mapping with MAPMAKER-QTL and by multiple liner regression using PROC GLM (SAS 1988).

RESULTS

Linkage maps calculations. The majority of the RAPD markers used in this study were previously classified as framework markers, i.e. their order was established with a likelihood support $\geq$1000:1 (Grattapaglia and Sederoff 1994). On the linkage maps, they are identified with bold type (Figures 1 and 2). When genotyping with RAPD markers, each arbitrary primer amplifies more than one segregating marker. So, in addition to the target framework markers, accessory markers are obtained by default. Most of the markers used were framework markers (100 of 165 in *E. grandis* and 83 of 166 in *E. urophylla*), however data for accessory markers were also obtained and included in the analysis when assembling the linkage maps and carrying out the QTL analysis. A LOD score of 5.0 and maximum θ= 0.30 were set as linkage thresholds for grouping markers. Map distances in centimorgans were calculated using Kosambi's mapping function. Orders of marker loci in each linkage group were established using a matrix correlation method implemented by MAPMAKER. The orders obtained compared very closely to the orders resulting from simulated annealing performed by GMENDEL. In the maps presented the order for the framework markers (in bold type) conforms to the 1000:1 likelihood support. The locus order support was relaxed to a 100:1 when builiding the maps to include all markers.

The framework marker orders of these maps (based on 122 meiosis), are well conserved when compared to our previously published maps (based on 62 meiosis). In *E. grandis*, there were four cases involving a switch in the order of two adjacent markers: Z18_1630 and R16_820 on group 8; R16_730 and N7_1322 on group 6; P8_1350 and R15_1650 on group 2 and U19_800 and R20_1080 on group 4. There were three

cases of a single marker out of the original order: marker X15_600 on group 5, marker N13_533 on group 7 and marker Y15_740 on group 9. In *E. urophylla*, there were three cases of switch in order of adjacent markers : markers B7_1549 and A18_509 on group 3; marker pairs G5_304 and G2_1444 and K12_631 and P8_570 on group 11. One triplet of markers was inverted in orientation (markers X4_300, T12_1500 and Z16_1480) on group 5. There were five cases of a single marker out of order: marker L17_560, on group 2, Z16_344 on group 6, K3_290 on group 8, X17_2100 on group 9 and M4_1027 on group 10.

In *E. urophylla*, the original marker linkage grouping remained the same. In *E. grandis*, however, one case of breakage and two cases of mergers of previously reported linkage groups were observed. Groups 8 and 12 merged into one. Group 11 was split in two pieces (designated 11a and 11b) which in turn were linked to groups 9 and 13. Although the total number of linkage groups was reduced from 14 to 12, it is still one more than the expected number based on the haploid number of chromosomes in *Eucalyptus grandis* (n=11). The mergers of groups 8 and 12 and groups 11 and 13 had been suggested previously (Grattapaglia and Sederoff 1994). However LOD scores for linkages were below the adopted threshold for grouping (LOD<5.0) and thus these groups were kept apart. This study, involving a larger number of meiosis, increased power and precision to the linkage analysis and resulted in a more likely hypothesis of linkage groups for *E. grandis*. Significant LOD scores for linkage among several markers in the merged groups were observed. For group 11, the increased sample size apparently resolved a case of spurious linkage that had kept the group together originally.

Quantitative traits. The two parents of the original cross used in this QTL experiment could not be evaluated for the vegetative propagation traits. Therefore the trait values for the F1 progeny individuals cannot be compared to the parental generation. It is known, however, that stump sprout cuttings of *E. urophylla* typically root at significantly higher percentages than *E. grandis* and that both species have similar sprouting ability and response to tissue culture (Bertolucci, unpublished results; Grattapaglia *et al.* 1987). Interspecific F1 hybrids typically display intermediate behavior for these traits at the family mean level. However, because the individuals crossed are highly heterozygous, the F1 is genetically heterogeneous, and a significant level of genetic variation exists. This genetic variation was explored in the QTL mapping experiment. Extreme phenotypes with traits values above two phenotypic standard

deviations from the mean were observed for all three traits and below the mean for all traits with the exception of %ROOT that displayed a right skewed distribution (Figure 3).

The frequency distributions of phenotypes for the three traits showed an approximately normal continuous variation (Figure 3). Mean, standard deviation and sample size for each trait are also presented. Sample sizes for phenotypic measurements were slightly less than the total number of individuals genotyped, as a result of loss of individuals either due to contamination in the tissue culture experiment or loss of plants in the greenhouse trials. One individual that did not sprout enough to yield operational cuttings but remained alive following the cut back of the stock plant was included in the analysis for #CUTT with zero trait value. This individual, for which no cuttings were available for establishing the rooting evaluation, was not included in the rooting measurements. Therefore the sample size dropped from n=97 to n=96 when measuring %ROOT. There were two individuals that had on the average less than five cuttings per plant. Because %ROOT estimated for these individuals was based on such a small sample of cuttings, QTL analyses were performed both including and excluding these individuals. No significant differences were observed in the results (data not shown), and therefore those individuals were kept in the data set. The majority of individuals had an average of six or more cuttings per plant. Therefore the estimated %ROOT correspond to the average percent rooting across the two plants, based on an average of six or more cuttings each, i.e. a total of 12 or more cuttings per individual genotype.

For all traits, the analyses were performed on the untransformed phenotypic data. When using MAPMAKER-QTL the data were log transformed to more closely fit a normal distribution and the analyses repeated. However, even for %ROOT, which displayed a right skewed distribution, the analyses of log transformed data did not alter any of the results. Therefore all results presented are for untransformed data.

Phenotypic correlations estimated among traits were not significant at $\alpha=0.05$. These were: FWS x #CUTT r= 0.17; FWS x %ROOT r= 0.004; #CUTT x %ROOT r=0.07.

QTL analysis. QTL analyses were performed using two different interval mapping methods. MAPMAKER-QTL (MMQ) was used to estimate QTL parameters and test statistics at 2 cM interval within every marker bracket and to select the most probable location for the QTL as the location that maximized the likelihood ratio. QTLSTAT (QST) employs non-linear least-squares to estimate the QTL genotype means and test the hypothesis of "no QTL" versus the hypothesis of "one QTL" for every marker bracket. The

QTL is therefore assigned to a particular marker interval directly, without trying to estimate the most likely position within that interval.

The results of the two analyses generally agreed closely, with a few exceptions (see below). A total of ten putative QTLs were detected for micropropagation response (FWS), six for sprouting ability (#CUTT) and four for rooting ability (%ROOT) (Tables 1 to 4). In three instances, a genomic region was declared significant based on one analysis but did not reach the significant threshold for the other. In *E. grandis* a QTL was declared on group 3 based on a P=0.003, when the LOD peak was only 1.3. On group 1 a QTL was declared following a P=0.009 when the LOD score was only 1.5. In *E. urophylla* on the other hand a QTL was declared based on a LOD peak of 1.6 but the P value from QST was 0.014. Typically, LOD scores or Wald statistics above the significant thresholds adopted were observed along stretches involving more than one marker interval. When two linked LOD peaks were observed, with overlapping LOD 1.0 support interval, only one QTL was declared in the region at the location with the highest peak. When the LOD 1.0 supports did not overlap, a procedure suggested by Lander and Botstein (1989) was used. The position of one QTL was fixed and the chromosome scanned again for a linked QTL effect. Such a procedure was applied for *E. grandis* in two regions: on linkage group 1 for FWS and group 5 for %ROOT. Only on linkage group 1 for FWS, after controlling for each peak, sufficient evidence remained (Δ LOD= 2.9) to declare two linked peaks (Figure 1 and Tables 1 and 2).

In the QTL summaries we report the marker interval where either the highest LOD score estimated by MMQ or Wald statistic by QST were observed. For MMQ we also locate the 1.0 LOD support interval (Tables 1 to 4). The results of both analyses generally agree. However the marker bracket where the highest LOD score was located generally did not correspond exactly to the one with the highest Wald statistics. For a total of 20 QTL regions, in only eight (40%) did the marker bracket with the highest LOD score correspond to the bracket with the peak . statistics. Frequently (7 in 20, 35%) we found that the marker brackets where the LOD and Wald peaks were detected were adjacent, rather than overlapping, and shared a common marker. This was more common for *E. grandis* than *E. urophylla*. For example: QTLs for FWS on groups 1, 7 and 8/12 and for #CUTT on groups 1 and 13/11b in *E. grandis* were located to different but adjacent intervals in the two analyses (Tables 1 and 2). Note that in the majority of those cases the most likely position estimated by MMQ was exactly at the shared marker, i.e. QTL position 0.0 cM. Finally we also found cases of non-adjacent, but nearby marker

intervals (5 in 20, 25%). Different but nearby marker intervals were located for QTLs controlling #CUTT on groups 2 and 8/12 for *E. grandis* and FWS on group 5 and #CUTT on group 2 in *E. urophylla*. It is important to point out that in both cases, i.e. adjacent or nearby, the peak intervals determined by QST were within the 1.0 LOD support interval estimated by MMQ.

A somewhat different case of disagreement was found in *E. grandis*. On group 1, LOD peaks in two intervals were detected for FWS using MMQ. Although one interval (P10_1250 - J5_510) was also significant with QST (Wald=13.1), the highest Wald statistics peak detected in the region was in a nearby interval (Wald= 15.7 in interval K10_1106 - Y20_620 - Table 2). In this case, the nearby interval detected with QST was not included in the 1.0 LOD support interval estimated by MMQ. Although MMQ provided evidence for two linked effects, such a statement should be viewed with caution in light of the disagreement of the two analyses regarding the exact position of one of the effects. A larger sample size or a different analytical approach (Zeng 1994) could help resolve this issue.

Estimates of the proportion of phenotypic variation explained by each QTL were obtained from MMQ (Tables 1 and 3). Such proportions were also estimated with QST as the coefficient of determination (R^2) (Tables 2 and 4) for the single locus model from the least squares analysis of variance table by dividing the Sum of Squares for the QTL genotype by the total Sum of Squares. The two estimates proportionally agree, with one exception (see below). However, the proportions estimated as R^2 values are always larger than the estimates of % variation explained from MMQ. For FWS, individual QTLs explained between 6.8 and 12.3% of the variation, while the corresponding R^2 values ranged from 0.09 to 0.18, i.e between 9 and 18% of the variation was explained by each single QTL model. For %ROOT, MMQ estimates were from 7.3 to 21% and between 10 and 28% for QST. For #CUTT, MMQ estimates were between 7.3 and 10.8% while QST estimated the same individual effects between 8 and 38%. A major discrepancy was observed for a QTL controlling #CUTT on group 2 in *E. urophylla*. While MMQ estimate was 9.2%, the R^2 corresponded to 38% of the variation (Tables 3 and 4). Knapp *et al.* (1992) pointed out that estimates of R^2 obtained from nonsimultaneous single locus models can be significantly inflated by sampling bias. Therefore, the estimates of proportion of variation explained by single locus models should be viewed with caution.

Simultaneous multilocus estimates of the total proportion of phenotypic variation explained by the joint action of the putative QTLs mapped for each parental tree were

obtained by multipoint interval mapping with MMQ and by multiple linear regression using PROC GLM (SAS 1988) (Table 5). Note that the simple arithmetic sum of the individual effects estimated (Tables 1 to 4) would always be significantly larger than the multipoint estimates. For example if we summed all the effects for FWS in *E. grandis* we would have 56.2% which is larger than 41.6% estimated by multipoint interval mapping. Similarly in *E. urophylla* we would find 35.7% compared to 25.2% by multipoint QTL mapping. Estimates of the total % phenotypic variation explained by the joint action of all putative QTLs mapped in both parents were also obtained by multipoint interval mapping and multiple linear regression (Table 5). The estimates obtained by linear regression are generally smaller than those obtained by interval mapping. This was also observed in a maize QTL mapping experiment (Doebley and Stec 1993). This might be due to the higher efficiency of the interval method in QTL mapping (Lander and Botstein 1989) and the fact that in the multiple linear regression procedure, all those observations that have missing genotypic data are deleted from the analysis, resulting in considerably smaller sample sizes especially when many markers are fitted into the model.

As more than one measurement of the traits were made on each individual, we were able to partition the total phenotypic variance into variance within and variance between individuals. Repeatability for each trait was estimated as the ratio between the variance between individuals and the total variance (Table 5). As pointed out by Falconer (1989), the repeatability sets an upper limit to the degree of genetic determination and to the heritability. By weighing the multipoint interval mapping estimates of total phenotypic variation by the repeatabilities we arrived to an upper limit estimate of the proportion of genetic variation explained by the QTLs mapped (Table 5).

Least square means of the alternative QTL genotypes and their associated standard deviations were estimated with QST (Tables 2 and 4). Note that for all the putative QTLs detected, variances of the alternative QTL genotype classes were generally equal and close in value across QTLs within traits. Higher values of the within-QTL class variances were observed in two QTLs for FWS: on group 1 in *E. grandis* and group 5 in *E. urophylla*. In both cases, these higher variances corresponded to QTLs where large differences in mean trait value between the two alternative QTL genotypes (Δ_1) were observed. In our QTL mapping experiment, these differences ranged from 0.46 to 2.1 phenotypic standard deviations (σ_p) (Tables 2 and 4). Estimates of Δ_1 were

between 0.5 and 0.7 for the majority (65%) of the QTLs detected. A potentially more interesting estimate from the breeding standpoint is the difference in mean trait value between the family mean and the favorable QTL genotype (Δ_2) (Tables 2 and 4). These differences ranged from 0.25 to 1.07 σ_p and typical values were between 0.2 and 0.4 σ_p.

The QTL detected with the highest LOD score (5.8) was for %ROOT in *E. urophylla*. The closest linked marker (R8_570) is shown segregating in a sample of F_1 individuals (Figure 4). The effect of the substitution of this linked RAPD marker resulted in the doubling of the rooting percentage (from 23.4 to 48.3%) or an increase in 12% rooting above the family mean (Δ_2 =0.58 σ_p) (Table 4). QTLs of large effects for FWS were detected for both parents (on group 1 for *E. grandis* and group 5 for *E. urophylla*), and in both cases the difference between the favorable allele and the family mean were estimated to be above 1.0 σ_p (Tables 2 and 4). For #CUTT, the QTL detected with the highest LOD score was on group 13/11b, with the most likely position estimated to be exactly at marker N13_2169 (Figure 5). However, in spite of the high LOD score and percent variation explained (10.8%) estimated by MMQ, the effective contribution of this region to the trait does not result in a significant shift from the family mean. The difference between the alternative QTL genotypes (Δ_1) is 0.63 σ_p, however the difference between the favorable QTL genotype and the family mean (Δ_2) is only 0.12 σ_p. It seems that estimates of Δ_1 can be misleading. Even though a significant shift in average trait value is caused by the allelic substitution at the QTL, the final trait value can still be within the average of the family. Estimates of Δ_2 seem to be more useful than both Δ_1 and % variance explained as indicators of the relative importance of the QTL detected, as they translate into more meaningful values from the standpoint of marker assisted selection.

In both species, there were two cases where the LOD 1.0 support interval of QTLs detected for FWS and #CUTT overlapped: on groups 1 and 8/12 in *E. grandis* and groups 1 and 2 in *E. urophylla*. Both traits involve multiple shoot formation from dormant buds, differing in the fact that in FWS, shoot formation is stimulated *in vitro* with the action of cytokinin while in #CUTT it relies exclusively on the intrinsic physiological ability to break dormancy of resting juvenile buds. It seems reasonable to suggest that these two traits - although not significantly correlated in this experiment - should share some common QTLs. Although pleiotropic gene action might be a possibility for these QTLs, at this point we cannot distinguish between pleiotropy and tight linkage of different QTLs. However,

the sharing of some QTL regions could be interpreted as an indirect biological validation for these QTLs.

DISCUSSION

Linkage map construction. In this study, linkage maps of RAPD markers that segregated in the pseudo-testcross configuration were used to locate quantitative trait loci. These single-tree linkage maps had been previously constructed based on the cosegregation analysis of markers for 62 individuals. This sample size was increased for the QTL analysis reported here. Maps constructed *de novo* based on 122 individuals displayed a well conserved linear order of the framework marker loci. Counting each marker out of the original order as an event (e.g. a switch in order of two adjacent markers counts as one event), there were a total of 7 order changes out of 100 markers in *E. grandis* (7%) and 10 changes in 83 markers in *E. urophylla* (12%). Therefore, on the average, only about 10% of the markers had their orders changed when increasing the sample size almost two-fold (from 62 to 122) and most of the changes were relatively unimportant order switches of adjacent markers.

Keats *et al.* (1991) pointed out that the 1000:1 support threshold suggested as a guideline for building framework linkage maps was conservative and would only prove adequate from empirical studies. Our results show that even when a relatively limited sample size (n=62) is used to build linkage maps, the adoption of the 1000:1 support to include markers into a more likely framework order results in a robust map, whose order is essentially the same as though twice as many progeny were used. These results also indicate that a two-step approach would be adequate to optimize the extensive genotyping work necessary in QTL mapping experiments. In the first step, segregation data for a large number of markers (>250) would be gathered only for a subset (~60) of the mapping population. Preliminary framework maps with a 1000:1 support for order would be constructed for both parents. Then in a second step, an extended set of progeny would be genotyped only for a selected group of evenly spaced framework markers, followed by a finer search with all markers available in potential regions of interest.

QTL mapping. Similarity between regression analysis and interval mapping analysis in QTL mapping has been observed previously (e.g. Stuber *et al.* 1992; Doebley

and Stec 1991). In this study we compared two interval mapping methods. With the exception of three genomic regions, declared significant by one method and only close to the threshold by the other, both interval mapping methods yielded very similar results in terms of QTL detection. The significance thresholds adopted (LOD 1.6 and P=0.01) were generally comparable. This was not surprising considering that both methods use only slightly different algorithms. Darvasi *et al.* (1993) estimated numerically that at a marker spacing of 10 cM and 11 markers per chromosome, a LOD score 1.53 - corresponding to a per-marker type I error rate of 0.0084 - ensures a 0.05 per-chromosome type I error. With infinite number of markers, the LOD threshold would have to increase to 1.96. These estimates were obtained under a true backcross model. In our study, marker spacings along both maps were similar, 10.6 cM for *E. grandis* and 9.5 cM for *E. urophylla*. The number of marker per chromosome averaged 14 for *E. grandis* and 15 for *E. urophylla*. Therefore, the stringency adopted to declare a QTL in this study seems satisfactory. However, although the analyses were carried out under a backcross model, our system does not meet the assumption of a true backcross from inbred lines. It is not clear at this point how much this violation would affect the thresholds adopted for QTL detection.

Power for QTL detection was limited due to the small sample size available (Figure 3). In a true backcross, a sample size of at least 500 individuals would be necessary to achieve an average power of 0.64 for detecting a QTL with a standardized gene substitution effect d=0.25; for d=0.5 the power was always close to 1 (Darvasi *et al.* 1993). In spite of the small sample size, some gain in power was probably acheived in our experiment by using clonal replicates of the F1 individuals. Clonal replication essentially increases the heritability of the trait (Bradshaw and Foster 1992). Heritability was shown to play a crucial role in determining the magnitude of additive genetic variance at any QTL that can be detected as statistically significant (Lande and Thompson 1990). Knapp and Bridges (1990) argued that if all the additive genetic variance is accounted for by markers, an additional replication of a clone increases statistical power by an amount equivalent to adding another offspring genotype. Finally, Strauss *et al.* (1992) estimated that a sample size of 200 would be needed to detect half of the additive genetic variance at α=0.01 when the within-family trait heritability is 0.5 and five effective QTLs control the trait.

No estimates of heritability are available for the traits investigated. However we were able to estimate an upper limit to broad sense heritability by calculating

repeatability. In our experimental conditions repeatabilities were on the order of 0.4 to 0.6. Assuming that by clonal replication we increased our sample size from ~100 to a "virtual" sample size of 200, and using our estimates of repeatability as upper limit heritability, our experimental results agree with the theoretical arguments discussed above. Standardized gene substitution effects for the QTL detected in our study (Δ_1 Tables 2 and 4) were typically between 0.5 and 0.7 phenotypic standard deviations, and the smallest effect detected was 0.45. In other words, our mapping experiment was efficient to detect only relatively strong effects. The limited power was also reflected in the total proportion of phenotypic and genetic variation that could be accounted by the putative QTLs and appears to be commensurate with the estimated repeatability. For FWS, repeatability 0.59, 10 QTLs were detected accounting for 27.6% of the genetic variation; for %ROOT, repeatability 0.52, the four QTLs detected accounted for 17.2% while for #CUTT, repeatability 0.42, six QTLs accounted for 7.3%. At this point we could only speculate on the proportion of the genetic variation that is due to additive effects.

LOD 1.0 support intervals for QTL positions were typically around 30 to 50 cM. Discrepancies in the most likely QTL location within the marker interval were observed between the two interval mapping methods, although in 75% of the cases the most likely position was either in the same or in an adjacent interval. From simulation studies, Darvasi et al. (1993) concluded that the confidence intervals for QTL map location can be rather broad, in some cases essentially covering the whole chromosome, and relatively independent of marker density. For standardized gene substitution effects equal to 0.5, a sample size of 1000 individuals would be required to reach an 11 cM confidence interval. Taking these simulations into account and given the size of our experiment we would not expect to be able to precisely locate QTL beyond the level of assigning them to linkage groups. The QTL positions reported herein should be regarded as tentative. While the issue of precise QTL location would be crucial for map-based cloning efforts, it should not represent a significant obstacle for marker assisted breeding. The extreme markers bracketing the 1.0 LOD support interval could be used for ensuring successful selection for the favorable QTL allele.

Pseudo-testcross QTL mapping. In the "pseudo-testcross" mapping strategy we explored the high levels of heterozygosity of outbred individuals and the efficiency of the RAPD assay to uncover large numbers of informative mating configuration of genetic markers (Grattapaglia and Sederoff 1994). In this study we have extended the

combined use of the "pseudo-testcross" mapping strategy and RAPD markers to map the first QTLs for species of *Eucalyptus*. To our knowledge this is the first report of quantitative trait locus mapping using RAPD markers in trees, and the first attempt to understand the genetic architecture of commercially important traits related vegetative propagation by rooted cuttings and micropropagation.

The pseudo-testcross strategy is based on the selection of single-dose markers present in one parent and absent in the other. As a result, the genetic linkage maps are individual-specific, and no RAPD markers are in common between the two maps. This same concept extends into QTL mapping. The QTLs mapped in this study are individual-specific. At this point we cannot establish homologies of linkage groups or homologies of QTLs in the two maps. Such homologies will await the localization of common RAPD marker loci on both maps. Conservation of RAPD markers and their linkage relationships in maps of different individuals will depend on the presence of the same RAPD marker loci and their allelic state. In a previous study in *E. grandis*, we have determined that approximately 33% of mapped RAPD markers were conserved across individuals of widely distinct origins, and approximately half of those also segregated (Grattapaglia and Sederoff 1994). Furthermore the conservation of favorable marker/QTL associations across individuals will depend essentially on the extent of linkage disequilibrium between marker and trait loci in the population (see below).

