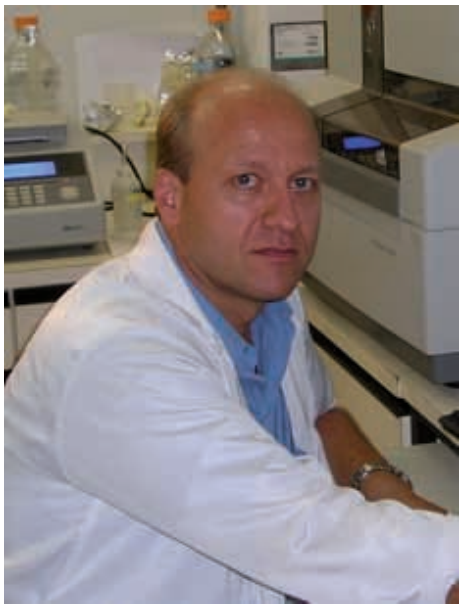
SECTION IV

Marker-assisted selection in forestry – case studies

CHAPTER 14

Marker-assisted selection in *Eucalyptus*

Dario Grattapaglia



SUMMARY

Planted *Eucalyptus* occupies globally more than 18 million hectares and has become the most widely planted hardwood tree in the world, supplying high-quality woody biomass for several industrial applications. In this chapter an overview is presented on the status and perspectives of marker-assisted selection (MAS) in species of *Eucalyptus*. After an introduction to the main features of modern eucalypt breeding and clonal forestry, some applications of molecular markers in support to operational breeding are presented. By reviewing the status of quantitative trait locus (QTL) mapping in *Eucalyptus*, the challenges and some realistic prospects for the application of MAS to improve relevant traits are outlined. With the expected availability of more powerful genomic tools, including a draft of the *Eucalyptus* genome, the main challenges in implementing MAS will be in phenotyping trees accurately, analysing the overwhelming amount of genomic data available and translating this into truly useful molecular tools for breeding.

INTRODUCTION

Planted *Eucalyptus* forests occupy globally more than 18 million hectares and have become the most widely planted hardwood forest tree in the world (FAO, 2001). *Eucalyptus* tree species used in production forestry are long-lived, evergreen species belonging to the angiosperm family Myrtaceae (Ladiges, Udovicic and Nelson, 2003). They are native to Australia and adjacent islands where they occur naturally from sea level to the alpine tree line, from high rainfall to semi-arid zones, and from the tropics to latitudes as high as 43° south (Eldridge *et al*., 1993; Ladiges, Udovicic and Nelson, 2003). Fast growth rates and a wide range of adaptability have contributed to the great interest that *Eucalyptus* species receive 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, pulp and paper manufacture as well as sawn wood. While *E. globulus* is the premier species for temperate zone plantations in Australia, Chile, Portugal and Spain, elite hybrid clones involving *E. grandis* and *E. urophylla* are used extensively by the pulp and paper industry in tropical regions or Brazil, China, the Democratic Republic of the Congo and South Africa because of their wood quality, rapid growth, canker disease resistance and high volumetric yield.

Planted *Eucalyptus* stands supply in a rational and efficient way, high-quality woody raw material that would otherwise come from native tropical forests. In the decades to come, the expansion of these "fibre farms" will likely be limited by the growth of crop plantations and by public opinion pressure. Increased productivity of forests and refinements in the quality of wood products by selective breeding will become of increasing strategic importance to the forest industry. Molecular tools based on the direct identification of useful variation at the DNA level are expected to provide new opportunities for the genetic manipulation of growth, form and especially wood properties of planted trees by marker-assisted selection (MAS) approaches.

Almost fifteen years have passed since the first experiments in molecular breeding of forest trees. The development of linkage maps and quantitative trait loci (QTL) information in trees was greatly accelerated by the advent of more accessible DNA marker techniques, new concepts in linkage mapping and novel strategies for advanced generation tree breeding. From the outset, many expectations were generated for fast and accurate methods for early marker-based selection in trees. Significant progress has been made and the knowledge gathered led to some short-term opportunities for the incorporation of genomic analysis in tree genetics and breeding. However, it also became clear that several challenges remained before more refined and higher impact applications could be implemented.

In this chapter, an overview is presented on the status of MAS in species of *Eucalyptus*. The term MAS is used in *latu sensu*, i.e. encompassing the several molecular techniques and approaches that offer potential to contribute to eucalypt breeding. Some recent reviews have detailed several aspects of *Eucalyptus* genome research including gene discovery, candidate gene mapping, functional genomics and physical mapping (Moran *et al*., 2002; Grattapaglia, 2004; Poke *et al*., 2005; Shepherd and Jones, 2005; Myburg *et al*., 2006). The focus of this chapter is a more applied one, attempting to link the realities of current

eucalypt breeding practice and the molecular tools available or in development. To set the stage for a realistic appraisal of MAS for *Eucalyptus*, a brief introduction is presented of the main features of modern eucalypt breeding and clonal forestry in order to provide a better understanding of the challenges and opportunities that lie ahead for cost-efficient molecular breeding. Following this section, some current low technological input applications of molecular markers in support of operational breeding are presented, such as the quantification of genetic diversity and relationships, the analysis of mating patterns and paternity in seed orchards and fingerprinting for quality assurance and quality control of clonal propagation. Within the framework of MAS for trait advancement, after reviewing the status of QTL mapping in *Eucalyptus*, the challenges and some realistic prospects for the application of MAS to improve relevant traits are outlined. Finally, with the expected availability of a draft of the whole *Eucalyptus* genome within the next years, a succinct summary is presented on the prospects of advancing genomic approaches for gene identification and subsequent application of MAS.

EUCALYPTUS BREEDING AND PLANTATION FORESTRY

Eucalyptus domestication

Eucalypts spread rapidly around the world following their discovery by Europeans in the late eighteenth century (Eldridge *et al.,* 1993). They were introduced into countries such as Brazil, Chile, France, India, Portugal and South Africa in the first quarter of the 1800s (Doughty, 2000) and rapidly adopted in forest plantations as their fast growth and good adaptability became known. During the nineteenth and twentieth centuries, large quantities of seeds were collected and distributed directly from Australia through a number of seed collection expeditions carried out both by government organizations and private forestry companies throughout the world.

