



Tansley review

Eucalyptus applied genomics: from gene sequences to breeding tools

Author for correspondence:

Dario Grattapaglia

Tel: +55 61 3448 4652

Fax: +55 61 3340 3624

Email: dario@cenargen.embrapa.br

Received: 23 October 2007

Accepted: 4 April 2008

Dario Grattapaglia^{1,2} and Matias Kirst³

¹Plant Genetics Laboratory, Embrapa – Genetic Resources and Biotechnology, Parque Estação Biológica, Brasília 70770-970 DF, Brazil; ²Graduate Program in Genomic Sciences and Biotechnology, Universidade Católica de Brasília – SGAN 916 módulo B, Brasília 70790-160 DF, Brazil; ³School of Forest Resources and Conservation, Graduate Program in Plant Molecular and Cellular Biology, and University of Florida Genetics Institute, University of Florida, PO Box 110410, Gainesville, FL 32611, USA

Contents

Summary	911	VI. Gene discovery and genetical genomics	916
I. Introduction	912	VII. Association mapping	918
II. <i>Eucalyptus</i> biology and domestication	913	VIII. Molecular breeding	920
III. <i>Eucalyptus</i> breeding and clonal forestry	913	IX. From gene sequences to breeding tools	922
IV. Marker-assisted management of genetic variation in breeding populations	914	X. Future developments and challenges	924
V. Genetic mapping and quantitative trait locus (QTL) analysis	915	Acknowledgements	926
		References	926

Summary

Key words: association genetics, *Eucalyptus*, genomics, molecular breeding, molecular markers, quantitative trait locus (QTL) mapping.

Eucalyptus is the most widely planted hardwood crop in the tropical and subtropical world because of its superior growth, broad adaptability and multipurpose wood properties. Plantation forestry of *Eucalyptus* supplies high-quality woody biomass for several industrial applications while reducing the pressure on tropical forests and associated biodiversity. This review links current eucalypt breeding practices with existing and emerging genomic tools. A brief discussion provides a background to modern eucalypt breeding together with some current applications of molecular markers in support of operational breeding. Quantitative trait locus (QTL) mapping and genetical genomics are reviewed and an in-depth perspective is provided on the power of association genetics to dissect quantitative variation in this highly diverse organism. Finally, some challenges and opportunities to integrate genomic information into directional selective breeding are discussed in light of the upcoming draft of the *Eucalyptus grandis* genome. Given the extraordinary genetic variation that exists in the genus *Eucalyptus*, the ingenuity of most breeders, and the powerful genomic tools that have become available, the prospects of applied genomics in *Eucalyptus* forest production are encouraging.

New Phytologist (2008) **179**: 911–929

© The Authors (2008). Journal compilation © *New Phytologist* (2008)

doi: 10.1111/j.1469-8137.2008.02503.x

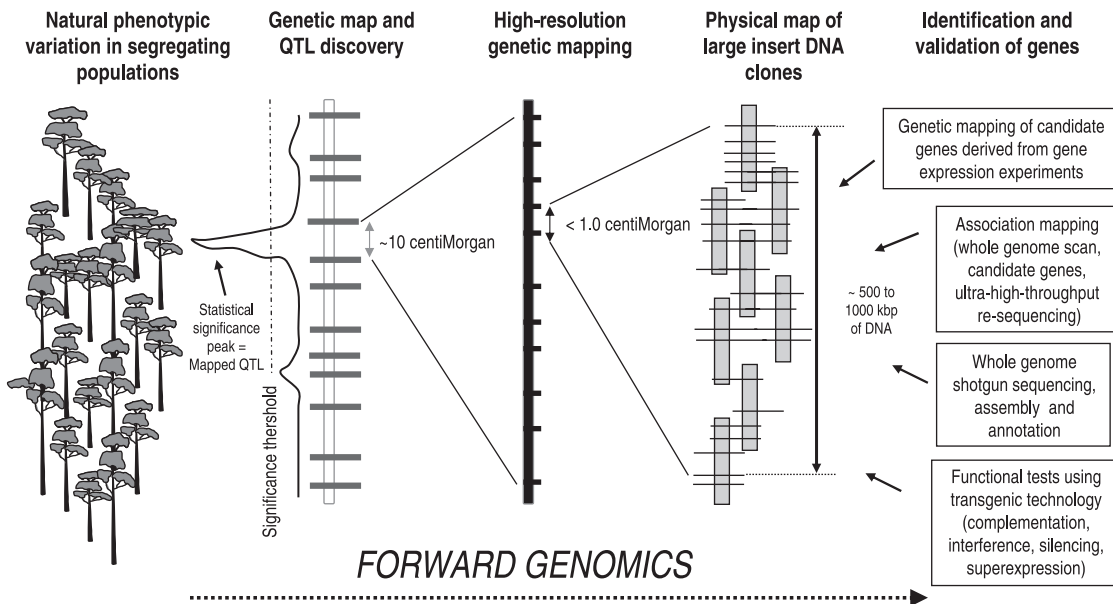


Fig. 1 Schematic representation of the forward genomics approach. Genetic mapping and quantitative trait locus (QTL) discovery are carried out in segregating populations specifically generated to display large amounts of natural phenotypic variation for breeding relevant traits. Major-effect QTLs are detected with a resolution of ~10 centiMorgans, defining target genomic segments involved in the control of the measured traits. Narrower genomic windows of < 1 cM are defined that can be anchored to physical maps of large insert clones such as Bacterial Artificial Chromosomes (BACs) covering ~500–1000 kbp of DNA by high-resolution mapping. Several reverse genomics approaches (boxes) can be used to identify, test and validate specific genes that underlie the target QTL.

I. Introduction

Intensive production forestry based on exotics began in the southern hemisphere *c.* 50 yr ago. Since then, the world forest industry has experienced a slow, steady, but now increasing shift of plantation forestry from the northern hemisphere to the tropics and subtropics. *Eucalyptus* species have been playing a significant role in this process (Kellison, 2001). High-productivity eucalypt forests supply high-quality raw material for pulp, paper, wood, and energy that would otherwise come from native tropical forests. The expansion of these 'fiber farms' may become limited by land needs for growth of food and other biofuel crops and, in some cases, by public pressure. Strategically, the increase in forest productivity and refinement of the quality of wood products through the use of genome-assisted breeding and transgenic technologies will become increasingly important to the forest industry.

While a number of genes affecting wood formation in forest trees have been intensively investigated and manipulated in recent years (reviewed by Bhalerao *et al.*, 2003; Boerjan, 2005), significant applications of transgenic technology to eucalypt production forestry are still to come. Challenges are also faced in molecular breeding applications of genomics. Seventeen years have passed since the first experiments in genetic mapping and molecular breeding of forest trees were described (Neale & Williams, 1991; Grattapaglia *et al.*, 1992).

From the outset, many expectations of fast and accurate methods for early marker-based selection for growth and wood properties in trees were generated. Considerable progress has been made but significant challenges still exist for the implementation of high-impact applications.

Reverse and forward genomics approaches have been pursued in *Eucalyptus* research. Reverse genomics operates by generating altered phenotypes from the manipulation of a given gene through transgenic technology or induced mutations. The forward genomics approach, that is, analysis of existing phenotypic variation to identify causal genetic variants, is based on the wide natural intra- and interspecific diversity that exists in *Eucalyptus*. The technologies involved in this latter approach include genetic mapping, quantitative trait locus (QTL) discovery, association genetics, physical mapping and genome sequencing, some of which will be discussed in this review (Fig. 1).

Recent reviews have described *Eucalyptus* genome research, including gene discovery, candidate gene mapping and functional genomics (Poke *et al.*, 2005; Myburg *et al.*, 2007). Here we discuss how the combination of the unique biology of the eucalypts and the current and emerging technologies, as well as the access to a draft of the *Eucalyptus* genome sequence, will allow a deeper understanding of complex quantitative traits and the translation of genomics into applied breeding tools. Three main points should be highlighted to summarize our main views.

- The wide phenotypic variation across *Eucalyptus* species has prompted remarkable gains through hybrid breeding. This same variation has been and will increasingly be a powerful resource for detecting major-effect QTLs in experimental hybrid populations as species-specific properties might be associated with high-frequency alleles fixed in contrasting species. The most significant gains from marker-assisted selection (MAS) may be achieved through the introgression of such valuable species-specific alleles into breeding populations.
- The broad genetic diversity of *Eucalyptus* natural populations is likely to contain an abundance of low-frequency alleles, some of which may be the next gems to be uncovered by a combined QTL/genetical genomics approach and used by breeders for MAS.
- The rapid improvements of genotyping and sequencing methods and major advances in bioinformatics will soon make it feasible to identify all the sequence variants that occur at moderate to high frequencies in any *Eucalyptus* breeding population by ultra-high-throughput shotgun sequencing of the genome of each individual tree. At this point the lack of well-replicated and accurately phenotyped experimental populations will be more limiting than genomic resources and technologies for an effective move from gene sequences to breeding tools.

II. *Eucalyptus* biology and domestication

The genus *Eucalyptus* includes over 700 species, some of which are the most widely planted hardwoods worldwide. They are long-lived, evergreen trees belonging to the angiosperm family Myrtaceae, which occurs predominantly in the southern hemisphere (Ladiges *et al.*, 2003). Eucalypts are native to Australia and islands to its north, occurring from sea level to the alpine tree line, from high-rainfall to semi-arid zones, and from the tropics to latitude 43° south (Ladiges *et al.*, 2003). Broadly speaking the eucalypts include species of the genera *Eucalyptus* L'Hérit., *Corymbia* Hill and Johnson, and *Angophora* Cav., although these last two (bloodwood taxa) warrant a phylogenetic separation from the *Eucalyptus* (nonbloodwood taxa) in the strict sense, based on molecular studies (e.g. Steane *et al.*, 2002). The latest taxonomic revision (Brooker, 2000) of the eucalypts recognizes over 700 species that belong to 13 main evolutionary lineages, still considering the bloodwood eucalypts as subgenera of *Eucalyptus*. Most species belong to the subgenus *Symphyomyrtus*, and it is mainly species from three sections of this subgenus that are used in plantation forestry such as *Eucalyptus grandis* and *Eucalyptus urophylla* (section *Transversaria*), *Eucalyptus globulus* (section *Maidenaria*) and *Eucalyptus camaldulensis* (section *Exsertaria*).

Eucalypts were rapidly adopted for plantation forestry following their discovery by Europeans in the 18th century (Eldridge *et al.*, 1993). They were introduced into India, France, Chile, Brazil, South Africa, and Portugal in the first quarter of the 1800s (Doughty, 2000) and quickly selected for

plantations as their remarkable growth and adaptability were realized. Several seed collection expeditions followed the first introductions, resulting in germplasm being broadly distributed throughout the world. Today several countries maintain large and diverse germplasm collections of *Eucalyptus* that include a broad range of provenances for the most widely planted species.

Eucalyptus are predominantly outcrossers (Moran *et al.*, 1989; Gaiotto *et al.*, 1997) and insect pollinated. Protandry, late-acting self-incompatibility barriers and, possibly, cryptic self-incompatibility mechanisms (Horsley & Johnson, 2007) are responsible for the preferential outcrossing system and strong selection against inbreeding (Potts & Savva, 1988). Interspecific hybridization in natural populations has been observed (Pryor & Johnson, 1971; Griffin *et al.*, 1988) and becomes relatively frequent in exotic conditions. A textbook case is the hybrid swarm derived from the Rio Claro arboretum established by Navarro de Andrade in Brazil c. 1904 with 144 *Eucalyptus* species (Campinhos & Ikemori, 1977; Brune & Zobel, 1981). Large commercial plantation forests were established with seeds from this arboretum in the 1960s. Most of them, however, had inferior quality because of the extensive segregation observed. However, some outstanding hybrid trees were selected that displayed superior growth, form, and disease resistance. Today, this property is successfully exploited by eucalypt breeders in tropical countries who take advantage of the naturally occurring genetic variation for growth, adaptability and wood properties among species to consolidate complementary traits in hybrid clones (de Assis, 2000).

III. *Eucalyptus* breeding and clonal forestry

The advent of industrially oriented eucalypt plantations in the 1960s and 1970s led to a formal approach to breeding in several countries (Eldridge *et al.*, 1993). Breeding of eucalypts for industrial plantation forestry developed rapidly in countries such as Brazil, South Africa, Portugal, and Chile. Target traits include volume growth, wood density, and pulp yield. Traits related to biotic and abiotic stresses are usually secondary targets that become more important when they impact the main traits. Following the standard concepts in tree breeding, large genetic gains have been obtained in the early stages of eucalypt domestication through species and provenance selection followed by individual selection and establishment of seed orchards. However, a major breakthrough in eucalypt plantation technology occurred in the 1970s with the plantation of the first commercial stands of selected clones derived from hardwood cuttings (Martin & Quillet, 1974; Campinhos & Ikemori, 1977). Since then, vegetative propagation has enhanced the genetic gains of breeding programs by capturing both additive and nonadditive effects into elite clones (Fig. 2).