QTL mapping in crop plants has usually relied on the availability of inbred lines that frequently were chosen to differ specifically with respect to QTLs affecting the traits of interest. When these inbred lines are crossed to produce segregating F_2 or BC families, a large amount of linkage disequilibrium is generated and quantitative trait associations with alternative marker genotypes can be readily measured. In outbred populations such as those of domestic animals and most forest trees, QTL mapping strategies involving inbred pedigrees are generally not applicable due to a significant genetic load and time constraints. For most traits of interest, populations are generally polymorphic at both the QTL and marker loci, and the degree of linkage disequilibrium that can be generated by crossing populations is limited. As pointed out by Soller (1991), in such populations, mapping can be based on the disequilibrium necessarily found within individual families within a single population. The pseudo-testcross QTL mapping approach explores precisely this source of disequilibrium by virtue of the specific coupling relationships between marker alleles and QTL alleles in the parents of the family. In this study we relied on the existing linkage disequilibrium within an interspecific

full-sib family to identify genetic factors controlling traits related to vegetative propagation in *Eucalyptus*. Maximum likelihood methods for QTL mapping in full-sib families have been presented, targeting specifically those cases where several unrelated families with few individuals are available (Knott and Haley 1992). As pointed out by those authors, where family sizes are large enough it may be possible to use least squares based methods to find marker-QTL linkages within single pedigrees without the need to accumulate evidence on individual markers across pedigrees. With increasing numbers of full sibs linkage phase can be accurately determined and power of QTL detection increases substantially. Furthermore, additional genotype data from grandparents obtained from three-generation pedigrees provides information only on phase of marker linkage in the parents. It has, however, little or no impact on the increase in mean test statistic (QTL detection power) unless only small families are available. In *Eucalyptus* as well as most forest tree species large full-sib families are available or can be readily produced, while three-generation pedigrees are typically rare.

QTL mapping in the pseudo-testcross configuration relies on within-family linkage disequilibrium to establish marker/trait associations. Separate QTL analyses are carried out for each parent of the cross under the conventional backcross model. For this reason, in the pseudo-testcross QTL analysis, dominant RAPD markers provide essentially the same amount of information as co dominant RFLP with the obvious advantage of speed in marker data gathering. Evidently, as in a true testcross, also in the pseudo-testcross no intralocus interactions such as dominance can be estimated. However, when compared to the conventional backcross model for QTL mapping, the pseudo-testcross differs in two main aspects. First, because of the undomesticated nature of the species for which this strategy is attractive, no prior genetic information is available about the parental genotypes and no planned "construction" of QTL genotypes is possible. Therefore the only QTLs that can potentially be detected are those that are heterozygous in the parents and where the differential effect between the alternative QTL alleles is relatively large. Knott and Haley (1992) pointed out that in outbreeding populations under selection, even in the best situation, i.e. when a QTL has only two alleles, on average 50% of parents do not produce segregating gametes at this locus.

Secondly, because of the likely heterozygosity at any QTL locus in both parents, the quantitative value of alternative marker genotypes is measured as the effect of one allelic substitution averaged over the potentially two alternative alleles inherited from

the other parent. If the genotype of one parent at the QTL locus is Q_1Q_2, and for the other parent Q_3Q_4, the QTL analysis essentially tests the difference between the average trait value of ($Q_1Q_3 + Q_1Q_4$) versus ($Q_2Q_3 + Q_2Q_4$) in the first parent and ($Q_3Q_1 + Q_3Q_2$) versus ($Q_4Q_1 + Q_4Q_2$) in the second parent. Also, the specific intralocus interactions that might take place and affect the final phenotype cannot be taken into acccount in the analysis. Therefore more genetic "noise" is present in the system since the effect of the QTL allele substitution is measured against a genetically heterogeneous background both at the locus as well as at the rest of the genome. Although we cannot predict to what extent, this "noise" should introduce a certain level of bias in the estimates of the magnitude of QTL effect and adversely affect the power of QTL detection. The substantial differences between the multipont estimates of total phenotypic variation and the arithmetic sum of the proportions of each individual QTL might partly be a result of such bias due to nonadditive sources of genetic variation.

Van Eck *et al.* (1994) compared qualitative and quantitative analysis of a known locus controlling tuber shape in a cross between heterozygous potato clones. Their results indicate that in the pseudo-tescross QTL analysis (analysis on each parent separately) the effect of a QTL allele in one parent might go unnoticed due to the masking effect of a stronger allele at the locus, contributed by the other heterozygous parent. To alleviate this problem they recommend using markers that detect different polymorphisms in both parents such as fully classified RFLPs segregating 1:1:1:1. However their results also show that the pseudo-testcross analysis correctly detected the *a priori* known QTL. The highest significance peak (lowest P value) was found exactly at the marker more closely linked to the QTL in the maternal parent, corroborating the previously determined position by the qualitative analysis. For QTL mapping they used analysis of variance and a small sample size (n=50), both conditions prone to have limited QTL detection power, particulary when an allele of weaker effect is under scrutiny. Had they used a larger progeny size and interval mapping methods, the missed effect could probably have been detected. However analyzing large progenies with RFLP markers becomes a very time consuming task. Furthermore only a small subset of RFLP markers will be fully classified in any particular cross. The speed and ease of RAPD markers in the pseudo-testcross configuration allows undertaking much larger genotype analysis both in sample size and number of markers compensating for the lower information content per marker locus. Highly polymorphic microsatellite markers would evidently be very powerful in this respect.

A potential problem that may arise in the pseudo-testcross QTL analysis is when spurious linkages occur between markers in both maps due to sampling variation in the genotypic data. If one of the markers involved in such linkages happens to be linked to a QTL in one of the parents, it may potentially lead to falsely declaring a QTL on the map of the other parent. In our study we specifically tested for this possibility, by analyzing jointly the marker data sets of both parents. At a relaxed LOD threshold (LOD 3.0) six spurious linkages were detected. Only in one case, a spurious linkage (LOD 3.1) between marker X15_1600 (*E. urophylla*, group 4) and markers I4_1360 and X15_1050 (*E. grandis*, group 1) could have led to a potentially erroneous QTL detection. The marker in *E. urophylla* is linked to the strongest QTL detected for %ROOT. A LOD 1.3 for %ROOT was estimated for the marker bracket in *E. grandis*. Although it did not reach the significance threshold adopted (LOD 1.6), regions such as those could accidentally be interpreted as QTL. Typically, such spurious linkages can be identified easily by the low LOD score and by the fact that they involve only one or two localized markers in the group, while true linkages always involve several if not all markers in the group. To further verify the spurious nature of these marker linkages or potential spurious QTL detection, the markers involved from both maps can be removed and the linkage and QTL analyses performed again.

Marker assisted selection for vegetative propagation in Eucalyptus. In spite of the low power for QTL detection stemming from the heterogeneous nature of the QTL mapping design and the limited sample size, we were able to map a number of putative QTLs controlling significant proportions of the phenotypic variation in vegetative propagation traits including micropropagation response. QTL controlling somatic embryogenesis (Armstrong *et al.* 1992) and *in vitro* androgenesis (Cowen *et al.* 1992) have been identified in maize. Recently, RAPD markers linked to two known genes that control somatic embryogenesis in alfalfa have also been mapped (Yu and Pauls 1993). The traits investigated in these studies were found to be under the control of a few genomic regions with large effects. For example a set of five markers (three of them linked) explained 82% of the phenotypic variance for percentage immature embryos forming embryogenic callus (Armstrong *et al.* 1992). A larger number of genes appears to be involved in the control of micropropagation response in *Eucalyptus*, as ten regions were identified controlling an estimated 52.5% of the phenotypic variance.

Our results suggest that stump sprouting and adventitious rooting response are also under the control of a number of genes. There is also an indication of a basic difference in genetic architecture between sprouting and rooting response. A major effect QTL for rooting was identified, while all the QTLs mapped for stump sprouting were of relatively small effect. However, given the interspecific nature of our pedigree, comparisions of the genetic architecture of the two traits might not be entirely legitimate. For these traits, only ~ 30% of the phenotypic variation could be explained by all the mapped regions, and given the relatively low repeatabilities estimated in our experimental conditions only a small proportion of the total genetic variation could be accounted for. A larger sample size and a gain in accuracy by increasing the number of measurements of each individual would substantially increase power for QTL detection. Furthermore, unlike the controlled conditions achieved for *in vitro* propagation assessments, seasonal effects have long been known to have strong influence both on sprouting and rooting responses in *Eucalyptus* (Fazio 1964; Blake 1972; Cremer 1973). It will be interesting to compare these results with those of experimental runs carried out in different times or locations. In spite of all this, it was interesting to see that the great majority of the explained variation for %ROOT could be accounted for by the three genomic regions inherited from the *E. urophylla* parent to the F_1. These included a strong QTL controlling an estimated >20% of the phenotypic variation. We also found that the *E. grandis* parent was responsible for most of the inherited variation in the ability to form shoots both *in vitro* and by stump sprouting. These results agree with the prior information available on the general behavior at the species level regarding these traits. *E. grandis* is the most widely and intensely bred and planted species due to its rapid growth and extensive adabtability. *E. grandis* is considered to sprout and micropropagate well (Hartney 1980) and it is frequently employed in hybridization programs to improve such traits in species such as *E. nitens* and *E. deglupta* (Zobel 1993). On the other hand, hybridization to *E. urophylla* typically improves the ability of *E. grandis* to form roots on cuttings.

High levels of intra and interspecific variation in vegetative propagation response have been observed in *Eucalyptus*. It is fairly common that superior families or individuals in volume production are identified, but success with rooted cuttings from them is prohibitively low for production purposes (Van Wyk 1985; Campinhos and Ikemori 1980). Typically large number of superior families and individuals have to be generated so that high selection intensities can be applied to obtain productive

genotypes that can also be vegetatively propagated. The manipulation of environmental factors known to affect rootability of cuttings has been undertaken with variable success (reviewed by Hartney 1980). However, to move vegetative propagation technology to a production level, clone-specific requirement become impractical, and a rather robust protocol is necessary. To achieve this goal, genetic manipulation of vegetative propagation response through interspecific hybridization has been the method of choice. Molecular marker assited breeding for vegetative propagation traits would be a highly desirable way to track the inheritance and segregation of important genomic regions on an individual basis. This should substantially accelerate the introgression of these traits into breeding populations, as well as facilitate the indirect pre-screening of individual clones reducing the number of individuals that need to undergo the time consuming assessments of sprouting and rooting response.

The pseudo-testcross QTL mapping strategy involves the construction of genetic linkage maps of molecular markers and the identification of QTLs for individual genotypes. This approach is particularly attractive because it mitigates the obstacle of linkage equilibrium faced by marker assisted breeding in outbred species such as forest trees. As pointed out by many, with linkage equilibrium, marker-trait associations established in one cross, would not hold in a second pedigree, because marker and QTL alleles would be randomly associated at the population level (Soller 1978; Beckmann and Soller 1983; Lande and Thompson 1990; Strauss *et al.* 1992). In this study we have shown that the within-family linkage disequilibrium can be used to identify genomic regions controlling quantitatively inherited traits related to vegetative propagation in *Eucalyptus*. With this approach one can contemplate the prospect of performing QTL analysis in any available full-sib family generated from undomesticated and highly heterozygous organisms such as forest trees. Another relevant aspect of our work is that QTL mapping seems feasible on two-generation pedigrees of the kinds commonly available in breeding programs. This is particularly important because it opens the possibility of using already existing families in retrospective QTL analyses allowing one to gather the necessary quantitative data in acceptable time.

Although the pseudo-testcross QTL mapping information is generated on an individual basis, the progressive accumulation of individual linkage maps with subsets of common markers among them will make obvious the relationships of QTL in different maps. This will eventually lead to the identification of general regions associated with

trait expression. However, even though such regions might be found, we expect that a multiplicity of QTL controlling economically important traits exist at the population level. This should be particularly true for genetically heterogeneous species such as forest trees. For a similar phenotypic expression, different QTL's can act in different ways in different individuals, depending on the inherent genetic background and the kinds of selective pressures that the individuals have been subject to. Even in highly domesticated crops such as maize and tomato, some experimental evidences exist on the issue of inconsistency of QTL expression across populations suggesting that heterogeneity of QTL is a probable hypothesis (Beavis *et al.* 1991; Tanksley and Hewitt 1988). In the single-tree QTL appproach adopted in this study, marker/trait associations are established at the individual level, and therefore substantial linkage disequilibria is expected to be maintained. Close linkages established between markers and QTL could be followed for several subsequent generations of selection and recombination. To optimize QTL mapping in each individual tree, larger progeny sizes than the ones used in this study would be required in the initial detection step to improve power for detection. In subsequent generations, however, the number of markers genotyped could be substantially reduced as only those particular marker segments containing the QTL's of interest would be tracked. Progeny sizes could then vary depending on the number of genomic regions targeted at selection to increase the probability of recovering genotypes with the correct QTL allele profiles.

ACKNOWLEDGMENTS

LITERATURE CITED

Armstrong CL, Romero-Severson J, Hodges TK (1992) Improved tissue culture response of an elite maize inbred through backcross breeding, and identification of chromosomal regions important for regeneration by RFLP analysis. Theor Appl Genet 84:755-762.

Beavis WD, Grant D, Albertsen M, Fincher R (1991) Quantitative trait loci for plant height in four maize populations and their associations with qualitative genetic loci. Theor Appl Genet 83:141-145.

Beckmann JS, Soller M (1983) Restriction fragment length polymorphisms in genetic improvement methodologies, mapping and costs. Theor Appl Genet 67: 35-43.

Blake TJ (1972) Studies on the lignotubers of *Eucalyptus* obliqua L'herit III. The effects of seasonal and nutritional factors on dormant bud development. New Phytol 71: 327-334.

Bradshaw HD Jr., Foster GS (1992) Marker-aided selection and propagation systems in trees: advantages of cloning for studying quantitative inheritance. Can J For Res 22:1044-1049.

Brandao LG (1984) The new eucalypt forest. Marcus Wallenberg Symposium, Proc # 1, Falun Sweden pp 3-15.

Campinhos E, Ikemori Y (1980) Mass production of *Eucalyptus* spp. by rooting cuttings. In: IUFRO Symposium "Genetic Improvement and Productivity of Fast-growing Trees. Sao Paulo, Brazil 16 pp.

Cowen NM, Johnson CD, Armstrong K, Miller M, Woosley A, Pescitelli S, Skokut M, Belmar S, Petolino JF (1992) Mapping genes conditioning *in vitro* androgenesis in maize using RFLP analysis. Theor Appl Genet 84: 720-724.

Cremer KW (1973) Ability of *Eucalyptus* regnans and associated evergreen hardwoods to recover from cutting or complete defoliation in different seasons. Aust For Res 6: 9-22.

Darvasi A, Weinreb A, Minke V, Weller JI, Soller M (1993) Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics 134: 943-951.

de Fossard RA (1974) Tissue culture of *Eucalyptus*. Aust For 37: 43-54.

Delwaulle JC (1985) Plantations clonales d'*Eucalyptus* hybridas an Congo. Rev Bois Forets Trop 208: 37-42.

Doebley J, Stec A (1993) Inheritance of the morphological differences between maize and teosinte: comparision of results for two F_2 populations. Genetics 134: 559-570.

Dudley JW (1993) Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. Crop Sci 33:660-668.

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

Falconer DS (1989) Introduction to quantitative genetics. Longman Scientific and Technical 3rd ed.

Fazio S (1964) Propagating *Eucalyptus* from cuttings. Comb Proc Internat Plant Propagators Soc 14: 288-290.

Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross mapping strategy and RAPD markers. Genetics (in press).

Grattapaglia D, Assis TF, Caldas LS (1987) Residual effects of benzyladenine and naphtalenacetic acid on the "in vitro" multiplication and rooting of eight *Eucalyptus* species. In: Abstracts of the Second National Symposium on Plant Tissue Culture. Brasilia, DF, Brazil. Abstr. # 12.

Grattapaglia D, Caldas LS, Machado MA, Assis TF (1990) Large scale micropropagation of *Eucalyptus* species and hybrids. In: Abstracts of the Seventh International Congress on Plant Tissue and Cell Culture. Amsterdam. Abstr. # A3-123.

Gupta PK, Mascarenhas AF (1987) *Eucalyptus*. In: Bonga JM, Durzan DJ (eds) Cell and tissue culture in forestry Vol. 3 pp 385-399.

Hartman HT, Kester D (1983) Plant propagation: principles and practices. Second edition Prentice Hall, Englewood Cliffs NJ.

Hartney VJ (1980) Vegetative propagation in the eucalypts. Aust For Res 10: 191-211.

Keats BJS, Sherman SL, Morton NE, Robson EB, Buetow KH, Cann HM, Cartwright PE, Chakravarti A, Francke U, Green PP, Ott J (1991) Guidelines for human linkage maps: an international system for human linkage maps (ISLM, 1990). genomics 9: 557-560.

Knapp SJ, Bridges WC (1990) Using molecular markers to estimate quantitative genetic parameters: power and genetic variances for unreplicated and replicated progeny. Genetics 126: 769-777.

Knapp SJ, Bridges WC, Liu BH (1992) Mapping quantitative trait loci using nonsimultaneous and simultaneous estimators and hypothesis tests. In: Beckmann J, Osborn TC (eds) Plant genomes: Methods for Genetic and Physical Mapping. Kluwer Academic Publishers, pp 209-238.

Knott SA, Haley CS (1992) Maximum likelihood mapping of quantitative trait loci using full-sib families. Genetics 132: 1211-1222.

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124: 743-756.

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185-199.

Lander, ES, Green P, Abrahamson J, Baarlow A, Daly MJ, Lincoln SE, Newburg L (1987) MapMaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174-181.

Liu BH, Knapp SJ (1990) GMENDEL: a program for Mendelian segregation and linkage analysis of individual or multiple progeny using log-likelihood ratios. J Hered 81: 407.

Murashige T, Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol Plant 15: 473-497.

SAS (1988) SAS/STAT User's Guide, version 6.03. SAS Institute, Cary, NC.

Soller M (1978) The use of loci associated with quantitative traits in dairy cattle improvement. Anim Prod 27: 133-139.

Soller M (1991) Mapping quantitative trait loci affecting traits of economic importance in animal populations using molecular markers. In: Schook LB, Lewin HA, Mclaren DG (eds) Gene-mapping techniques and applications. Marcel Dekker Inc. pp 21-50.

Strauss SH, Lande R, Namkoong G (1992) Limitations of molecular-marker aided selection in forest tree breeding. Can J For Res 22: 1050-1061.

Stuber, CW (1992) Biochemical and molecular markers in plant breeding. In: Dudley JW, Hallauer AR, Ryder M (eds) Plant breeding reviews Vol. 9. John Wiley & Sons, Inc., pp 37-61.

Stuber CW, Lincoln SE, Wolff DW, Helentjaris T, Lander ES (1992) Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics 132:823-839.

Tanksley SD, Hewitt JD (1988) Use of molecular markers in breeding for soluble solids in tomato - a re-examination. Theor Appl Genet 75:811-823.

Van Eck HJ, Jacobs JME, Stam P, Ton J, Stiekema WJ, Jacobsen E (1994) Multiple alleles for tuber shape in dipoid potato detected by qualitative and quantitative genetic analysis using RFLPs. genetics 137: 303-309.

Van Wyk G (1985) Tree breeding in support of vegetative propagation of *Eucalyptus grandis* (Hill) Maiden. S Afr For J 12:33-40.

Table 1. QTL summary for *Eucalyptus grandis* clone 44 as determined by interval mapping analysis using MAPMAKER-QTL. Listed are the locations and magnitudes of effect of QTL controlling traits related to vegetative propagation response.

TRAIT [a]	LINKAGE GROUP	MARKER INTERVAL	QTL [b] POSITION	LOD PEAK	1.0 LOD SUPPORT [c] INTERVAL	% VAR.[d] EXPL.
FWS	1	P10_1250 -J5_510	0.0	2.1	OFF END - P10_1250 - 8.0	9.6
	1	I4_1360 - Z3_830	0.0	2.3	16.0 - I4_1360 - Z3_830	9.1
	3	R20_1650 - U19_550	8.0	1.3 *	8.0 - R20_1650 - J7_560 - 8.0	6.0
	7	N15_1079 - U10_1500	0.0	1.9	6.0 - X12_1250 - OFF END	7.5
	10	N12_207 - J17_815	18.0	2.2	OFF END - J17_815 - 12.0	11.7
	8/12	N6_634 - Y16_550	4.0	2.7	16.0 - N6_534 - Y16_550	12.3
#CUTT	1	R9_670 - U16_930	0.0	1.5 *	4.0 - J5_887 - Z3_830 - 4.0	7.4
	2	D3_1746 - K10_913	0.0	1.6	20.0 - X2_1200 - D3_1746 - 14.0	7.3
	8/12	K9_534 - Y3_1350	0.0	1.7	10 .0 - R3_1200 - N6_634 - 14.0	7.8
	13/11	N13_2169 - P19_460	0.0	2.4	14.0 - X12_530 - P19_460 - 24.0	10.8
%ROOT	5	K12_954 - K15_567	0.0	1.8	4.0 - G5_1898 - K15_567 - 2.0	8.5

[a] FWS, fresh weight of micropropagated shoot clumps; #CUTT, number of operational stump sprout cuttings; %ROOT, percent rooting of cuttings

[b] Most likely QTL position corresponding to LOD peak, as estimated by MAPMAKER-QTL; cM distance from leftmost marker of interval

[c] Interval over which the position of the QTL is at most 10 times less likely than the most likley position estimated by MAPMAKER-QTL; from left to right: cM distance from the left , marker segment and cM distance to the right; OFF END= off the end of linkage group.

[d] Percent of the phenotypic variation explained, as estimated by MAPMAKER-QTL

* Significant in the QTLSTAT analysis (P≤ 0.01)

Table 2. QTL summary for *Eucalyptus grandis* clone 44 as determined by least squares interval mapping analysis using QTLSTAT. Listed are the locations and properties of QTL controlling traits related to vegetative propagation response.