Eucalyptus species have a mixed mating system, but are predominantly outcrossers and animal pollinated. High levels of outcrossing are maintained by protandry and various incomplete pre- and post-zygotic barriers to self-fertilization including strong selection against the products of inbreeding (Pryor, 1976). Although the major eucalypt subgenera do not hybridize in nature, hybridization among species within the same subgenus has been detected, often making separation of species difficult (Pryor and Johnson, 1971). Hybridization becomes more frequent in exotic conditions outside the natural species range. In fact, this property has been widely exploited by eucalypt breeders who take advantage of the naturally occurring genetic variation for growth and wood properties among species (de Assis, 2000). Several artificial hybrid combinations have been produced, although hybrid inviability tends to increase with increasing taxonomic distance between the parents (Griffin, Burgess and Wolf, 1988; Potts and Dungey, 2004).

In several countries the continued plantation from local seed sources gave rise to landraces adapted to the specific environment of the country (Eldridge *et al.*, 1993). Seed collections from such local exotic plantings of multiple species became common and where plantings occurred, F_1 hybrids were derived (Potts and Dungey, 2004). While several of these F_1 hybrids performed well, especially when deployed as clones, seed collection from hybrid stands often resulted in plantations that performed poorly in subsequent generations and were

extremely variable. A textbook case is the Rio Claro hybrid swarm in Brazil (Campinhos and Ikemori, 1977; Brune and Zobel, 1981), a eucalypt arboretum where Navarro de Andrade, the "father of eucalypts" in Brazil first introduced and planted a collection of 144 different *Eucalyptus* species between 1904 and 1909. Several of these species hybridized once the natural barriers to introgression were removed in the exotic habitat so that seeds collected from these stands were largely interspecific hybrids. Large commercial plantations were established in Brazil with seeds from this arboretum following fiscal incentives for reforestation granted by the government starting in 1966. Although some of the resulting forests were on average economically inferior, in these very variable stands some outstanding trees for growth, form and disease resistance derived from chance events of recombination were found. The advent of operational cloning techniques at the beginning of the 1980s allowed capturing the superiority of such hybrids that are still used today in some of the most productive eucalypt clonal plantations in the world.

The history of eucalypt breeding, which is short when compared with crop species, was detailed by Eldridge *et al.* (1993) and more recently reviewed and updated by Potts (2004). Some of the earliest breeding was undertaken by French foresters in Morocco in 1954–55 (Eldridge *et al*., 1993). The advent of industrially-oriented eucalypt stands in the 1960s led to a more formal approach to breeding with, for example, the establishment of the Florida *E. grandis* breeding programme in 1961 (Franklin, 1986), *E. globulus* breeding in Portugal in 1965–66 (Potts *et al*., 2004) and large provenance tests of *E. camaldulensis* in many countries (Eldridge *et al.,* 1993). However, a major breakthrough in eucalypt plantation technology occurred in the 1970s with the establishment of the first commercial stands of selected clones derived from hardwood cuttings in the Democratic Republic of the Congo (Martin and Quillet, 1974) followed by Aracruz in Brazil (Campinhos and Ikemori, 1977). At the same time, in many tropical countries such as Brazil and South Africa, efforts were intensified to establish extensive provenance/progeny trials of species such as *E. urophylla*, *E. grandis* and some others that belonged to the same subgenus *Symphyomyrtus* (Eldridge *et al*., 1993). These trials were established from open pollinated seed lots collected from selected trees in the wild and constituted the base populations for subsequent selective breeding in many countries. This initial effort, which was carried out typically by government forestry research institutions, was followed during the 1980s by more intensive collections by private organizations, targeting elite provenances identified in earlier collections as being more adapted for species such as *E. grandis, E. tereticornis* and *E. viminalis* (Eldridge *et al*., 1993).

Eucalyptus breeding and plantation forestry

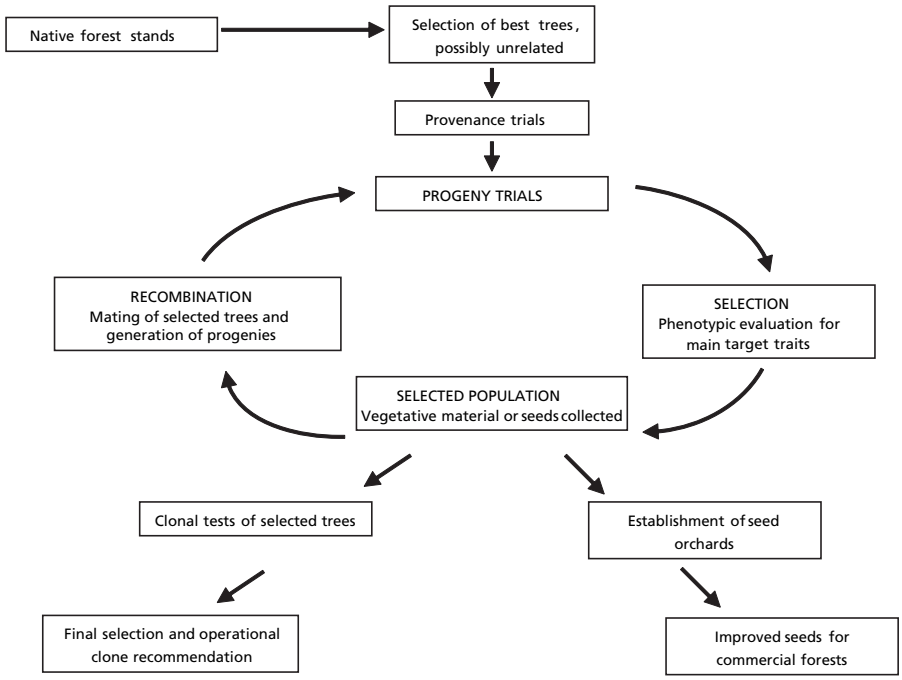
Eucalyptus plantation forestry species are well known for their fast growth, straight form, valuable wood properties, wide adaptability to soils and climates, and ease of management through coppicing (Eldridge *et al*., 1993; Potts, 2004). They are now planted in more than 90 countries where the various species are grown for products as diverse as sawn timber, poles, firewood, pulp, charcoal, essential oils, honey and tannin as well as for shade and shelter (Doughty, 2000). They are an important source of fuel and building material in rural communities in countries such