Clonal propagation and hybrid breeding have become a powerful combination of tools for the improvement of wood product quality. Segregation of high genetic diversity in

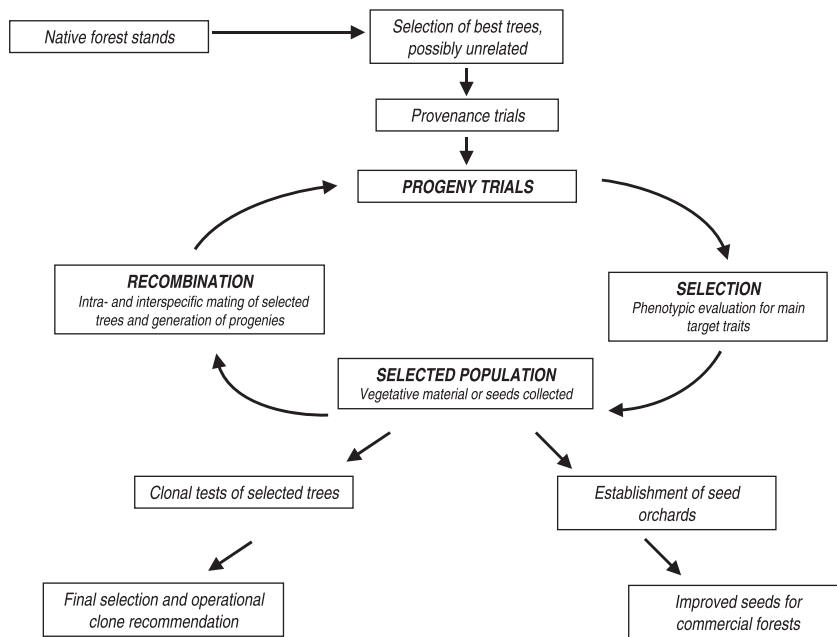


Fig. 2 Basic breeding scheme used in *Eucalyptus* breeding involving recurrent selection cycles from which elite trees are derived for the establishment of seed orchards and/or to be tested as potential commercial clones.

hybrid populations, combined with intensive within-family selection and clonal propagation, resulted in exceptional genetic gains for growth and adaptability to tropical conditions, wood with higher pulp yield and vegetative propagation capability. Eucalypt hybrids deployed as clones currently make up a significant proportion of the existing commercial plantations and are recognized as some of the most advanced genetic materials in forestry. In the last 15 yr, with the increasing realization that the actual 'pulp factory' is the tree, the breeding focus of pulp and paper companies has shifted from volume growth to wood quality. The objective is to improve pulp yield per hectare by reducing wood specific consumption (WSC), that is, the amount of wood in cubic meters necessary to produce one ton of pulp. Clonal forestry of *E. grandis* hybrids in the 1980s reduced WSC by 20% (Ikemori *et al.*, 1994). Second-generation clones derived from hybridization with *E. globulus*, the superior species for wood quality, has allowed a further reduction of 20% (de Assis *et al.*, 2005). Trees that yield more cellulose create gains along the entire production chain by generating savings from tree harvesting and transportation to chipping and pulping, while mitigating the expansion of the commercial forest land base and reducing effluent waste. It is in the context of a specialized, industrially oriented breeding program that fully exploits the power of hybrid breeding and clonal forestry that we discuss the prospects of genomics and molecular breeding of *Eucalyptus*. From the genetics perspective, however, eucalypts are still in the early stages of domestication when compared with crop species. Most eucalypt breeding programs are only a few generations removed from the wild and this fact has important implications when applying genomics approaches.

IV. 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. Nevertheless, molecular markers have been used to solve several questions related to the management of genetic variation, identity and relationship in breeding and production populations. The correct identification of clones is currently the most common application of molecular markers in *Eucalyptus* operational breeding and production forestry. Quality control of large-scale clonal plantation operations is crucial, especially in vertically integrated production systems where the pulp mill relies on the availability of wood from clones with specific wood properties. Plantation forest planning and propagation happen several years before wood consumption – therefore, mislabeling can seriously affect the whole production process. Correct clonal identity also has important implications in breeding procedures where mislabeled clones can significantly affect the expected gains from breeding. Several technologies have been used to resolve clonal identity in *Eucalyptus* such as random amplification of polymorphic DNA (RAPD) (Keil & Griffin, 1994), amplified fragment length polymorphism (AFLP) (Gaiotto *et al.*, 1997) or DNA microarray technology (Lezar *et al.*, 2004). However, co-dominant microsatellites have provided the most powerful method for the unique identification of *Eucalyptus* trees (Kirst *et al.*, 2005a; Ottewell *et al.*, 2005). Most of the microsatellites currently used are derived from dinucleotide repeats that provide powerful discrimination, but present challenges when multilocus profiles

need to be compared between laboratories. To this end we have developed a novel set of microsatellites for *Eucalyptus*, based on tetra- and pentanucleotide repeats, the gold standard established by human forensics. Selected markers of this kind display heterozygosities above 0.45 while providing significantly more robust multilocus genotype profiles (C. Sansaloni and D. Grattapaglia, unpublished).

Molecular markers have been used to characterize the range of genetic variation in germplasm banks of *Eucalyptus globulus*, thereby assisting in the design of new seed collections (Nesbitt *et al.*, 1995). Marker-based genetic distance data have been used to improve the structure of breeding populations (Marcucci-Poltri *et al.*, 2003) and seed orchards (Zelener *et al.*, 2005). Given the wide genetic diversity and multiple sources of germplasm for eucalypt breeding, choices have to be made as to which elite parents should be mated. Any means of predicting tree performance would be valuable for the breeder. Vaillancourt *et al.* (1995a) showed that the ability of RAPD-based genetic distance to predict heterosis was significant but accounted for less than 5% of the variation in specific combining ability (SCA) in *E. globulus* progenies. However, Baril *et al.* (1997) showed that a genetic distance based on RAPD markers with similar frequencies in two intercrossed *Eucalyptus* species successfully predicted the value of a hybrid cross with a global coefficient of determination of 81.6%. In a recent large-scale study, de Aguiar *et al.* (2007) showed that microsatellite-based genetic divergence was significantly correlated with SCA and hybrid family mean for volume growth. However, only 6.3% of the SCA was explained by the genetic distance, revealing that it is still of limited value for practical purposes.

Knowledge of outcrossing versus selfing rates in breeding populations is essential for maintaining adequate levels of genetic variability and continuous gains over generations. *Eucalypts* are preferentially outcrossed both in natural populations and in seed orchards (Moran *et al.*, 1989; Gaiotto *et al.*, 1997) although complex patterns of mating influenced by crop fecundity and orchard position of mother trees (Burczyk *et al.*, 2002) and high rates of pollination from outside the seed orchard (Chaix *et al.*, 2003) have been described. The ability to retrospectively select better combining parents using DNA paternity testing resulted in significant gains in volume growth in the forest planted with the improved seeds (Grattapaglia *et al.*, 2004a).

V. Genetic mapping and QTL analysis

1. Genetic markers and maps

The first *Eucalyptus* linkage maps were built with combinations of several hundred RAPD and AFLP markers (Grattapaglia & Sederoff, 1994; Verhaegen & Plomion, 1996; Marques *et al.*, 1998; Myburg *et al.*, 2003) together with restriction fragment length polymorphism (RFLP), isozymes, expressed sequence

tags (ESTs), genes and more recently some microsatellites (e.g. Byrne *et al.*, 1995; Bundock *et al.*, 2000; Gion *et al.*, 2000; Brondani *et al.*, 2002; Thamarus *et al.*, 2002). Genetic mapping was carried out mostly using a pseudo-testcross (Grattapaglia & Sederoff, 1994) with results of limited value for inter-experimental sharing of linkage and QTL data. Until mid 2006, only 137 autosomal microsatellite markers had been published for species of *Eucalyptus*, including 67 from *E. globulus*, *E. nitens*, *E. sieberi*, and *E. leucoxylon* (Byrne *et al.*, 1996; Steane *et al.*, 2001; Glaubitz *et al.*, 2001; <http://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 *et al.*, 2002). Recently a microsatellite only consensus map covering at least 90% of the estimated recombining genome of *Eucalyptus* was reported. This map has 234 mapped loci on 11 linkage groups, an observed length of 1568 cM and a mean distance between markers of 8.4 cM (Brondani *et al.*, 2006). Microsatellite transferability across intensively planted species of the subgenus *Symphyomyrtus* varies between 80 and 100% depending on the section to which the species belongs. Transferability is approx. 50–60% for species of different subgenera such as *Idiogenes* and *Monocalyptus* and 25% for the related genus *Corymbia* (Kirst *et al.*, 1997). Microsatellite comparative maps have also shown that genome homology across species of the same subgenus is high, as is marker collinearity along linkage maps (Marques *et al.*, 2002). The generalized use of an increasingly large set of interspecific transferable markers and consensus mapping information will allow faster and more detailed investigation of QTL synteny among species, validation of QTLs and expression QTLs across variable genetic backgrounds, and positioning of a growing number of candidate genes co-localized with QTLs, to be tested in association mapping experiments.

2. QTL analysis and validation

QTL mapping in *Eucalyptus* has invariably found major-effect QTLs for all traits considered, in spite of the generally limited experimental precision, the lack of pre-designed pedigrees that maximize phenotypic segregation, and the relatively small segregating populations. The success in detecting those major-effect QTLs can be explained by the undomesticated nature and wide genetic heterogeneity of eucalypts, and the fact that most QTL mapping experiments were carried out in interspecific populations with contrasting gene pools. QTLs 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). QTLs that regulate traits related to vegetative propagation ability have also been detected (Grattapaglia *et al.*, 1995; Marques *et al.*, 1999), as well as a major QTL for early flowering (Missiaggia *et al.*, 2005). In addition, QTLs for insect resistance, essential oil traits and terpenes were mapped (Shepherd *et al.*, 1999;

Henery *et al.*, 2007). A major QTL for *Puccinia psidii* rust resistance was found and mapped in *E. grandis* (Junghans *et al.*, 2003) and later positioned on a microsatellite consensus map (N. Bueno and D. Grattapaglia, unpublished). Two major-effect QTLs for resistance to another fungal epidemic caused by *Mycosphaerella* were recently mapped in *E. globulus* (Freeman *et al.*, 2008), providing further evidence for oligogenic control of fungal disease resistance in *Eucalyptus*. Major QTLs were also found for harvest age traits such as volume growth, wood specific gravity, bark thickness and stem form (Grattapaglia *et al.*, 1996; Verhaegen *et al.*, 1997; Kirst *et al.*, 2004, 2005b; Thamarus *et al.*, 2004).

In spite of a relatively large number of QTLs mapped in *Eucalyptus*, the application of QTL information for directional selection is still an unfulfilled promise. This is largely a consequence of some intrinsic properties of the genus, including (a) its recent domestication and hence the high genetic heterogeneity and linkage equilibrium of breeding populations; (b) the biological barrier that has been encountered to the development of inbred lines to allow a more precise understanding of the architecture of quantitative traits, and (c) the absence of simply inherited traits that could be immediately and more easily targeted. Furthermore, the limited resolution and allelic coverage of mapping experiments have represented a major obstacle. QTL studies have typically identified broad genomic regions which are likely to comprise several hundred genes or *cis* regulatory elements and, therefore, represent only a small step closer to the identification of the causative polymorphism. Larger population sizes and marker densities could effectively improve the resolution by an order of magnitude, but even under the best scenario, intervals of less than a few centi-Morgans would hardly be obtained. Fine-mapping to narrow the QTL interval is difficult to implement in eucalypts because of the impossibility of inbreeding and thus isogenizing lines differing exclusively in the target segment.

Another limitation faced by QTL studies in eucalypts is the restricted number of alleles that are evaluated per cross/study. Because only the genetic variation in the two parental lines can be evaluated in a single cross, alleles that are relevant for the trait variation may either not be sampled or not be variable in the population. Low-frequency, high-value alleles are highly likely not to be captured in these populations. Furthermore, the map positions of most QTLs discovered to date in *Eucalyptus* cannot be compared because they were mapped onto RAPD or AFLP maps, although this picture has recently started to change. Where transferable markers were also mapped, a preliminary comparison of QTL locations has been made. Marques *et al.* (2002) showed that putative QTLs for vegetative propagation traits were located on homeologous linkage groups in different *Eucalyptus* species. Thamarus *et al.* (2004) reported that QTLs for wood properties traits could be detected in two related full-sib families of *E. globulus*, and Marques *et al.* (2005) were able to verify high-effect QTLs for adventitious rooting also in related *E. globulus* pedigrees.

Freeman *et al.* (2008) identified two major-effect QTLs for *Mycosphaerella* resistance on a microsatellite map – the two QTLs could be validated in a second F₂ family and one QTL in a third F₂ family. Recently, comparative QTL mapping across three unrelated pedigrees of *E. grandis* and *E. urophylla*, as well as a comparison of QTLs mapped in *E. globulus*, revealed a number of syntenic QTLs for cellulose yield, lignin content and different but correlated fiber traits (D. Grattapaglia, unpublished). These results demonstrate the feasibility of validating QTLs and identifying target genomic regions for subsequent investigations. However, the lack of a genome sequence for a *Eucalyptus* species complicates the identification of genes and/or noncoding regulatory variants underlying a QTL, although this difficulty may soon be mitigated by the sequencing of the *E. grandis* genome.