TRAIT [a]	LINKAGE GROUP	MARKER INTERVAL	P VALUE	WALD [b] STATISTICS	R^2 [c]	GENOTYPE MEANS ±S.D. [d] (+)	(-)	Δ_1 [e]	Δ_2 [f]
FWS	1	K10_1106 - Y20_620	0.0001	15.7	0.17	269.8 ±25.4	438.5 ±23.8	2.10	1.02
	1	X15_1050 - I4_1360	0.002	10.1	0.10	329.1 ±13.1	374.4 ±14.3	0.57	0.22
	3	R20_1650 - U19_550	0.003	9.2	0.09	379.3 ±14.4	335.1 ±13.4	0.55	0.28
	7	X12_1250 - N15_1079	0.0007	11.9	0.11	376.8 ±12.4	334.1 ±14.7	0.54	0.25
	10	N12_207 - J17_815	0.003	8.6	0.13	378.1 ±12.9	334.8 ±14.6	0.55	0.26
	8/12	Y16_550 - Y17_515	0.0002	14.1	0.11	383.4 ±14.4	335.9 ±12.6	0.60	0.33
#CUTT	1	J5_887 - R9_670	0.009	7.1	0.13	11.4 ±1.0	14.2 ±1.0	0.55	0.32
	2	M10_800 - X2_1200	0.01	6.5	0.11	13.4 ±1.3	11.1 ±1.2	0.45	0.15
	8/12	R3_1200 - Z18_1630	0.001	11.0	0.24	14.4 ±1.0	10.9 ±1.0	0.70	0.36
	13/11	X12_530 - N13_2169	0.003	9.5	0.14	13.2 ±1.0	10.0 ±1.0	0.63	0.12
%ROOT	5	K12_954 - K15_567	0.005	8.3	0.10	27.6 ±5.2	41.4 ±4.4	0.61	0.27

[a] FWS, fresh weight of micropropagated shoot clumps; #CUTT, number of operational stump sprout cuttings; %ROTT, percent rooting of cuttings

[b] A Wald statistics of 10.0 is approximately equal to P = 0.001

[c] Percent of the phenotypic variation explained, estimated as the non-linear regression R^2 (SS_{marker}/SS_{total}) using QTLSTAT

[d] Estimates of genotype means for the alternative RAPD marker-linked QTL alleles; (+) presence of the RAPD band; (-) absence of the RAPD band

[e] Difference between alternative QTL genotypes expressed in phenotypic standard deviations

[f] Difference between the favorable QTL genotype and the population mean expressed in phenotypic standard deviations

* Significant in the MAPMAKER-QTL analysis (LOD ≥ 1.6)

Table 3. QTL summary for *Eucalyptus urophylla* clone 28, as determined by interval mapping analysis using MAPMAKER-QTL. Listed are the locations and magnitudes of effect of QTL controlling traits related to vegetative propagation response.

TRAIT [a]	LINKAGE GROUP	MARKER INTERVAL	QTL[b] POSITION	LOD PEAK	1.0 LOD SUPPORT [c] INTERVAL	% VAR.[d] EXPL.
FWS	1	Q13_980 - X15_700	8.0	1.6	18.0 - Q13_980 - OFF END	6.8
	2	K19_1780 - L17_560	4.0	2.6	2.0 -G2_584 - A11_1439	10.2
	5	U16_2080 - U19_1450	0.0	2.0	16.0 - U16_2080 - K15_500 - 6.0	8.0
	6	X12_340 - W9_950	2.0	1.6	2.0 - Z16_344 - OFF END	10.7
#CUTT	1	W20_530 - K9_941	10.0	1.7	10.0 - W20_530 - U7_1100 - 4.0	8.9
	2	A11_1439-X12_740	0.0	2.0	4.0 - B6_810 - X12_740	9.2
%ROOT	4	R8_570 - X15_1600	14.0	5.8	OFF END - X15_1600 - 20.0	21.0
	8	K3_290 - N15_485	0.0	1.9	2.0 - U13_450 - K3_290 - 18.0	8.6
	9	G5_451 - W11_650	0.0	1.7	W12_1500 - OFF END	7.3

[a] FWS, fresh weight of micropropagated shoot clumps; #CUTT, number of operational stump sprout cuttings; %ROOT, percent rooting of cuttings

[b] Most likely QTL position corresponding to LOD peak, as estimated by MAPMAKER-QTL; cM distance from leftmost marker of interval

[c] Interval over which the position of the QTL is at most 10 times less likely than the most likley position estimated by MAPMAKER-QTL; from left to right: cM distance from the left , marker segment and cM distance to the right; OFF END= off the end of linkage group.

[d] Percent of the phenotypic variation explained, as estimated by MAPMAKER-QTL

Table 4. QTL summary for *Eucalyptus urophylla* clone 28 as determined by least squares interval mapping analysis using QTLSTAT. Listed are the locations and properties of QTL controlling traits related to vegetative propagation response.

TRAIT[a]	LINKAGE GROUP	MARKER INTERVAL	P VALUE	WALD [b] STATISTICS	R^2 [c]	GENOTYPE MEANS ±S.D [d] (+)	(-)	Δ_1 [e]	Δ_2 [f]
FWS	1	X15_700 - M4_1228	0.014 *	6.0	0.09	341.2 ±13.4	377.8 ±13.6	0.46	0.26
	2	K19_1780 - L17_560	0.0005	12.7	0.11	382.6 ±13.4	329.2 ±13.6	0.67	0.32
	5	K15_500 -U20_420	0.0006	12.4	0.18	442.3 ±24.3	291.0 ±20.2	1.90	1.07
	6	X12_340 - W9_950	0.004	8.6	0.15	340.4 ±13.0	385.8 ±15.3	0.57	0.36
#CUTT	1	W20_530 - K9_941	0.01	6.5	0.08	11.3 ±1.2	13.9 ±1.0	0.52	0.26
	2	X7_1750 - J7_452	0.001	11.3	0.38	16.8 ±1.7	8.8 ±1.5	1.58	0.83
%ROOT	4	R8_570 - X15_1600	0.0000	25.4	0.28	48.3 ±4.2	23.4 ±4.7	1.09	0.58
	8	K3_290 - N15_485	0.007	7.4	0.11	40.8 ±4.5	28.7 ±5.0	0.53	0.25
	9	Y13_390 - G5_451	0.004	8.8	0.10	42.4 ±4.4	29.0 ±4.9	0.59	0.32

[a] FWS, fresh weight of micropropagated shoot clumps; #CUTT, number of operational stump sprout cuttings; %ROOT, percent rooting of cuttings

[b] A Wald statistics of 10.0 is approximately equal to P = 0.001

[c] Percent of the phenotypic variation explained, estimated as the non-linear regression R^2 (SS_{marker}/SS_{total}) using QTLSTAT

[d] Estimates of genotype means for the alternative RAPD marker-linked QTL alleles; (+) presence of the RAPD band; (-) absence of the RAPD band

[e] Difference between alternative QTL genotypes expressed in phenotypic standard deviations

[f] Difference between the favorable QTL genotype and the population mean expressed in phenotypic standard deviations

* Significant in the MAPMAKER-QTL analysis (LOD ≥ 1.6)

Table 5. Summary of the pseudo-testcross QTL analysis for traits related to vegetative propagation response: multipoint estimates of the % phenotypic variation explained by the mapped QTL (obtained by interval mapping using MAPMAKER-QTL and linear models using PROC GLM (SAS)), repeatabilities and % genetic variation explained.

	FWS		#CUTT		%ROOT	
	E. grandis	*E. urophylla*	*E. grandis*	*E. urophylla*	*E. grandis*	*E. urophylla*
# PUTATIVE QTL MAPPED	6	4	4	2	1	3
% PHENOTYPIC VARIATION Multipoint MapMaker-QTL	41.6	25.2	22.9	14.7	8.5	26.3
% PHENOTYPIC VARIATION Multipoint PROC GLM (SAS)	32.2	15.4	13.5	4.2	7.9	28.5
TOTAL % PHENOTYPIC VARIATION Multipoint MapMaker-QTL		52.5		28.2		32.6
TOTAL % PHENOTYPIC VARIATION Multipoint PROC GLM (SAS)		46.8		17.4		33.0
REPEATABILITY [a]		0.59		0.42		0.52
% GENETIC VARIATION [b]		31.0		11.8		17.0

[a] Estimated as the ratio between the variance between individuals and the total variance.
[b] Estimated as the product between the total phenotypic variation estimated by multipoint interval mapping (MMQ) and the trait repeatability

Figure 1. Quantitative Trait Locus (QTL) map of traits related to vegetative propagation response in *Eucalyptus grandis* clone 44. Linkage maps of RAPD markers were constructed using MAPMAKER (LOD 5.0 θ=0.30) and GMENDEL (P value=0.0001 θ=0.30). RAPD markers in bold type were classified as framework markers (ordered with log-likelihood support ≥ 1000:1) while the remaining markers were ordered with support ≥ 100:1. Bars to the left of linkage groups correspond to the 1.0 LOD support intervals for the location of the QTL (i.e. the interval over which the QTL position is at most 10 times less likely than the most likely position). Arrows indicate the most likely position (highest LOD peak) estimated with MAPMAKER-QTL.

GROUP 1

- P10_1250
- J5_510
- X15_850
- Z19_510
- R7_480
- V7_400
- K10_1106
- Y20_620
- J5_887
- R9_670
- U16_930
- Y3_470
- X15_1050
- I4_1360
- Z3_830
- V7_450
- R13_480
- K8_2000

GROUP 2

- U19_500
- U20_1000
- L5_711
- P8_1350
- G12_591
- X12_620
- L17_533
- K10_549
- J17_3335
- R15_1650
- M10_800
- X2_1200
- K10_835
- D3_1746
- K10_913

GROUP 3

- K4_230
- Y20_760
- R8_1100
- R20_1650
- U19_550
- K16_727
- N20_1850
- M5_600
- Y15_1220
- J7_560
- K19_304
- J1_924
- V7_770
- R15_210
- G13_1160
- X4_1850
- P11_1800

GROUP 4

- U19_800
- R20_1080
- R20_2700
- X1_720
- K19_1498
- F4_377
- K9_884
- I4_915
- Y13_750
- N15_629

GROUP 5

- M12_759
- R13_430
- X1_1450
- X7_580
- W9_1830
- A2_1691
- X15_600
- M10_1300
- M6_981
- Y16_1400
- Y17_1500
- H7_2350
- K4_480
- A20_1068
- U15_600
- B6_780
- T12_220
- G5_1898
- K12_954
- K15_567
- G12_1231
- P11_1450
- Y14_1400
- A10_635
- I4_1800

GROUP 6

- U20_900
- Z18_900
- J5_1584
- P10_650
- R16_730
- B6_515
- U20_850
- X7_1100
- W12_650
- Y17_280
- N7_1322
- M5_961
- Y14_2100
- A11_2374
- P10_2600
- B7_1275
- R3_330
- L16_389
- W12_1350
- F4_472

GROUP 7

- K9_1660
- A11_920
- A11_980
- H7_1100
- X17_620
- N13_833
- X12_1250
- N15_1079
- U10_1500
- N6_420
- Z19_1300

GROUP 8/12

- N20_1750
- D8_1990
- U13_1650
- R3_1200
- Z18_1630
- V2_1050
- N13_3121
- R16_820
- J17_332
- K9_534
- Y3_1350
- J17_285
- N6_634
- Y16_550
- Y17_515
- U19_260
- R7_570

GROUP 9/11a

- L16_285
- T8_525
- R7_750
- X9_650
- D3_748
- P10_530
- N13_540
- Y15_650
- Z7_1580
- Y15_740
- J13_1370
- Y20_550

GROUP 10

- N12_207
- J17_815
- R9_1300

GROUP 13/11b

- Q13_540
- Q14_970
- R3_1030
- W20_840
- X12_630
- N13_2169
- P19_460
- B6_759
- N13_580
- Q13_1150
- G13_1570

GROUP 14

- Y17_1300
- A19_682
- R20_1150
- P2_700
- K12_480
- Y9_360

Scale (left of GROUP 7): 20, 40, 60, 80, 100, 120, 140, 160, 180, 200 — **200 cM**

Legend:

FRESH WEIGHT OF MICROPROPAGATED SHOOTS (FWS)

\# STUMP SPROUT CUTTINGS (#CUTT)

% ROOTING OF CUTTINGS (%ROOT)

Figure 2. Quantitative Trait Locus (QTL) map of traits related to vegetative propagation response in *Eucalyptus urophylla* clone 28. Linkage maps of RAPD markers were constructed using MAPMAKER (LOD 5.0 θ=0.30) and GMENDEL (P value=0.0001 θ=0.30). RAPD markers in bold type were classified as framework markers (ordered with log-likelihood support ≥ 1000:1) while the remaining markers were ordered with support ≥ 100:1. Bars to the left of linkage groups correspond to the 1.0 LOD support intervals for the location of the QTL (i.e. the interval over which the QTL position is at most 10 times less likely than the most likely position). Arrows indicate the most likely position (highest LOD peak) estimated with MAPMAKER-QTL.
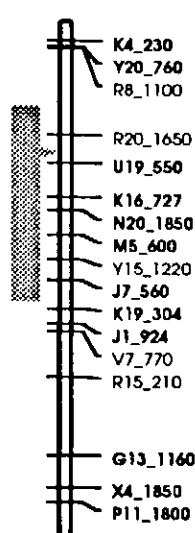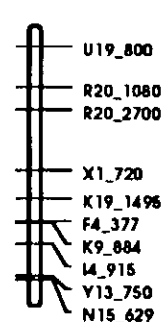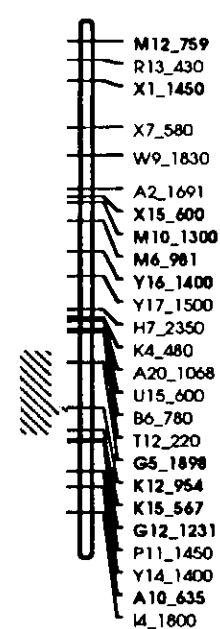
GROUP 1
- I4_340
- J7_808
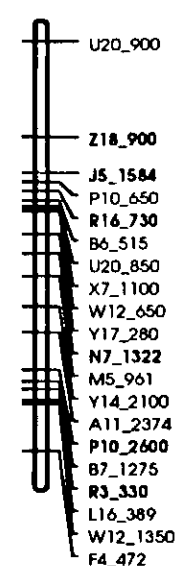- W20_530
- K9_941
- P8_970
- R20_370
- K9_1500
- U7_1100
- Q13_980
- X15_700
- M4_1228

GROUP 2
- J11_1450
- N8_1229
- R3_280
- N14_515
- H7_2100
- G2_584
- G2_519
- K19_1780
- L17_560
- J11_420
- X9_520
- N6_945
- X7_1750
- J7_452
- J13_477
- U20_290
- J5_490
- B6_810
- A11_1439
- X12_740
- A19_2448
- A20_2127
- N14_779
- T12_370
- D8_1342
- R8_700

GROUP 3
- N8_626
- Q14_780
- L16_613
- X3_850
- X7_700
- M12_483
- B7_1549
- A18_509
- X12_600
- R13_1120
- N7_983
- K14_495
- X15_396

GROUP 4
- R8_570
- X15_1600
- U15_210
- X2_1400
- Q1_480
- A10_1673
- R15_870
- P19_630
- W20_660
- K16_610

GROUP 5
- U16_1070
- R8_1050
- P13_310
- P13_1060
- T12_250
- R7_1450
- U16_2080
- U19_1450
- D8_487
- L5_368
- Y20_390
- R16_870
- F4_1442
- Z18_370
- K15_500
- U20_420
- R8_420
- U13_320
- M10_340
- X4_300
- X2_420
- T12_1500
- Z16_1480

GROUP 6
- X4_680
- X12_1600
- R7_450
- K20_535
- U13_870
- Z3_550
- A11_635
- D3_508
- Y15_600
- Z3_280
- N12_976
- Y13_1700
- Y3_1900
- L17_656
- N20_980
- J11_1315
- R16_1600
- Z16_344
- X9_850
- X4_1000
- X12_340
- W9_950
- W12_780

GROUP 7
- W11_1700
- G14_1710
- L5_506
- L8_1635
- U15_1570
- W11_1280
- Z19_760
- U7_800
- M12_612
- W9_1450
- K14_1078
- Y20_1200
- A11_1040
- A2_985
- Y9_320
- P19_550
- U15_520
- U15_870
- I4_704
- G14_927
- G14_507
- K12_1575
- U19_1590

GROUP 8
- B6_1581
- R15_480
- X17_800
- Z6_520
- W11_820
- U13_450
- K3_290
- N15_485
- U20_450
- V2_1300
- N8_412

GROUP 9
- X17_2100
- A2_461
- N15_1541
- Q1_1100
- W12_1500
- Y13_390
- G5_451
- W11_650
- X3_1250

GROUP 10
- N14_1028
- R3_1750
- N20_1050
- N13_397
- N12_1229
- M4_1027

GROUP 11
- K14_1329
- K10_339
- Y9_1050
- G5_304
- G2_1444
- K15_1400
- M4_906
- R8_460
- N8_440
- K12_631
- P8_570

Scale:
20
40
60
80
100 **200 cM**
120
140
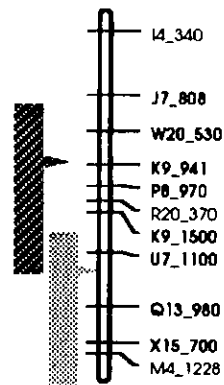160
180
200

Legend:
- FRESH WEIGHT OF MICROPROPAGATED SHOOTS (FWS)
- # STUMP SPROUT CUTTINGS (#CUTT)
- % ROOTING OF CUTTINGS (%ROOT)

Figure 3. Frequency distributions for vegetative propagation traits in the interspecific F1 family used for QTL mapping. (A) fresh weight of micropropagated shoots (FWS) in mg; (A) number of stump sprout cuttings (#CUTT); (C) percent rooted cuttings (%ROOT). Mean (x), standard deviation (σ) and sample size used in the QTL analysis are indicated besides the histograms.

A

x = 357.1
σ = 79.7
n = 118

NUMBER OF INDIVIDUALS

180 220 260 300 340 380 420 460 500 540

FRESH WEIGHT OF MICROPROPAGATED SHOOTS - FWS (mg)

