Southern Forests 2008, 70(2): 69–75
Printed in South Africa — All rights reserved

SOUTHERN FORESTS
EISSN 2070–2639
doi: 10.2989/SOUTH.FOR.2008.70.2.2.530

Perspectives on genome mapping and marker-assisted breeding of eucalypts

D Grattapaglia

EMBRAPA Genetic Resources and Biotechnology, CP 02372-70770-900 Brasília, Distrito Federal, Brazil
and Genomic Sciences Program, Universidade Católica de Brasília, SGAN 916 Módulo B Brasília, Distrito Federal, Brazil
E-mail: dario@cenargen.embrapa.br

In *Eucalyptus* and forest tree breeding in general, the generalised application of molecular markers for directional selection is still an unfulfilled promise. In highly heterogeneous eucalypts, while conventional quantitative trait loci (QTL) mapping has revealed useful markers that are currently exploited in within-family selection tactics, only a more direct linkage disequilibrium mapping approach will likely uncover population-wide applicable marker–trait associations. Due to the very low-range linkage disequilibrium seen in *Eucalyptus*, a whole-genome association approach does not seem immediately feasible. Association studies based on candidate genes have been started in *Eucalyptus*, but only small proportions of the variation have been accounted for by such genes to be exciting news to breeders. Selection of candidate genes for direct manipulation or association studies based on their presumed biochemical role is not an easy task even for well-defined phenotypes and/or known metabolic pathways. Going from phenotypes to genes by a forward genomics approach based on an integrative expression–QTL mapping route, should prove to be a powerful way to choose target genes for marker-assisted selection. In this context at least two possibilities have recently emerged. The first one is to use high-performance genotyping technologies that would allow sufficient throughput and low cost for association genetic analysis of thousands of genes at a time. The second one would be to have access to a whole genome sequence so that candidate genes in a fine mapping interval delimited by flanking markers could be mined, reannotated and then analysed in association mapping. A fully public draft of the *E. grandis* genome will be sequenced by the US Department of Energy within the next few years following a proposal submitted by the International *Eucalyptus* Genome Network (EUCAGEN). As this genome project advances and more powerful tools become accessible, the true challenge for understanding and manipulating the complex nature of important traits in *Eucalyptus* will depend to a large extent on our ability to accurately phenotype trees, analyse the overwhelming amount of genomic data and translate this into truly useful molecular tools for breeding.

Keywords: *Eucalyptus*, genomics, molecular breeding

Introduction

Intensive production forestry based on exotics began in the Southern Hemisphere about 50 years ago. Since then, the world forest industry has experienced a slow but steady and now increasing shift of plantation forestry from the Northern Hemisphere to the southern tropics and subtropics and to the warmer, temperate climates of New Zealand, Chile and South Africa. *Eucalyptus* species have been key players in this process. High-productivity eucalypt forests have supplied, in a rational and efficient way, high-quality woody raw material for pulp, paper and energy. Planted forests have had an important role as substitution forests for woody biomass that would otherwise come from native tropical forests. However, it is clear that the expansion of these 'fibre farms' will likely be limited by the growth of food and biofuels crops and, in some cases, by public opinion pressure. Increased forest productivities and refinements in the quality of wood products by genome-assisted breeding and transgenic technologies will become increasingly strategic to the forest industry.

While a number of genes involved in lignin composition have been intensively investigated and manipulated in recent years, high-impact applications of transgenics in eucalypt production forestry are still to come at the same time that biosafety challenges persist. Challenges are also faced by molecular breeding and marker-assisted selection (MAS), the theme of this brief presentation. Twenty years have passed since the first demonstrations that major-effect quantitative trait loci (QTLs) could be mapped with molecular markers in plants and almost 15 years since the first experiments in genetic mapping of forest trees. Many expectations of fast and accurate methods for early marker-based selection for wood properties in trees were generated. Significant progress has been made and some short-term applications have been incorporated into tree genetics and breeding. However, it also became clear that several challenges are still ahead before more refined and higher impact applications can be implemented. While the reverse genomics approach aims to determine the

phenotype that results from mutating a given gene through transgenic technology, the forward genomics approach, i.e. going from the analysis of existing phenotypic variation for wood properties to the causal genetic variants, is based on the wide natural intra- and inter-specific variation that exists in *Eucalyptus*. The key technologies involved in this approach include genetic mapping, QTL discovery, physical mapping and genome sequencing. In this paper, I briefly report on some of the advances we have made in the GENOLYPTUS project, the Brazilian Network of *Eucalyptus* Genome Research, and discuss some of the challenges and opportunities that exist for implementing marker-assisted selection in *Eucalyptus*. More in-depth reviews on some of these topics have been published elsewhere (Poke *et al*., 2005; Grattapaglia, 2007; Myburg *et al*., 2007).