VI. Gene discovery and genetical genomics

1. EST sequencing

Since the first *Eucalyptus* gene sequence was deposited in Genbank 14 yr ago (Feuillet *et al.*, 1993), the number of public sequences has increased very slowly compared with other crops or major tree species. The reason has been the industrially oriented, proprietary nature of the genomic research carried out in *Eucalyptus*. About 10 yr ago, tens of thousands of sequences were generated by Genesis in New Zealand and later incorporated by the US-based company Arborgen (M. Hinchee, pers. comm.). Dupont, in collaboration with the former Australian company Forbio, generated a database of approx. 14 000 ESTs from *E. grandis* c. 1997 (S. Tingey, Dupont Crop Genetics, pers. comm.) that were later annotated and used in expression QTL studies (Kirst *et al.*, 2004). Other mostly private EST databases for *Eucalyptus* were built between 2002 and 2004 in the Genolyptus project (Grattapaglia, 2004; Grattapaglia *et al.*, 2004b) and the ForEST project (Vicentini *et al.*, 2005) in Brazil and in Japan (S. Sato, unpublished). The different species planted around the world and the hybrid breeding system adopted in *Eucalyptus* have driven a distinctive multi-species approach to EST sequencing efforts.

A slight increase in the number of publicly available sequences started in 2003 with some small-scale cDNA microarray experiments carried out in *Eucalyptus*, which led to the publication of a few thousand sequences (Kirst *et al.*, 2004; Paux *et al.*, 2004; Foucart *et al.*, 2006). During 2007 the French group led by J. Grima-Pettenati deposited a set of ~12 000 sequences, mostly derived from *E. globulus*, while a Chilean research group deposited another 8743 sequences. As of 31 December 2007 there were only 24 698 EST sequences available in Genbank out of the 43 664 deposited nucleotide sequences for *Eucalyptus*. It is expected, however, that with the breakthrough advancements of speed and dramatic cost reductions brought about by pyrosequencing technologies

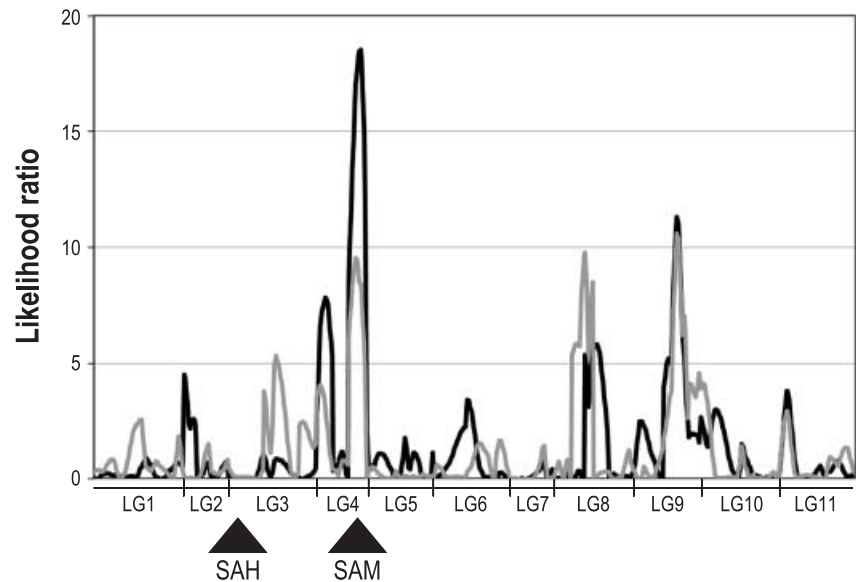


Fig. 3 Likelihood ratio (LR) profiles of the quantitative trait locus (QTL) analysis of the gene expression of S-adenosylmethionine synthase (SAM; black line) and adenosylhomocysteinase (SAH; gray line) measured in a segregating population of *Eucalyptus globulus* and a hybrid of *Eucalyptus grandis* × *E. globulus*. The genetic map linkage groups (LG1–LG11) are displayed on the x-axis. The genetic map positions of the genes (black triangles) suggest that SAM is *cis*-regulated, while SAH is controlled in *trans* by three loci (gray lines; peaks ~10) located in different genomic regions.

(Margulies *et al.*, 2005) the size of public *Eucalyptus* EST databases will very soon become much larger and more diverse. Recently, the generation of more than 1 million *E. grandis* reads with average lengths of 100–200 bp from three runs of Genome Sequencer 20 and FLX Systems (454 Life Sciences Corporation) was reported, generating a preliminary assembly with approx. 29 000 contigs (*E. Novaes et al.*, unpublished). As such large numbers of sequences become public, there will be a stimulus for the publication of existing private EST databases and a significant increase in available resources.

2. Genetical genomics

In an attempt to partially bypass some of the limitations to detecting genes underlying QTLs, Kirst *et al.* (2004) used a genetical genomics approach. This approach was originally proposed by Jansen & Nap (2001) and relies on analyzing gene expression measured by microarrays as a quantitative trait, using QTL analysis. By combining the QTL analysis of gene expression and trait in a segregating population, loci significantly associated with both phenotype and transcription of specific genes can be identified. Such ‘double-QTLs’ identify potential candidate genes for the regulation of the trait. This approach was applied in *Eucalyptus*, with the expression of ~2700 genes putatively involved in cell wall formation, lignin and cellulose metabolism, cell growth, and protein targeting being monitored. The key role of some lignin biosynthesis genes was confirmed, and some other unexpected new genes with major effect were also identified as highly correlated with volume growth. In a subsequent study, it was also shown that expression QTLs could explain up to 70% of the transcript level variation in one gene, and hotspots with co-localized expression QTLs were identified for over 800 genes, suggesting coordinated genetic regulation.

However, the gene expression data also demonstrated a lack of conservation of the genetic architecture of transcript abundance regulation in different genetic backgrounds, indicating that many different loci could be involved in modulation of transcription of these genes in a complex and variable network of gene expression control (Kirst *et al.*, 2005b). For MAS, such ‘master expression QTL’ may be very promising targets for tree breeding as they appear to be key regulators of a broad range of genes and, possibly, phenotypic traits. The identification of such large-effect genes by genetical genomics could also reveal important positional leads to nearby noncoding regulatory variants responsible for quantitative variation, although the challenges both in understanding and in assigning such variants to a measurable fraction of the phenotypic variation of a trait will be significant in *Eucalyptus*.

A limitation of the genetical genomics approach in *Eucalyptus* is the lack of a full genome sequence, or at least a genetic map including large numbers of mapped genes. This information is required to test whether the genetic control of gene expression is *cis*- or *trans*-acting. Therefore, unless the genes have been genetically mapped previously (Fig. 3), it is not possible to distinguish between these two modes (*cis/trans*) of regulation. Genetical genomics studies are also constrained by the same limitations as traditional QTL studies in terms of population and allelic diversity range surveyed. However, unlike traditional QTL studies, the genetical genomics strategy permits the rapid identification of candidate genes that underlie a QTL, assuming that transcription profiles are available for the segregating populations. Genetical genomics may also be a particularly valuable approach for detection of genes that are important for hybrid breeding, allowing identification of genes where alternative alleles are fixed in one or both hybrid species. For these loci, the same QTL is expected to segregate consistently in multiple, independent crosses involving the

two parental species. These genes are primary targets for breeders attempting to create a favorable combination of alleles in tree improvement programs.

VII. Association mapping

Association genetics or linkage disequilibrium (LD) mapping differs from traditional QTL analysis primarily at the level of resolution and allele diversity captured. The approach relies on the genetic analysis of a population with unknown ancestry, but which at some point in time shared a common (unknown) ancestor (Cardon & Bell, 2001). Higher resolution is achieved because lineages derived from the common ancestor may extend for many generations, therefore reducing linkage disequilibrium to neighboring loci. By working with a large sample of unrelated individuals, the breadth of alleles and genes that contribute to the trait variation can also be captured. Fundamentally, the finding of a statistically significant correlation between alleles and phenotype identifies a locus that is a candidate for regulating the trait or a locus in LD with the causal polymorphism. In the few *Eucalyptus* genes analyzed to date, significant LD has been found for only a few hundred base pairs (Santos, 2005; Thumma *et al.*, 2005; M. Kirst, unpublished; D. Faria and D. Grattapaglia, unpublished) (Fig. 4). If LD decays rapidly (i.e. within a few hundred bases), a significant association would suggest that one or more genes in the close vicinity of the polymorphic site, that is, within a few hundred base pairs, are implicated in the trait variation.

The successful identification of single nucleotide polymorphisms (SNPs) and haplotypes associated with microfibril angle in *E. globulus* by Thumma *et al.* (2005) demonstrated the feasibility of association genetic studies in *Eucalyptus*. Reports in other forest tree species are becoming increasingly common, particularly in pines (Gonzalez-Martinez *et al.*, 2006, 2007) and other conifers, where QTL/positional cloning approaches are prohibitive because of the large genome size. In *Eucalyptus* and other woody angiosperms with moderate genome size, association genetics represents one alternative to QTL mapping and positional cloning. However, it is questionable how effective association genetics will be for detecting allelic combinations that are explored in hybrid breeding, which is currently the basis of the most successful tree improvement programs in *Eucalyptus* and several woody angiosperms. While overdominance appears to be a significant driving force behind heterosis in these hardwoods, it is unclear if association genetics will be able to identify interspecific allelic combinations that produce superior phenotypes. Therefore, it will become important to evaluate the use of association genetics in populations that combine pure species and hybrid individuals. Unfortunately, interspecific hybridization is one of the driving forces in the creation of LD. As a result, significant associations detected in an F_1 hybrid are meaningless as there will be complete LD among all the loci inherited from

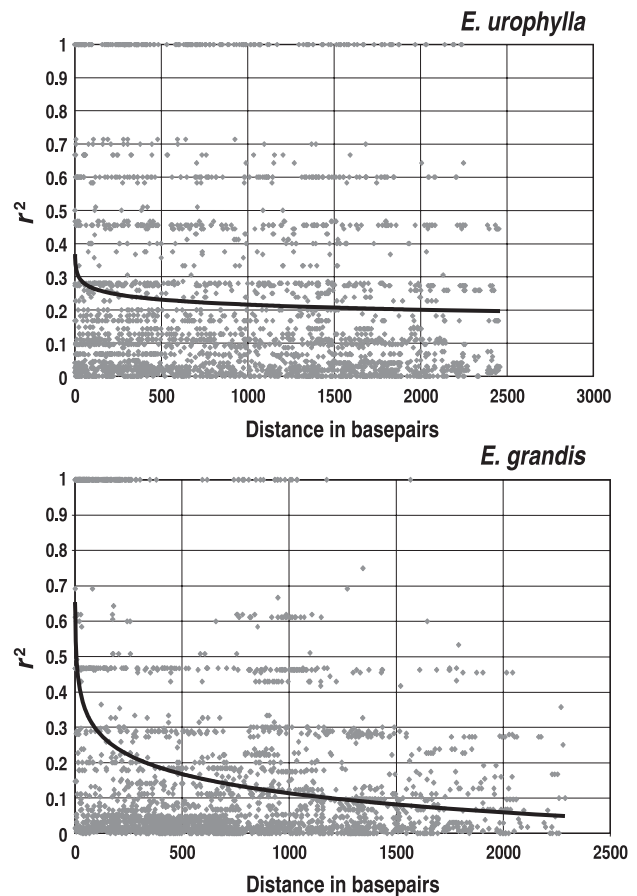


Fig. 4 Rapid decay of intragenic linkage disequilibrium estimated as r^2 in the cinnamyl alcohol dehydrogenase (*cad*) gene in two different *Eucalyptus* species (D. Faria and D. Grattapaglia, unpublished).

each parental species. These issues will probably be of limited importance in conifer association studies, as interspecific hybrid breeding is uncommon and heterosis is not as prevalent as in hardwoods.

The application of association genetics in *Eucalyptus* and trees in general implies the need to detect polymorphisms in candidate genes that can be tested for association and to assess the level of LD around these polymorphisms. However, focus on candidate genes may become unnecessary in the near future as new high-throughput genotyping technologies are developed for discovery of polymorphisms in large numbers of genes and their testing for associations at a limited cost (see Section X: Future developments and challenges). Also essential is the assessment of any genetic structure that may exist in the association population as this may lead to detection of spurious, false-positive associations (Knowler *et al.*, 1988; Pritchard *et al.*, 2000). In this section we discuss the potential for carrying out association genetic studies in *Eucalyptus* species, in natural and breeding populations, and the expected pitfalls and limitations of the application of this strategy.

1. Nucleotide diversity and linkage disequilibrium

Initial estimates of nucleotide diversity in a few *Eucalyptus* species suggest that levels of polymorphism are adequate for association genetic studies. Nucleotide diversity is similar to or higher than that detected in maize (*Zea mays*) and *Populus*. In the analysis of partial sequences of S-adenosylmethionine synthase, ferulate 5-hydroxylase and cinnamyl-alcohol dehydrogenase in ~16 genotypes of *E. globulus* we detected values for nucleotide diversity (π) of < 1%, in coding regions and untranslated regions (UTRs) (M. Kirst, unpublished). A twofold higher nucleotide diversity was seen in a 440-bp sequence of the Caffeoyl CoA O-methyltransferase gene for *E. grandis* ($\pi = 0.00356$) when compared with *E. globulus* ($\pi = 0.00168$), possibly reflecting the wider geographical distribution of the former species (Santos, 2005). Thumma *et al.* (2005) described similar levels of diversity in cinnamyl-CoA reductase in *E. globulus* and also a rapid decay of LD, attaining an r^2 of 0.2–0.3 in a few hundred to a thousand nucleotides. With such rapid LD decay, the resolution should be adequate for identification of specific genes as regulators of phenotypic traits in *Eucalyptus* association studies. Despite this expectation, it is now clear from work in the model plant *Arabidopsis thaliana* that LD decay and nucleotide diversity are highly variable in the genome, reflecting evolutionary and natural selection effects at specific loci (Clark *et al.*, 2007; Kim *et al.*, 2007).