B

x = 12.6
σ = 5.1
n = 97

NUMBER OF INDIVIDUALS

0 2 4 6 8 10 12 14 16 18 20 22 24

 # STUMP SPROUT CUTTINGS - #CUTT

C

x = 35.1
σ = 22.8
n = 96

NUMBER OF INDIVIDUALS

0 10 20 30 40 50 60 70 80 90 100

%ROOTING OF CUTTINGS - %ROOT
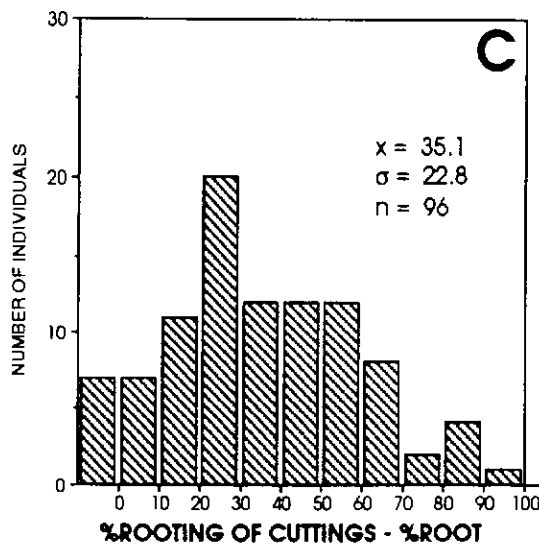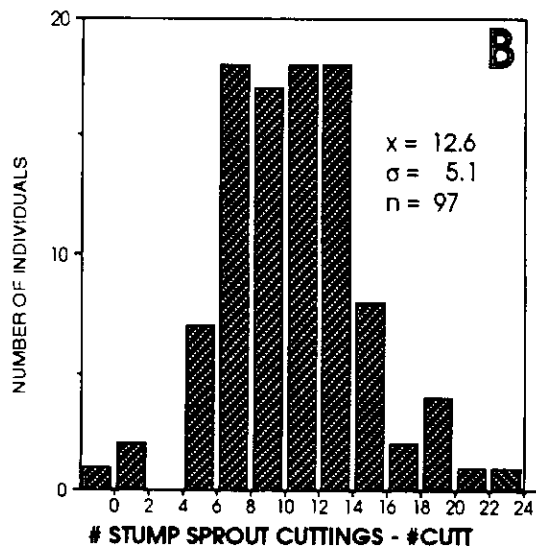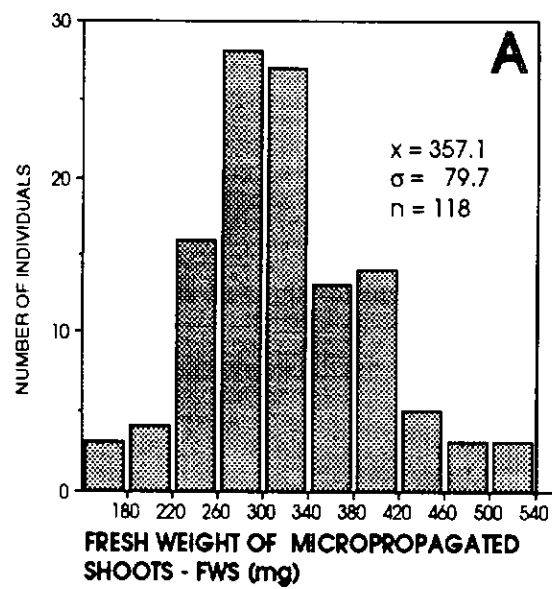
Figure 4. Segregation of RAPD marker R8_570 from *Eucalyptus urophylla* clone 28. R8_570 was found to cosegregate in coupling with a QTL controlling % rooting of cuttings (LOD 5.8, P=0.000 - Tables 3 and 4). Single-locus estimates of the proportion of the phenotypic variance explained by this QTL were 21% (MAPMAKER-QTL) and 28% (QTLSTAT). Last lane in top panel and first lane in bottom panel are 1 Kb ladder size standards. From left to right, top panel shows the RAPD profile for 31 F1 progeny; bottom panel shows an additional 29 progeny and the two parents, *E. grandis* clone 44 and *E. urophylla* clone 28. Arrows indicate the segregating marker.
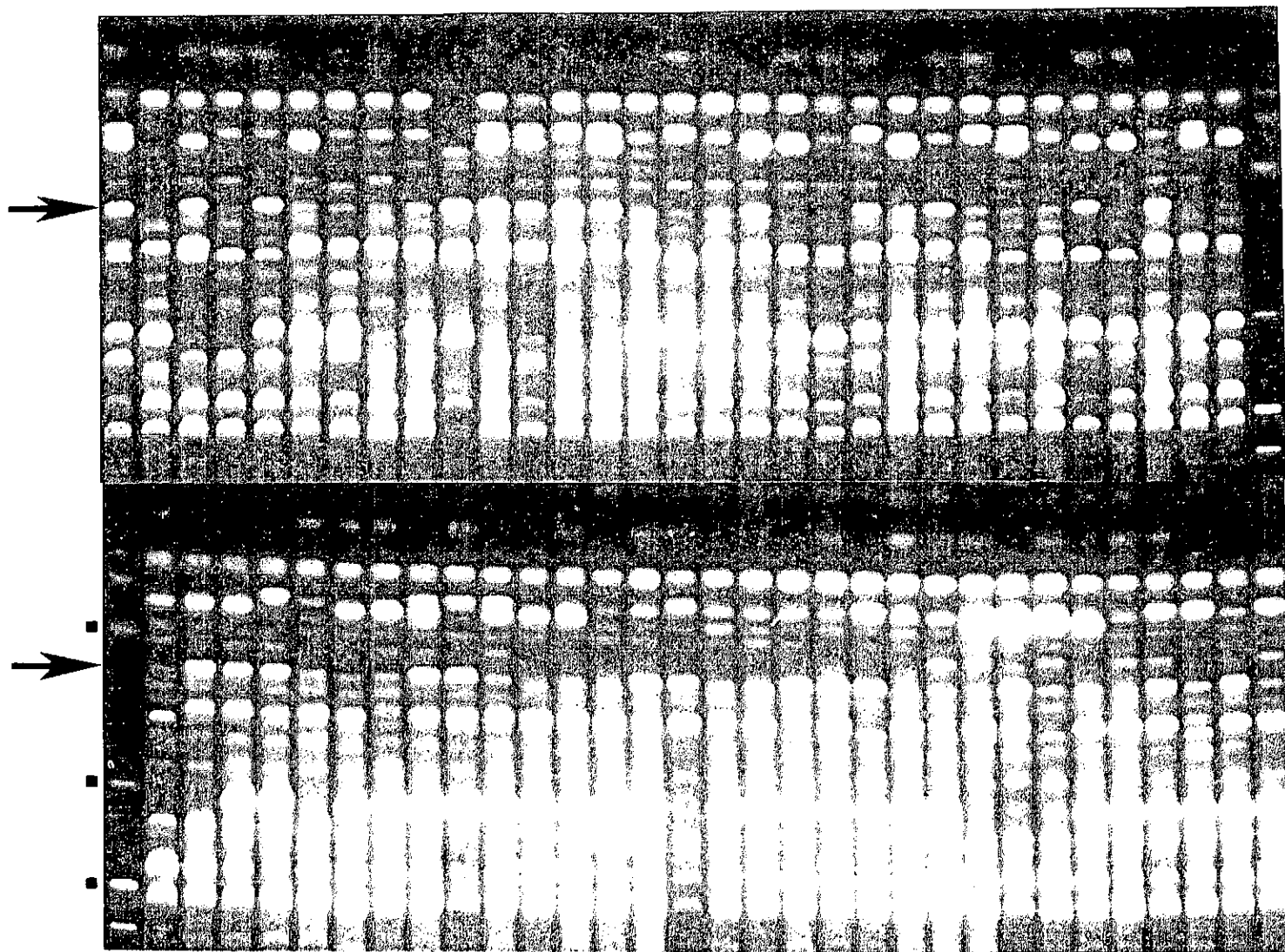
Figure 5. Segregation of RAPD marker N13_2169 from *Eucalyptus grandis* clone 44.
N13_2169 was found to cosegregate in coupling with a QTL controlling number of stump sprout cuttings (LOD 2.4, p= 0.003, Tables 1 and 2). Single-locus estimates of the proportion of the phenotypic variance explained by this QTL were 10.8% (MAPMAKER-QTL) and 14% (QTLSTAT). Last lane in top panel and first lane in bottom panel are 1 Kb ladder size standards. From left to right, top panel shows the RAPD profile for 30 F1 progeny; bottom panel shows an additional 28 progeny and the two parents, *E. urophylla* clone 28 and *E. grandis* clone 44. Arrows indicate the segregating marker.
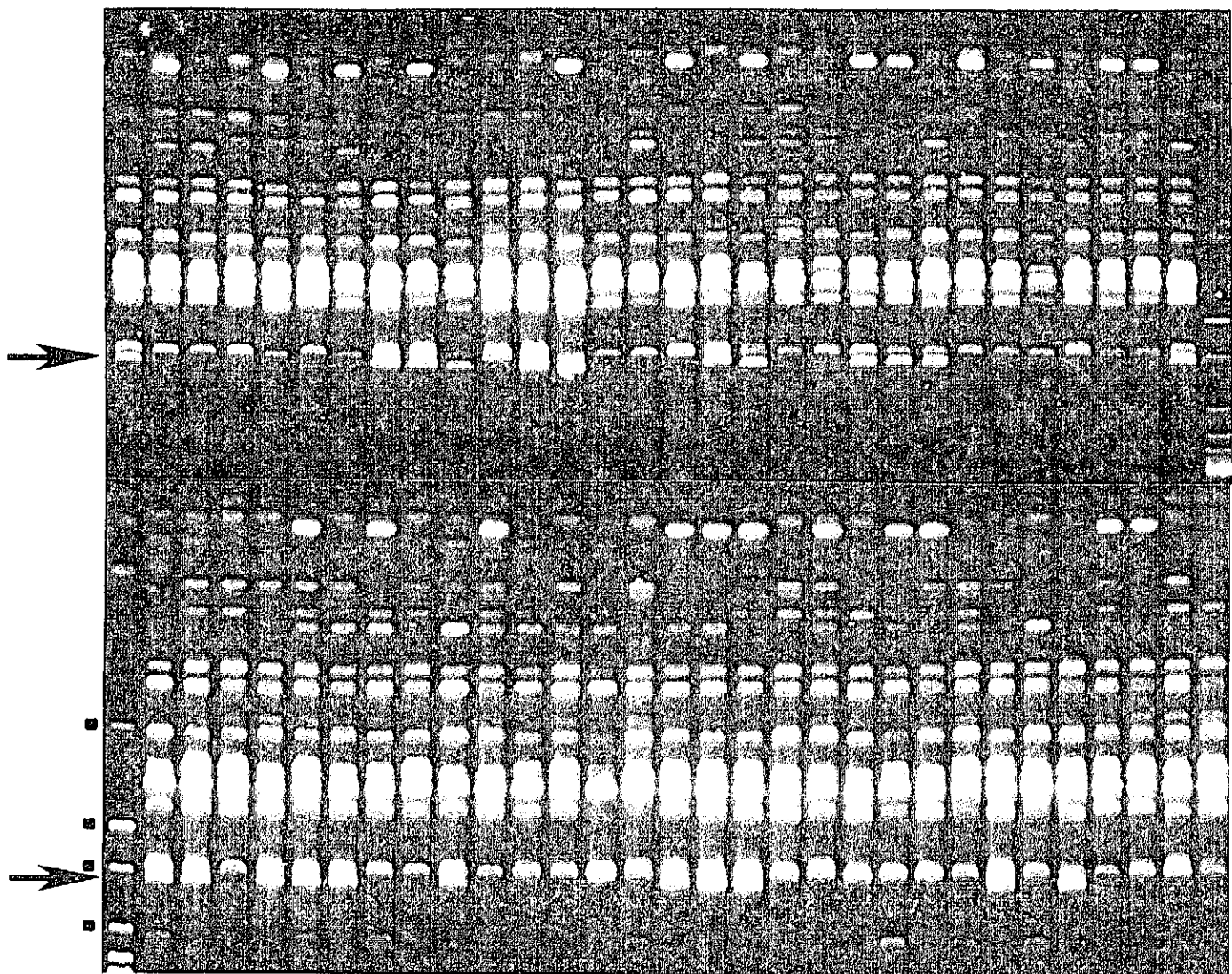
Figure 6. Segregation of RAPD marker Y16_550 from *Eucalyptus grandis* clone 44. Y16_550 was found to be linked in coupling (θ= 0.15) with a QTL controlling fresh weight of micropropagated shoots (FWS) (LOD 2.7, p=0.0002, Tables 1 and 2). Single-locus estimates of the proportion of the phenotypic variance explained by this QTL were 12.3% (MAPMAKER-QTL) and 11% (QTLSTAT). Last lane in top panel and first lane in bottom panel are 1 Kb ladder size standards. From left to right, top panel shows the RAPD profile for 30 F1 progeny; bottom panel shows an additional 28 progeny and the two parents, *E. urophylla* clone 28 and *E. grandis* clone 44. Arrows indicate the segregating marker.
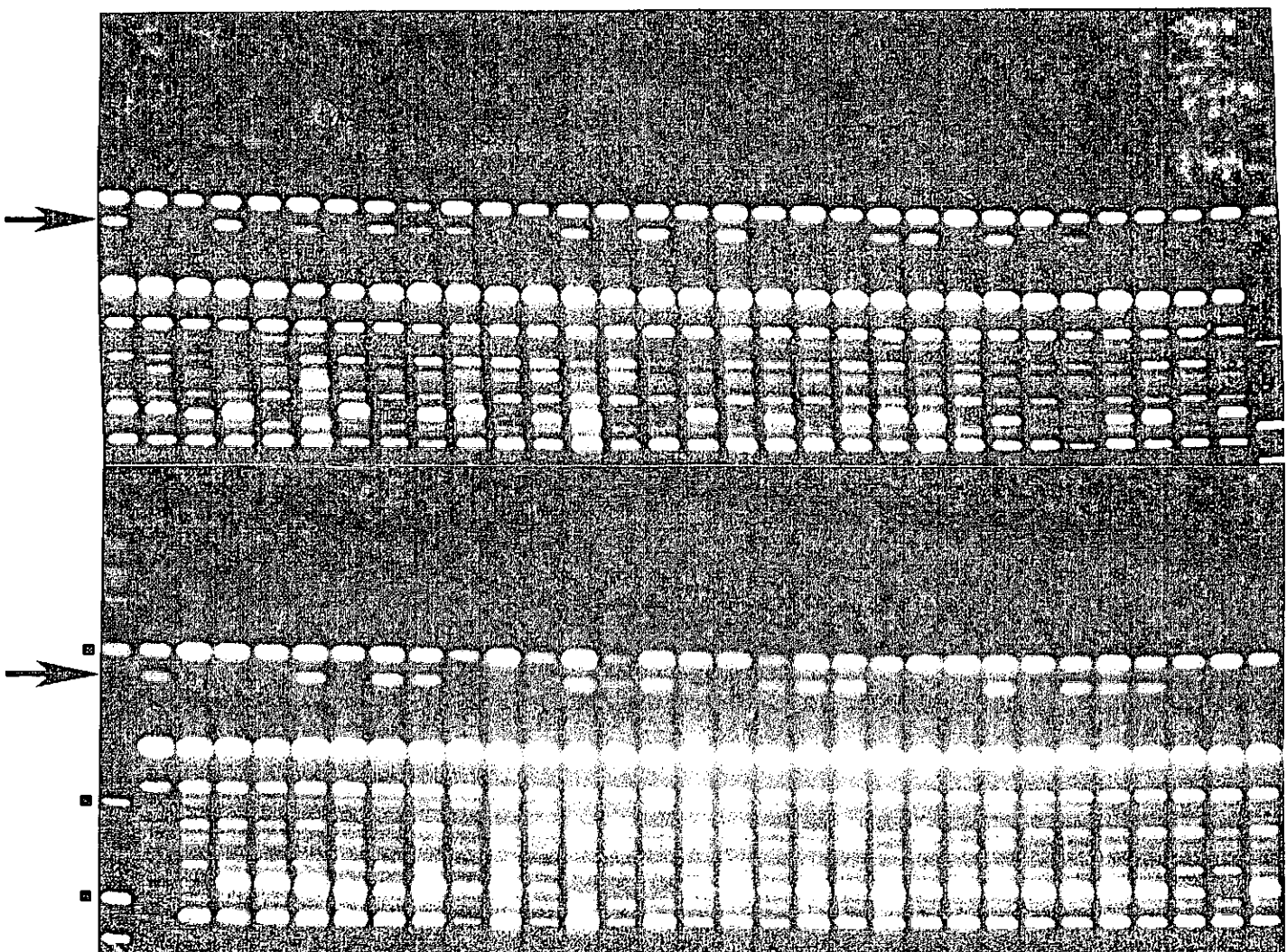
ABSTRACT

We performed a retrospective QTL mapping study, using a maternal open pollinated half-sib family of *Eucalyptus grandis* at harvest age. QTL mapping within-half sib families explores the within family linkage disequilibrium and requires that the common parent tree be heterozygous both at the quantitative trait locus and linked marker locus. Traits of economical importance in forest productivity were investigated: (1) circumference at breast height (CBH); (2) wood specific gravity (WSG); (3) percent dry weight of bark (%BARK) and (4) cellulose pulp yield (%PULP). A sequential QTL mapping approach was applied to CBH using a total sample of 1085 half-sibs. Multiple bulks segregant analysis was efficient to protect against false positives but power for detection was low. Still, even in a heterogeneous genetic background and for a trait of low heritability, the BSA approach was successful in detecting a QTL later confirmed by interval mapping. Selective genotyping detected two out of the three QTLs detected by full scale QTL mapping. Selective genotyping in a half-sib is complicated by the unknown contribution of QTL alleles from the pollen pool. Non-simultaneous QTL mapping on a sample of 300 half-sibs was carried out by linear regression and two interval mapping methods. A total of 12 QTLs were detected controlling an estimated (multipoint) 11% of the phenotypic variation for CBH (3 QTLs), 22% for WSG (5 QTLs) and 12% for %BARK (3 QTLs). Least square means of the alternative QTL genotypes (d) were estimated between 0.4 and 0.5 phenotypic standard deviations (sp). Overlapping QTLs for CBH, WSG and %BARK were observed where either one or more genetic loci with pleiotropic effect or a cluster of linked genes control all three traits. Simultaneous QTL mapping by multivariate regression detected a significant case of digenic epistasis ($P<0.01$) between unlinked QTLs, that increased the total proportion of the phenotypic variance explained in CBH from 11 to 15% . Our results indicate the existence of major genes involved in the quantitative expression of economically important traits related to forest productivity in *Eucalyptus grandis*. These findings shed some light on the architecture of quantitative traits in forest trees and may have some important implications for the planning of future QTL mapping experiments and eventually for the contemplation of marker assisted tree breeding. Epistatic interactions amongst QTLs might prove of considerable importance in the architecture of quantitative traits and the advancement of selection in breeding populations of forest trees.

INTRODUCTION

The major obstacle in forest tree improvement is the time necessary to complete a breeding generation. Current practice has relied almost exclusively on the analysis of phenotypes at rotation (harvest) age. For most traits of commercial value, early selection of individual trees only becomes efficient half-way through rotation, which is still several years time, even for fast-growing species of *Eucalyptus* in tropical climate. Tree breeding is made even more difficult by the changes that occur during the transition from juvenility to maturity. Wood properties change during growth and maturation. Wood specific gravity is only adequately expressed at the phenotypic level after the tree has produced several growth rings, and early height growth is usually a poor predictor of volume growth at rotation. Methods to improve the efficiency of early selection would be of considerable value to increase the genetic gain per unit time. Many of the complications of phenotypic analysis in forest tree breeding could be alleviated by the direct identification of genotypes using a diagnostic system based on molecular markers co segregating with the traits of interest.

Most forest trees are undomesticated plants with high levels of genetic diversity. The most advanced tree breeding programs are only in the first few generations of selection in contrast with common crops where cultivar development has been conducted for several centuries. Besides estimates of genetic parameters, little is known about the genetic architecture and mode of inheritance of most traits in forest trees. Quantitative trait dissection will allow the examination of many hypothesis that have been the subject of debate for several years, including the very concept of a discrete number of unitary genetic effects (QTLs) controlling quantitative traits in heterogeneous forest populations (Strauss *et al.* 1992)

Genetic linkage maps of molecular and biochemical markers have offered a valuable tool to investigate the genetic architecture of polygenic traits and potentially assist in the manipulation of the genetic factors controlling such traits on an individual basis, through marker assisted selection (MAS) and recombination of genotypes displaying desired QTL combinations. Studies in recent years have used molecular marker maps to investigate the inheritance of quantitative traits. Results to date support the existence of a few major genes controlling large proportions of the total variation in a wide range of quantitatively inherited traits. In plants, these studies have been restricted to a few annual crop plants (reviewed by Stuber 1992; Dudley 1993). QTL

mapping studies in plants have typically been performed using segregating populations derived from crosses between divergent inbred lines. It is not clear to what extent these results are relevant to predict the ability to detect QTL in allogamous forest tree populations and apply marker assisted breeding.

In recent years, the exciting perspective of using molecular markers as an early selection tool in forest tree breeding has stimulated the undertaking of Quantitative Trait Loci (QTL) mapping experiments in some of the main forest tree species such as *Pinus taeda* and *Populus trichocarpa x deltoides* (Williams and Neale 1992; Bradshaw and Foster 1992). In an attempt to follow the models in crop plants, the approach taken has involved the use of three-generation pedigrees and RFLP markers. However, these kinds of pedigrees are not the ones typically available in most tree breeding programs, and are difficult to obtain due to significant genetic load and time constraints. On the other hand, two-generation pedigrees involving crosses between highly heterozygous parents are commonly available or can be promptly produced in the form of full and half-sib families for the vast majority of forest tree species. Because the parents of such crosses are outbred and highly heterozygous, the F1 is genetically heterogeneous and segregation is expected for quantitative traits.

In view of the long generation times of forest trees, retrospective QTL mapping experiments using existing two-generation pedigrees represent a key approach to develop QTL mapping information in acceptable time to contemplate the incorporation of MAS in operational forest tree breeding. In a previous study we described the combined use of RAPD markers and the pseudo-testcross marker configuration in a full-sib family of *Eucalyptus* as a general strategy for the construction of single individual genetic linkage maps in outbred forest trees as well as in any highly heterozygous sexually reproducing living organism (Grattapaglia and Sederoff 1994). We then extended the use of the pseudo-testcross strategy and explored the within-family linkage disequilibrium to locate QTLs controlling significant proportions of the phenotypic variation in vegetative propagation traits (Grattapaglia *et al.* 1994). In this study we carry out a retrospective QTL mapping study, using a maternal open pollinated half-sib family at rotation age (6.5 years). If a parent tree is heterozygous at a marker locus, its half-sibs can be partitioned into two groups, those that received the chromosomal segment marked by one allele and those that received the homologous segment with the alternative marker allele. If the tree is also heterozygous at a closely linked QTL, the groups of half-sibs will also differ with respect to the quantitative trait. This

half-sib approach for QTL mapping studies was originally recommended within groups of paternal half-sibs in domestic animals (Geldermann 1975). Its value has been investigated by simulations (Soller and Genizi 1978; Haley 1991) while experimental results in cattle (Geldermann et al. 1985; Beever et al. 1990) demonstrated the detection of significant marker-associated quantitative effects. The objectives of the present study included: (1) test the power of the within half-sib QTL mapping approach in a forest tree species; (2) attempt to locate genomic regions controlling quantitatively inherited traits with varying heritabilities in fast growing tropical Eucalyptus ; (3) test a sequential approach to QTL mapping by (a) bulk segregant analysis (BSA), (b) selective genotyping and (c) full-scale co segregation analysis, and compare the efficiency of these strategies for QTL detection.

MATERIAL AND METHODS

Plant Material. The experimental material consisted of an open pollinated maternal half-sib family of an elite clone of Eucalyptus grandis (clone 44, Coffs Harbor provenance, Australia - selection from a Zimbabwe seed source). Seeds were obtained from a managed seed orchard at Aracruz Florestal S.A., Brazil where the E. grandis clone is used as the only female clone, and 25 E.urophylla clones are used as pollinators. The ratio of pollinators to maternal clonal ramets was 3:1 and the spatial distribution was such that the maternal crowns were surrounded by pollinator trees. Pollination in Eucalyptus is predominantly entomophilous. Cross pollination was enhanced by the establishment of bee cages in the orchard, and a 800 m strip of native forest isolated the orchard from other Eucalyptus stands. The maternal E. grandis clone was characterized as highly self-incompatible by controlled self pollination. Therefore the seeds collected in the orchard are derived from outcrossing events with very high probability. Furthermore, seedlings derived from self-pollination (estimated at less than 3%) are easily identified and roughed in the nursery stage (Ikemori and Campinhos 1983). Commercial production plantations have been established with this seed lot at Aracruz Florestal over the past 8 years, thus a few million plants were available for this study. We selected and delimited a square area of approximately 1 hectare in an homogeneous forest stand planted in 1986 at a spacing of 3 m between rows x 2 m between plants. At time of trait evaluation, the stand was at harvest age (6.5

years). Even terrain, uniform soil type and minimal number of missing trees were the main factors considered in the selection of the experimental area.

Traits measured. The following growth and wood quality traits of commercial value were measured: (1) circumference at breast height (1.3 m from the ground) (CBH); (2) wood specific gravity (WSG); (3) percent dry weight of bark (%BARK) and (4) cellulose pulp yield (%PULP). CBH was measured on a total of 1085 standing trees and recorded in cm using a measuring tape. The trees were felled to allow leaf sample collection for DNA analysis. Total height growth was measured on 50 felled trees to estimate height x diameter correlation. For the first 300 individuals, a 4-5 cm thick wood disk was sampled at breast height for measurement of wood quality traits. Wood disks were kept immersed into water until time of WSG measurement. WSG determinations were made for the entire wood disk by a gravitometric method. The wet weight of the water saturated wood disk was measured. Then, disk volume was determined as the ratio between the weight of the displaced water and the specific gravity of water at $26^{\circ}C$ (996.6 kg/m^3). Dry weight of the disk was obtained following oven drying at $105^{\circ}C$ for a minimum of 48 hours. WSG were calculated as the ratio between the dry weight (Kg) and the volume of the disk (m^3). Throughout this procedure measurements were made separately for the bark and the solid wood so that WSG with and without bark were obtained. Percent bark was calculated as the ratio between the dry weight of bark and the total dry weight of the disk.

For a sample of 164 individuals, a micro pulping technique was used to estimate percent cellulose pulp yield. A sample consisting of 15 grams of uniformly prepared wood chips (1.0-1.2 x 5-11 x 30-50 mm) were taken from the center part of the dried wood disks. Wood chip samples were subject to alkaline cooking concentrations corresponding to 160 - 180 kg of active alkali (expressed as NaOH) per ton of dry wood chip, i.e. 16 to 18% active alkali. Cooking was performed at $170^{\circ}C$ during two hours into rotatory mini-digestors at a 5:1 ratio of white licor to wood chip. Following digestion, the dark licor was rinsed off, and the wood fibers in suspension were desintegrated in a blender until they became pulp. The pulp was then washed, filtered under vacuum and dried at $105^{\circ}C$ for 48 hours. Pulp yield was calculated as the ratio between the dry weight of pulp and the initial weight of wood chips.

RAPD marker genotyping. Leaf samples from a total of 1085 individuals were collected immediately following tree falling and kept on ice. RAPD assay conditions, marker identification and scoring were performed as described elsewhere (Grattapaglia and Sederoff 1994). A linkage map of RAPD markers had been previously constructed for the common maternal parent (*E. grandis* clone 44) using a pseudo-testcross strategy in a full-sib cross to a single *E.urophylla* tree, clone 27. The RAPD markers used in this study were previously determined to segregate in the pseudo-testcross configuration, i.e. they were heterozygous in the *E. grandis* clone and absent (null) in *E.urophylla* clone 27. This clone was also part of the group of pollinators "plus" trees in the orchard, originally selected from the same "Rio Claro" landrace. Therefore an assumption was made that all the markers found to be absent in *E.urophylla* clone 27 would also be absent (or at very low frequency for the "band present" RAPD allele) in the remaining pollinator clones, so that the markers to be used in the QTL study uniquely identified the maternal gametic contribution to the half-sib progeny. After testing a total of 20 markers on a subset of six *E.urophylla* clones, it was verified that all genotypes, without exception, were null for the screened markers. A further confirmation of the validity of this assumption was obtained later in the study by testing departures from the expected 1:1 segregation ratio on the half-sib marker data using a χ^2 test.

Bulk Segregant Analysis (BSA). The DNA pooling technique originally proposed by Arnheim *et al.* (1985) and later adopted by Michelmore *et al.* (1991) was employed as a first step to identify genomic regions controlling CBH. A total of 1085 plants were used for the analysis. Individuals that were 1.7 σ_p above and below the mean for CBH were selected to compose the bulked DNA samples. To minimize false positives and ambiguous results, four replicate bulks of 10 individuals each, at each tail of the distribution, were constructed by mixing equimolar amounts of genomic DNA. The individuals within each tail were not randomly assigned to bulks. Rather, the 40 individuals in each tail were ranked and sequentially assigned to the bulks. A total of eight bulks, four "low" and four "high" together with the maternal genomic DNA and a pooled DNA sample of six paternal individuals as controls, were screened with a total of 130 framework markers whose position on the map had been previously determined (Grattapaglia and Sederoff 1994). RAPD polymorphisms between the two groups of

bulks were interpreted as indicative of marker-trait association and further tested by full scale co segregation analysis (see below).