as China, Ethiopia, India, Peru and Viet Nam. However, it is the increasing global demand for short fibre pulp that has driven the massive expansion of eucalypt plantations and accompanying breeding practices throughout the world during the twentieth century (Turnbull, 1999). Their high fibre content relative to other wood components, coupled with the uniformity of fibres relative to other angiosperm species, has led to high demand for eucalypt pulp for coated and uncoated free-sheet paper, bleach board, sanitary products (fluff pulp), and to a lesser extent for top liners on cardboard boxes, corrugating medium, and as a filler in long fibre conifer products such as newsprint and containerboard (Kellison, 2001). In the last ten years, the development of new wood drying and sawing technologies has also increased interest in using plantation eucalypts for sawn wood, veneer, medium density fibreboard and as extenders in plastic and moulded timber (Kellison, 2001).

An FAO report estimated a total of 17.9 million ha of planted *Eucalyptus* worldwide with India as the largest planter with over 8 million ha followed by Brazil with 3 million (FAO, 2001). The majority of plantations consists of only a few eucalypt species and hybrids. The most important are *E. grandis, E. globulus, E. urophylla* and *E. camaldulensis*, which together with their hybrids account for about 80 percent of the plantation area, followed by *E. nitens, E. saligna, E. deglupta, E. pilularis, Corymbia citriodora* and *E. teriticornis* (Eldridge *et al.,* 1993; Waugh, 2004). Market favourites for pulpwood are *E. grandis, E. urophylla* and their hybrids in tropical and subtropical regions and *E. globulus* in temperate regions.

Although eucalypt breeding is currently a very dynamic and technically advanced operation carried out mainly by several private companies, eucalypts should be seen as still in their domestication infancy when compared with crop species, with most breeding programmes only one or two generations removed from the wild. However, with the combination of ample genetic variation both at the intra and interspecific levels and the ability to clone elite genotypes, eucalypts have quickly become among the most advanced genetic material in forestry. Breeding of eucalypts has moved faster in countries such as Brazil, Chile, Portugal and South Africa that adopted *Eucalyptus* for industrial plantation forestry. Most eucalypt breeding programmes worldwide are focused on genetically improving trees for industrial pulpwood production (Borralho, 2001; Kanowski and Borralho, 2004). The target traits of most breeding programmes include volumetric growth per hectare, wood density and pulp yield (Borralho, Cotterill and Kanowski, 1993). Traits such as pest and disease resistance and adaptability to abiotic stresses such as frost, drought or wind are usually secondary targets that become important when they have an impact on one or more of the main traits. Following the standard concepts in tree breeding, large genetic gains have been obtained in the early stages of eucalypt domestication, simply through species and provenance selection followed by individual selection and establishment of clonal or seedling seed orchards or clonal propagation of elite selections for direct deployment (Eldridge *et al.,* 1993; Kanowski and Borralho, 2004; Potts, 2004). Subsequent population improvement has also demonstrated significant genetic gain through recurrent selection in an open-pollinated breeding population coupled with open or controlled pollinated populations of the most elite selections or specialized

FIGURE 1
Basic breeding scheme used in *Eucalyptus* breeding involving recurrent selection cycles from which elite parent trees are derived for the establishment of seed orchards, as well as individual trees to be tested and later recommended as clones



breeds (Potts, 2004). For species that are easily propagated vegetatively, such as *E. grandis*, *E. urophylla* and several of their hybrids, clonally propagated breeding populations have enhanced gains by allowing the capture of additive and non-additive genetic effects (Figure 1).

Clonal forestry of *Eucalyptus*

After more than 25 years following the introduction of clonal forestry of *Eucalyptus* (Campinhos, 1980; Brandão, Campinhos and Ikemori, 1984), this forest production system is now perfectly integrated into the strategies and plans of advanced generation breeding programmes. Clonal propagation and hybrid breeding have constituted an extremely powerful combination of tools for the rapid improvement of the quality of wood and wood products. While the first hybrid clones were selected based on large-scale screening of high-yielding spontaneous hybrids resistant to diseases (such as the eucalypt canker), today clones are being derived increasingly from deliberate interspecific hybrid production strategies (Figure 2). Eucalypt hybrids, involving two or more species deployed as clones, currently make up a significant proportion of eucalypt plantation forestry, particularly in the tropics and subtropics. In a recent survey of clonal forestry in Brazil, for example, considering all the large and medium-sized companies, the area planted with clones corresponded

FIGURE 2
One-stop pollination of *Eucalyptus*

(A) Elite parent trees, kept as grafts in indoor insect-proof orchards, are induced to flower with growth regulators in approximately 12 to 15 months. (B) Flowers to be used are still closed; open protandric flowers are discarded. (C) Flowers are cut open before anthesis with a nail cutter. (D) Pollen from the other parent is deposited directly at the base of the style and no bag protection is needed as the greenhouse is kept free of insects. (Photographs courtesy of Teotônio F. de Assis)

to more than 1 008 000 ha, involving 362 different clones at a rate of 2 to 40 clones per company, and a range of 10 to 34 000 ha per clone (mean 4 150 ha). The annual introduction of new clonal plantations to support expansion of forest-based industrial production is in the order of 238 000 ha, with a mean of 1 820 ha per clone (de Assis, Rezende and Aguiar, 2005).

An important paradigm shift in eucalypt breeding for pulp and paper began in the 1990s with the increasing realization

that the actual "pulp factory" is the tree. Particularly in vertically integrated pulp production systems, as highly productive clonal forests with over 40 m^3/ha/yr became the standard (Binkley and Stape, 2004), the focus shifted quickly from volume growth to wood quality with the objective of improving pulp yield per hectare by reducing wood specific consumption (WSC), i.e. the amount of wood in cubic metres necessary to produce one tonne of pulp. Trees that yield more cellulose generate savings all the way from tree harvesting, transportation, chipping and pulping while mitigating the need for an accelerated expansion of the forest land base.