The studies described above have been limited to a few *Eucalyptus* species and loci. It is unclear if such levels of polymorphism and LD will also be found in all commercially relevant species, including those that occur as small populations, or that may have originated from relatively recent admixture events. Small effective population size contributes to lower nucleotide diversity (as a result of genetic drift) and admixture leads to higher LD. Selective sweeps, particularly around loci that may be essential for adaptation of some *Eucalyptus* species to particular biotic and abiotic stresses, also lead to high levels of LD, at least in some areas of the genome. Although typically localized, these regions of high LD may be precisely the ones that are being targeted by breeders for gene identification through genetic association analysis (e.g. for freezing tolerance).

Another factor that creates LD is population admixture, but a similar effect occurs in the event of species hybridization followed by speciation. Interspecific hybridization leads to complete LD among alleles fixed in the two species, in the first generation. Although hybridization does not occur among species in different subgenera (Griffin *et al.*, 1988), natural hybrids from species within a subgenus can be observed in nature and are relatively common (Griffin *et al.*, 1988; Potts & Dungey, 2004). In the event of one *Eucalyptus* species having originated from a relatively recent hybridization event, one may expect significantly higher levels of LD among loci. The rate of LD decay will be dependent on the frequency of recombination and the number of generations, but is likely to

remain high for many generations. These effects may also be detected in commercial *Eucalyptus* breeding programs which frequently rely on the transgressive segregation of hybrids.

It has been suggested that historical interspecific hybridizations between tree species could be useful for detecting genes related to adaptation using admixture mapping. This possibility was evaluated in *Populus*, primarily because of the availability of well-characterized hybrid zones, where *Populus alba* and *Populus tremula* have been hybridizing for many generations (Lexer *et al.*, 2007). While it is a useful approach in those circumstances, we believe that admixture mapping will have minor utility in *Eucalyptus* genomics. The primary reason is that, although natural hybrids have been observed, it is unclear whether they are common and widespread enough to be useful for admixture mapping. Alternatively, artificially managed admixed populations could be created. However, even with advanced breeding techniques that shorten rotation cycles to only a few years, creating such a population would require several generations to achieve adequate LD for pinpointing genes that control quantitative traits. Furthermore, admixture mapping depends on the availability of genome-wide panels of markers informative for ancestry between species. Microsatellite allele frequency differentials and estimates of F_{st} amongst provenances within species (Steane *et al.*, 2006) or even between tropical and temperate plantation species that belong to different taxonomic sections such as *E. grandis* and *E. globulus* (Kirst, 1999; D. Faria and D. Grattapaglia, unpublished) are typically modest. It may therefore be challenging to develop a robust and large set of ancestry-informative markers for *Eucalyptus*, especially between species or provenances that are closely related. Considering the rapid development in technology for DNA sequencing and SNP genotyping, we expect that within the time period needed to develop such AIM, whole-genome association analysis based on several hundred thousand markers will be far more efficient in defining genes that regulate complex traits.

2. Population structure – natural and breeding populations

Population genetics studies in *Eucalyptus* have demonstrated that most species have very limited population structure. Low F_{st} values (~0.10) have been reported with nuclear biparentally inherited markers in *E. grandis*, *E. globulus* and *E. urophylla*, which are species with wide geographic distributions and large population sizes (House & Bell, 1994; Kirst, 1999; Steane *et al.*, 2006; D. Faria and D. Grattapaglia, unpublished). Even between two geographically separated populations of *E. grandis* located in Atherton (Queensland) and Coffs Harbor (New South Wales) the F_{st} was estimated at 0.06 based on 33 microsatellites. Similarly, a low F_{st} (0.05) was detected between two disjunct *E. globulus* provenances, one continental (Victoria) and one from Flinders Islands (D.

Faria and D. Grattapaglia, unpublished). Limited structure among populations is expected in these species considering their outcrossing nature and wide pollen and seed dispersal. As a consequence, identification of spurious associations as a result of population structure, which has plagued studies in humans and some crop species (Knowler *et al.*, 1988; Pritchard *et al.*, 2000), has typically not been considered a major concern in *Eucalyptus* association studies. However, exceptions may exist and species and populations will have to be considered individually. For instance, population structure measured on the basis of the J_{LA} region of the maternally inherited chloroplast DNA indicates much higher levels of population differentiation in *E. urophylla* ($G_{ST} = 0.58$) than in *E. grandis* ($G_{ST} = 0.30$) (Jones *et al.*, 2006; Payn *et al.*, 2007). *Eucalyptus urophylla* is widely distributed in disjunct populations that occupy the islands of Indonesia and Timor. Population isolation may contribute to inbreeding, and, in the long term, population differentiation (Hartl & Clark, 1997). Other unknown factors related to the history of the natural populations (e.g. genetic bottlenecks and migration) may also contribute to higher population differentiation and are often unknown.

Association populations created from natural stands may represent a good source of germplasm for association studies. However, most of these studies will be carried out in the context of existing breeding programs developed by companies and research institutions. Many of these breeding programs are derived from seed collected and distributed directly from Australia in the late 19th century, or from species and provenance tests. The population structure – if any – of the material utilized in these programs is in most cases unknown, but could be significant. A lack of source data also suggests that some inbreeding may be taking place within these programs without the breeders' knowledge. For instance, a STRUCTURE (Pritchard *et al.*, 2000) analysis and assignment test of a group of elite clones used in operational plantation forestry in Brazil showed that most of them were assigned with high confidence to *E. grandis* and *E. urophylla* with approx. 50% probability for each parental species, corroborating their expected very recent hybrid origin (D. Faria and D. Grattapaglia, unpublished). Failure to recognize those differences in an association study will lead to identification of false-positive associations.

VIII. Molecular breeding

1. Perspectives of marker-assisted selection

Considering the rapid decay of LD reported for *Eucalyptus*, the marker-trait association detected in one segregating population cannot be assumed to hold in a different pedigree. As a result, MAS based on long-range marker-trait association detected by QTL analysis may only be useful for within-family selection. This is in principle a limitation, considering that most *Eucalyptus* breeding programs attempt to evaluate a

broad diversity of families to capture valuable allelic combinations in the progeny. However, *Eucalyptus* breeding strategies vary broadly according to the target species or hybrid, the possibility of deploying clones, and the amount of resources available to the breeder, so that within-family early marker-assisted selection might in fact be a viable and valuable tool.

A reasonable premise is that MAS will only be justifiable when the program has already reached a relatively high degree of sophistication, fully exploiting all the accessible breeding and propagation tools. Breeding for hybrid performance combined with clonal propagation of selected individuals is currently the method of choice for extracting new elite clones (Rezende & de Resende, 2000; Potts, 2004; Potts & Dungey, 2004; Bison *et al.*, 2007). Progeny trials, together with expanded single-family plots where larger numbers (> 1000) of full-sibs are deployed per family, are used for very intensive within-family selection. Vegetative propagules are rescued from selected trees and used for establishing clonal trials. This strategy involves a significant amount of time and effort being devoted to clonal testing before effective recommendations can be made regarding new, operational clones. At the same time this breeding scheme generates large amounts of LD, a favorable condition for MAS in forest trees (Strauss *et al.*, 1992). Furthermore, numerous reports have indicated that dominance variation is a key component to explain the superiority of tropical hybrid *Eucalyptus* clones, especially for growth (Bouvet & Vigneron, 1996; de Assis, 2000; Rezende & de Resende, 2000). This and other nonadditive sources of genetic variation can be captured by vegetative propagation in *Eucalyptus*, a process that is not yet fully operational in conifers. Therefore, genomic segments containing favorable alleles or allele combinations segregating within families could be efficiently tagged with microsatellite markers in strong LD with the actual causative polymorphisms in a QTL mapping stage, and later used for within-family MAS selection for superior individuals (O'Malley *et al.*, 1994). QTL linked markers would allow very early selection (seedling stage), reducing the time required to identify elite trees, especially for traits related to wood properties, and reducing the number of trees to be selected, propagated, and advanced to clonal trials (Grattapaglia, 2008; Fig. 5). Given their relatively short rotations and the ability to capture nonadditive genetic variation, eucalypts in fact may be the first forest tree species where this kind of MAS scheme could be applied.

Although the costs of genotyping have dropped significantly in recent years, DNA analysis is still costly. The most likely cost-beneficial application of MAS in *Eucalyptus* will be for traits that provide significant added value to the final product, such as branching habit (for solid wood) or wood chemical and physical traits, or for traits that facilitate clonal deployment, such as adventitious rooting or somatic embryogenesis response. Among all possible quality traits, the preference would be for those that display medium to high heritabilities but where phenotype assessment is difficult, expensive, or

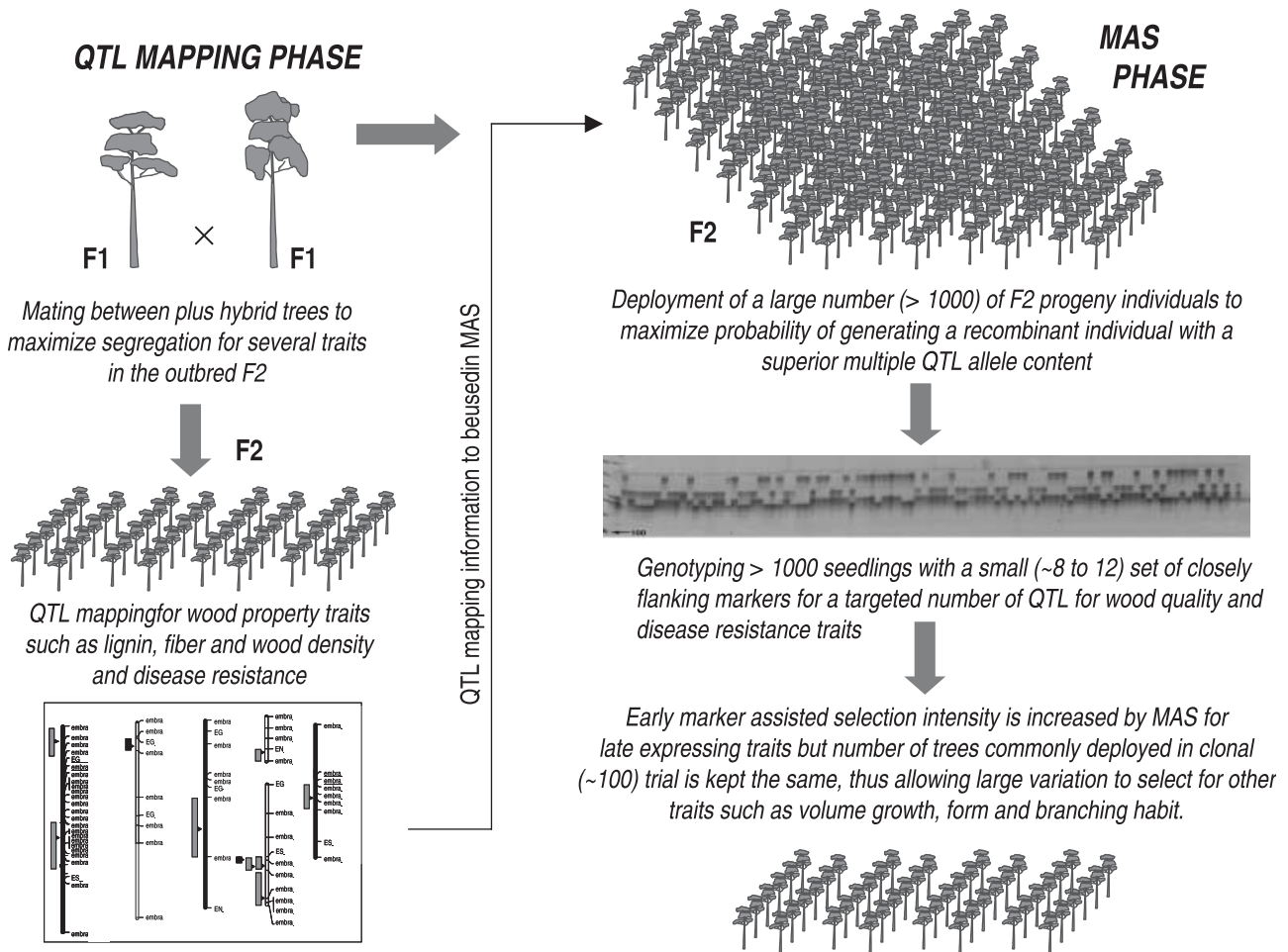


Fig. 5 Strategy for early within-family marker-assisted selection (MAS) of superior (plus) trees in a *Eucalyptus* clonal selection program. Quantitative trait locus (QTL) mapping phase: QTL mapping is carried out and flanking markers in strong linkage disequilibrium with favorable alleles or allelic combinations at major-effect QTLs are identified within-family. MAS phase: QTL linked markers are genotyped in an expanded number of progeny individuals of the same family to increase within family selection intensity at very early age for late expressing traits such as wood properties (modified from fig. 10 in Grattapaglia, 2008; with kind permission from Springer Science).

requires waiting until the tree reaches maturity. Wood quality traits typically require the tree to start accumulating mature wood and involve relatively lengthy procedures for phenotypic evaluation in the laboratory. These kinds of traits could be interesting targets for MAS in *Eucalyptus*, given that the costs of genotyping are sufficiently competitive, and precision is high when compared with direct phenotype measurements. It is important to point out, however, that with the recent developments of fast sampling and indirect wood chemistry measurements, the potential gain will be realized only on the basis of the time savings provided by very early selection.