Selective genotyping From the same population of 1085 individuals, ninety six progeny with extreme phenotypes for CBH (48 in each tail) were chosen for selective genotyping (Lander and Botstein 1989). The two groups included the 80 individuals used to compose the bulks in the BSA step. These individuals were genotyped for a total of 52 evenly spaced markers in the *E. grandis* clone 44 map (Figure 2), including the markers found to be polymorphic in the BSA screening. A $\chi 2$ test was carried out to test departures form the expected 1:1 segregation ratio within each extreme group. Significant departures at $\alpha=0.05$ were taken as an indication of marker-trait association, and later tested by full scale co segregation analysis (see below).

Full scale co segregation analysis. A total of 300 individuals for which data was available for CBH, WSG and %BARK were genotyped for 81 RAPD markers corresponding to 77 marker loci (Figure 6). A subset of these individuals (n=164) had also been evaluated for %PULP. The sample of 300 used for the full-scale co segregation analysis was random, composed by the first 300 trees measured in the experimental plot. Evidently, some of the individuals used for the BSA and SG experiments were also part of the sample of 300, to maintain its integrity as a random sample. The RAPD markers used in this analysis were selected from a total of 241 RAPD markers in the original *E. grandis* clone 44 map, trying to maximize genome coverage with the minimal number of PCR reactions. Although all the RAPD markers used in this study have been mapped previously (Grattapaglia and Sederoff 1994), a genetic map was calculated from the genotypic data *de novo* and checked for consistency with the previously reported map. Linkage relationships among markers were determined using MAPMAKER (Lander *et al.* 1987). To allow the detection of linkage of RAPD markers in repulsion phase the data set was duplicated and recoded. LOD 5.0, and maximum $\theta=0.40$ were used as linkage thresholds for grouping markers. The software program GMENDEL (Liu and Knapp 1990) with threshold P value= 0.0001 and $\theta=0.40$ was also used for the linkage analysis.

QTL mapping analysis was performed by single factor regression using GLM (SAS 1988) and by two interval mapping methods implemented by MAPMAKER-QTL (MMQ) (Lander and Botstein 1989) and QTLSTAT (QST) (Knapp *et al.* 1992). QTL analyses were

performed under a backcross model. A LOD score threshold of 1.5 or nominal significance level of p=0.02 were used to declare the presence of a linked QTL in the interval. With this stringency and given the number of markers per chromosome used, a per-chromosome false positive rate of approximately 5% was ensured as estimated numerically by Darvasi et al. 1993. For each LOD peak, the 1.0 LOD support intervals were determined. For all detected QTLs, the percentage of variance explained as estimated by both interval mapping methods, least square means for the alternative QTL genotypes and the shifts in trait value in phenotypic standard deviations were also reported. When linked QTLs with no overlapping 1.0 LOD support intervals were detected, the locus with highest LOD score was fixed and the chromosome scanned again for the linked effect. Multipoint estimates of the total variation explained by the main effects of the mapped QTLs were obtained by interval mapping with MAPMAKER-QTL and by multivariate linear regression using GLM (SAS 1988). Multilocus regression models were built, including unlinked main effects and two-way interaction terms to test for significant digenic epistasis.

RESULTS

Quantitative traits. With the exception of %BARK, that showed a right skewed distribution, the frequency distributions of phenotypes for the four traits showed an approximately normally distributed continuous variation (Figure 1). Mean, standard deviation and sample sizes used in the full scale co segregation analysis for each trait are also presented. Phenotypic correlations were estimated among traits within the family studied (Table 1). A significant positive correlation was observed between CBH and WSG (r= 0.28), a strong negative correlation between CBH and %BARK and between WSG and %BARK. As expected, phenotypic correlation between CBH and height growth was high (r= 0.95). Results presented for CBH should also be regarded valid for height growth and therefore could be interpreted as representative of the relevant commercial trait, i.e. final volume growth of the tree.

Detection of selfed individuals and paternal marker alleles. In spite of the high self-incompatibility of the common maternal parent and the careful seed orchard management and nursery roughing (see Material and Methods), we tested our mapping population for the presence of progeny individuals derived from self

pollination. We were able to carry out a screening test for selfs using a group of four RAPD marker pairs from different linkage groups that had been previously determined to be co dominant alleles at the same locus by DNA hybridization experiments (Grattapaglia and Sederoff 1994). These were: Y15_740/Y15_760; Y17_525/Y17_515; A10_635/A10_562; A11_980/A11_920 (Figure 6). In a selfing event, 50% of the selfed individuals will show both alleles on the gel, and can therefore be detected as selfs. With four independent markers, the probability of detecting a selfed individual is $1-(1/2)^4 = 0.9375$. We screened the mapping population (n=300) as well as the extreme individuals used both for the BSA and the selective genotyping experiments. We found 11 putative selfed individuals, i.e. a rate of 3.6%. These individuals were eliminated from the experiment and substituted by true half-sibs previously screened with our set of co dominant markers.

Besides contamination due to selfs, we were also concerned with the potential presence of paternal marker alleles that would be erroneously classified as maternal alleles in the analysis. We screened a subset of E.urophylla clones for 20 markers heterozygous in the maternal parent and did not detect any "band present" allele. An indirect examination of our assumption of uniqueness of the maternal marker alleles was performed by conducting a $\chi 2$ test for the expected 1:1 segregation ratio on the genotypic data for the 300 individuals. Significant distortions of marker loci with an excess of the "band present" allele versus the absence could indicate the presence of the marker allele in the pollen pool at high frequencies. We found a total of 11 marker loci significantly distorted at $\alpha=0.05$. Of these, six were significant at $\alpha=0.01$ and located in two clusters, one on group 6/9 around marker U20_900 and the other on group 8/12 around marker K9_534. At $\alpha=0.05$, four such distortion would be expected due to chance alone. The distortions at $\alpha=0.01$ could indicate the presence of homologous paternal marker alleles. However, because these distortions are clustered, such an hypothesis seems unlikely. In addition half of the distorted markers displayed an excess of null alleles which could not be explained by paternal contribution. Rather, because of the relatively large sample size analyzed, these distortions might have a biological basis such as pre or post zygotic selection in the maternal parent, against some unfavorable linked allele in the region.

Bulk Segregant Analysis. From a total of 130 markers screened only two markers showed a clear polymorphic pattern between the high and low CBH bulks. Although

two markers were detected, they should be regarded as one event, because they are linked on group 7 at approximately 25 cM (A11_980 and U10_1500). The most clear polymorphism was obtained with marker U10_1500. It was present in the all the high CBH bulks and absent in all the low bulks (Figure 3a). Later, in the selective genotyping step, a significant difference in the frequency of marker U10_1500 was observed between the high and low groups confirming the nature of the BSA polymorphism (Figure 3b, Table 2). Upon full scale co segregation analysis, the genomic region on group 7 involving U10_1500 was deemed significant for CBH as well as for other correlated traits (Tables 3 and 6) (see discussion below).

For 10 markers, ambiguous BSA polymorphisms were observed, when, within each group of bulks, the marker was present or absent for 3 out of four bulks. One of these markers was Y17_1500 on group 5 that was later determined to be linked to a major QTL region affecting CBH, WSG and %BARK. Not surprisingly, the sensitivity of the BSA approach for quantitative traits is clearly lower than the one obtained when tagging genes controlling simply inherited disease resistance or morphological traits. Our results with replicate bulks indicate that the lack of sensitivity resides more in false negatives rather than false positives. We assayed, *a posteriori*, individual samples that composed our bulks for markers later found to be in QTL regions (data not shown). We observed that frequently one individual carrying the RAPD band was enough to "contaminate" an otherwise null genotype bulk. This was particularly common for RAPD markers that were efficiently amplified resulting in bands of high intensity.

Selective Genotyping. A subset of evenly spaced markers covering an estimated 85% of the maternal genome was used to carry out the selective genotyping step (Figure 2). Genotype data for the 52 markers screened were used to perform $\chi2$ tests for segregation distortion in the two extreme groups. The initial expectation was that a putative marker-trait association would be indicated by a significant distortion from the expected 1:1 segregation ratio in both extreme phenotype groups but in different directions. The marker allele in excess in the "high" extreme would be linked in coupling to the favorable QTL allele. However, after performing the analysis we did not observe this pattern of segregation distortion. Rather, in the cases where a segregation distortion was observed, we found that markers were distorted in the "low" phenotype group but segregated accordingly in the "high" group (Table 2). As a result, we also observed that the t test for difference between means of the two marker classes on the extreme

phenotypes were significant only when the segregation distortion in the "low" group was strong (e.g. markers U10_1500 and A11_980, Table 2). Marker R4_1300 (Figure 4) in spite of the comparatively low significance level of the t test on the extreme phenotypes (P =0.04) was later found to be significantly associated with CBH.

The most obvious explanation for the lack of agreement between the expected and observed patterns of segregation distortion in the selective genotyping experiment is the contribution of paternal alleles at the same as well as other QTLs (Figure 5). Suppose a situation where a maternal QTL and RAPD marker are linked in coupling, i.e. the favorable allele Q linked to the (+) genotype for the marker. The simple expectation would be that an excess of individuals with the (+) genotype will be found in the "high" phenotype extreme and a correspondingly lower number of (+) individuals in the "low" group. However, favorable alleles at the same QTL exist in the pollen pool, and these are linked to null alleles for the RAPD marker locus (Qi -). If a half-sib individual that has received the unfavorable maternal allele q, receives a favorable paternal allele Qi at the same locus, it is reasonable to think that the resulting phenotype could be placed in the high end of the distribution depending upon the magnitude of effect of the paternal allele. These individuals would therefore be included in the "high" phenotype group upon selective genotyping in a half-sib family. As they have a null genotype for the marker, they will inflate the null genotype class and mask the expected segregation distortion. For a favorable maternal allele linked in repulsion to the (+) RAPD marker the situation would be analogous with the difference that the paternal favorable alleles combined to (q +) maternal haplotypes would inflate the frequency of (+) genotypes in the high tail and again mask the distortion.

The extent to which incoming favorable paternal alleles can mask a segregation distortion due to marker-QTL linkage in the maternal parent will depend essentially on the gene frequency for these QTL alleles Qi in the pollen pool as well as their magnitude of effect relative to the favorable maternal allele. In our experiment, all the $\chi 2$ tests for segregation distortion in the high tail group were non significant, including those for markers later found to be associated with QTL (Table 2). This result strongly suggests that favorable QTL alleles with important effect on CBH are relatively frequent in the E.urophylla pollen pool. Similar explanation would apply to the "low" group. An unfavorable QTL allele from the pollen pool could move an individual with the favorable maternal haplotype (Q +) to the low tail, inflating the frequency of (+) genotypes and masking a distortion. Our data suggests however, that such event

should be more rare if the maternal QTL allele is truly favorable . In our experiment significant distortions were observed for markers in the "low" phenotype group with an excess of the RAPD allele later determined to be linked in coupling to the unfavorable maternal allele. This was the case for all the distortions observed, including for markers U10_1500 and R4_1300 (Figures 3 and 4) later found to be associated with QTLs.

To confirm putative marker-trait linkages detected by segregation distortion in the selective genotyping step, we genotyped a random sample of 204 individuals. A $\chi 2$ test was used to verify potential pre-existing segregation distortions for the markers, and a t test was performed on the difference between means for the two marker classes. Following this analysis two markers were found to be significantly associated with a quantitative trait difference in CBH. These were R4_1300 on group 5 and A11_980 on group 7. Marker U10_1500, in spite of the strong segregation distortion ($\chi 2 = 27.0$), the significant t test on the extreme phenotypes (P= 0.000) and the relatively short map distance from A11_980 (~ 25 cM) did not yield a significant t test on the random sample. The putative marker-trait associations for markers R15_1650 and R10_840 were also dismissed based on the non-significant t test. However in the case of R15_1650, a nearby marker interval comprising markers K10_835 and X2_1200 was later deemed significant by linear ANOVA and interval mapping on a larger sample of individuals (n=300) (Tables 3 and 6). It is worth mentioning that the marker X2_1200 was screened in selective genotyping but did not yield any indication of trait association (Table 2). In conclusion, the selective genotyping experiment detected two out of the three genomic regions significantly associated with CBH expression as detected by full scale co segregation analysis. For the undetected QTL region on group 2, the selective genotyping essentially resulted in a false negative.

Linkage map construction. To perform a full scale co segregation analysis of all traits on all RAPD markers, we reconstructed a genetic linkage map for *E. grandis* clone 44 using its open pollinated half-sib family. The map presented is based on the pseudo-testcross (1:1) segregation of 81 RAPD markers (77 loci since four markers pair are co dominant) from 300 meiotic events. Forty-nine of those markers were previously classified as framework markers, i.e. their order was established with a likelihood support $\geq 1000:1$ (Grattapaglia and Sederoff 1994). When genotyping with RAPD markers, each arbitrary primer amplifies more than one segregating marker. Primers and markers used in this study were selected to maximize the target genome coverage with the minimal

number of RAPD assays, to allow the genotyping of a relatively large progeny size in acceptable time. A total of 34 primers were used in this study to amplify 81 markers. Only 10,200 PCR reactions were necessary to genotype 300 individuals for all the markers. At a rate of 400 reactions/day this would require approximately one person per month time.

A LOD score of 5.0 and maximum $\theta=0.40$ were set as linkage thresholds for grouping markers. Map distances in centimorgans were calculated using Kosambi's mapping function. Orders of marker loci in each linkage group were established using a matrix correlation method implemented by MAPMAKER. The 1000:1 likelihood support criterion was used to establish the final linear order of marker loci. Due to the relatively large sample size and even spatial distribution of the markers selected for the QTL analysis, the orders obtained promptly conformed to the 1000:1 likelihood support criterion recommended for building framework maps (Keats *et al.* 1991). The LOD 3.0 support was tested by permuting the order of all adjacent triplets of markers. Finally, the orders obtained compared very closely to the orders resulting from simulated annealing performed by GMENDEL.

The framework marker order of the map (based on 300 meiosis), is well conserved when compared to our previously published maps of *E. grandis* clone 44. When comparing framework map orders obtained for *E. grandis* we observed 7% of order changes when we increased sample size from 62 to 122 meiosis within the same pedigree. Most of these changes were simple order switches of adjacent markers (Grattapaglia *et al.* 1994). In this study we had the opportunity to compare linkage maps of the same individual tree constructed using two independent pedigrees, a full-sib and a half-sib, that shared a common maternal parent. Comparisons were possible for a total of 64 markers that were mapped in this study as well as on the previously published map based on n=122 in a full-sib cross. There were six cases involving a switch in the order of two adjacent markers: markers Y16_550 and Y17_515 on group 8/12; R20_2700 and R20_1080 on group 4; N15_629 and K9_884 also on group 4; Q13_1150 and G13_1570 on group 11/13; A11_980 and K9_1660 on group 7; Y17_1300 and R20_1150 on group 14. There were two cases of markers out of order: K9_534 on group 8/12; marker X1_720 on group 4. On group 2 a block of markers (K10_549 to R15_1650) was inverted in orientation in relation to the rest of the markers in the group. Overall we observed 9 order changes in 64 comparisons, i.e. a rate of 14% order difference.

The linkage grouping obtained in this experiment was essentially the same as the one obtained previously from the full-sib pedigree with n=122, confirming the breakage of group 11, the linkage of the two pieces to groups 9 and 13, and the merger of groups 8 and 12. However, in this experiment we observed a merger of group 6 and group 9/11. Although the recombination fraction between the two groups was high (0.38), the LOD score was greater than 5.0. With this merger, we achieved 11 linkage groups, corresponding to the expected number based on the haploid number of chromosomes in *Eucalyptus grandis* (n=11).

Full scale co segregation QTL analysis. For all traits, QTL analyses were performed and results are presented for untransformed phenotypic data. The analysis of normalized (log transformed) data for %BARK did not alter any of the results. QTL analysis was initially performed using single factor regression analysis. A total of 33 out of 308 (4 traits x 77 markers) F tests were found to be significant at $\alpha= 0.02$, and of these, 23 at $\alpha=0.01$ (Table 3). The percent of significant tests is very similar to those observed in QTL mapping experiments using crosses between inbred lines of plants (Kennard *et al.* 1994; Nodari *et al.* 1993) as well as in half-sib families of cattle (Geldermann *et al.* 1985). Several of the significant markers associated with trait expression were linked on the same linkage group. Counting only genetically independent associations, i.e. one per linkage group per trait, the linear analysis detected a total of 12 QTL regions: three for CBH, five for WSG and four for %BARK. A significant associations at $\alpha=0.03$ was found for %PULP on group 6/9 (Table 3). The great majority of the QTLs mapped to four linkage groups. Overlapping QTLs for CBH, WSG and %BARK were observed on group 5 where either one or more genetic loci with pleiotropic effect or a cluster of linked genes control all three traits. Overlapping of QTL positions was also observed for CBH and WSG on groups 2 and 7 and for WSG and %BARK on group 6/9.

Interval QTL mapping analysis was performed using two different approaches for which software packages were available. MAPMAKER-QTL (MMQ) 3.0 was used to estimate QTL parameters and test statistics at 2 cM interval within every marker bracket and to select the most probable location for the QTL as the location that maximized the likelihood ratio. QTLSTAT estimates the least-squares QTL genotype means and tests the hypothesis of "no QTL" versus the hypothesis of "one QTL" for every marker bracket. The QTL is therefore assigned to a particular marker interval directly, using non-linear least squares, without trying to estimate the most likely position within that interval.

The results of the two interval mapping analyses generally agreed with each other as well as with the results of linear regression (Tables 4 and 5) with some exceptions (see below). In a similar way as in the regression analysis, LOD scores or Wald statistics above the significant thresholds adopted were observed along stretches involving adjacent marker intervals. In the summary of the least squares interval mapping (Table 5), the highest peaks of Wald statistics for non-adjacent intervals are reported. In the MAPMAKER-QTL analysis, when two linked LOD peaks were observed, with overlapping LOD 1.0 support interval, only one QTL was declared in the region at the location with the highest peak. When the LOD 1.0 supports did not overlap, a procedure suggested by Lander and Botstein (1989) was used. The position of one QTL was fixed and the chromosome scanned again for a linked QTL effect. Such a procedure was applied for CBH, WSG and %BARK to test the likelihood of linked QTL effects detected on group 5 (Table 4). For WSG, upon fixing the QTL in the first interval (X1_1450 - R4_1300), a raise in LOD = 1.72 was observed at the position of the second putative QTL (interval Y17_1500 - V7_1200). When the second QTL was fixed, a raise in LOD= 2.1 was observed at the first QTL. This result is suggestive of two linked QTL effects for WSG on group 5. Such a proposition is further strengthened by the fact that both regions were detected as significant in the QST analysis (Table 5). For CBH the two QTLs were mapped exactly within the same two intervals (Table 4). However the evidence for two linked effects is weaker: the raise in LOD was 1.7 at the first QTL when fixing the effect of the second QTL, but only 1.2 at the second QTL when fixing for the first. Evidence for two QTLs for CBH on group 5 gets even weaker when we observe that only one QTL was detected by least squares interval mapping (Table 5). Finally for %BARK the result of the procedure is that only one QTL exists, as a flat likelihood surface was obtained when the first QTL (X1_1450 -R4_1300) was fixed.

Disagreements on QTL detection were observed mainly between the QST results (Table 5) and the results of linear regression (Table 3) and MAPMAKER-QTL (Table 4). QTLs for CBH and WSG on group 2, WSG on group 7 and %PULP on group 6/9 were detected both by linear regression and interval mapping using MAPMAKER-QTL but not by least squares interval mapping at $\alpha= 0.02$. On the other hand, MAPMAKER-QTL did not detect a QTL for %BARK on group 14 while both linear regression and QST did. In the QTL summaries we report the marker interval where either the highest LOD score estimated by MMQ or Wald statistic by QST were observed. For MMQ we also locate the most likely position of the QTL within the interval and the 1.0 LOD support. Marker

brackets where the highest LOD score was located generally did not correspond exactly to the one with the highest Wald statistics. For the seven QTLs detected by both methods, in only three did the marker bracket with the highest LOD score correspond to the bracket with the peak Wald statistics. In the other five cases we found that the marker brackets where the LOD and Wald peaks were detected were adjacent, rather than overlapping, shared a common marker and were within the 1.0 LOD support interval determined by MMQ. We obtained similar results when mapping QTLs controlling vegetative propagation in a full-sib cross of *Eucalyptus* (Grattapaglia *et al.* 1994).

An interval mapping analysis with MMQ was carried out on a data set combining the random sample of 300 individuals with the 96 extreme individuals used in the selective genotyping experiment. As 11 individuals from the random sample had been also part of the extreme groups, a final sample size of 385 individuals was used in the analysis. A total of 24 common markers were used in the joint analysis, distributed on groups 2, 3, 5,7, 8/12 and 14. Linear regression models were not used in this combined analysis since they would tend to overestimate effects due to the biased selection of progeny (Lander and Botstein 1989). Besides the QTLs on group 5, the combined analysis also detected a QTL for CBH on group 7 (Table 6), confirming the result obtained originally in the linear regression analysis on the random sample (Table 3). Moreover in the same interval on group 7, QTL effects for %BARK and %PULP were also mapped. No new QTL region was detected by the combined analysis.

Estimates of the proportion of phenotypic variation explained by each QTL were obtained from MMQ (Table 4). Such proportions were also estimated with QST as the coefficient of determination (R^2) (Table 5) from the least squares analysis of variance table by dividing the Type I Sum of Squares for the QTL genotype by the total Sum of Squares in non simultaneous interval QTL mapping. The two estimates proportionally agree with the exception of the QTL for CBH on group 5 detected as two peaks with MMQ (%VAR. EXPL. of 4.0 and 6.6) and one peak with QST (R^2= 0.20). This disagreement might be partly due to the imprecision in the estimation of the exact location for the QTL. Proportions of phenotypic variance estimated as R^2 values were typically larger than the estimates of % variation explained from MMQ which in turn were larger than those estimated by linear regression (data not shown). For all traits, individual QTLs explained between 5 and 15% of the phenotypic variation, with the exception of the QTL for CBH on group 5 where the R^2 = 0.20 (20%). These estimates

obtained by non-simultaneous testing should however be viewed with caution (Knapp *et al.* 1992).

Least square means of the alternative QTL genotypes and their associated standard deviations were estimated with QST (Table 5). For all the putative QTLs detected, variances of the alternative QTL genotype classes were generally equal and close in value across QTLs within traits. Estimates of the difference in mean trait value between the two alternative QTL genotypes (Δ_1) ranged from 0.26 to 0.51 phenotypic standard deviations (σ_p) and most of the values were around 0.4 (Table 5). Differences in mean trait value between the family mean and the favorable QTL genotype (Δ_2) ranged from 0.10 to 0.27 σ_p and typical values were around 0.2 σ_p.