Genetic resources for Eucalyptus *applied genomics*

While public genomic resources, including a complete genome sequence, will become available in the very near future, biological resources and precise phenotyping represent the real limitation of many applied genomic projects. Especially in forest trees, where generation times and phenotype assessment can take years, the availability of ideal experimental populations should be one of the main targets in any genomic project. The driving principle we have adopted in the GENOLYPTUS project is that there is ample genetic variation within the genus *Eucalyptus*, and more specifically within the subgenus *Symphyomyrtus*, to allow profound genetic modification of the current planting stock in Brazil. For example, *E. globulus* contrasts with commonly used tropical species such as *E. grandis, E. urophylla* and *E. camaldulensis*, for it displays a number of wood properties extremely interesting to industry. *Eucalyptus globulus* germplasm stands out as a very rich source of genetic variation for all the target wood traits and therefore a key resource for eucalypt genomic research, especially for the pulp and paper industries. It is well known by breeders and wood technologists that *E. globulus* displays the best combination of wood properties for pulp and paper among the commercially planted *Eucalyptus* species, resulting in a high pulp yield requiring approximately 25% less wood than *E. grandis* to produce the same ton of cellulose. Therefore, even at slower growth rates due to its temperate origin, *E. globulus* wood is today the preferred raw material by the mills, generating a pulp that has increasingly been seen as a distinct and superior product by the market.

A number of experimental data from hybridisation experiments in Brazil are already available to clearly demonstrate that the introgression of temperate *E. globulus* alleles into tropical hybrid breeding programs, coupled to clonal propagation of selected individuals, will result in significant reductions in wood-specific consumption (de Assis 2000). Most Brazilian breeders are currently investing heavily based on the potential impact that the use of *E. globulus* could have in their programs. It is now just a matter of time and systematic investment for such gains to be realised in the mill. Several trials have been established in recent years and elite hybrid clones of *E. globulus* with *E. grandis* and *E. urophylla* that consolidate outstanding growth and wood properties in tropical conditions were selected and will soon make up the bulk of the clonal forests for pulp and paper in Brazil. This same view was also adopted in the construction of the biological resources for genomic research in the GENOLYPTUS project. Over 20 intra- and inter-specific families involving different *Eucalyptus* species were generated and are currently being used for QTL mapping and gene expression work. By establishing a rich resource of genetic variation resulting from hybridisation we hope to contribute to uncovering the genetic causes that make the *E. globulus* wood so different from the wood of *E. grandis*.

QTL mapping for wood properties traits

Molecular marker maps have been successfully used to detect major-effect QTLs in *Eucalyptus* for wood properties at rotation age traits, such as volume growth, wood specific gravity, bark thickness and stem form, and for several wood traits such as pulp yield at Kappa 18, alkali consumption at Kappa 18, basic density, oven-dry lignin content, extractives-free lignin content, extractives content, cellulose content, heat content, fibre length and fibre coarseness using near-infrared (NIR) analysis of wood core samples (reviewed in Myburg *et al*., 2007). With the recent development of more comprehensive genetic maps built with transportable, multiallelic microsatellite markers (Brondani *et al*., 2006), inter-pedigree QTL validation efforts have started to identify genomic regions controlling traits of interest for MAS consistently. In the GENOLYPTUS project we have consolidated the construction of multiple genetic maps and the detection of QTLs for several wood properties. We have recently shown the possibility of performing comparative QTL position analysis between independent experiments. Genetic maps and QTL analyses have been carried out for three independent genetically unrelated families and mapping is currently ongoing for four more families. Comparative QTL mapping across these pedigrees as well as to QTLs, and candidate genes mapping carried out in *E. globulus* by other research groups, revealed a number of syntenic QTLs for cellulose yield, lignin content and for different but correlated fibre traits as well as candidate genes for the lignification pathway. These are exciting results for *Eucalyptus* as they revealed the first QTL validation data and demonstrated the power of using a higher density of microsatellites for QTL validation, directed search for allelic variants at QTLs in multiple pedigrees and thus allowing precise determination of target genomic regions for gene discovery and MAS (Missiaggia *et al*., 2005).