2. Breeding using transgenic technology

Transgenic technology is undoubtedly a powerful complementary tool available to the molecular breeder. Considering that industrial *Eucalyptus* forests are almost exclusively clonal,

transgenics are likely to have an increasing role not only in wood quality improvement but in resolving problems related to pest and pathogen susceptibility and/or abiotic stress tolerance (e.g. tolerance to frost and drought) that might limit the expansion or survival of existing plantations, as in the case of annual crops. The introduction of genes that confer traits that do not display variation within the *Eucalyptus* gene pool, or are impossible to produce by natural recombination processes, might radically modify the ways in which forests are planted or in which forest products are derived (Grattapaglia, 2008).

The development of efficient transgenic technologies for *Eucalyptus* would represent a key step for functional genomics studies. Gene tagging through insertional mutagenesis approaches is a particularly appealing method in heterozygous trees that cannot be easily selfed, as it would generate dominant phenotypes. In *Eucalyptus*, the use of such an approach as a

practical tool for genome-wide gene characterization and breeding faces the same kinds of logistical and biological obstacles as found in *Populus*, but these might be overcome by large-scale collaborative efforts (Busov *et al.*, 2005). To mitigate the biological and biochemical limitations to the study of wood formation in trees using mutant phenotypes, *in vitro* wood formation systems can be employed to introduce transgenes transiently or stably into growing wood-producing tissue. Such a system has been developed in *Eucalyptus* (Spokevicius *et al.*, 2005) and recently used to show that β -tubulin determines cellulose microfibril orientation during xylogenesis in *E. globulus* (Spokevicius *et al.*, 2007).

In spite of the recognized economic importance of *Eucalyptus* in world forestry, very little has been published on transgenic experiments in species of the genus; most of this work has been summarized by Poke *et al.* (2005). While *Eucalyptus* tissue can be transformed by *Agrobacterium tumefaciens*, major difficulties are faced in the regeneration step. Several reports have documented the production of transformed callus, tissue, and root organs; however, reports on transformed plants are scarce (Machado *et al.*, 1997; MacRae & van Staden, 1999). A marked genotype effect has been observed on the efficiency of regeneration and, consequently, stable transformation. This fact has prompted several groups to first identify *Eucalyptus* 'lab rats,' that is, easily regenerable genotypes, and only then to develop improved protocols to generate large numbers of independent transformation events (Tournier *et al.*, 2003). These 'lab rats' are not yet available in the public domain, and this is also the case for most of the transformation protocols. However, recently, Chen *et al.* (2006) described a basic *Agrobacterium*-mediated genetic transformation protocol employing organogenesis for the production of transgenic plants using *Eucalyptus camaldulensis*. Efficient proprietary transformation protocols have also been developed in Japan, where an *E. camaldulensis* 'lab rat' has been used (Kawazu *et al.*, 2003), as well as in different laboratories in the USA, where *E. grandis* and *E. urophylla* have been used (M. Hinchee and V. Chiang, pers. comm.).

The current information on *Eucalyptus* transgenesis points to a very promising future as far as the technical possibility of generating stably transformed *Eucalyptus* plants is concerned. However, some strategic issues regarding the adoption of transgenic technology for wood quality manipulation in *Eucalyptus* should be taken into account, including (a) the magnitude of the gain and cost/benefit relationship obtained by manipulating lignification or cellulose genes when compared with the exploitation of the wide natural genetic variation in the genus; (b) the specific biosafety and intellectual property issues relevant to transgenic eucalypts and the time and investment necessary to resolve them in order to actually be able to plant transgenic trees on a large scale; (c) the speed at which breeding programs generate new and better clones for adaptive traits (growth, wood properties, pest resistance, clonability, etc.) compared with the time needed for regulatory approval

of every new transgenic clone; (d) the lifespan of a patent in local regulation as compared with the time needed to effectively make returns on the patent from the planted forest before the patent goes into the public domain; (e) the market issues that the company has to consider in adopting transgenics, in relation to both public perception and forest certification processes. All these and other issues will have to be carefully considered, including the fact that, just as occurred in annual crops such as soybean (*Glycine max*), maize, and cotton (*Gossypium* spp.), the use of transgenics could produce a major technological divide and become a necessary condition for the continuing competitiveness of forest-based industry worldwide (Grattapaglia, 2008).

IX. From gene sequences to breeding tools

The successful application of gene sequence information to *Eucalyptus* breeding has not yet been demonstrated, but this may change in the near future. The expected completion of the genome sequence of *E. grandis* by 2010 and the use of complementary or novel quantitative genomic approaches (e.g. genetical genomics and association genetics) may rapidly lead to such applications, much as has been the case in *A. thaliana* and some agricultural crops. The first example of the identification of SNPs associated with wood microfibril angle was reported in *E. globulus* (Thumma *et al.*, 2005) and SNPs have also been identified for several wood property and physiological traits in pines (Gonzalez-Martinez *et al.*, 2006, 2007). Not surprisingly, the identification of significant associations has generally occurred for traits with high heritability. In none of these cases, however, has the biological cause of the SNP–trait association been demonstrated, or the gene function verified by alternative approaches.

The strategy to be adopted for successful identification and introgression of valuable alleles discovered through genomic analysis into *Eucalyptus* breeding programs will be dependent primarily on (1) the successful identification of polymorphisms associated with traits of interest, (2) the frequency of superior alleles in the base breeding population, and (3) their phenotypic effect. We have discussed current and new approaches to the detection of genetic variants of interest (see Sections VI and VII on QTL and association mapping). Here we explore likely scenarios regarding the other two factors – allele frequency and phenotypic effect – considering the life-history characteristics of *Eucalyptus*. In the last sections we discuss how current technology and future developments may change the outcome of genomic research targeted to assist *Eucalyptus* breeding in the next few decades.

1. Superior alleles with high frequency in the breeding population

In plants, association genetic studies have typically only targeted the analysis of SNP loci where the frequency of the minor allele is relatively high (> 10%). Lower frequencies, much like

small population sizes, severely hamper the statistical power of detection of association, particularly for traits with low heritability (Long & Langley, 1999). SNPs identified in association with quantitative phenotypes will probably be of small effect, explaining less than 5% of the phenotypic variance, much as has been observed previously in loblolly pine (*Pinus taeda*) (Gonzalez-Martinez *et al.*, 2007) and recently by Thumma *et al.* (2005) in *E. globulus*. Therefore, it is unlikely that traditional breeders will, at least in the short term, make selection decisions based on such limited gain. Because of its recent domestication coupled to the capacity for clonal selection, every breeding cycle in *Eucalyptus* still produces gains that are almost one order of magnitude higher for growth traits (Snedden & Verry, 2004). As a result, at least initially, marker-assisted breeding will only serve to complement traditional breeding. We anticipate that this scenario will remain unchanged until 'whole-genome' association approaches – where almost all polymorphisms with an effect on a phenotype can be detected – are developed.

2. Low-frequency alleles of large effect

It has been generally considered that alleles at low frequencies (< 5%) in the breeding population will be of limited use for MAS. Such alleles may not significantly shift population means and therefore do not have great relevance at the population level – very few individuals carry them, particularly in a homozygous state. However, increasing evidence from human association genetic studies indicates that rare alleles can have a tremendous impact in an individual's phenotype (Romeo *et al.*, 2007; Topol & Frazer, 2007). This effect will probably be accentuated in *Eucalyptus*. *Eucalyptus* species are generally outcrossers, and have large effective population sizes, a broad geographic distribution and wide pollen and seed distributions. Therefore, rare mutants are frequently maintained at low frequencies in the population for a large number of generations, particularly if they are selectively neutral. The broad genetic and phenotypic diversity of *Eucalyptus* natural populations is likely to contain an abundance of unique haplotypes and low-frequency alleles; some of these alleles may be the gems that breeders are seeking for MAS.

A recent survey of a relatively small number of *Eucalyptus* haplotypes identified almost 300 large-effect SNPs – that is, polymorphisms that lead to the introduction or removal of a STOP codon – in transcribed sequences (E. Novaes, unpublished). Therefore, the largest breeding gains from MAS may be achieved not through the continuous incorporation of frequent alleles of small effect, but rather through the identification of rare, high-value alleles. The potential impact of such alleles in tree breeding programs is exemplified by a rare loblolly pine mutant, the cinnamyl alcohol dehydrogenase *cad-null* allele, which has been associated in some studies with the superior character of that genotype in terms of growth and

wood properties (Mackay *et al.*, 1997; Ralph *et al.*, 1997; Wu *et al.*, 1999). In species with the life history of *Eucalyptus*, where the probability of maintaining rare alleles in the population is high, it is possible that the majority of the superior phenotype detected in some individuals may be attributable to such uncommon allelic variants. Obviously, identifying those alleles and increasing their frequency to usable levels in a breeding population will be challenging. Furthermore, in a *Eucalyptus* hybrid breeding/clonal deployment situation, where the magnitude of nonadditive variation becomes important, the issue will be not only the detection of rare alleles of large effects but also the detection of rare allelic combinations both within and between loci. Those alleles and allele combinations are unlikely to be uncovered through association genetics because very few individuals will carry them, but they may be detected by an integrated sequential QTL analysis and genetical genomics approach. Nonetheless, capturing all the rare alleles with important additive and dominance effects in a population would require a QTL/genetical genomic analysis of a large number of crosses in a diallel mating design (which is too costly with the current technology). An alternative would be to create crosses that involve a subset of the breeding population that encompasses the majority of the genetic variation in the species, and carry out analysis of only those populations. This approach is currently being pursued in maize by a Nested Association Mapping (NAM) strategy where, in parallel to the development and analysis of an association population, 26 recombinant inbred line populations were generated. Each population was derived from the cross of a 'reference' common parent to each one of 26 diverse founder lines that best represents the genetic variability of the association population (Yu *et al.*, 2008). One of the discoveries expected from this series of QTL populations is the detection of alleles that may be unique to one of a few maize lines, which segregate in the progeny and have an impact on the phenotypic variation. Although not all important rare alleles with significant additive effects will be detected using this approach, it will permit evaluation of how frequently these alleles are missed in an association population. Maize has many similarities with *Eucalyptus* and other tree species in terms of high nucleotide diversity and low LD, largely because of rare alleles.

3. Incorporation of species-specific high-value alleles

Different *Eucalyptus* species are recognized for their superior characteristics in terms of growth, wood quality and resistance to biotic and abiotic stress. For instance, *E. nitens* is highly tolerant to frost damage; *E. grandis* has superior adaptability and growth properties, and *E. globulus* has better wood properties. These species-specific properties are likely to be associated with alleles that are fixed or at very high frequencies in each species and may provide a selective advantage in the natural range of the species for individuals carrying them. In the short term it is possible that the most significant gains

from MAS may be achieved through the introgression of valuable species-specific alleles into the breeding populations. As described previously (QTL analysis and genetical genomics), those alleles of large effect may be relatively easy to detect in the analysis of hybrid segregating populations. However, association genetic approaches may be hampered by spurious associations resulting from population structure. Although genetic structure is relatively modest among *Eucalyptus* species, there will be very significant genetic differentiation among species for those genes that have become fixed in any one species. Any allele fixed in one species may be identified as significantly associated with traits that are superior in that species.

The incorporation of species-specific superior alleles in any breeding program is also likely to be problematic, at least through traditional breeding. In the first generation following hybridization one should expect complete LD among alleles that are fixed in the two species and contribute to the hybrid genome. LD decays as a function of the recombination rate, effective population size and number of generations. Recombination rates are not constant in the genome (Kim *et al.*, 2007), and some evidence suggests that lower rates occur in hybrids because of recombination suppression in parts of the genome that have diverged (Bradshaw & Stettler, 1994; Bradshaw *et al.*, 1994). Therefore, transgenic approaches may be the best approach for rapidly and efficiently incorporating these alleles into the breeding program.

X. Future developments and challenges

One of the major goals of genomics is to describe all levels of genetic information – from DNA sequence, to mRNA, to protein – and develop predictive models that define phenotype based on genetic, epigenetic, developmental and environmental properties. Therefore, a complete description of these networks would also imply full knowledge of the contribution of epigenetic and environmental factors to the phenotype. We are still a long way from achieving this goal for any phenotype in any organism, but technological developments, particularly in genome sequencing and transcription analysis, have advanced tremendously since the turn of the century. Comprehensive databases of gene expression have already been created for *Populus*, based on whole-genome microarrays and ESTs (Sjödín *et al.*, 2006). This and other plant transcriptome databases (www.arabidopsis.org/portals/expression) allow any researcher to obtain gene expression information for almost any gene in a broad spectrum of tissues, developmental stages, growth conditions and genotypes. Similar advances in DNA sequencing technology may translate into the description of the genome sequence for any individual with a moderate genome size, such as *Eucalyptus*, during the next decade.