Multi-QT Locus modeling. Simultaneous QTL mapping was carried out by fitting multivariate regression models with the markers detected by non simultaneous mapping using linear regression (Table 8). This procedure also allowed unbiased hypothesis testing for the markers included in the model using Type III Sum of Squares as suggested by Knapp *et al.* 1992. The proportion of phenotypic variance explained by each QTL were also estimated, however because they are non additive can be misleading (Knapp *et al.* 1992). All the markers used to build the models were retained after simultaneous testing with the exception of marker R4_1300 for WSG (Table 8). Interestingly, R4_1300 was included in the model together with a linked marker (V7_1200) based on the evidence provided by interval mapping with MMQ for the existence of two linked QTLs on the group. Therefore, results from simultaneous mapping suggest that only one QTL for WSG is located on group 5. It should be pointed out that although simultaneous QTL mapping is effective for removing false positive QTLs, it does not overcome the problem of false negatives (Knapp *et al.* 1992). The number of mapped QTLs for all traits should evidently be regarded as minimal due also to reasons related to detection power (see below).

Multilocus models provided estimates of the total proportion of phenotypic variance explained by the joint action of the mapped QTLs. Similar estimates were obtained by multipoint interval mapping with MMQ, and by multivariate analysis (Table 7). R^2 values were obtained initially for multivariate models including all the significant markers detected (Table 3). However, multiple linked QTLs are difficult to discriminate from a single effect detected by linkage to several markers. Therefore more conservative R^2 values were also obtained with models including only genetically

independent markers detected with the highest significance level. Reductions of ~25% in R^2 values were observed (Table 3). Still, over 10% of the phenotypic variance could be accounted for by the QTLs mapped for CBH and %BARK and twice as much (22%) for WSG.

To test for digenic epistasis, we carried out multivariate regression analysis including all the main effects and two-way interactions of unlinked markers detected in the non simultaneous linear analysis. Significant effects were tested using Type III and Type IV Sum of Squares to account both for cases of unbalanced data and missing cells (Knapp *et al.* 1992). With the exception of CBH (see below) no significant two-way interactions were detected at α =0.01. For CBH a highly significant interaction was detected between QTLs on groups 5 and 7 (P = 0.0097) (Table 9). Type III Sum of Squares are presented which were equal to Type IV, as the data for these markers contained no missing cell. The QTL detected on group 5 (marker Y17_1500) positively interacts with the QTL on group 7 (marker N15_1079). The P value observed for the second interaction term involving Y17_1500, although not significant, (P= 0.0531) suggests that Y17_1500 might also have an effect on the QTL detected on group 2 (marker K10_835). However, another intriguing result was that upon fitting a model including interaction terms, marker K10_835 was deemed not significant. This result indicates that a two-locus model including markers Y17_1500, N15_1079 and their interaction could account for 15% of the phenotypic variance in CBH. The interaction term was responsible for an almost 50% increase in the phenotypic variance explained by the model.

DISCUSSION

Half-sib linkage mapping. Half-sib linkage and QTL mapping in outbred species depends essentially on the ability to uniquely discriminate the maternal allelic contribution to the offspring. Half-sib mapping has been applied to several conifers taking advantage of their unique haploid biology (Conkle 1980; Grattapaglia *et al.* 1991; Tulsieram *et al.* 1992; Nelson *et al.* 1993). Genetic segregation of heterozygous marker loci can be readily examined in the haploid megagametophytes. However for most outbred species such as domestic animals and forest trees direct analysis of haploid tissue is impractical, and half-sib analysis has to performed at the diploid level. As pointed out by Soller (1991) when diallelic markers are utilized it will not be possible in many cases to determine which of the common parent allele was transferred to a

particular progeny. When both markers are at equal frequencies in the population, it will be possible to identify the transferred allele for only half to three-quarters of the progeny, depending on whether the other parents are also scored for the markers. The situation is improved by the use of multiallelic markers such as VNTR or microsatellites. Each marker allele will be individually rare and it should be possible to trace it from the common parent to offspring in virtually all instances. However, such markers require a large and costly 'up front ' development of specific VNTR probes or specific microsatellite primers pairs.

In this study we used RAPD markers in combination with an open pollinated maternal half-sib family to construct a genetic linkage map of 77 RAPD marker loci, for a single tree of *Eucalyptus* grandis. Marker genotype analysis was carried out on diploid tissue of half-sib individuals. The RAPD markers used were previously determined to be heterozygous for the amplified RAPD allele in the maternal parent , and absent, or at very low frequencies (null genotype) in the paternal *E.urophylla*. This informative mating configuration was made possible by the high sensitivity of the RAPD assay to single base changes, allowing the unique identification of the maternal alleles in the offspring. Paradoxically, the fact that RAPD are dominant markers and detect only one allele at a locus facilitates the occurrence of such configurations, because the necessary null genotypes of the paternal parents actually correspond to undetected alleles at the RAPD loci. It has to be pointed out that the inter specific nature of the pedigree favored the detection of such DNA polymorphism between the maternal and pollen parents. However we believe that this strategy should be applicable also at the intra specific level. Earlier we found that 33% of the mapped markers in the *E. grandis* individual used in this study were present in a second individual of the same species and distinct origin (Grattapaglia and Sederoff 1994). Although limited, this estimate suggests that about two thirds of the RAPD markers should be useful for half-sib mapping at the intra specific level.

The number of useful RAPD markers for half-sib mapping could be significantly increased by pre-screening random primers for useful polymorphisms. Such a pre-screening approach is possible because the RAPD assay is technically simple and fast to perform. Two possibilities exist. Pre-screening primers on the DNA of all the pollen parents involved in the cross (if available) and a sample of 6 to 8 progeny of a full-sib cross of the maternal parent with one of the pollen parents. With this arrangement pseudo-testcross heterozygous markers from the maternal parent are identified in the full-sib

cross and their frequency is surveyed in the remaining pollen parents. Alternatively, an indirect approach would involve direct genotyping of the maternal parent and a sample of half-sib individuals of adequate size for detecting segregation distortions from the expected 1:1. Haley (1991) discussed this approach to verify the 1:1 segregation for VNTR alleles from sires heterozygotes for the marker when mated to homozygous dams. To discriminate between hypothesis of segregation ratio 1:1 to 3:1 (frequency of RAPD marker in pollen pool p= 0.5) a sample size of 80 would provide power of 0.9 at α=0.001 (nominal α=0.05 for each $\chi2$ test). However the frequency of the RAPD marker in the pollen pool will likely be lower. For example if p=0.2, the alternative segregation ratio will be 3:2 and to reach equivalent power in the $\chi2$ test several hundred progeny would be required. Therefore RAPD markers at low frequencies in the pollen pool causing subtle segregation distortions would not be detected in this approach unless huge sample sizes were analyzed. In this case, the direct screening of all paternal individuals becomes more efficient if a precise estimate of the extent of genotype misclassifications due to paternal alleles is required.

With this study we had the opportunity to compare linkage maps of the same individual tree constructed using two independent pedigrees, a full-sib and a half-sib, that shared the common maternal parent. Grouping of markers remained essentially the same but the larger sample size was efficient in allowing the detection of linkage between previously unlinked groups of markers and the coalescing of the map into the expected number of linkage groups based on the haploid chromosome number. A 14% difference in locus order was observed between the half-sib map and the previously published map. It is very unlikely that any biological reason underlies such differences. Therefore we are left to conclude that sampling bias will introduce a level of uncertainty in the final locus order, even when building framework maps at likelihood support $\geq$1000:1 and increasing the sample size (from 122 to 300 in this study). It will be interesting to see the rate of locus order changes with a larger sample size obtained by pooling the data sets of the full-sib and half-sib experiments.

QTL mapping using extreme phenotypes. One of our objectives was to test a sequential QTL mapping approach starting with selective sampling techniques and compare their relative efficiency. The efficiency of the BSA approach for a quantitative traits was clearly lower than the one reported when tagging genes controlling simply inherited disease resistance or morphological traits (Michelmore *et al.* 1991; Chaparro *et*

al. 1993). Recently, the BSA approach in combination with RAPD markers and dihaploid lines was used to identify a major QTL controlling milling energy in barley (Chamers *et al.* 1993). In our study, the multiple bulks segregant analysis adopted was very efficient to protect against false positives with a consequent reduction in power for detection. It is remarkable, however that even in a highly heterogeneous genetic background such as a half-sib family, and for a quantitative trait of low individual level heritability such as CBH (0.31) (Rezende and Bertolucci 1993) , the BSA approach was successful in detecting a genomic region later confirmed to contain a QTL.

The identification of a QTL by BSA depends essentially on the precision with which individual genotypes from the extremes of the trait distribution can be selected to compose the bulks. To increase this precision, the sample size used in the BSA step involved a total of 1085 individuals. However, the QTL later detected as being of major effect on CBH (on group 5) was not detected by BSA, due to the sensitivity of the PCR assay for the specific marker in the region. In conclusion if a very large sample size is already available and the phenotype determination is quick and inexpensive, BSA followed by genotyping a random sample of individuals might be a very useful way to quickly identify one or some of the genomic regions controlling the trait of interest. However, because the detection depends on the molecular characteristic of the marker screened BSA does not seem to be adequate to dissect quantitative traits, i.e. identify all the genetic factors controlling the trait. A less stringent screening than the one adopted in this study might be useful in this respect. However, the large number of individual assays necessary to screen the false positives might overrule the intended advantage of BSA as a short cut for QTL mapping.

Two out of the three genomic regions later detected as containing QTLs for CBH were identified by selective genotyping. However there is at least one major complicating factor when performing selective genotyping analysis in a half-sib. This is the unknown contribution of QTL alleles from the pollen pool, both in frequency and magnitude. On a random sample of half-sibs, this contribution is assumed random. However, when only extreme phenotypes are analyzed, particular full-sib families within the half-sib family could be significantly over represented in one or both of the extreme groups, biasing the expected segregation ratios for the maternal markers in case of linkage to a QTL. To be detected, the maternal QTL allele has to be of major effect so that its expression will cause a detectable shift in trait value independent on the paternal alleles. In this situation, a segregation distortion can be detected in the "low"

group as it was the case in our experiment. In conclusion, our results suggest that if the maternal QTL is of sufficient magnitude to be detected, more information for linkage will be provided by the "low" phenotype group. Therefore in a half-sib, the effort should be directed to genotyping a larger sample of individuals at the low end of the distribution instead of the approach of equal sample sizes at both ends. With a larger sample size, more confidence would also be associated with the detected distortions, minimizing false positives. Our empirical results agree with the expectation that selective genotyping and sequential sampling will be most efficient in QTL studies aimed at investigating differences between lines or strains that differ in some single outstanding trait (Darvasi and Soller 1992; Motro and Soller 1993).

Selective genotyping was applied exclusively to CBH. Trait determination for CBH is quick and inexpensive and an almost unlimited sample size was available for the study. Thus a sample size of 1085 individuals was measured and used to extract the two extreme tails each containing sufficient individuals to carry out the planned segregation distortion tests. Lander and Botstein (1989) point out that selective genotyping should substantially increase efficiency whenever growing and phenotyping additional progeny requires less effort than completely genotyping individuals. However selective genotyping becomes less efficient as more than one trait is analyzed simultaneously, unless the proportion selected is very small, since this would require essentially scoring the entire enlarged population for the markers (Lebowitz *et al.* 1987). Similarly, very little can be gained by sequential sampling procedures (Motro and Soller 1993). We did not apply selective genotyping to other traits for the reasons mentioned above. WSG and %PULP determinations are relatively costly, and an enlarged sample size could not be justified on the basis of selective genotyping inasmuch as new extreme phenotype groups would have to be extracted and genotyped. However, WSG and %BARK were also measured for the selected "high" and "low" groups extracted on the basis of CBH to allow an analysis of combined (extreme+random) data by interval mapping. Furthermore, measuring these traits for the "high" group was justified on the basis of breeding objectives. Phenotypic correlations among these traits within the extreme groups were higher than those obtained on random samples (data not shown). *A posteriori* analysis for these quantitative traits using a t test on markers A11_980 and R4_1300 also revealed significant associations between these markers and quantitative differences in WSG and %BARK (data not shown).

Within half-sib QTL mapping in forest trees. QTL mapping within-half sib families explores the within family linkage disequilibrium between marker and trait loci. It requires that the common parent tree be heterozygous both at the quantitative trait locus and linked marker locus. Half-sibs from a heterozygous tree can be partitioned into two groups, those that received the chromosomal segment containing one marker allele and those that received the homologous segment with the alternative allele. If the tree is also heterozygous at a closely linked QTL, the groups of half-sibs will also differ with respect to the quantitative trait. The difference in means of the trait values in the two classes characterized by maternal gametes is expected to be proportional to $\delta(1-2\theta)$, where δ is the phenotypic difference between half-sibs of the two genotypes and θ is the map distance in Morgans between the QTL and marker locus. With interval mapping the power for detecting QTLs is slightly increased and the term $(1-2\theta)$ can be replaced by the square root of $(1-\theta^*)$ where θ^* is the recombination frequency between two markers of the interval (Zeng, personal communication).

Haley (1991), using maximum likelihood methods, estimated that in a backcross of a sire, heterozygous for a major gene, to a group of homozygous dams, a power of QTL detection equal to 0.67 would be expected for n=400 half-sibs at α=0.001, δ =0.5 and θ=0.1. At θ=0.05 power would increase to 0.92. For δ=0.25, maximum power, at θ=0.0, would be 0.24. In other words half-sib mapping will only have sufficient power to detect genes of relatively large effects. We can use the expression of Haley (1991) to arrive to the approximate power expected with our sample size at α=0.001. For a maximum map distance between a marker and QTL of 10 cM, n=300 would achieve a power of approximately 0.5 to detect QTLs with δ=0.5, and power of 1.0 for QTLs with δ=1.0. For the majority of the QTLs detected in our experiment , the least-square estimates of the difference between marker genotype classes were between 0.4 and 0.5 (Δ_1, Table 5). Equating these values to $\delta(1-2\theta)$ and correcting for the increased power with interval mapping, we arrive to values of δ between 0.42 and 0.52, i.e. $\Delta_1 \sim \delta$. From these results and given the power predictions discussed above, we can suggest that it is unlikely that QTLs with effects > δ=1.0 were missed in our analysis, however we might have missed approximately half of the QTLs with effects δ=0.5, and evidently the majority of those with $\delta\leq$0.5. However, because we used interval mapping, our power was slightly higher than the one estimated with Haley's expression.

Weller *et al.*(1990) estimated that in a half-sib experiment, QTLs with δ=0.1 will be very hard to map even with several thousand progenies, but that QTLs with δ=0.3 would

be readily mapped in experiments with 1000-2000 offspring. To increase power for detection they proposed a 'granddaughter' design where progeny tests of half-sib sires from a single elite sire are evaluated in order to decrease the error variance of quantitative trait evaluation. Although power can be significantly increased, particularly for traits of low heritability, this design involves three generations. Furthermore, for any number of daughters, greater power is achieved with fewer sires and more daughters per sire than with more sires with fewer daughters per sire. This design is particularly attractive for animals where large half-sib families from a single sire are difficult to obtain, and highly specialized pedigrees involving multiple generations are available. In forest trees, on the other hand, large half-sib families are available and three generation pedigrees are difficult to obtain.

Recently, we have examined QTL mapping in a full-sib cross between heterozygous individuals (Grattapaglia et $al.$ 1994). Standardized phenotypic difference between marker genotype classes (Δ_1) for the detected QTLs were typically between 0.5 and 0.7, and a few QTLs with $\Delta_1 > 1.5$ were detected. Sample size was only around 100, however two measurements per genotype were made reducing the error variance of trait evaluation. It is difficult to draw a precise comparison between the traits evaluated in the two experiments, however only QTLs of relatively large effects were detected in both. As expected, when we compare the values of Δ_1 and the total proportions of phenotypic variance explained, the within full-sib approach is superior to the half-sib design. However, both depend on the heterozygosity of the QTL and linked marker. While the heterozygosity of the marker should not represent a limitation with RAPD, no prior information is available on the heterozygosity of the QTL. In outbreeding populations, when a QTL has only two alleles at equal frequencies, assuming Hardy-Weinberg equilibrium, on average 50% of parents do not produce segregating gametes at the locus. Therefore the QTL will be segregating only in 75% of the full-sib families and 50% of half-sib families. If gene frequencies are unequal these proportions will be less. However when multiple QTL alleles are present, a larger proportion of individuals will be heterozygous at the QTL. For example with 4 alleles at equal frequencies, in 97.6% of the full-sib and 75% of the half-sib families, the QTL will be segregating. Highly heterogeneous undomesticated populations such as those of forest trees are likely to have multiple alleles at QTLs. Multiple alleles at QTLs have been reported in potato, also an outbred and highly heterozygous species (Van Eck et $al.$ 1994).Therefore it is expected that in trees, the vast majority of QTLs can potentially be detected with a

relatively high probability by analyzing almost any cross. The critical issue will be the relative magnitude of effects of the segregating alleles and the nature of their interactions. Evidently as populations undergo selection, QTL detection is expected to become more difficult as the frequency of favorable alleles with comparable effects increase. However in highly heterogeneous populations we expect to find not only multiple alleles but also multiple QTLs. Thus opportunity will exist to continually uncover important discrete genetic factors controlling traits of interest.

Genetic architecture of quantitative traits in forest trees. To our knowledge this is the first report of quantitative trait locus mapping in trees for traits at harvest age, and the first attempt to understand the genetic architecture of economically important productivity related traits in *Eucalyptus* . Our findings shed some light on the architecture of quantitative traits in forest trees and may have some important implications for the planning of future QTL mapping experiments and eventually for the contemplation of marker assisted breeding. By applying full scale co segregation analysis on a random sample of 300 half-sib individuals, we were able to detect genomic regions significantly associated with the quantitative trait expression of economically important traits in *Eucalyptus* grandis. Given the magnitude of the phenotypic variation explained by the joint action of those regions (Table 7) these results indicate that, genes or groups of closely linked genes with relatively large phenotypic effects ($\delta \sim 0.5$) are involved in the control of the quantitative traits investigated, and can be identified by a half-sib mapping approach.

If we consider that no effects with $\delta > 1.0$ were missed and we were able to detect half of the effects with $\delta = 0.5$ (power of 0.5, see previous section), we may speculate on the total proportion of the phenotypic variance that can be attributed to major genes. Not surprisingly there seems to be good concordance with the expectations based on the relative magnitude of trait heritability. For CBH, a trait of low to intermediate heritability (11% of the phenotypic variation explained, Table 7) a maximum of ~22% of the phenotypic variation in the half-sib examined is controlled by the action of half a dozen major QTL alleles from the maternal parent. This number improves for WSG, a highly heritable trait, where at least four and potentially five independent QTLs were detected accounting for 21% of the phenotypic variation. For WSG we could therefore predict that about 10 major maternal genes are responsible for ~ 42% of the phenotypic variation. Control of %BARK by major genes would be

somewhere between 20 and 40%. For %PULP QTL detection power was considerably lower, as the sample size used was only 164 individuals and the trait heritability is low (0.23) (Demuner and Bertolucci 1992). To detect all of the major genes ($\delta \geq 0.5$) involved in the control of such traits in the *Eucalyptus* pedigree examined, a large experiment involving at least a thousand individuals would be necessary. However only the maternal contribution to the trait could possibly be estimated because in the analysis the paternal contribution can not be accounted for and is assumed to have a normal distribution with an average effect of zero.

If one wishes to dissect the genetic architecture of a quantitative trait - i.e. defining the majority of the discrete factors controlling the trait - in *Eucalyptus* and probably in any forest tree, very large experiments will be needed because several QTLs of minor effects will be present. In this respect, half-sib strategies do not seem adequate. Sample sizes required will be exceedingly large to account for the fact that the quantitative value associated with alternative marker genotypes is measured as the effect of one allelic substitution averaged over all the alleles inherited from the pollen parents. A relatively large number of full-sib families with a few hundred individuals each, would be most efficient to achieve a reasonable probability that most of the major QTL will be found segregating and adequate detection power is achieved.

By fitting multiple QTLs models and their two-way interactions we detected a case of digenic epistasis between unlinked QTLs, that significantly improved the total proportion of the phenotypic variance explained by the detected QTLs. It is worth to point out that we uncovered a significant interaction despite the small number that could be tested in this experiment. QTL mapping studies offer a valuable tool to investigate this source of non-additive genetic variation. Although data from classical genetic testing does not support this notion, epistatic interactions amongst QTLs might prove of considerable importance in the architecture of quantitative traits and the advancement of selection in breeding populations. We have no biological premise on which to base the epistatic effect found in our experiment. However it is interesting to note that at the location of the major QTL involved in the interaction (on group 5), QTLs for CBH, WSG and %BARK were also mapped. It would not be surprising to find that a major gene located in that region is involved not only in the control of all the growth traits analyzed, but also affects the final trait expression by epistatic interactions with other genes. An important regulatory locus would meet such functions.

Marker assisted selection (MAS) in Eucalyptus . Similarly to the pseudo-testcross approach within full-sib families, QTL mapping within half-sib families involves the construction of individual-specific genetic linkage maps and detection of major, individual-specific QTL alleles. In spite of the difficulties for detection of minor QTLs , this approach is particularly attractive because it mitigates the obstacle of linkage equilibrium faced by marker assisted breeding in outbred forest trees (Soller 1978; Beckmann and Soller 1983; Lande and Thompson 1990; Strauss *et al.* 1992). To establish homologies of linkage groups or homologies of QTLs in maps from different elite trees, common marker loci will be needed. The conservation of favorable marker/QTL associations across individuals will depend essentially on the extent of linkage disequilibrium between marker and trait loci in the population. The progressive accumulation of individual linkage maps with subsets of common markers among them will indicate the relationships of QTL in different maps and allow to test for linkage disequilibrium between marker and trait loci in small breeding populations. The existence of generalized genomic regions associated with trait expression could also be verified. Alternatively we may find that a multiplicity of QTL controlling economically important traits exist at the population level. This would give even greater opportunity for the application of MAS.

Two steps are generally necessary for the implementation of MAS: location of QTL and manipulation of them with genetic markers in subsequent generations of selection and recombination. We have shown that the linkage disequilibrium within a half-sib family allows one to identify genomic regions with significant effects on quantitatively inherited traits of relevance to forest productivity in *Eucalyptus* . With this approach one can consider the prospect of performing QTL analysis in any existing half-sib family block plantations in retrospective QTL analyses allowing one to gather the necessary quantitative data on elite individuals in acceptable time. In the single-tree QTL approach adopted, marker/trait associations are established at the individual level, and therefore substantial linkage disequilibria is expected to be maintained. Close linkages established between markers and QTL could be followed for several subsequent generations of selection and recombination. After QTLs for individual trees have been detected in relatively large experiments, the number of markers genotyped in subsequent generations derived from those trees could be substantially reduced as only those particular marker segments containing the QTLs of interest would be tracked. Progeny sizes could then vary depending on the number of genomic regions targeted

at selection to increase the probability of recovering genotypes with the correct QTL allele profiles.