Gene discovery

Up until the announcement of the completion of the genome sequence of *Populus trichocarpa* (Tuskan *et al*., 2006), gene discovery in trees followed the general route of sequencing only the expressed portions of the genome, called expressed sequence tags (ESTs). EST sequencing quickly generates a large index of partial genes for the organism of interest making them available in organised collections of clusterised sequences for further molecular investigation. Partial gene sequences generated have had multiple applications including: (1) the identification of genes and gene families involved in the control of target traits; (2) the identification of new molecular markers, such

as microsatellites and single-nucleotide polymorphisms (SNPs) for mapping; (3) supplying sequence information or biological reagents to build microarrays for large-scale gene expression studies; and (4) supplying sequences or genes for transgenic experiments. Efforts to build EST databases for *Eucalyptus* started about 10 years ago by the now-extinct Fletcher Challenge Corporation in New Zealand, when several tens-of-thousand sequences were generated. With advancements in sequencing technology the size of the databases quickly became an order of magnitude larger with hundreds-of-thousands of sequences. The different species planted around the world and the hybrid breeding system adopted in *Eucalyptus* has driven a distinctive multi-species approach to the EST sequencing efforts. For example, in the GENOLYPTUS project we have completed an initial database of over 125 000 EST sequences derived from 20 different cDNA libraries from four *Eucalyptus* species with an increased focus on xylem transcripts. Several-thousand cDNA clones were sequenced from *E. globulus, E. grandis, E. pellita* and *E. urophylla* xylem and phloem libraries derived from a number of individuals for each species to tackle both gene and SNP discovery both at the inter- and intra-specific levels. Data mining of this database has revealed that it contains all the known genes for lignin and cellulose metabolism as well as several other genes for cell wall structural proteins that have been described as involved in the control of chemical wood properties (Pasquali *et al*., 2005). It is expected, however, that with the breakthrough advancements of speed and dramatic cost reductions brought about by pyrosequencing technologies, the size of public *Eucalyptus* EST databases will very soon become much larger and diverse. Recently, Novaes *et al*. (2007) reported on the generation of more than 1 million *E. grandis* sequence reads with average lengths of 100–200 bp from only three runs of a Genome Sequencer 20 and FLX Systems (454 Life Sciences Corporation), generating a preliminary assembly with approximately 29 000 contigs. As such large numbers of sequences are made public it will certainly drive the publication of all the existing private EST databases and a radical jump in the number of available sequences will happen, consolidating a very valuable source of multi-species sequence information for eucalypt genomic research.

Analysis of expression QTLs

The main focus of gene expression studies in trees until now has been the elucidation of the metabolic pathways that determine wood formation. Gene expression has been analysed in different stages of the lignification process, from meristematic cells all the way to maturation and programmed cell death. It has been seen that genes that code for enzymes involved in lignin and cellulose biosynthesis, as well as a number of transcription factors and other genes that regulate wood formation, operate in a very well defined, rigorous stage-specific way. Most studies have been carried out in poplar using microarrays that represent the full transcribed genetic complement of the species. This approach has revealed genes that are over- or under-expressed in specific moments of wood development. However, proving the cause–effect relationship between such genes and the phenotypic variation observed is a much more complex task that requires additional experiments. Microarrays are therefore seen today as a very effective but only exploratory way to identify key genes to understand and manipulate wood formation.

A more functional and forward genomics approach to the study of gene expression has been the integration between genetics and genomics based on the analysis of gene expression in parallel to an underlying Mendelian framework of genetic mapping and QTL discovery. This approach allows the colocalisation between: (1) expression QTLs, i.e. QTLs identified that explain observed differences in expression levels of specific genes on the array; (2) QTLs identified for wood quality traits; and (3) the actual position of the gene on the genetic map. This approach was pioneered in trees in an *E. grandis* × *E. globulus* backcross pedigree by monitoring the expression of 2 700 genes putatively involved in cell wall formation, lignin and cellulose metabolism, cell growth and protein targeting. The key role of some lignin biosynthesis genes was confirmed and some other new unexpected genes with major effects were discovered highly correlated with volume growth as well (Kirst *et al*., 2004). However, in a subsequent study, expression data also showed that the lack of conservation of the genetic architecture of transcript abundance regulation in different genetic backgrounds indicates that many different loci could be involved in modulation of transcription of these genes, and that there is a complex and variable network of gene expression control (Kirst *et al*., 2005a).