The US National Institute of Health request for proposals in revolutionary genome sequencing technologies (The \$1000 Genome) was launched in 2004. The 'goal of this

initiative is to reduce costs by at least four orders of magnitude, so that a mammalian-sized genome could be sequenced for approximately \$1000' within the next 10 yr. Since then, novel sequencing methods have led to a reduction in cost of at least one order of magnitude (Margulies *et al.*, 2005). Already, sequencing of an average *Eucalyptus* genome (~600 Mbp, 12X coverage) could be achieved using 454 Life Sciences (GS FLX) for less than \$1 million. This cost is expected to be rapidly reduced further through use of the established technology. In the short term, approaches based on single-base addition methods (SBA) promise to deliver over several Gbp of sequence at a cost of a few thousand dollars (http://marketing.appliedbiosystems.com/mk/get/SOLID_KNOWLEDGE_LANDING). Other methods still in development may be able to deliver complete genome sequences of relatively large genomes within the next decade. If the \$1000 human genome becomes feasible it can be expected that sequencing any *Eucalyptus* genome will be achievable at some fraction of that cost, creating the ideal scenario for the effective integration of molecular technologies into classical breeding programs.

As a first evaluation of these new sequencing technologies, our laboratory has recently completed the sequencing of 1 million short expressed sequence tags derived from a pool of tissues collected from 21 partially unrelated genotypes of *E. grandis* (E. Novaes, unpublished). The sequencing approach was based on massively parallel pyrosequencing, which can produce several hundred thousand short sequencing reads (~100–250 nucleotides) in a single assay. The purpose of our specific study was the discovery of genes and simultaneous identification of SNPs based on sequence alignments among unrelated haplotypes (Fig. 6). One limitation of this study was that LD and nucleotide diversity could not be estimated because the identity of each EST could not be related back to each *E. grandis* genotype. However, one may envision a scenario where cDNA material from individuals is sequenced either in separate reactions or utilizing some barcoding system that identifies the origin of each sequence (Parameswaran *et al.*, 2007), permitting the assessment of individual haplotypes.

SNP discovery and validation in *Eucalyptus* is currently limited to identifying SNPs *in silico* in EST assemblies and resequencing in relevant germplasm. However, standard approaches used to produce automated, unsupervised alignments (e.g. PHRAP; Lee & Vega, 2004) and to detect polymorphisms (e.g. POLYPHRED and POLYBAYES; Nickerson *et al.*, 1997; Marth *et al.*, 1999) are very limited when applied to highly diverse species such as *Eucalyptus*. To address the limitation of these software, new bioinformatics tools that are designed specifically for aligning and identifying SNPs in genetically diverse species are being developed (PINESAP; J. L. Wegrzyn, pers. comm.). Also, although SNPs are highly abundant in *Eucalyptus*, the lack of inbred lines or haploid tissue has rendered the detection of polymorphisms by resequencing almost intractable because of the very high frequency of indels which complicate direct sequencing from PCR without

(a)

Reference nucleotide (position)	A (158)	G (200)	T (258)	C (287)	C (431)	A (605)
# reads (forward)	27	29	18	14	13	35
# reads (reverse)	29	41	48	47	26	42
Alternative nucleotide	G	A	C	A	T	C
# reads (forward)	6	6	10	10	4	8
# reads (reverse)	4	9	12	7	12	7

(b)

Reference sequence	Start	Strand	Sequence	End
Reference sequence	584	+	CC-ATGG-AACG-TACGACTTCATCATT-CGT-GGAT-GCCGACAAGGA	625
EPE3D2102IPD9F	70	-	CCTATGG-AACG-TACGACTTCATCATT-CGT-GGAT-GCCGACAAGGA	28
EPE3D2101A5YSB	137	+	CC-ATGG-AACG-TACGACTTCATCATT-CGT-GGAT-GCCGACAAGGA	178
EPE3D2102JB6TU	137	+	CC-ATGG-AACG-TACGACTTCATCATT-CGT-GGAT-GCCGACAAGGA	177
EPE3D2102G9XCJ	137	+	CC-ATGG-AACG-TACGACTTCATCATT-CGT-GGAT-GCCGACAAGGA	178
EPE3D2101B7Z57	137	+	CC-ATGG-AACG-TACGACTTCATCATT-CGT-GGAT-GCCGACAAGGA	178
EPE3D2102GRHSL	136	-	CC-ATGG-AACG-TACGACTTCATCATT-CGT-GGAT-GCCGACAAGGA	95
EPE3D2101B9176	110	+	CC-ATGG-AACG-TACGACTTCATCATT-CGT-GGAT-GCCGACAAGGA	151
EPE3D2101EYG9P	136	-	CC-ATGG-AACG-TACGACTTCATCATT-CGT-GGAT-GCCGACAAGGA	95
EPE3D2101AWNSN	102	-	CC-ATGG-AACG-TACGACTTCATCATT-CGT-GGAT-GCCGACAAGGA	61
EPE3D2101BK40P	195	-	CC-ATGG-AACG-TACGACTTCATCATT-CGT-GGAT-GCCGACAAGGA	154
EZH8CPV02HE68U	48	-	CC-ATGGG--CG-TACGACTTCATCATT-CGT-GGAT-GCCGACAAGGA	8
EB6U72G01B7TNW	39	+	CC-ATGG-AACG-TACGACTTCATCATT-CGT-GGAT-GCCGACAAGGA	81
EPE3D2102HF0Y	11	+	CC-ATGG-AACG-TACGACTTCATCATT-CGT-GGAT-GCCGACAAGGA	53
EZH8CPV01BU497	96	-	CC-ATGG-AACG-TACGACTTCATCATT-CGT-GGAT-GCCGACAAGGA	55
EPE3D2102GBLRU	1	+	ATGGGAACGGTACGACTTCATCATT-CGTAGGAT-GCCGACAAGGA	43

Fig. 6 Single nucleotide polymorphism (SNP) discovery by massively parallel sequencing of normalized cDNA from unrelated *Eucalyptus* trees. (a) SNPs identified in the *Eucalyptus grandis* homologs of the lignin gene caffeoyl-CoA 3-O-methyltransferase. The reference nucleotide at an SNP is defined by the most frequently observed allele. In this analysis, at least two reads containing the reference and alternative SNP nucleotide, in both forward and reverse directions, were required to declare the presence of a SNP. (b) Alignment of 15 sequences (eight forward and seven reverse sequences) that contain the alternative alleles for the SNP located in position 605 (box). The gray shading indicates other variable positions that have not been declared polymorphic because of the limited number of sequences to support them.

cloning. With the trend towards genome-wide association analyses, there is an immediate need for robust, inexpensive, high-density and high-throughput markers. An approach that would allow parallel genotyping of the *Eucalyptus* gene space for sequence polymorphisms is highly desirable. Given the very high levels of sequence polymorphism in the *Eucalyptus* genome, Diversity Array Technology (DArT) (Jaccoud *et al.*, 2001; Wenzl *et al.*, 2006) could provide several thousand markers enriched for low-copy regions to allow a first move toward selection on a genome-wide scale (Meuwissen *et al.*, 2001) in specifically tailored small MAS *Eucalyptus* breeding populations. Unpublished data from an interspecific (*E. grandis* × *E. globulus*) expression profiling experiment carried out in the Genolyptus project (G. Pasquali, unpublished) on a 385 000 50mer Nimblegen oligoarray platform, representing 21 403 *Eucalyptus* spp. unigenes, was explored for putative single feature polymorphisms (SFPs) (Borevitz *et al.*, 2003). By analyzing the contrasts of least square means for probe vs. species interaction we found 32 077 probes in 10 781 genes that differentially hybridized to *E. globulus* and *E. grandis*, and are thus candidate segregating markers in crosses between these two species. Based on these exploratory data we recently validated the inheritance and segregation of SFPs using a pseudo-test-cross configuration in two segregating families

and a microarray with 15 180 short probes covering 1518 candidate genes involved in wood formation (10 probes/gene). A total of 1137 probes representing 658 genes were deemed polymorphic between the parents and segregated in the progeny sets, representing an efficiency of 43% for genetically mapping genes (D. Grattapaglia, unpublished). The expansion of this experiment will allow linkage mapping of several thousand genes and provide opportunities to verify co-localization with QTLs, thus supplying positional candidates to be tested in association genetics experiments.

Within the next decade we can expect the improvement of current SNP discovery, genotyping and sequencing methods, to the point that it will be feasible to identify all the sequence variants that occur at moderate to high frequencies in any *Eucalyptus* breeding population. The strategy will probably rely on ultra-high-throughput shotgun sequencing of each individual tree genome as has been proposed for human genomes (Mardis, 2008). With current technologies, assembling would still require a high-quality reference sequence and deep enough coverage so that the two homologous chromosomes of a heterozygous tree can be tentatively discriminated and separately assembled. Currently in progress, the complete sequence of the first *Eucalyptus* genome (*E. grandis*, clone BRASUZ1) will represent a milestone achievement and

serve well as a first reference sequence for future genomic undertakings. At that time the main limitation to moving from genotypes to phenotypes will probably be the availability of appropriate material, sufficiently replicated across field sites and precisely phenotyped for traits of interest. Therefore, for short-rotation woody crops such as *Eucalyptus*, it is essential that development of these populations be initiated within the next few years. Research institutions and industry with such materials readily available will potentially achieve the greatest gains when the technology becomes available. Given the extraordinary genetic variation that exists in the genus *Eucalyptus*, the ingenuity of most breeders and the powerful genomic tools that have become available, the outlook for applied genomics in *Eucalyptus* forest production is encouraging.

Acknowledgements

DG acknowledges support from the Brazilian Ministry of Science and Technology, the participating forestry companies in the Genolyptus project and the Brazilian National Research Council (CNPq) as well as the students, research collaborators and breeders for continued discussions and scientific input. Special thanks are due to Danielle Faria for providing access to yet unpublished work. MK acknowledges support from the US Department of Energy and National Science Foundation and the Consortium for Plant Biotechnology Research, as well as Evandro Novaes and Derek Drost for assistance with the preparation of the data presented here.

References

- de Aguiar MS, Ferreira DF, Aguiar AM, Bison O, Rezende GDSP, Grattapaglia, D. 2007. Potencial de híbridos entre clones-élite de eucalipto por meio de marcadores microssatélites. *Pesquisa Agropecuária Brasileira* 42: 1007–1012.
- de Assis TF. 2000. Production and use of *Eucalyptus* hybrids for industrial purposes. In: Nikles DG, ed. *Proceedings of the QFRI/CRC Workshop on Hybrid Breeding and Genetics of Forest Trees*. Brisbane, Australia: Department of Primary Industries, 63–74.
- de Assis TF, Rezende GDSP, Aguiar AM. 2005. Current status of breeding and deployment for clonal forestry with tropical eucalypt hybrids in Brazil. *International Forestry Review* 7: 61.
- Baril CP, Verhaegen D, Vigneron P, Bouvet JM, Kremer A. 1997. Structure of the specific combining ability between two species of *Eucalyptus*. I. RAPD data. *Theoretical and Applied Genetics* 94:796–803.
- Bhalerao R, Nilsson O, Sandberg G. 2003. Out of the woods: forest biotechnology enters the genomic era. *Current Opinion in Biotechnology* 14: 206–213.
- Bison O, Ramalho MAP, Rezende GDSP, Aguiar AM, de Resende MDV. 2007. Combining ability of elite clones of *Eucalyptus grandis* and *Eucalyptus urophylla* with *Eucalyptus globulus*. *Genetics and Molecular Biology* 30: 417–422.
- Boerjan W. 2005. Biotechnology and the domestication of forest trees. *Current Opinion in Biotechnology* 16: 159–166.
- Borevitz JO, Liang D, Plouffe D, Chang HS, Zhu T, Weigel D, Berry CC, Winzeler E, Chory J. 2003. Large-scale identification of single-feature polymorphisms in complex genomes. *Genome Research* 13: 513–523.
- Bouvet JM, Vigneron P. 1996. Variance structure in *Eucalyptus* hybrid populations. *Silvae Genetica* 45: 171–177.
- Bradshaw HD, Stettler RF. 1994. Molecular genetics of growth and development in *Populus*. 2. Segregation distortion due to genetic load. *Theoretical and Applied Genetics* 89: 551–558.
- Bradshaw HD, Villar M, Watson BD, Otto KG, Stewart S, Stettler RF. 1994. Molecular genetics of growth and development in *Populus*. 3. A genetic linkage map of a hybrid poplar composed of RFLP, STS, and RAPD markers. *Theoretical and Applied Genetics* 89: 167–178.
- Brondani RPV, Brondani C, Grattapaglia D. 2002. Towards a genus-wide reference linkage map for *Eucalyptus* based exclusively on highly informative microsatellite markers. *Molecular Genetics and Genomics* 267: 338–347.
- Brondani RPV, Brondani C, Tarchini R, Grattapaglia D. 1998. Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. *Theoretical and Applied Genetics* 97: 816–827.
- Brondani RPV, Williams ER, Brondani C, Grattapaglia D. 2006. A microsatellite-based consensus linkage map for species of *Eucalyptus* and a novel set of 230 microsatellite markers for the genus. *BMC Plant Biology* 6: 20.
- Brooker MIH. 2000. A new classification of the genus *Eucalyptus* L'Her. (Myrtaceae). *Australian Systematic Botany* 13: 79–148.
- Brune A, Zobel BJ. 1981. Genetic base populations, gene pools and breeding populations for *Eucalyptus* in Brazil. *Silvae Genetica* 30: 146–149.
- Bundock PC, Hayden M, Vaillancourt RE. 2000. Linkage maps of *Eucalyptus globulus* using RAPD and microsatellite markers. *Silvae Genetica* 49: 223–232.
- Burczyk J, Adams WT, Moran GF, Griffin AR. 2002. Complex patterns of mating revealed in a *Eucalyptus regnans* seed orchard using allozyme markers and the neighbourhood model. *Molecular Ecology* 11: 2379–2391.
- Busov V, Fladung M, Groover A, Strauss S. 2005. Insertional mutagenesis in *Populus*: relevance and feasibility. *Tree Genetics and Genomes* 1: 135–142.
- Byrne M, Marquezgarcia MI, Uren T, Smith DS, Moran GF. 1996. Conservation and genetic diversity of microsatellite loci in the genus *Eucalyptus*. *Australian Journal of Botany* 44: 331–341.
- Byrne M, Murrell JC, Allen B, Moran GF. 1995. An integrated genetic linkage map for *Eucalyptus* using RFLP, RAPD and isozyme markers. *Theoretical and Applied Genetics* 91: 869–875.
- Byrne M, Murrell JC, Owen JV, Kriedemann P, Williams ER, Moran GF. 1997a. Identification and mode of action of quantitative trait loci affecting seedling height and leaf area in *Eucalyptus nitens*. *Theoretical and Applied Genetics* 94: 674–681.
- Byrne M, Murrell JC, Owen JV, Williams ER, Moran GF. 1997b. Mapping of quantitative trait loci influencing frost tolerance in *Eucalyptus nitens*. *Theoretical and Applied Genetics* 95: 975–979.
- Campinhos E, Ikemori YK. 1977. Tree improvement program of *Eucalyptus* spp.: preliminary results. In: *Third World Consultation on forest tree breeding*. Canberra, Australia: CSIRO, 717–738.
- Cardon LR, Bell JL. 2001. Association study designs for complex diseases. *Nature Reviews Genetics* 2: 91–99.
- Chaix G, Gerber S, Razafimaharo V, Vigneron P, Verhaegen D, Hamon S. 2003. Gene flow estimation with microsatellites in a Malagasy seed orchard of *Eucalyptus grandis*. *Theoretical and Applied Genetics* 107: 705–712.
- Chen ZZ, Ho CK, Ahn IS, Chiang VL. 2006. *Eucalyptus*. *Methods in Molecular Biology* 344: 125–34.
- Clark RM, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P, Warthmann N, Hu TT, Fu G, Hinds DA et al. 2007. Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* 317: 338–342.
- Doughty RW. 2000. *The eucalyptus: a natural and commercial history of the gum tree*. Baltimore, MD, USA: The Johns Hopkins University Press.
- Eldridge K, Davidson J, Harwood C, van Wyk G. 1993. *Eucalypt domestication and breeding*. Oxford, UK: Clarendon Press.