The contemplation of marker assisted breeding should be done on a case-by-case basis (Williams and Neale 1992). In the context of *Eucalyptus* breeding the prospects are positive. Hybrid breeding combined to clonal propagation of selected individuals are increasingly being used. Large amounts of linkage disequilibrium are generated by hybridization and substantial amounts of non-additive genetic variation can be captured by vegetative propagation. These were seen as favorable conditions for MAS (Weller and Fernando 1991; Strauss *et al.* 1992). Individual trees of specialized breeding populations could be efficiently QTL mapped by analyzing the performance of their offspring both full-sibs and half-sibs. From a cross involving one or two QTL mapped trees, efficiency of within-family selection for superior individuals could be significantly improved by the introduction of marker information. Marker information would also be very useful for selecting specific pairs of parents that display complementary QTLs.

In a population of *Eucalyptus urophylla* at the same age and environmental conditions as those in our experiment, a within family, individual tree heritability of 0.31 was estimated for CBH (Rezende and Bertolucci 1993). In our experiment, two QTLs detected for CBH and their interaction accounted for 15% of the phenotypic variance in CBH (Table 9). This translates into an estimated 48% of the additive genetic variance (Soller *et al.* 1976). When the proportion of the additive genetic variance explained by the marker loci exceeds the heritability of the character, selection on the markers alone is more efficient than selection on the individual phenotype (Smith 1967). Plotting our results on the efficiency curves presented by Lande and Thompson (1990), individual tree selection for CBH based on markers would have an efficiency ~1.5 that of phenotypic selection alone.

Besides increasing accuracy of selection at the individual level for low heritability traits, efficiency could be significantly improved by decreasing the generation interval. Furthermore indirect selection for traits difficult to score could also benefit from marker assisted breeding. In these situations not only low heritability traits could be considered but also high heritability traits. WSG is a trait of high heritability. However even in fast growing *Eucalyptus* , WSG requires a few years to reach full expression. Combined early selection for CBH and WSG using markers would allow a significant increase in selection intensity within families at a much younger age and save considerable time in

the establishment of field trials of the individuals tested as clones. With the advent of more cost and time efficient marker technologies and the demonstration of existence of major genes for quantitative traits in trees, the prospects for the use of MAS in forest trees are promising, contrary to earlier expectations (Strauss *et al.* 1992). However they should not be overstated until data is accumulated on realized comparative gains from marker assisted selection versus phenotypic selection.

ACKNOWLEDGMENTS

LITERATURE CITED

Arnheim N, Starnge C, Erlich H (1985) Use of pooled DNA samples to detect linkage disequilibrium of polymorphic restriction fragments and human disease: studies of the *HLA* class II loci. Proc Natl Acad Sci USA **82**:6970-6974.

Beckmann JS, Soller M (1983) Restriction fragment length polymorphisms in genetic improvement methodologies, mapping and costs. Theor Appl Genet **67**: 35-43.

Beever JE, George PD, Fernando RL, Stormont CJ, Lewin HA (1990) Associations between genetic markers and growth and carcas traits in a paternal half-sib family of Angus catle. J Anim Sci **68**: 337-344.

Bradshaw HD Jr., Foster GS (1992) Marker-aided selection and propagation systems in trees: advantages of cloning for studying quantitative inheritance. Can J For Res **22**:1044-1049.

Chalmers KJ, Barua UM, Hackett CA, Thomas WTB, Waugh R, Powell W (1993) Identification of RAPD markers linked to genetic factors controlling the milling energy requirements of barley. Theor Appl Genet **87**: 314-320.

Chaparro J X, Werner DJ, O'Malley DM, Sederoff RR (1994) Targeted mapping and linkage analysis of morphological, isozyme and RAPD markers in peach. Theor Appl Genet **87**: 805-815.

Conkle, M. T. 1981 Isozyme variation and linkage in six conifer species. In: Proceedings of the Symposium on Isozymes in North American Forest Trees and Forest Insects. Pacific Southwest Exp. St. USDA Genet Tech Rep PSW-48. pp. 11-17.

Darvasi A, Soller M (1992) Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. Theor Appl Genet **85**: 353-359.

Darvasi A, Weinreb A, Minke V, Weller JI, Soller M (1993) Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics **134**: 943-951.

Demuner BJ, Bertolucci FLG (1992) Tree breeding: a new approach through evaluation of genetic and phenotypic parameters of eucalypt wood and pulp. In: Proceedings of the 26th Congress of the Brazilian Pulp and Paper Industries.

Dudley JW (1993) Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. Crop Sci **33**:660-668.

Geldermann H (1975) Investigations on inheritance of quantitative characters in animals by gene markers. I. Methods. Theor Appl Genet **46**: 319-330.

Geldermann H, Pieper U, Roth B (1985) Effects of marked chromosome sections on milk performance in cattle. Theor Appl Genet **70**: 138-146.

Grattapaglia D, Bertolucci FLB, Sederoff R (1994) Genetic mapping of quantitative trait loci controlling vegetative propagation in *Eucalyptus* grandis and *E.urophylla* using a pseudo-testcross mapping strategy and RAPD markers. (Chapter VI, his dissertation)

Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *E.urophylla* using a pseudo-testcross mapping strategy and RAPD markers. Genetics (in press).

Grattapaglia D, Wilcox P, Chaparro JX, O'Malley D, McCord S, Whetten R, McIntyre L, Sederoff R (1991) A RAPD map of loblolly pine in 60 days. Third International Congress of Intl. Soc. for Plant Mol. Biol. abs. 2224.

Haley CS (1991) Use of DNA fingerprints for the detection of major genes for quantitative traits in domestic species. Anim Genetics **22**: 259-277.

Ikemori Y, Campinhos E Jr. (1983) Produção de sementes de *Eucalyptus* grandis x *Eucalyptus* urophylla por polinização aberta: resultados preliminares. Silvicultura **8**: 306-308.

Keats BJS, Sherman SL, Morton NE, Robson EB, Buetow KH, Cann HM, Cartwright PE, Chakravarti A, Francke U, Green PP, Ott J (1991) Guidelines for human linkage maps: an international system for human linkage maps (ISLM, 1990). Genomics **9**: 557-560.

Kennard WC, Slocum MK, Figdore SS, Osborn TC (1994) Genetic analysis of morphological variation in *Brassica oleracea* using molecular markers. Theor Appl Genet **87**: 721-732.

Knapp SJ, Bridges WC, Liu BH (1992) Mapping quantitative trait loci using nonsimultaneous and simultaneous estimators and hypothesis tests. In: Beckmann J, Osborn TC (eds) Plant genomes: Methods for Genetic and Physical Mapping. Kluwer Academic Publishers, pp 209-238.

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics **124**: 743-756.

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121**:185-199.

Lander, ES, Green P, Abrahamson J, Baarlow A, Daly MJ, Lincoln SE, Newburg L (1987) MapMaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics **1**:174-181.

Lebowitz RJ, Soller M, Beckmann JS (1987) Trait-based analysis for the detection of linkage between marker loci and quantitative trait loci in crosses between inbred lines Theor Appl Genet **73**: 556-562.

Liu BH, Knapp SJ (1990) GMENDEL: a program for Mendelian segregation and linkage analysis of individual or multiple progeny using log-likelihood ratios. J Hered **81**: 407.

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.

Motro U, Soller M (1993) Sequential sampling in determining linkage between marker loci and quantitative trait loci. Theor Appl. Genet **85**: 658-664.

Nelson, C. D., W. L. Nance and R. L. Doudrick, 1993 A partial genetic linkage map of slash pine (*Pinus elliotti* Engelm var. *elliottii*) based on random amplified polymorphic DNA's. Theor. Appl. Genet. **87**: 145-151.

Nodari RO, Tsai SM, Gilbertson RL, Gepts P (1993) Towards an integrated linkage map of common bean. II. Development of an RFLP-based linkage map. Theor Appl Genet **85**: 513-520.

Rezende GDSP, Bertolucci FLG (1993) Uso da seleção combinada na determinacão da eficiência da seleção precoce em progênies de meios-irmãos de *Eucalyptus urophylla*. In: Proceedings of the 1st Panamerican Forestry Congress.

SAS (1988) SAS/STAT User's Guide, version 6.03. SAS Institute, Cary, NC.

Smith C (1967) Improvement of metric traits through specific genetic loci. Anim Prod **9**: 349-358.

Soller M (1978) The use of loci associated with quantitative traits in dairy cattle improvement. Anim Prod **27**: 133-139.

Soller M (1991) Mapping quantitative trait loci affecting traits of economic importance in animal populations using molecular markers. In: Schook LB, Lewin HA, Mclaren DG (eds) Gene-mapping techniques and applications. Marcel Dekker Inc. pp 21-50.

Soller M, Genizi A (1978) The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. Biometrics **34**:47-55.

Soller M, Genizi A, Brody T (1976) On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. Theor Appl Genet **47**:35-39.

Strauss SH, Lande R, Namkoong G (1992) Limitations of molecular-marker aided selection in forest tree breeding. Can J For Res **22**: 1050-1061.

Stuber, CW (1992) Biochemical and molecular markers in plant breeding. In: Dudley JW, Hallauer AR, Ryder M (eds) Plant breeding reviews Vol. 9. John Wiley & Sons, Inc., pp 37-61.

Tulsieram L K, Glaubitz JC, Kiss G, Carlson J (1992) Single tree genetic linkage mapping using haploid DNA from megagametophytes. Bio/Technology **10**: 686-690.

Van Eck HJ, Jacobs JME, Stam P, Ton J, Stiekema WJ, Jacobsen E (1994) Multiple alleles for tuber shape in dipoid potato detected by qualitative and quantitative genetic analysis using RFLPs. Genetics **137**: 303-309.

Weller JI, Fernando RL (1991) Strategies for the improvement of animal production using marker-assisted selection. In: Schook LB, Lewin HA, Mclaren DG (eds) Gene-mapping techniques and applications. Marcel Dekker Inc. pp 305-328.

Weller JI, Kashi Y, Soller M (1990) Power of "daughter" and "granddaughter" designs for mapping of quantitative traits in dairy cattle using genetic markers. J Dairy Sci **73**: 2525-2532.

Table 2. Results of selective genotyping analysis for CBH: χ^2 on marker data for segregation distortion and t tests for difference between means for CBH of the two genotypic classes.

MARKER	GROUP	TESTS ON EXTREME PHENOTYPES [a]			TESTS ON RANDOM SAMPLE (n=204)	
		χ^2 1:1 (high tail)	χ^2 1:1 (low tail)	P VALUE (t test)	χ^2 1:1	P VALUE (t test)
R15_1650	2	2.7	5.3	0.3	1.96	0.28
X2_1200	2	0.0	0.0	0.24	0.98	0.12
R4_1300	5	0.0	6.8	0.04	1.62	0.004 *
U10_1500	7	0.0	27.0	0.000	1.16	0.3
A11_980	7	1.5	24.0	0.000	0.02	0.02 *
R10_840	8	2.0	5.3	0.2	0.08	0.33

[a] $\chi^2 \geq 3.84$ denotes significant segregation distortion at $\alpha=0.05$
* significant marker-CBH association detected by selective genotyping

Table 3. Summary of significant marker-trait associations (P Values≤ 0.02) detected by simple linear regression (non simultaneous QTL mappping) using GLM (SAS) . Multipoint estimates of total phenotypic variation were obtained by fitting multivariate regression models (see Table 8 for multivariate models).

MARKER	LINKAGE GROUP	CBH	WSG	%BARK	%PULP
K10_835	2	**0.002**	**0.01**		
K10_549		0.02			
J17_3335		0.02			
X1_1450	5	0.001	0.002	0.01	
R4_1300		0.02	**0.000**	**0.001**	
X15_600			0.04	0.02	
K1_1000			0.000	0.008	
M6_961			0.000	0.007	
Y16_1500		0.02	0.004	0.003	
Y17_1500		**0.000**	0.002	0.02	
V7_1200			**0.000**	0.02	
J13_1370	6/9		0.02		
Y15_760			**0.007**		
Y15_650				0.008	
P10_530				**0.004**	
U20_900					0.03 ns
Z18_900					0.04 ns
N15_1079	7	**0.007**	0.004		
A11_980			**0.000**		
Z18_1630	8/12			**0.02**	
R20_1150	14			**0.01**	
Y20_760	3		**0.02**		
R^2 (ALL MARKERS)		0.14	0.30	0.16	-
R^2 **(MARKERS IN BOLD TYPE)**		0.11	0.21	0.12	-

ns = not significant

Table 4. QTL summary for *Eucalyptus grandis* clone 44 as determined by interval mapping analysis using MAPMAKER-QTL. Listed are the locations and magnitudes of effect of QTL controlling growth and wood quality traits.

TRAIT [a]	LINKAGE GROUP	MARKER INTERVAL	QTL [b] POSITION	LOD PEAK	1.0 LOD SUPPORT [c] INTERVAL	% VAR.[d] EXPL.
CBH	2	K10_835 - X2_1200	0.0	1.7	20.0 - K10_835 - X2_1200	3.9
	5	X1_1450 - R4_1300	2.0	3.1	OFF END - X1_1450 - 26.0	6.6
	5	Y17_1500 - V7_1200	0.0	2.4	Y16_1500 - Y17_1500 - 14.0	4.0
WSG	2	K10_835 - X2_1200	0.0	1.9	14.0 - K10_835 - 6.0	3.4
	5	X1_1450 - R4_1300	20.0	4.7	28.0 - R4_1300 - 2.0	10.2
	5	Y17_1500 - V7_1200	12.0	4.2	22.0 - V7_1200 - 12.0	8.4
	6/9	Y15_760 - Y16_650	14.0	3.1	4.0 - Y15_760 - 30.0	9.1
	7	K9_1660 - A11_980	10.0	3.7	16.0 - A11_980 - OFF END	6.1
%BARK	5	X1_1450 - R4_1300	20.0	3.0	28.0 - R4_1300 - X15_600	7.2
	5	M6_961 - Y16_1500	12.0	2.3	K1_1000 - Y17_1500 - 24.0	4.4
	6/9	Y15_650 - P10_530	16.0	3.3	6.0 - Y15_650 - P10_530 - 3.0	9.2
	14	R20_1150 - Y17_1300	0.0	0.9 *	-	1.5
%PULP	6/9	U20_900 - Z18_900	14.0	1.5	30.0 - U20_900 - P10_650	9.9

[a] CBH, circumference at breast height; WSG, wood specific gravity; %BARK, percentage dry weight of bark ; %PULP, percentage cellulose pulp yield
[b] Most likely QTL position corresponding to LOD peak, as estimated by MAPMAKER-QTL; cM distance from leftmost marker of interval
[c] Interval over which the position of the QTL is at most 10 times less likely than the most likley position estimated by MAPMAKER-QTL; from left to right: cM distance from the left , marker segment and cM distance to the right; OFF END= off the end of linkage group.
[d] Percent of the phenotypic variation explained, as estimated by MAPMAKER-QTL
* Significant in the QTLSTAT analysis (P≤ 0.02)

Table 5. QTL summary for *Eucalyptus grandis* clone 44 as determined by least squares interval mapping analysis using QTLSTAT. Listed are the locations and properties of QTL controlling growth and wood quality traits

TRAIT [a]	LINKAGE GROUP	MARKER INTERVAL	P VALUE	WALD [b] STATISTICS	R^2 [c]	GENOTYPE MEANS ±S.D. [d] (+)	(-)	Δ_1 [e]	Δ_2 [f]
CBH	2	K10_549 - J17_3335	0.031 *	4.7	0.09	58.9 ± 1.2	55.0 ± 1.3	0.26	0.10
	5	Y16_1500 -Y17_1500	0.001	12.4	0.20	60.0 ± 1.1	53.8 ± 1.3	0.42	0.18
WSG	2	K10_549 - K10_835	0.115 *	2.5	0.03	473.9 ± 3.8	486.0 ± 4.9	0.34	0.18
	5	X1_1450 - R4_1300	0.001	11.1	0.12	470.9 ± 3.1	486.6 ± 3.1	0.44	0.20
	5	K1_1000 - M6_961	0.000	19.2	0.09	470.2 ± 2.9	488.4 ± 2.9	0.51	0.25
	5	Y16_1500 - Y17_1500	0.002	9.5	0.12	485.2 ± 2.9	471.6 ± 3.2	0.38	0.16
	6/9	Y15_650 - P10_530	0.004	8.2	0.08	472.6 ± 3.0	487.4 ± 3.5	0.41	0.22
	7	U10_1500 - K9_1660	0.121 *	2.4	0.06	483.3 ± 2.9	476.4 ± 3.1	0.19	0.10
%BARK	5	X1_1450 - R4_1300	0.003	9.3	0.12	10.19 ± 0.18	9.36 ± 0.18	0.41	0.26
	5	K1_1000 - M6_961	0.005	8.2	0.07	10.10 ± 0.17	9.40 ± 0.17	0.35	0.22
	6/9	Y15_650 - P10_530	0.000	13.4	0.09	10.17 ± 0.16	9.23 ± 0.19	0.47	0.25
	14	R20_1150 - Y17_1300	0.02	5.3	0.12	9.50 ± 0.16	10.20 ± 0.22	0.35	0.27
%PULP	6/9	U20_900 - Z18_900	0.058 *	3.7	0.09	51.28 ± 0.17	50.85 ± 0.23	0.29	0.11

a CBH, circumference at breast height; WSG, wood specific gravity; %BARK, percentage dry weight of bark; %PULP, percentage cellulose pulp yield

b A Wald statistics of 10.0 is approximately equal to P = 0.001

c Percent of the phenotypic variation explained, estimated as the non-linear regression R^2 (SS_{marker}/SS_{total}) using QTLSTAT

d Estimates of genotype means for the alternative RAPD marker-linked QTL alleles; (+) presence of the RAPD band; (-) absence of band

e Difference between alternative QTL genotypes expressed in phenotypic standard deviations

f Difference between the favorable QTL genotype and the population mean expressed in phenotypic standard deviations

* Significant in the MAPMAKER-QTL analysis (LOD ≥ 1.6)

Table 6. Results of the QTL analysis on the combined data set (n=385). Data for the random sample and the extreme phenotypes used in the selective genotypig analysis were analyzed jointly for the marker data in common, by interval mapping analysis using MAPMAKER-QTL.

TRAIT [a]	LINKAGE GROUP	MARKER INTERVAL	QTL[b] POSITION	LOD PEAK	1.0 LOD SUPPORT [c] INTERVAL	% VAR.[d] EXPL.
CBH	5	X1_1450 - R4_1300	4.0	3.2	OFF END - X1_1450 - 24.0	7.5
	7	K9_1660 - A11_980	12.0	3.9	10.0 - K9_1660- A11_980 - OFF END	5.8
WSG	5	X1_1450 - R4_1300	20.0	4.1	OFF END - R4_1300 - 36.0	8.3
	7	K9_1660 - A11_980	16.0	5.9	4.0 - A11_980 - OFF END	9.0
%BARK	5	X1_1450 - R4_1300	24.0	4.0	26.0 - R4_1300 - 24.0	8.0
	7	K9_1660 - A11_980	10.0	1.7	15.0 - K9-1660 - OFF END	2.9
%PULP	7	K9_1660 - A11_980	10.0	2.1	10.0 - K9_1660 - OFF END	7.4

[a] CBH, circumference at breast height; WSG, wood specific gravity; %BARK, percentage dry weight of bark; %PULP, percentage cellulose pulp yield

[b] Most likely QTL position corresponding to LOD peak, as estimated by MAPMAKER-QTL; cM distance from leftmost marker of interval

[c] Interval over which the position of the QTL is at most 10 times less likely than the most likley position estimated by MAPMAKER-QTL; from left to right: cM distance from the left , marker segment and cM distance to the right; OFF END= off the end of linkage group.

[d] Percent of the phenotypic variation explained, as estimated by MAPMAKER-QTL

Table 7. Summary of the half-sib QTL analysis for growth and wood quality traits following non simultaneous QTL mapping.

	CBH	WSG	%BARK	%PULP
# PUTATIVE QTL MAPPED [a]	3 [b]	5 [c]	3	1
% PHENOTYPIC VARIATION Multipoint MapMaker-QTL	13.7	24.7	12.6	9.9
% PHENOTYPIC VARIATION Multivariate Regression GLM (SAS)	10.9 [d]	21.2	11.6	-

[a]   QTLs mapped were counted as putative only if detected by at least two of the three analytical methods used
[b]   this number would be 2 if results of simultaneous QTL mapping fitting multivariate models with digenic epistasis are considered (Table 9)
[c]   this number would be 4 if results of simultaneous QTL mapping are considered (see Table 8)
[d]   this number would be 15.0 if results of multivariate models with digenic epistasis are considered (Table 9)

Table 8. Simultaneous multilocus models for CBH WSG and %BARK. Summarized analysis of variance tables obtained by multivariate linear regression of the markers declared significantly associated with trait expression by non simultaneous tests. GLM (SAS) was used to fit a model with the simple effects.

Dependent Variable : CBH

Marker	DF	Type III SS	F	P Value	R^2
Y17_1500	1	2270.239078	11.89	0.0007	0.045
N15_1079	1	1284.325440	6.72	0.0101	0.025
K10_835	1	2100.744406	11.00	0.0011	0.040

R^2 0.108501

Dependent Variable: WSG

Marker	DF	Type III SS	F	P Value	R^2
Y20_760	1	3961.04449	3.55	0.0608	0.013
A11_980	1	18048.83536	16.18	0.0001	0.06
V7_1200	1	5942.47625	5.33	0.0219	0.02
R4_1300	1	2723.07481	2.44	0.1196 ns	
Y15_760	1	7610.33572	6.82	0.0096	0.025
K10_835	1	12811.09652	11.49	0.0008	0.042

R^2 0.212567

Dependent Variable: BARK

Marker	DF	Type III SS	F	P Value	R^2
X1_1450	1	15.50369596	4.33	0.0385	0.018
P10_530	1	27.06907336	7.57	0.0065	0.031
R20_1150	1	23.64744748	6.61	0.0108	0.026
Z18_1630	1	19.53259833	5.46	0.0204	0.022

R^2 0.116351

Panel A: Histogram of NUMBER OF INDIVIDUALS vs CIRCUMFERENCE AT BREAST HEIGHT - CBH (cm), with x = 57.4, σ = 14.8, n = 300.