Clonal forestry of *E. grandis* x *E. urophylla* selected clones in the 1980s was able to reduce WSC from 4.9 to 4.0 m^3/tonne of pulp (Ikemori, Penchel and Bertolucci, 1994). However, it is now well known by breeders that *E. globulus* has the best combination of wood properties for pulp and paper among the commercially planted *Eucalyptus* species, resulting in a high pulp yield requiring approximately 25 percent less wood to produce the same tonne of cellulose. While only 3.0 m^3 of *E. globulus* wood are required per tonne of pulp, 4 m^3 are needed from selected *E. grandis. E. globulus* has a very adequate wood density in the range of 550 kg/m^3, the longest fibre length and the largest content of holocellulose and pentosans of any other intensively planted species (Sanchez, 2002). *E. globulus*, however, is much more demanding on soil fertility, is not adapted to tropical temperatures, is slower growing and more difficult to propagate clonally than *E. grandis*. In the last ten years, based on the very successful pioneering experiences in Brazil led by Teotônio de Assis, several breeding programmes in tropical countries have started an intensive effort to introgress superior *E. globulus* pulp traits into the tropical and subtropical high yielding genetic backgrounds of *E. grandis* and *E. urophylla*. Given the very high genetic diversity that segregates in such crosses together with the technical possibility of practising intensive within-family selection and clonal propagation, this effort has resulted in the development of exceptional trees that combine superior growth and adaptability to tropical conditions, higher pulp yielding wood and easy propagation using minicutting/hydroponics technology (de Assis, 2000, 2001; Figure 3). A new wave of clonal forestry is therefore starting that will most likely result in another significant jump in the quality of *Eucalyptus* forests.

It is therefore in the context of a highly specialized industrially-oriented breeding programme that fully exploits the power of hybrid breeding and clonal forestry that one needs to discuss the prospects of MAS in *Eucalyptus*. Understanding the fundamental differences between *E. grandis* and *E. globulus* at the molecular level to exploit better the natural allelic variation that exists in the genus has been the starting point.

MARKER-ASSISTED MANAGEMENT OF GENETIC VARIATION IN BREEDING POPULATIONS

The use of genome information for the practice of directional selection of superior genotypes still represents a challenge that depends on further and more refined experimental work (see below). Nevertheless, molecular markers can be used immediately to solve several questions related to the management and identification of genetic variation in breeding and production populations. These applications can be useful essentially to any breeding programme independently of its stage of develop-

FIGURE 3
Selection and clonal propagation of elite trees by the minicutting technology

(A) Elite trees are selected, juvenile sprouts are induced by partial bark stripping while keeping the tree alive. (B) Operational mother plants in hydroponic sand beds from where apical minicuttings are harvested for propagation. (C) Minicuttings are rooted without any use of growth regulators in controlled environment greenhouses. (D) High productivity clonal forest stands that reach over 60 m^3/ha/year.

ment. Although isozyme markers were initially used for these purposes (Moran and Bell, 1983), DNA polymorphisms provide an enhanced level of resolution both at the locus level with much higher expected heterozygosity values and at the genome level with greater coverage. DNA markers provide a powerful tool to quantify existing levels of genetic variation in breeding and production populations of forest trees. Molecular markers can be used to estimate the extent of genetic divergence between individuals selected to compose such populations and resolve several issues of individual identity even at high levels of relatedness, including varietal protection and the verification of alleged parentage in open pollinated breeding systems. Some operational applications of molecular markers for management of genetic variation in *Eucalyptus* are outlined below.

Identification of elite clones

The correct identification of clones is currently the most common application of molecular markers in *Eucalyptus* operational

breeding and production forestry. This application is nowadays routinely used by several forest companies in Australia, Brazil, Portugal, South Africa and Spain. Quality control and quality assurance of large-scale clonal plantation operations become crucial aspects in forestry, especially in vertically integrated production systems where the pulp mill plans on the availability of clones with specific wood properties at specific times. Given the scale of such operations that frequently have to feed plantation programmes of several thousand hectares per year, (i.e. several million seedlings), mislabellings can seriously affect the expected production. Correct clonal identity has also important implications in several breeding procedures such as seed orchard management or controlled pollination programmes affecting the expected gains of breeding cycles.

Several technologies are available today to resolve questions of clonal identity in *Eucalyptus.* Dominant markers such as random amplified polymorphic DNA (RAPD) or amplified fragment length polymorphism (AFLP) have been used for clonal fingerprinting of eucalypts (Keil and Griffin, 1994; Nesbitt *et al.,* 1997; Costa e Silva and Grattapaglia, 1997). Dominant markers are, however, very limited in their ability to establish conclusively the identity of two redundant individual trees due to artefact polymorphisms. Dominant markers can be used to establish that two individuals are not the same, but the statement that two individuals are identical is usually only approximate and no formal test statistics can be attached to this assertion. The high degree of multi-allelism and the very clear and simple co-dominant Mendelian inheritance of microsatellites provide an extremely powerful system for the unique identification of individuals for fingerprinting purposes and parentage testing particularly when the individuals are expected to be related. Kirst *et al.* (2005a) demonstrated the high resolving power of this class of markers in *Eucalyptus*. A breeding population of 192 individuals of *E. grandis* was genotyped with a set of six highly polymorphic microsatellites. The number of alleles detected ranged from 6 to 33 with an average of 19.8±9.2 and the expected heterozygosity averaged 0.86±0.11. Using three loci all 192 genotypes could be readily discriminated. The combined probability of identity (i.e. the probability of two individuals having the same multilocus genotype) considering all six loci was less than one in 2 000 million. Similarity coefficients estimated from microsatellite data were much smaller, thus more discriminative, than those usually obtained in similar studies with RAPD and AFLP markers. In common with human forensic DNA analysis, the standard method for clonal identification in eucalypts today is based on multiplexed, multicolour fluorescent analysis of microsatellite markers sized in an automatic sequencer. The identity of samples is declared based on a maximum likelihood ratio where the likelihood of observing those genetic data conditional on the hypothesis of the two samples being derived from the same clone is compared with the alternative hypothesis, i.e. that the two samples are derived from different clones. Furthermore, the repeatability and precision of multilocus genotype determination allows correct comparisons across laboratories and at different times.