- Feuillet C, Boudet AM, Grima-Pettenati J. 1993. Nucleotide sequence of a cDNA encoding cinnamyl alcohol dehydrogenase from *Eucalyptus*. *Plant Physiology* 103: 1447.
- Foucart C, Paux E, Ladouce N, San-Clemente H, Grima-Pettenati J, Sivadon P. 2006. Transcript profiling of a xylem vs phloem cDNA subtractive library identifies new genes expressed during xylogenesis in *Eucalyptus*. *New Phytologist* 170: 739–752.
- Freeman JS, Potts BM, Vaillancourt RE. 2008. Few Mendelian genes underlie the quantitative response of a forest tree, *Eucalyptus globulus*, to a natural fungal epidemic. *Genetics* 178: 563–571.
- Gaiotto FA, Bramucci M, Grattapaglia D. 1997. Estimation of outcrossing rate in a breeding population of *Eucalyptus urophylla* s.t. Blake with dominant RAPD and AFLP markers. *Theoretical and Applied Genetics* 95: 842–849.
- Gion J-M, Rech P, Grima-Pettenati J, Verhaegen D, Plomion C. 2000. Mapping candidate genes in *Eucalyptus* with emphasis on lignification genes. *Molecular Breeding* 6: 441–449.
- Glaubitiz JC, Emebiri LC, Moran GF. 2001. Dinucleotide microsatellite from *Eucalyptus sieberi*: Inheritance, diversity, and improved scoring of single-base differences. *Genome* 44: 1041–1045.
- Gonzalez-Martinez SC, Ersoz E, Brown GR, Wheeler NC, Neale DB. 2006. DNA sequence variation and selection of tag SNPs at candidate genes for drought-stress response in *Pinus taeda* L. *Genetics* 172: 1915–1926.
- Gonzalez-Martinez SC, Wheeler NC, Ersoz E, Nelson DR, Neale DB. 2007. Association genetics in *Pinus taeda* L. I. Wood property traits. *Genetics* 175: 399–409.
- Grattapaglia D. 2004. Integrating genomics into *Eucalyptus* breeding. *Genetics and Molecular Research* 3: 369–379.
- Grattapaglia D. 2008. Genomics of *Eucalyptus*, a global tree for energy, paper, and wood. In: Moore P, Ming R, eds. *Genomics of tropical crop plants*. New York, NY, USA: Springer, 259–298.
- Grattapaglia D, Alfenas AC, Coelho ASG, Bearzoti E, Pappas GJ, Pasquali G, Pereira G, Colodette J, Gomide JL, Bueno J *et al.* 2004b. Building resources for molecular breeding of *Eucalyptus*. In: Borralho NMG, Pereira JS, Marques C, Coutinho J, Madeira M, Tone M, eds. *Proceedings of the International IUFRO Conference – Eucalyptus in a changing world*. Aveiro, Portugal: RAIZ Instituto de Investigação de Floresta e Papel, 20–32.
- Grattapaglia D, Bertolucci FLG, Penchel R, Sederoff R. 1996. Genetic mapping of quantitative trait loci controlling growth and wood quality traits in *Eucalyptus grandis* using a maternal half-sib family and RAPD markers. *Genetics* 144: 1205–1214.
- Grattapaglia D, Bertolucci FLG, Sederoff R. 1995. Genetic mapping of QTLs controlling vegetative propagation in *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross mapping strategy and RAPD markers. *Theoretical and Applied Genetics* 90: 933–947.
- Grattapaglia D, Chaparro J, Wilcox P, Mccord S, Werner D, Amerson H, Mckeand S, Bridgwater F, Whetten R, O'malley D, *et al.* 1992. Mapping in woody plants with RAPD markers: applications to breeding in forestry and horticulture. In: *Proceedings of the Symposium Applications of RAPD Technology to Plant Breeding*. Minneapolis, MN, USA: Crop Science Society of America, American Society of Horticultural Science, American Genetic Association, 37–40.
- Grattapaglia D, Ribeiro VJ, Rezende, GDSP. 2004a. Retrospective selection of elite parent trees using paternity testing with microsatellite markers: an alternative short term breeding tactic for *Eucalyptus*. *Theoretical and Applied Genetics* 109: 192–199.
- Grattapaglia D, Sederoff R. 1994. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. *Genetics* 137: 1121–1137.
- Griffin AR, Burgess IP, Wolf L. 1988. Patterns of natural and manipulated hybridisation in the genus *Eucalyptus* L'Herit: a review. *Australian Journal of Botany* 36: 41–66.
- Hartl DL, Clark AG. 1997. *Principles of population genetics*. Sunderland, MA, USA: Sinauer Associates.
- Henery ML, Moran GF, Wallis IR, Foley WJ. 2007. Identification of quantitative trait loci influencing foliar concentrations of terpenes and formylated phloroglucinol compounds in *Eucalyptus nitens*. *New Phytologist* 176: 82–95.
- Horsley TN, Johnson SD. 2007. Is *Eucalyptus* cryptically self-incompatible? *Annals of Botany* 100: 1373–1378.
- House APN, Bell JC. 1994. Isozyme variation and mating system in *Eucalyptus urophylla* S.T. Blake. *Silvae Genetica* 43: 167–176.
- Ikemori YK, Penchel RM, Bertolucci FLG. 1994. Integrating biotechnology into *Eucalyptus* breeding. In: *Proceedings of the International Symposium on Wood Biotechnology*. Tokyo, Japan: TAPPI, Japan Wood Research Society and Nippon Paper Industries, 79–84.
- Jaccoud D, Peng K, Feinstein D, Kilian A. 2001. Diversity arrays: a solid state technology for sequence independent genotyping. *Nucleic Acids Research* 29: e25.
- Jansen RC, Nap JP. 2001. Genetical genomics: the added value from segregation. *Trends in Genetics* 17: 388–391.
- Jones ME, Shepherd M, Henry RJ, Delves A. 2006. Chloroplast DNA variation and population structure in the widespread forest tree, *Eucalyptus grandis*. *Conservation Genetics* 7: 691–703.
- Jungthans DT, Alfenas AC, Brommonschenkel SH, Oda S, Mello EJ Grattapaglia D. 2003. Resistance to rust (*Puccinia psidii* Winter) in *Eucalyptus*: mode of inheritance and mapping of a major gene with RAPD markers. *Theoretical and Applied Genetics* 108: 175–80.
- Kawazu T, Doi K, Kondo K. 2003. Process for transformation of mature trees of *Eucalyptus* Plants. US Patent 6,563,024 B1.
- Keil M, Griffin AR. 1994. Use of random amplified polymorphic DNA (RAPD) markers in the discrimination and verification of genotypes in *Eucalyptus*. *Theoretical and Applied Genetics* 89: 442–450.
- Kellison RC. 2001. Present and future uses of eucalyptus wood in the world. *Proceedings of the IUFRO International Symposium 'Developing the eucalypt of the future'*. Valdivia, Chile: INFOR Instituto Forestal (published on CD).
- Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, Ecker JR, Weigel D, Nordborg M. 2007. Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nature Genetics* 39: 1151–1155.
- Kirst M. 1999. Desenvolvimento e aplicação de sistemas de genotipagem multiloco, baseados em microsatélites para análise de espécies do genero *Eucalyptus*. MSc thesis, Universidade Federal de Viçosa, Viçosa, Brazil.
- Kirst M, Basten CJ, Myburg A, Zeng Z-B, Sederoff R. 2005b. Genetic architecture of transcript level variation in differentiating xylem of *Eucalyptus* hybrids. *Genetics* 169: 2295–2303.
- Kirst M, Brondani RPV, Brondani C, Grattapaglia D. 1997. Screening of designed primer pairs for recovery of microsatellite markers and their transferability among species of *Eucalyptus*. In: *Proceedings of IUFRO Conference Eucalyptus Genetics and Silviculture*. Colombo, Brazil: EMBRAPA, 167–171.
- Kirst M, Cordeiro CM, Rezende GD, Grattapaglia D. 2005a. Power of microsatellite markers for fingerprinting and parentage analysis in *Eucalyptus grandis* breeding populations. *Journal of Heredity* 96: 161–166.
- Kirst M, Myburg AA, De Leon JP, Kirst ME, Scott J, Sederoff R. 2004. Coordinated genetic regulation of growth and lignin revealed by quantitative trait locus analysis of cDNA microarray data in an interspecific backcross of *Eucalyptus*. *Plant Physiology* 135: 2368–2378.
- Knowler WC, Williams RC, Pettitt DJ, Steinberg AG. 1988. Gm3-5,13,14 and type-2 diabetes-mellitus – an association in American-Indians with genetic admixture. *American Journal of Human Genetics* 43: 520–526.
- Ladiges PY, Udovicic F, Nelson G. 2003. Australian biogeographical connections and the phylogeny of large genera in the plant family *Myrtaceae*. *Journal of Biogeography* 30: 989–998.
- Lee WH, Vega VB. 2004. Heterogeneity detector: finding heterogeneous positions in Phred/Phrap assemblies. *Bioinformatics* 20: 2863–2864.