Recently, in the framework of the GENOLYPTUS project we have successfully tested a transcriptome-wide oligoarray of 398 000 60-mer probes representing all the c. 21 000 unique genes discovered in the sequencing work. The pilot experiment showed that the oligoarray platform selected (Nimblegen Systems) is extremely robust providing 100% consistency in signal across probe replicates as well as between biological replicates (i.e. different trees of the same clone). Gene expression in differentiating xylem was compared intra- and inter-specifically between *E. grandis* and *E. globulus*. The preliminary analyses showed that the number of genes differentially expressed both between individuals within species and between species is large with some interesting genes emerging that have prompted new ongoing experiments. These results point to a greater complexity of the genetic control of wood formation at least when measured at the adult stage. Further experiments are now planned using this first version of the GENOLYPTUS array where an expression-QTL mapping analysis in segregating populations will be carried out as a strategy to identify interesting candidate genes for association mapping experiments.

Marker-assisted selection

Twenty years have passed since the first demonstrations that QTLs for major effects could be mapped with DNA markers in plants (e.g. Paterson *et al*., 1988) and several reviews have described the potential benefits and caveats of MAS in the plant genetics literature (e.g. Tanksley, 1993; Dekkers and Hospital, 2002). However, large-scale operational MAS is still largely restricted to

very few crops and for very specific applications. Maize is probably the best example, where the financial returns on hybrid seed development, coupled to the ability to fully control germplasm, has prompted large-scale investments in MAS by the private sector based on high-throughput SNP genotyping platforms and whole-genome selection strategies (Meuwissen *et al*., 2001). Based on a detailed understanding of the molecular architecture of quantitative traits, current applications include yield-oriented advanced backcross QTL systems, whole-genome selection strategies (Meuwissen *et al*., 2001) as well as accelerated line conversion following trait introgression by marker-assisted backcrossing. In *Eucalyptus* and forest tree breeding in general, the application of molecular markers for directional selection is still an unfulfilled promise. This is largely due to: (1) the recent domestication of tree crops and hence the wide genetic heterogeneity and LE of breeding populations; (2) the inability to develop inbred lines at least on a short-term basis to allow a more precise understanding of genetic architecture of quantitative traits; (3) the absence of simply inherited traits that could be immediately and more easily targeted; and finally (4) to the very limited number of scientists actually working on forest trees genomics.

Eucalyptus breeding programs vary broadly according to several aspects including the target traits and environments, the target species or hybrid, the possibility of deploying clones and the amount of resources available to the breeder. However, from the standpoint of integrating MAS, a reasonable premise is that this will only be a justifiable option when the breeding program has already reached a relatively high level of sophistication, fully exploiting all the accessible breeding and propagation tools. Advanced breeding programs that aim at elite clone selection involve a significant amount of time and effort being devoted to clonal testing before effective recommendations can be made concerning new clones for operational plantations. Progeny trials, together with expanded single-family plots where larger numbers of full-sibs per family are deployed, are used to allow intensive within-family selection based on all the available information. This selection is generally carried out at half-rotation age based on growth performance and on a preliminary assessment of wood specific gravity using indirect non-destructive techniques. Vegetative propagules are then rescued from selected trees either by coppicing, sequential grafting or *in vitro* techniques, multiplied and then used for the establishment of clonal tests.

This breeding scheme generates large amounts of linkage disequilibrium (LD) by hybridisation and substantial amounts of non-additive genetic variation can be captured by vegetative propagation, favourable conditions for MAS in forest trees (Strauss *et al*., 1992). Favourable alleles at QTLs segregating within families could be efficiently tagged with microsatellite markers in LE with the actual functional polymorphisms and used for marker-assisted within-family selection for superior individuals (O'Malley *et al*., 1994). QTL-linked markers could be used to carry out early selection thus reducing the time necessary to carry out the first selection, especially for traits related to wood properties, and at the same time reducing the number of trees to be selected, propagated and advanced all the way to clonal trials. This MAS scheme is currently being tested in the GENOLYPTUS project. Therefore in the context of molecular breeding, given their relatively short rotations and the possibility of deploying clones to capture non-additive genetic variation, it is reasonable to state that eucalypt is a forest tree crop for which MAS has good prospects of application. The cost of scoring molecular markers dictates that the most likely application of MAS in *Eucalyptus* will be for traits that provide significant added value to the final product. Within all possible quality traits, the option would be for those that display medium to high heritabilities but where phenotype assessment is difficult, expensive or requires waiting until the tree reaches maturity. Wood quality traits typically require the tree to start accumulating late 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 developments of fast sampling and indirect wood chemistry measurements based on NIR spectroscopy (e.g. Schimleck *et al*., 1996), the potential gain will only be realised on the basis of the time savings provided by very early selection.