Varietal protection

Following publication of the varietal protection law in Brazil, specific instructions for protecting *Eucalyptus* clones were published in 2002 by the Ministry

of Agriculture of Brazil based on a set of validated morphological descriptors. To the best of the author's knowledge, this is the only country today that has formalized such descriptors, which include 36 morphological traits of leaves, flowers, bark and fruit as well as wood density. Although these descriptors generally satisfy the basic requirements of stability and low environmental influence, they are still difficult to evaluate, especially those related to mature traits in flowers and fruits. Furthermore, it is common that clones are related by common ancestry making their discrimination even more difficult. The high power of discrimination coupled with the general acceptance of DNA technology by eucalypt breeders in Brazil resulted in the inclusion of molecular markers as additional descriptors (Grattapaglia *et al*., 2003). The inclusion of DNA markers represented a remarkable advance that Brazil made in the international landscape of varietal protection of forest trees. Currently all requests for clonal protection are accompanied by a multilocus DNA profile (DNA fingerprint) of 15 to 20 microsatellite markers that were recommended based on several aspects such as robustness, polymorphic information content and general availability in the public domain. The perspective for the following years points to an increased number of applications for clone protection by forest companies in view of the outstanding value of elite eucalypt clones for the maintenance of competitiveness of the forestry-based industry. It can be expected that DNA markers will add a significant power of resolution for distinctness, uniformity and stability (DUS) tests in varietal protection of eucalypt clones, especially when closely related individuals are under scrutiny in legal disputes over clonal property.

Characterization of breeding populations

Breeding populations can be characterized by quantifying the levels and organization of genetic variation within and between breeding groups, sublines and progenies. These data can immediately be used to improve the structure of breeding populations, infuse new material and decide on selection, enrichment or elimination of germplasm entries. In the incomplete pedigree systems frequently used in eucalypts, marker-based systems have been used to monitor the levels of random genetic variation throughout the different cycles of a breeding programme thus allowing much greater flexibility and control over the rate of reduction of genetic variability. For example, RAPD markers were successfully used to characterize the wide range of genetic variation in a germplasm bank of *E. globulus* and thereby assist in the designing of further seed collections (Nesbitt *et al.,* 1995). Gaiotto and Grattapaglia (1997) estimated the distribution of genetic variability within and between open pollinated families of a long-term breeding population of *E. urophylla,* and proposed a selection strategy within and between families for incomplete pedigreed populations based on the incorporation of genetic diversity measures. Marcucci-Poltri *et al*. (2003) used AFLP and microsatellite markers to obtain quantitative estimates of genetic diversity in a *E. dunnii* breeding population selected for fitness to subtropical and cold environments. Molecular data were used to design a clonal seed orchard using the nine most divergent pairs of genotypes, thereby retaining 95.2 percent of the total number of alleles from the 140 polymorphic AFLP loci and the four microsatellite loci analysed. In a subsequent study, Zelener *et al.* (2005) selected *E. dunni* trees using trait selection index

and genetic diversity measures estimated from AFLP and microsatellites. Genetic differentiation estimates consistently showed low differentiation among provenances and great differentiation among families suggesting that orchard design should be based on individual or family selection rather than on provenance selection.

Mating and deployment designs based on genetic distance

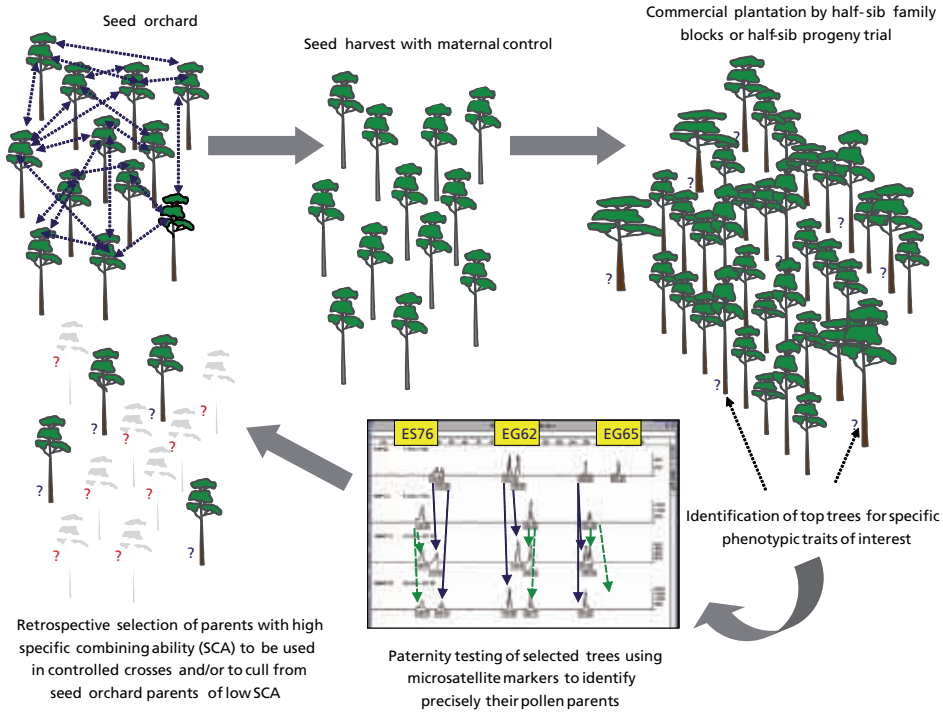
Given the wide genetic diversity and multiple sources of available germplasm for eucalypt breeding, choices typically have to be made as to which elite parents should be mated. Some selection based on the individual's own performance or on pedigree information is used before including it in a mating design. Any means of predicting tree performance deserves attention. One of the "holy grails" of molecular breeders has been the ability to predict progeny performance accurately based on distance estimates among parents from genetic marker data. Vaillancourt *et al.* (1995a) used genetic distances based on RAPD markers to predict heterosis in *E. globulus* progenies. The ability of genetic distance to predict heterosis was significant but accounted for less than 5 percent of the variation in specific combing ability. Baril *et al.* (1997) used the structure of RAPD genetic diversity within and between *E. grandis* and *E. urophylla* to work out prediction equations for the tree trunk volume of individual hybrids at 38 months. Surprisingly, this study showed that a genetic distance based on RAPD markers with similar frequencies in the two species successfully predicted the value of a cross. Through this model, the distance calculated between species explained the general combining ability and the specific combining ability of volume growth with a global coefficient of determination of 81.6 percent. RAPD markers were used to recommend more divergent crosses in a reciprocal recurrent selection programme for hybrid breeding in Brazil (Ribeiro, Bertolucci and Grattapaglia, 1997). A set with the 20 most and 20 least divergent crosses between populations was recommended. Matings between more divergent individuals will potentially allow segregation to be maximized in the resulting progenies and transgressive segregants to be recovered and used as clones.