- Lexer C, Buerkle CA, Joseph JA, Heinze B, Fay MF. 2007. Admixture in European *Populus* hybrid zones makes feasible the mapping of loci that contribute to reproductive isolation and trait differences. *Heredity* 98: 74–84.
- Lezar S, Myburg AA, Berger DK, Wingfield MJ, Wingfield BD. 2004. Development and assessment of microarray-based DNA fingerprinting in *Eucalyptus grandis*. *Theoretical and Applied Genetics* 109: 1329–1336.
- Long AD, Langley CH. 1999. The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. *Genome Research* 9: 720–731.
- Machado LO, de Andrade GM, Cid LPB, Penchel RM, Brasileiro ACM. 1997. Agrobacterium strain specificity and shooty tumour formation in eucalypt *Eucalyptus grandis* × *E. urophylla*. *Plant Cell Reports* 16: 299–303.
- MacKay JJ, O'Malley DM, Presnell T, Booker FL, Campbell MM, Whetten RW, Sederoff RR. 1997. Inheritance, gene expression, and lignin characterization in a mutant pine deficient in cinnamyl alcohol dehydrogenase. *Proceedings of the National Academy of Sciences, USA* 94: 8255–8260.
- MacRae S, van Staden J. 1999. Transgenic *Eucalyptus*. In: Bajaj YPS, ed. *Biotechnology in agriculture and forestry*. Heidelberg, Germany: Springer, 88–114.
- Marcucci-Poltri SN, Zelener N, Rodriguez Traverso J, Gelid P, Hopp H. 2003. Selection of a seed orchard of *Eucalyptus dunnii* based on genetic diversity criteria calculated using molecular markers. *Tree Physiology* 23: 625–632.
- Mardis ER. 2008. The impact of next-generation sequencing technology on genetics. *Trends in Genetics* 24: 133–41.
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z *et al.* 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437: 376–380.
- Marques CM, Araujo JA, Ferreira JG, Whetten R, O'Malley DM, Liu BH, Sederoff RR. 1998. AFLP genetic maps of *Eucalyptus globulus* and *E. tereticornis*. *Theoretical and Applied Genetics* 96: 727–737.
- Marques CM, Brondani RPV, Grattapaglia D, Sederoff R. 2002. Conservation and synteny of SSR loci and QTLs for vegetative propagation in four *Eucalyptus* species. *Theoretical and Applied Genetics* 105: 474–478.
- Marques CM, Carocha VJ, Pereira de Sá AR, Oliveira MR, Pires AM, Sederoff RR, Borralho NMG. 2005. Verification of QTL linked markers for propagation traits in *Eucalyptus*. *Tree Genetics and Genomes* 1: 103–108.
- Marques CM, Vasquez-Kool J, Carocha VJ, Ferreira JG, O'Malley DM, Liu BH, Sederoff RR. 1999. Genetic dissection of vegetative propagation traits in *Eucalyptus tereticornis* and *E. globulus*. *Theoretical and Applied Genetics* 99: 936–946.
- Marth GT, Korfi I, Yandell MD, Yeh RT, Gu Z, Zakeri H, Stitzel NO, Hillier L, Kwok PY, Gish WR. 1999. A general approach to single-nucleotide polymorphism discovery. *Nature Genetics* 23: 452–456.
- Martin B, Quillet J. 1974. The propagation by cuttings of forest trees in the Congo. *Bois et Forêts des Tropiques* 155: 15–33.
- Meuwissen TH, Hayes BJ, Goddard ME. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819–1829.
- Missiaggia AA, Piacuzzi AL, Grattapaglia D. 2005. Genetic mapping of Eef1, a major effect QTL for early flowering in *Eucalyptus grandis*. *Tree Genetics and Genomes* 1: 79–84.
- Moran G, Bell JC, Griffin AR. 1989. Reduction in levels of inbreeding in a seed orchard of *Eucalyptus regnans* F. Muell, compared with natural populations. *Silvae Genetica* 38: 32–36.
- Myburg AA, Griffin AR, Sederoff RR, Whetten RW. 2003. Comparative genetic linkage maps of *Eucalyptus grandis*, *Eucalyptus globulus* and their F1 hybrid based on a double pseudo-backcross mapping approach. *Theoretical and Applied Genetics* 107: 1028–1042.
- Myburg AA, Potts B, Marques CM, Kirst M, Gion JM, Grattapaglia D, Grima-Pettenati J. 2007. *Eucalyptus*. In: Kole C, ed. *Genome mapping and molecular breeding in plants, Vol. 7: Forest trees*. New York, NY, USA: Springer, 115–160.
- Neale DB, Williams CG. 1991. Restriction fragment length polymorphism mapping in conifers and applications to forest genetics and tree improvement. *Canadian Journal of Forest Research* 21: 545–554.
- Nesbitt KA, Potts BM, Vaillancourt RE, West AK, Reid JB. 1995. Partitioning and distribution of RAPD variation in a forest tree species, *Eucalyptus globulus* (Myrtaceae). *Heredity* 74: 628–637.
- Nickerson DA, Tobe VO, Taylor SL. 1997. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Research* 25: 2745–2751.
- O'Malley D, Sederoff R, Grattapaglia D. 1994. Methods for within family selection in woody perennials using genetic markers. United States Patent and Trademark Office – Pat #6,054,634. www.uspto.gov.
- Ottewell KM, Donnellan SC, Moran GF, Paton DC. 2005. Multiplexed microsatellite markers for the genetic analysis of *Eucalyptus leucocoxylon*, Myrtaceae and their utility for ecological and breeding studies in other *Eucalyptus* species. *Journal of Heredity* 96: 445–451.
- Parameswaran P, Jalili R, Tao L, Shokralla S, Gharizadeh B, Ronaghi M, Fire AZ. 2007. A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. *Nucleic Acids Research* 35: e130, doi:10.1093/nar/gkm760.
- Paux E, Tamasloukht M, Ladouce N, Sivadon P, Grima-Pettenati J. 2004. Identification of genes preferentially expressed during wood formation in *Eucalyptus*. *Plant Molecular Biology* 55: 263–80.
- Payn KG, Dvorak WS, Myburg AA. 2007. Chloroplast DNA phylogeography reveals the island colonisation route of *Eucalyptus urophylla* (Myrtaceae). *Australian Journal of Botany* 55: 673–683.
- Poke FS, Vaillancourt RE, Potts BM, Reid JB. 2005. Genomic research in *Eucalyptus*. *Genetica* 125: 79–101.
- Potts BM. 2004. Genetic improvement of eucalypts. In: Burley J, Evans J, Youngquist JA, eds. *Encyclopedia of forest science*. Oxford, UK: Elsevier Science, 1480–1490.
- Potts BM, Dungey HS. 2004. Hybridisation of *Eucalyptus*: Key issues for breeders and geneticists. *New Forests* 27: 115–138.
- Potts BM, Savva M. 1988. Self-incompatibility in *Eucalyptus*. In: Knox RB, Singh MB, Troini L, eds. *Pollination '88*. Melbourne, Australia: School of Botany, University of Melbourne, 165–170.
- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
- Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. 2000. Association mapping in structured populations. *American Journal of Human Genetics* 67: 170–181.
- Pryor LD, Johnson LAS. 1971. *A classification of the eucalypts*. Canberra, Australia: Australian National University Press.
- Ralph J, MacKay JJ, Hatfield RD, O'Malley DM, Whetten RW, Sederoff RR. 1997. Abnormal lignin in a loblolly pine mutant. *Science* 277: 235–239.
- Rezende GDSP, de Resende MDV. 2000. Dominance effects in *Eucalyptus grandis*, *Eucalyptus urophylla* and hybrids. In: Dungey HSD, Nikles DG, eds. *Hybrid breeding and genetics of forest trees*. Brisbane, Australia: Queensland Department of Primary Industries, 93–100.
- Romeo S, Pennacchio LA, Fu YX, Boerwinkle E, Tybjaerg-Hansen A, Hobbs HH, Cohen JC. 2007. Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. *Nature Genetics* 39: 513–516.
- Santos SN. 2005. Genes de lignificação em *Eucalyptus*: estrutura e diversidade genética dos genes 4cl e ccoaoamt. MSc thesis, Universidade Católica de Brasília, Brasília, Brazil (available online at http://biblioteca.universia.net/html_bura/ficha/params/id/3275859.html).
- Shepherd M, Chaparro JX, Teasdale R. 1999. Genetic mapping of monoterpene composition in an interspecific eucalypt hybrid. *Theoretical and Applied Genetics* 99: 1207–1215.

- Sjödín A, Bylesjö M, Skogström O, Eriksson D, Nilsson P, Rydén P, Jansson S, Karlsson J. 2006. UPSC-BASE – *Populus* transcriptomics online. *Plant Journal* 48: 806–817.
- Snedden CL, Verryin SD. 2004. A comparative study of predicted gains for selection from a cloned breeding population and the implications for deployment. In: Tomé M, ed. *Proceedings of the IUFRO Conference 'Eucalyptus in a Changing World'*. Aveiro, Portugal: RAIZ, Instituto Investigação de Floresta e Papel, 137–144.
- Spokevicius AV, Southerton SG, MacMillan CP, Qiu D, Gan S, Tibbits JFG, Moran GF, Bossinger G. 2007. B-tubulin affects cellulose microfibril orientation in plant secondary fibre cell walls. *Plant Journal* 51: 717–726.
- Spokevicius AV, Van Beveren K, Leitch MA, Bossinger G. 2005. Agrobacterium-mediated in vitro transformation of wood-producing stem segments in eucalypts. *Plant Cell Reports* 23: 617–624.
- Steane DA, Conod N, Jones RC, Vaillancourt RE, Potts BM. 2006. A comparative analysis of population structure of a forest tree, *Eucalyptus globulus* (Myrtaceae), using microsatellite markers and quantitative traits. *Tree Genetics and Genomes* 2: 30–38.
- Steane DA, Nicolle D, Vaillancourt RE, Potts BM. 2002. Higher-level relationships among the eucalypts are resolved by ITS-sequence data. *Australian Systematic Botany* 15: 49–62.
- Steane DA, Vaillancourt RE, Russell J, Powell W, Marshall D, Potts BM. 2001. Development and characterization microsatellite loci in *Eucalyptus globulus* (Myrtaceae). *Silvae Genetica* 50: 89–91.
- Strauss SH, Lande R, Namkoong G. 1992. Obstacles to molecular-marker-aided selection in forest trees. *Canadian Journal of Forest Research* 22: 1050–1061.
- Thamarus K, Groom K, Murrell J, Byrne M, Moran G. 2002. A genetic linkage map for *Eucalyptus globulus* with candidate loci for wood, fibre and floral traits. *Theoretical and Applied Genetics* 104: 379–387.
- Thamarus KA, Groom K, Bradley A, Raymond CA, Schimleck LR, Williams ER, Moran GF. 2004. Identification of quantitative trait loci for wood and fibre properties in two full-sib pedigrees of *Eucalyptus globulus*. *Theoretical and Applied Genetics* 109: 856–864.
- Thumma RF, Nolan MF, Evans R, Moran GF. 2005. Polymorphisms in cinnamoyl CoA reductase (CCR) are associated with variation in microfibril angle in *Eucalyptus* spp. *Genetics* 171: 1257–1265.
- Topol EJ, Frazer KA. 2007. The resequencing imperative. *Nature Genetics* 39: 439–440.
- Tournier V, Grat S, Marque C, El Kayal W, Penchel R, de Andrade G, Boudet AM, Teulieres C. 2003. An efficient procedure to stably introduce genes into an economically important pulp tree (*Eucalyptus grandis* × *Eucalyptus urophylla*). *Transgenic Research* 12: 403–411.
- Vaillancourt RE, Potts BM, Manson A, Eldridge T, Reid JB. 1995b. Using RAPD's to detect QTLs in an interspecific F2 hybrid of *Eucalyptus*. In: Potts BM, Borralho NMG, Reid JB, Cromer RN, Tibbits WN, Raymond CA, eds. *Proceedings of the CRC/IUFRO Conference Eucalypt plantations, improving fibre yield and quality*. Hobart, Australia: CRC for Temperate Hardwood Forestry, 430–433.
- Vaillancourt RE, Potts BM, Watson M, Volker PW, Hodge GR, Reid, JB, West, AK. 1995a. Prediction of heterosis using RAPD generated distances between *Eucalyptus globulus* trees. In: Potts BM, Borralho NMG, Reid JB, Cromer RN, Tibbits WN, Raymond CA, eds. *Proceedings of the CRC/IUFRO Conference Eucalypt plantations, improving fibre yield and quality*. Hobart, Australia: CRC for Temperate Hardwood Forestry, 455–456.
- Verhaegen D, Plomion C. 1996. Genetic mapping in *Eucalyptus urophylla* and *E. grandis*. RAPD markers. *Genome* 39: 1051–1061.
- Verhaegen D, Plomion C, Gion JM, Poitel M, Costa P, Kremer A. 1997. Quantitative trait dissection analysis in *Eucalyptus* using RAPD markers. 1. Detection of QTL in interspecific hybrid progeny, stability of QTL expression across different ages. *Theoretical and Applied Genetics* 95: 597–608.
- Vicentini R, Sasaki FT, Gimenes MA, Maia IG, Menossi M. 2005. In silico evaluation of the *Eucalyptus* transcriptome. *Genetics and Molecular Biology* 28: 487–495.
- Wenzl P, Li H, Carling J, Zhou M, Raman H, Paul E, Hearnden P, Maier C, Xia L, Caig V *et al.* 2006. A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. *BMC Genomics* 7: 206.
- Wu RL, Remington DL, MacKay JJ, McKeand SE, O'Malley DM. 1999. Average effect of a mutation in lignin biosynthesis in loblolly pine. *Theoretical and Applied Genetics* 99: 705–710.
- Yu J, Holland JB, McMullen MD, Buckler ES. 2008. Genetic design and statistical power of Nested Association Mapping in maize. *Genetics* 178: 539–551.
- Zelener N, Poltri SN, Bartoloni N, Lopez CR, Hopp HE. 2005. Selection strategy for a seedling seed orchard design based on trait selection index and genomic analysis by molecular markers: a case study for *Eucalyptus dunnii*. *Tree Physiology* 25: 1457–1467.



About New Phytologist

- *New Phytologist* is owned by a non-profit-making **charitable trust** dedicated to the promotion of plant science, facilitating projects from symposia to open access for our Tansley reviews. Complete information is available at www.newphytologist.org.
- Regular papers, Letters, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication 'as-ready' via *Early View* – our average submission to decision time is just 29 days. Online-only colour is **free**, and essential print colour costs will be met if necessary. We also provide 25 offprints as well as a PDF for each article.
- For online summaries and ToC alerts, go to the website and click on 'Journal online'. You can take out a **personal subscription** to the journal for a fraction of the institutional price. Rates start at £135 in Europe/\$251 in the USA & Canada for the online edition (click on 'Subscribe' at the website).
- If you have any questions, do get in touch with Central Office (newphytol@lancaster.ac.uk; tel +44 1524 594691) or, for a local contact in North America, the US Office (newphytol@ornl.gov; tel +1 865 576 5261).