Panel B: Histogram of NUMBER OF INDIVIDUALS vs WOOD SPECIFIC GRAVITY - WSG (Kg/m^3), with x = 479.6, σ = 35.7, n = 300.
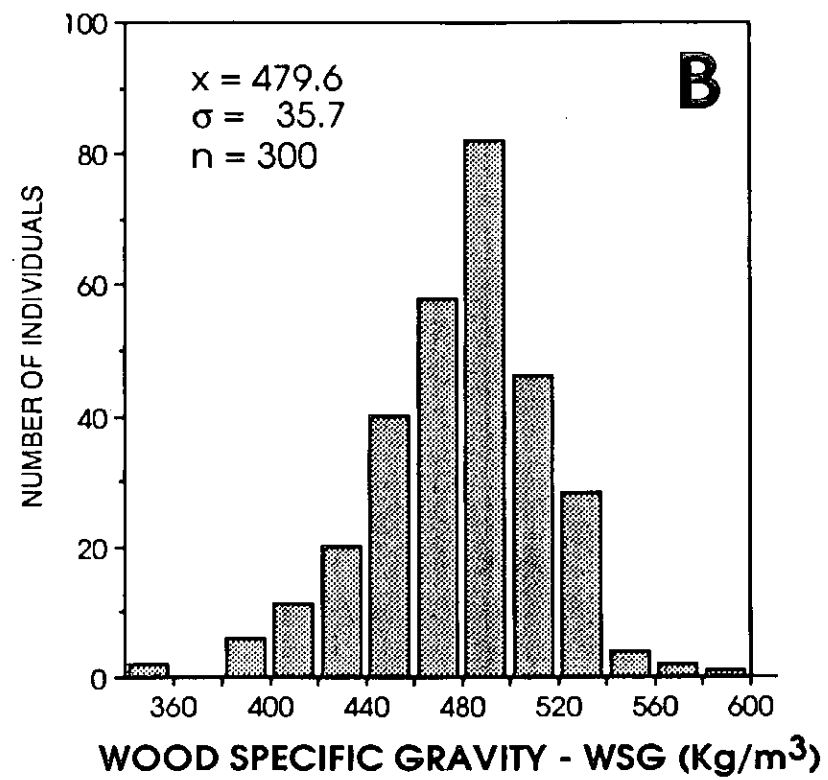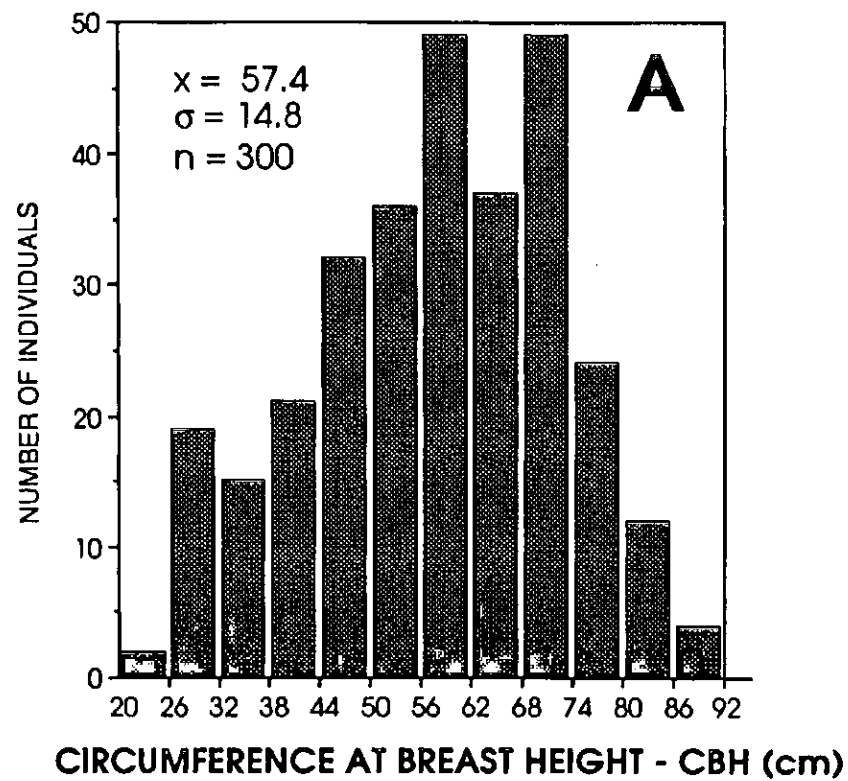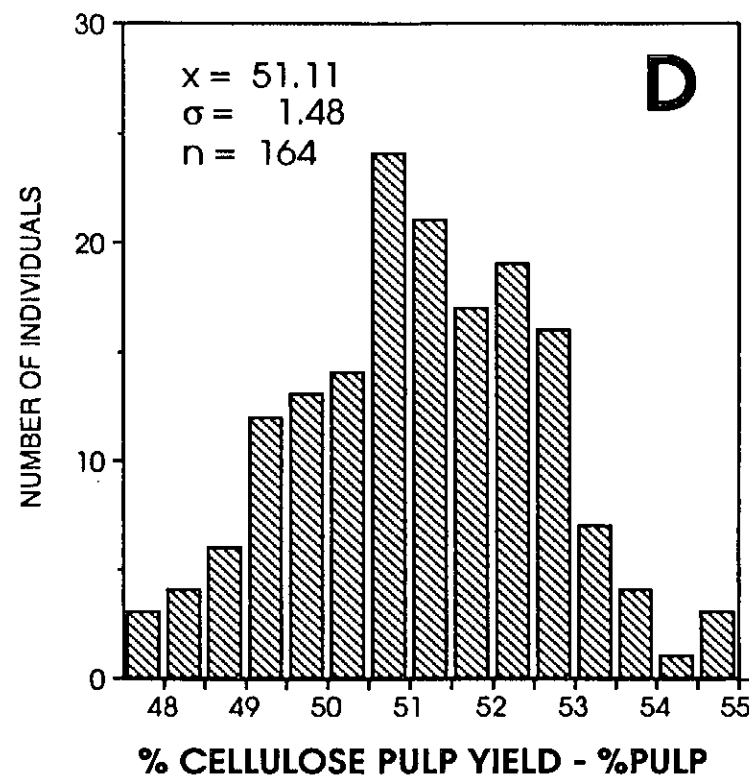
Figure 1. Continued

Figure 2. Spatial distribution of RAPD markers used in the selective genotyping experiment. Markers were selected based on their map location estimated in the original linkage map built for the maternal parent *E. grandis* clone 44 using a pseudo-testcross strategy (Grattapaglia and Sederoff 1994).
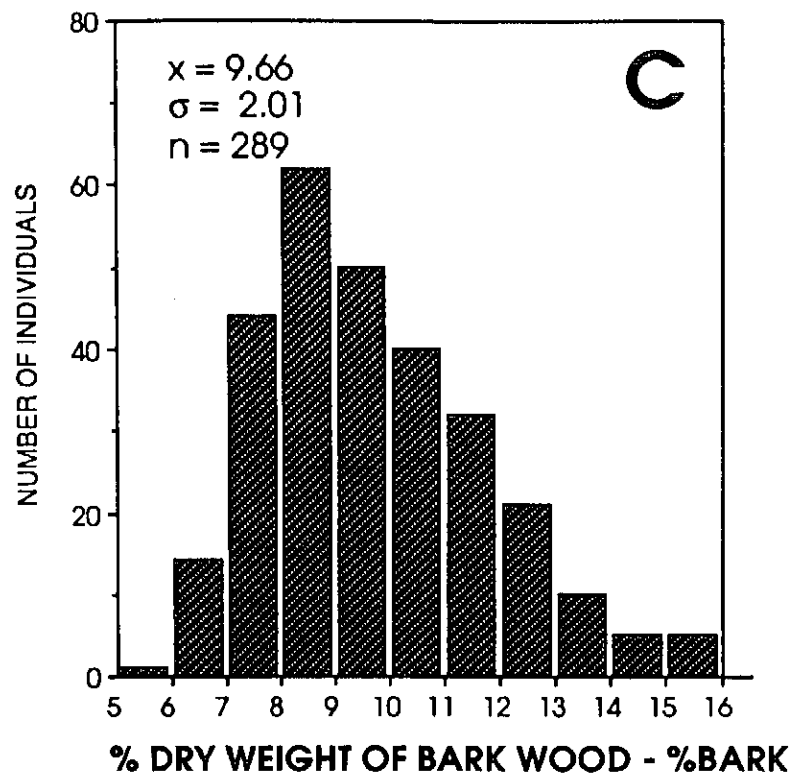
GENETIC LINKAGE MAP OF
Eucalyptus grandis clone 44
Grattapaglia & Sederoff 1994

Marker screened in selective genotyping

Figure 3. Results of Bulk Segregant Analysis for CBH. (A) from left to right: 1 Kb size standard, 4 "low" CBH bulked DNA samples , 4 "high" CBH bulked DNA samples; polymorphic marker U10_1500 between the two groups of replicate bulks is indicated by arrow. (B) gel showing the same marker U10_1500 (indicated by arrows) genotyped for individual samples in the selective genotyping experiment. From left to right, top to bottom, 1 Kb size standard, 48 individuals from the "high" extreme of CBH distribution (> 1.7 σp), 1 Kb size standard, 48 individuals from the "low" extreme of CBH distribution (< 1.7 σp), 1 Kb size standard. Note the 1:1 segregation (24 (+):24(-)) for the marker in the "high" group and the distorted segregation (7 (+):41(-)) (χ2=24.0) in the low group. Dots along the side indicate from top to bottom standard fragments of size 1600 bp and 1018 bp.
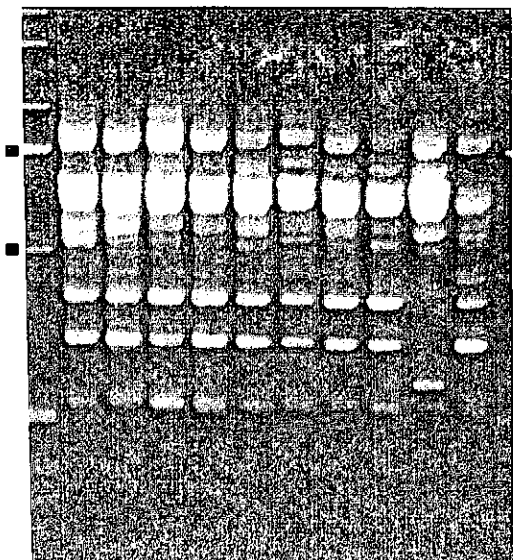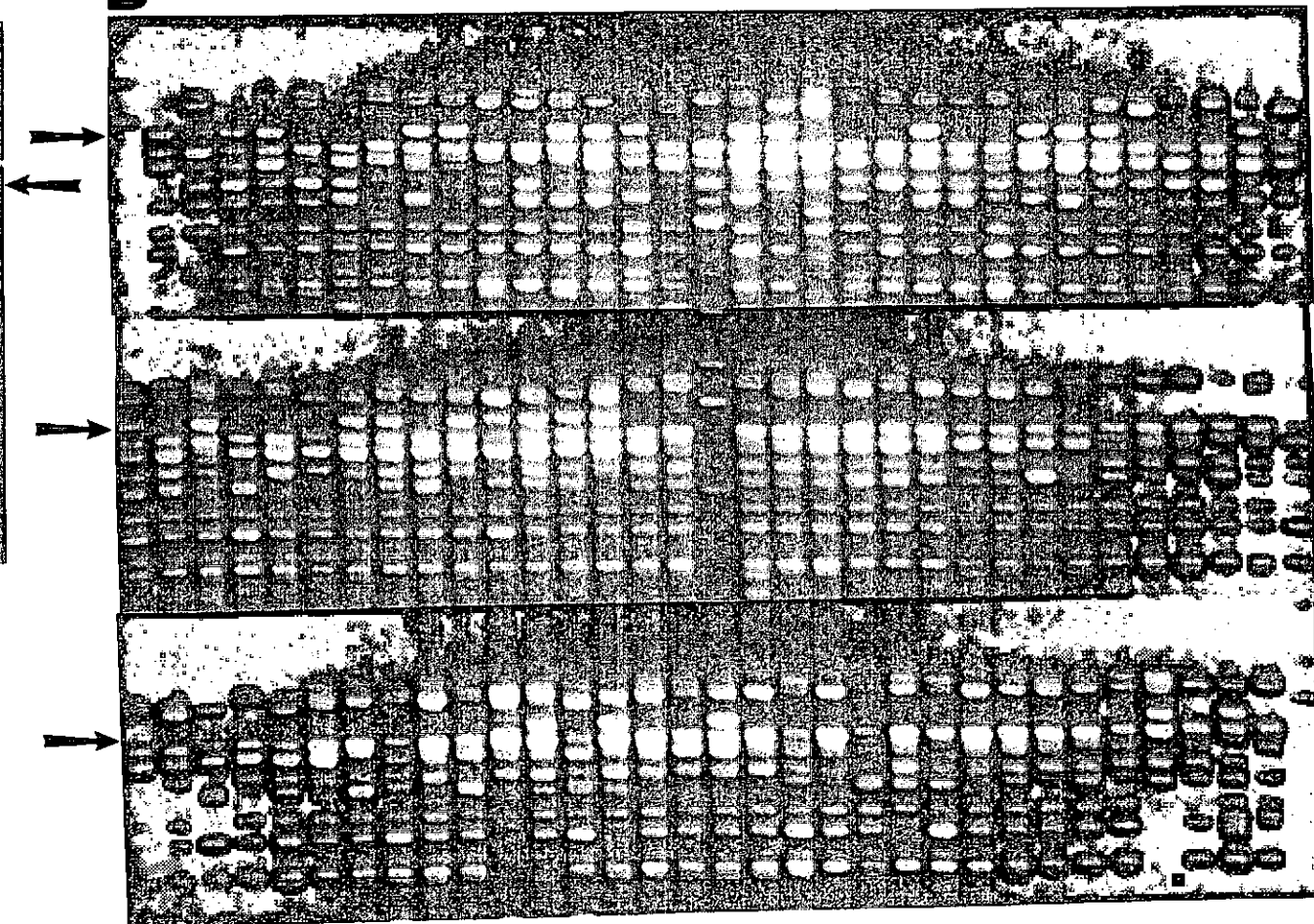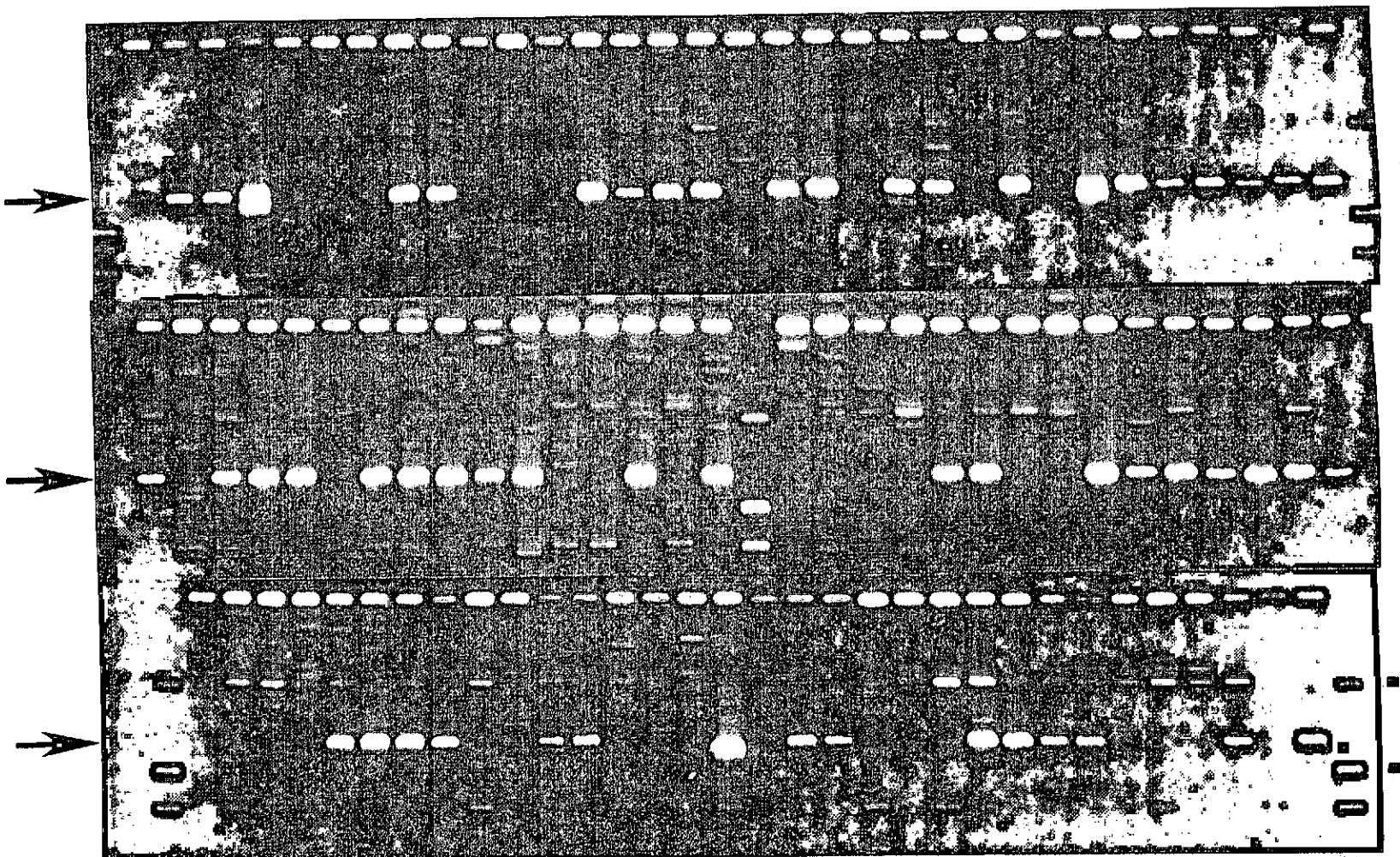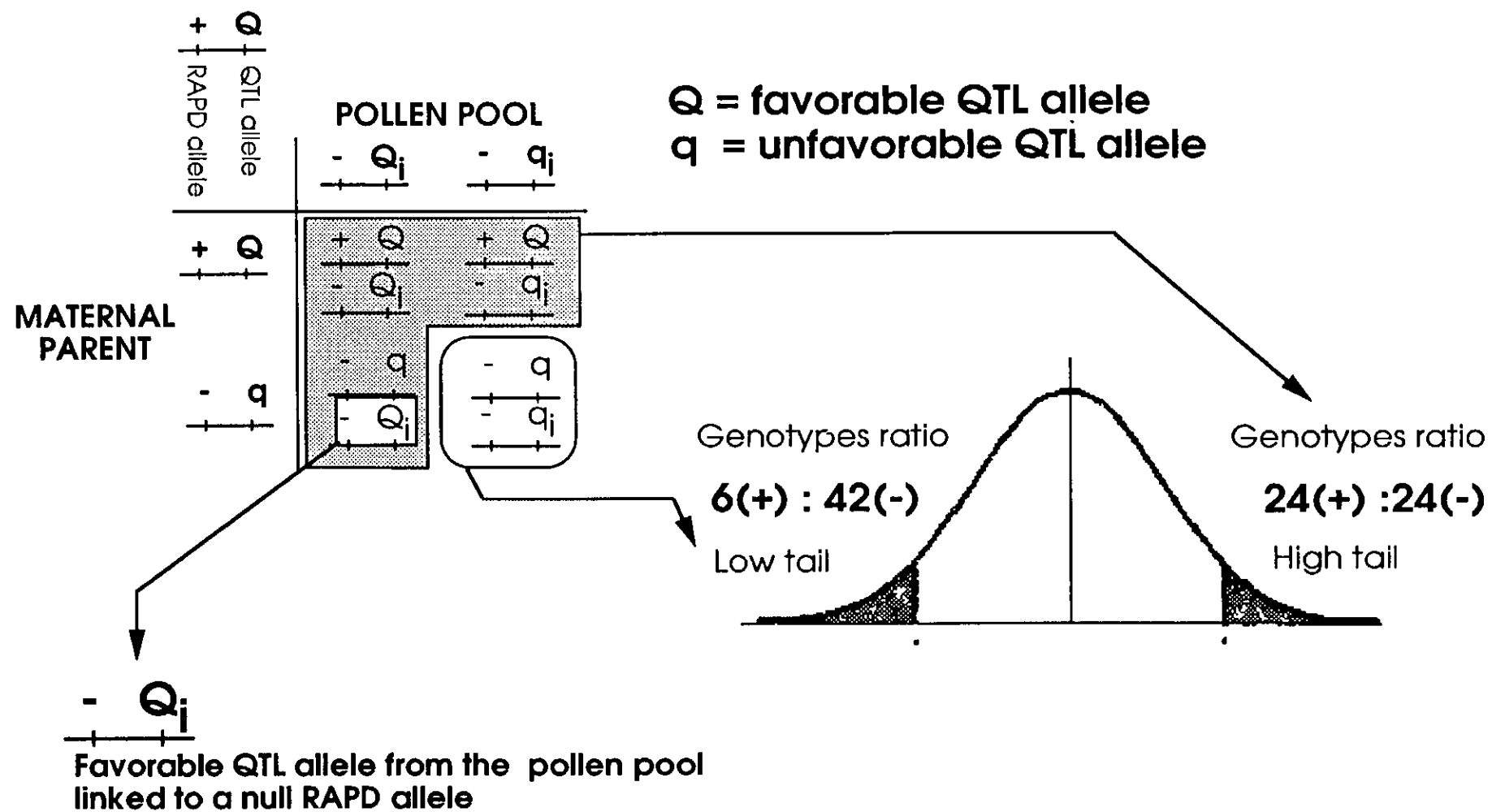
Figure 6. Quantitative Trait Locus (QTL) map of growth and wood quality traits in *Eucalyptus grandis* clone 44. Linkage maps of RAPD markers were constructed using MAPMAKER (LOD 5.0 q=0.30) and GMENDEL (P value=0.0001 q=0.30). RAPD markers were ordered with log-likelihood support $\geq$ 1000:1. Bars to the left of linkage groups correspond to the 1.0 LOD support intervals for the location of the QTL (i.e. the interval over which the QTL position is at most 10 times less likely than the most likely position). Arrows from bars indicate the most likely position (highest LOD peak) estimated with MAPMAKER-QTL. Bar without arrow for %BARK on group 14 corresponds to a QTL not detected by MAPMAKER-QTL, thus no position was estimated within interval. Asterisks indicate QTLs that were not retained when simultaneous QTL mapping was performed (see Tables 8 and 9 and text for details).

GROUP 1 · GROUP 2 · GROUP 3 · GROUP 4 · GROUP 5 · GROUP 6/9

cM

20
40
60
80
100
120
140
160
180
200

GROUP 1
V7_450
X15_1050
Y20_620
K10_1106
V7_400
X15_850
J17_1800

GROUP 2
K10_913
K10_835
X2_1200
K10_549
J17_3335
K10_2700
R15_1650
U20_1000

GROUP 3
Y20_760
Y15_1220
K19_304
J13_407
G13_1150

GROUP 4
R20_2700
R20_1080
K1_1349
K19_1495
X1_720
K10_1645
K19_448
N15_629
K9_884

GROUP 5
X1_1450
R4_1300
X15_600
K1_1000
M6_961
Y16_1500
Y17_1500
V7_1200
A10_635/562

GROUP 6/9
Y20_550
J13_1370
Y15_760/740
Y15_650
P10_530
U20_900
Z18_900
P10_650
U20_850
A11_2374
P10_2600

■ CIRCUMFERENCE AT BREAST HEIGHT (CBH)

▓ WOOD SPECIFIC GRAVITY (WSG)

▨ % WEIGHT OF BARK WOOD (%BARK)

▧ % CELLULOSE PULP YIELD (%PULP)

GROUP 7
Y15_340
U10_1500
N15_1079
K9_1660
A11_980/920

GROUP 8/12
R10_840
Y16_550
Y17_525/515
K9_534
J17_285
Y3_1350
X1_1080
N13_3121
Z18_1630
U13_1650
Y3_1800

GROUP 10
J17_815
A7_550
K1_714
A16_489

GROUP 13/11
Q13_1150
G13_1570
N13_2169
Y15_1600
Q13_540

GROUP 14
R20_1150
Y17_1300
P2_700