RAPD data were used to quantify relatedness among elite eucalypt clones for deployment purposes. As the history of selective breeding in eucalypts is very recent, little, if any, pedigree information is typically available. Furthermore, clonal plantations of *Eucalyptus* generally involve only a few superior genotypes of unknown origin. Costa e Silva and Grattapaglia (1997) used RAPD markers to quantify the genetic relatedness among a group of 15 elite clones. Comparative similarity analyses showed that there was significantly more genomic variation in the group of clones than both within and between unrelated half-sib families from a single species. Data on genetic similarity among clones were also used to propose a deployment strategy in a "genetic mosaic", i.e. avoiding planting more genetically related clones side by side in contiguous forest blocks. This proposed strategy was based on the premise that related clones share a common origin and ancestry, have been subject to similar evolutionary selective pressures, and therefore share common susceptibility/tolerance alleles at pest and pathogen defence loci.

Mating system and paternity in breeding populations

Open pollinated breeding by controlling exclusively the maternal progenitor and

FIGURE 4
Retrospective selection strategy of elite parents based on paternity testing of progeny individuals displaying superior performance using microsatellite markers



half-sib progeny testing is still common practice in some eucalypt breeding programmes. It is a low cost option that allows good estimation of the breeding value of maternal parents. Nevertheless, a large amount of genetic variation is usually encountered within half-sib families and selection intensity within families is limited by the number of individuals usually deployed in a progeny test. Knowledge of outcrossing versus selfing rates is essential for maintaining adequate levels of genetic variability for continuous gains. A number of studies have shown that eucalypts are preferentially outcrossed both in natural populations as well as seed orchards. Isozyme markers were originally used (Moran, Bell and Griffin, 1989), but other types of markers now provide a much higher level of resolution. Outcrossing rate in an open pollinated breeding population of *E. urophylla* was estimated at 93 percent using RAPD markers, indicating predominant outcrossing and maintenance of adequate genetic variability within families (Gaiotto, Bramucci and Grattapaglia, 1997). A complex pattern of mating was described in a *E. regnans* seed orchard in Australia where gene dispersal was influenced by crop fecundity and orchard position of mother trees with approximately 50 percent of effective pollen gametes coming from males more than 40 metres away from mother trees (Burczyk *et al.*, 2002). In a

detailed mating system study in a *E. grandis* orchard in Madagascar, the outcrossing rate was found to be 96.7 percent but a pollination rate from outside the seed orchard of 39.2 percent was estimated based on six microsatellite markers (Chaix *et al*., 2003).

The ability to determine paternity precisely using DNA markers was recently proposed as a short-term breeding tactic for *Eucalyptus*. The conventional way to drive modifications in old forest tree seed orchards is to establish progeny trials involving each parent tree and then evaluate its contribution to the performance of the progeny by estimating its general and specific combining ability (GCA and SCA). Grattapaglia, Ribeiro and Rezende (2004) successfully applied an alternative retrospective parent selection tactic based on paternity testing of superior offspring. After identifying seed mixtures, selfed individuals and offspring sired by pollen parents outside the orchard, one particular pollen parent was found to have sired significantly more high-yielding progeny trees. Based on these results, low reproductive success parents were culled from the orchard and management procedures were adopted to minimize external pollen contamination. A significant difference ($p < 0.01$) in mean annual increment was observed between forest stands produced with seed from the orchard before and after selection of parents and revitalization of the orchard. An average realized gain of 24.3 percent in volume growth was obtained from the selection of parents as measured in forest stands at age two to four years. The marker-assisted tree breeding tactic efficiently identified top parents in a seed orchard and resulted in an improved seed variety. It should be applicable for rapidly improving the quality of output from seed orchards especially when the breeder is faced with an emergency demand for improved seeds (Figure 4).

MOLECULAR BREEDING

Molecular markers and maps for *Eucalyptus*

In the last ten years a number of studies have reported genetic maps for *Eucalyptus* built from combinations of several hundred RAPD, AFLP or RFLP markers (Grattapaglia and Sederoff, 1994; Verhaegen and Plomion, 1996; Marques *et al.,* 1998; Myburg *et al.,* 2003), together with RFLP, isozymes, EST, genes and some microsatellites (e.g. Byrne *et al*., 1995; Gion *et al.,* 2000; Bundock, Hayden and Vaillancourt, 2000; Thamarus *et al.,* 2002; Brondani *et al.,* 2002). In contrast to crop species where mapping populations are designed based on contrasting inbred lines, map construction in eucalypts has relied on available pedigrees drawn from operational breeding programmes. These pedigrees generally involve only the highly heterozygous parents and their F_1 progeny, either full-sibs of half-sibs. Genetic mapping has therefore been carried out using a pseudo-testcross strategy, analysing dominant markers present in one parent and absent in the other (Grattapaglia and Sederoff, 1994). Maps are therefore individual-specific and cannot be aligned or integrated as such unless other markers common to both maps are also used. Consequently, although some genome maps of eucalypts have been constructed, the use of the linkage information tends to remain restricted to the pedigree employed as the mapping population, limiting the interexperimental sharing of linkage mapping and QTL data generated.

It is now well accepted that true advancements in QTL validation across pedigrees for the eventual practice of MAS in *Eucalyptus*, will strongly depend on the availability of

higher-throughput, higher polymorphism typing systems such as microsatellites, organized in dense genetic maps (Brondani *et al.,* 1998; Thamarus *et al.,* 2002). Only 137 autosomal microsatellite markers have been published to date for species of *Eucalyptus*, including 67 from *E. globulus, E. nitens, E. sieberi* and *E. leucoxyon* (Byrne *et al.,* 1996; Steane *et al*., 2001; Glaubitz, Emebiri and Moran, 2001; www.ffp.csiro.au/tigr/molecular/eucmsps.html; Ottewell *et al.*, 2005) and 70 from *E. grandis* and *E. urophylla* (Brondani *et al*., 1998; Brondani, Brondani and Grattapaglia, 2002). Recently a set of 35 chloroplast DNA microsatellites was developed based on the full cp-DNA sequence of *E. globulus* (Steane, Jones and Vaillancourt, 2005). Microsatellite transferability across species of the subgenus Symphyomyrtus, which includes all the most widely planted species, varies between 80 and 100 percent depending on the section to which they belong. It still remains around 50 to 60 percent for species of different subgenera such as Idiogenes and Monocalyptus and goes down to 25 percent for the related genus Corymbia (Kirst *et al.,* 1997). Microsatellite comparative mapping data have also shown that genome homology across species of the same subgenus Symphyomyrtus is very high, not only in terms of microsatellite flanking sequence conservation, but also marker order along linkage maps (Marques *et al.,* 2002). Although some tens of microsatellites have been mapped on existing RAPD and AFLP framework maps (Brondani, Brondani and Grattapaglia, 2002; Marques *et al.,* 2002; Thamarus *et al.,* 2002), the genus *Eucalyptus* still lacks a more comprehensive genetic map widely useful for molecular breeding practice. To fill this gap, a novel set of 230 new microsatellites has recently been developed and a consensus map assembled covering at least 90 percent of the recombining genome of *Eucalyptus.* This map has 234 mapped loci on 11 linkage groups, an observed length of 1 568 cM and a mean distance between markers of 8.4 cM (Brondani *et al.,* 2006). This represents an important step forward for *Eucalyptus* comparative genomics, opening stimulating perspectives for evolutionary studies and molecular breeding applications. The generalized use of an increasingly larger set of interspecific transferable markers and consensus mapping information will allow faster and more detailed investigations of QTL synteny among species, validation of QTL and expression-QTL across variable genetic backgrounds, and positioning of a growing number of candidate genes co-localized with QTL, to be tested in association mapping experiments.

QTL mapping in *Eucalyptus*

Following the construction of linkage maps, several groups have reported the identification of genomic regions that have a significant effect on the expression of economically important traits in *Eucalyptus*. QTL mapping experiments have, without exception, found a few major effect QTL for all traits considered in spite of the limited experimental precision, the lack of pre-designed pedigree to maximize phenotypic segregation, and the relatively small segregating populations evaluated. This can be explained by the undomesticated nature and wide genetic heterogeneity of eucalypts added to the fact that most QTL mapping experiments were carried out in interspecific populations thus taking advantage of contrasting gene pools. QTL for juvenile traits such as seedling height, leaf area and seedling frost tolerance have been mapped (Vaillancourt *et al*., 1995b; Byrne *et al.,* 1997a, b), while traits related to vegetative

propagation ability such as adventitious rooting, stump sprouting and *in vitro* shoot multiplication have also been detected (Grattapaglia, Bertolucci and Sederoff, 1995; Marques *et al.,* 1999), as has a major QTL for early flowering (Missiaggia, Piacezzi and Grattapaglia, 2005). In addition, QTL for insect resistance and essential oil traits were mapped (Shepherd, Chaparro and Teasdale, 1999) and recently a major QTL for *Puccinia psidii* rust resistance with quasi Mendelian inheritance was found and mapped in *E. grandis* (Junghans *et al.,* 2003). Major QTL were also found for rotation age traits such as volume growth, wood specific gravity, bark thickness and stem form (Grattapaglia *et al*., 1996; Verhaegen *et al*., 1997; Thamarus *et al*., 2004; Kirst *et al.,* 2004, 2005b).

Although QTL of relatively large effects have been detected for growth traits, when it comes to potential application in MAS the best opportunities for QTL mapping are those related to specialized wood properties that have a direct impact on industrial processes. These traits are usually difficult to measure both because they require destructive whole stem sampling and because they are traits that are expressed late. Myburg (2001) demonstrated the application of indirect, high-throughput phenotyping of wood quality traits in *Eucalyptus* by near infrared reflectance spectroscopy (NIRS) for QTL mapping in a hybrid *E. grandis* x *E. globulus* backcross population. Approximately 300 individuals that had been previously genotyped with AFLP markers were analysed by NIRS, and predictions made for pulp yield, alkali consumption, basic density, fibre length and coarseness, and several wood chemical properties (lignin, cellulose and extractives). A variety of molecular marker classes and pedigree types were used in these experiments. QTL were detected in F_1, inbred or outbred F_2 and half-sib families with or without clonal replicates. Also looking at wood quality traits, Thamarus *et al.* (2004) used novel high-throughput and traditional methods to quantify wood density, fibre length, pulp yield and microfibril angle (MFA) in two full-sib families of *E. globulus* that shared a common parent. Pulp yield and cellulose content were determined by NIRS, and MFA was quantified by SilviScan. Except for fibre length, QTL for all traits could be detected in both populations, including three QTL in common genetic regions on both crosses for wood density, one for pulp yield and one for MFA. The proportion of phenotypic variation explained by the QTL identified in both crosses ranged from 3.2 to 15.8 percent.

Recently QTL analysis of transcript levels of lignin-related genes showed that their mRNA abundance is regulated by two genetic loci co-localized with QTL for growth, suggesting that the same genomic regions are regulating growth, lignin content and composition (Kirst *et al.,* 2004). In a subsequent study, Kirst *et al*. (2005b) showed that one identified expression QTL explained up to 70 percent of the transcript level variation for over 800 genes and that hotspots with co-localized expression QTL were identified on single tree AFLP typically containing genes associated with specific metabolic and regulatory pathways, suggesting coordinated genetic regulation. The correlation of gene expression profiles in segregating progeny can also extend knowledge about genes involved in these pathways. Complementary DNAs representing previously uncharacterized or hypothetical genes, whose transcript levels are strongly correlated with those of genes with known functions, may be associated with the same pathway or biological process.

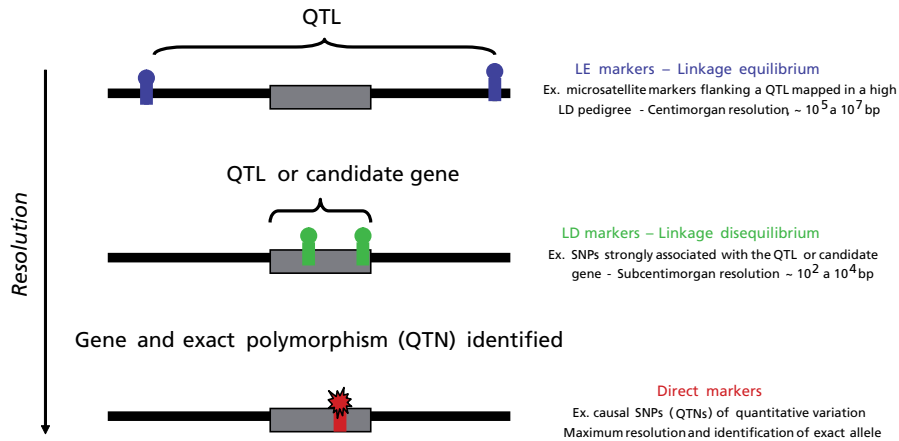
Similarly, new functions can tentatively be assigned to previously characterized genes that had not been described in the context of revealing pleiotropic action of these genes. However, a major limitation in this type of study in *Eucalyptus* is the lack of a completed genome sequence because, without this, the relative locations of large numbers of genes and their expression QTL cannot be determined. This information is required to assess whether the genetic control of gene expression variation is in the cis- or trans-form for each gene. In the context of MAS in *Eucalyptus*, a better understanding of such "master expression QTL" that apparently control cascades of gene expression of important biochemical pathways may be very promising targets for detailed characterization in association mapping experiments to uncover relevant polymorphisms to be used in molecular breeding practice.

In summary, although the number of reports detecting QTL in *Eucalyptus* has grown and these have become increasingly sophisticated, the large majority of mapped QTL have been localized on RAPD or AFLP maps. Consequently, it is impossible to compare positions of QTL for the same or correlated traits, seriously limiting the long-term value of such mapping for MAS. Exceptions are QTL studies where transferable markers such as a few microsatellites (Marques *et al*., 2002; Thamarus *et al.,* 2004) or candidate genes (Gion *et al.,* 2000; Thamarus *et al.,* 2004) were also mapped so that it is at least possible to make a rough preliminary comparison of QTL locations at the linkage group level. Especially in the genus *Eucalyptus* where breeders worldwide take advantage of interspecific genetic variation for wood properties and disease resistance through hybridization, the recent availability of a robust, genus-wide genetic map with highly transferable microsatellite markers (Brondani *et al*., 2006) should stimulate improved genomic undertakings including QTL validation across pedigrees, co-localization of QTL and candidate genes for guiding association mapping experiments, positional cloning of QTL and eventually MAS.

MAS in *Eucalyptus*

Twenty years have passed since the first demonstrations that QTL for major effects could be mapped with molecular markers (Stuber *et al*., 1980; Paterson *et al.,* 1988; Lander and Botstein, 1989), and several reviews have described the potential benefits and caveats of MAS in the plant genetics literature (e.g. Tanksley, 1993; Beavis, 1998; Young, 1999; Mauricio, 2001; Dekkers and Hospital, 2002). Yet, large-scale operational MAS is still largely restricted to very few crops and for very specific applications. Maize is probably the best example, where the financial returns on hybrid seed development coupled with the ability to control germplasm fully, has prompted large-scale investments in MAS by the private sector based on high-throughput single nucleotide polymorphism (SNP) genotyping platforms. Based on a detailed understanding of the molecular architecture of quantitative traits, current applications include yield oriented advanced backcross QTL (AB-QTL) systems as well as accelerated line conversion following trait introgression by marker-assisted backcrossing (MABC). In *Eucalyptus* and forest tree breeding in general, the application of molecular markers for directional selection is still an unfulfilled promise. This is largely due to: the recent domestication of tree crops and hence the wide genetic heterogeneity of breeding populations; the inability to develop inbred lines at least on a short-term basis to allow a

FIGURE 5
Classification of three different types of marker-trait associations relevant to *Eucalyptus* MAS (see text for details)



more precise understanding of genetic architecture of quantitative traits; the absence of simply inherited traits that could be immediately and more easily targeted; and finally to the very limited number of scientists actually working on forest trees.

If applying MAS in other intensively studied crops besides maize is already a significant challenge; this challenge is even more difficult and complex for *Eucalyptus* and forest trees in general as it pre-supposes: (i) the manipulation of polygenic traits with variable heritabilities in breeding populations with a heterogeneous genetic base and in linkage equilibrium; (ii) its incorporation in breeding schemes that involve altering the frequencies of favourable alleles through recurrent selection in large populations; and (iii) dealing with age x age trait correlations, and late expressing phenotypes (Grattapaglia, 2000). In applying MAS for forest trees, more will likely be learned from experiences in livestock (Dekkers, 2004; Chapter 10) than from annual crop plants, with the added advantage that gains can be quickly realized by large-scale cloning of selected individuals. In this context, the categorization of three different levels of marker-trait association described by Dekkers (2004) are relevant to trees: (a) direct markers, i.e. loci that code for the functional mutation; (b) linkage disequilibrium (LD) markers: loci that are in population-wide LD with the functional mutation; (c) linkage equilibrium (LE) markers: loci that are in population-wide LE with the functional mutation in outbred populations (Figure 5). In forest trees, besides the recent encouraging discovery of an LD marker for MFA in *Eucalyptus* (Thumma *et al.*, 2005), only LE marker-trait associations have been described. LE markers have been readily detected on a genome-wide basis by analysing large full-sib families with sparse marker maps allowing the detection of most QTL of moderate to large effects. For the other two types of marker-trait association, it is only now that the first association mapping experiments are being started to uncover