Association mapping

With the rapid advancement of genome projects generating a large amount of sequence information and SNP (one-letter variations in the DNA sequence that contribute to differences among individuals) data, plant genomics has experienced a growing interest in an alternative approach for the identification of genes underlying quantitative traits. The new model is based on the possibility of investigating sequence variation directly into genes and not at linked markers. This approach exploits candidate gene sequence variation and relies on the existence of LD (non-random association between alleles at linked loci) between detectable sequence SNP and quantitative trait nucleotide (QTN) polymorphisms that ultimately determine the patterns of phenotypic variation (Neale and Savolainen 2004).

In considering MAS for forest trees, more will likely be learned from experiences in livestock (Dekkers, 2004) than from annual crop plants, with the added advantage in trees, however, that gains can be quickly realised by large-scale cloning of selected individuals. In this context, the categorisation of three different levels of marker-trait associations described by Dekkers (2004) are relevant to trees: (1) direct markers, i.e. loci that code for the functional mutation; (2) LD markers, i.e. loci that are in population-wide linkage disequilibrium with the functional mutation; and (3) LE markers, i.e. loci that are in population-wide LE with the functional mutation in outbred populations.

In forest trees, besides the recent encouraging discovery of an LD marker for microfibril angle in *Eucalyptus* (Thumma *et al*., 2005), only LE marker–trait associations have been described. LE markers have been readily detected on a genome-wide basis by analysing large full-sib families with sparse marker maps allowing the detection of most QTL of moderate to large effects as discussed above. For the

other two types of marker-trait association, in pines only now the first association mapping experiments are being started to uncover LD markers, i.e. polymorphisms that are sufficiently close to the functional mutation (Neale and Savolainen, 2004; González-Martínez *et al*., 2007). The challenge is considerable, however, as LD in outcrossing forest trees such as pines decays very rapidly, in general within 1 000 bp (Neale and Savolainen, 2004), and similar behaviour has been seen in the few *Eucalyptus* genes analysed to date with significant LD extending for only a few hundred base-pairs (Kirst *et al*., 2005b; Thumma *et al*., 2005; Faria *et al*., 2006). Genome-wide association studies for LD marker-trait discovery in trees will require very high SNP marker densities that are currently still impracticable, so that the only alternative left is a candidate gene approach. Finally, direct markers (i.e. polymorphisms that code for the functional mutations) would be the most valuable and directly applicable in breeding. However, they are the most difficult to detect because causality is very difficult to prove unless very high penetrance Mendelian inheritances are tackled.

From the operational point of view, the candidate gene approach has the advantage that once a major effect gene is determined and validated, MAS could then be practised directly on the gene and therefore would not rely on the need for strong association (LD) between the marker allele and the favourable allele at the gene of interest. The challenge, however, is the correct selection of candidate genes. This is not an easy task and every effort should be made to maximise the probability of choosing the proper genes. The choice of candidate genes is an elusive target for the majority of phenotypes relevant to forest trees. It requires knowledge of biochemistry, physiology and development that is generally not available even for well-defined phenotypes and/or known metabolic pathways. Testing the role of a candidate gene can be carried out by a conventional cosegregation analysis in structured segregating populations where the gene is used as a marker in the attempt to relate the sequence polymorphism in the gene with variation in the quantitative trait. Allelic variation at the gene is defined by haplotypes, comprising a number of SNPs. The majority of SNPs have no effect, but some cause subtle differences in the final effect of the gene and hence the phenotype. Significant differences in phenotypic means among candidate gene haplotype classes should identify candidate gene alleles with the greatest effect on the trait of interest. Another approach to test and validate candidate genes is to look for SNP–phenotype associations in germplasm collections or natural populations involving contrasting phenotypes. The objective, again, is to correlate the distribution of candidate gene genotypes in the form of DNA sequences and relevant phenotypes.

Breeding by transgenic technology

In the context of molecular breeding, transgenic technology is undoubtedly a very powerful complementary tool available to the breeder. Considering that industrial *Eucalyptus* forests are almost exclusively clonal, transgenics will most likely have an increasing role not only in wood quality improvement but mainly for resolving problems related to pest and pathogen susceptibility 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 impossible to be attained by the natural recombination processes (e.g. frost tolerance) might radically modify how and where eucalypt forests are planted or that forest products are derived. However, some strategic issues in the adoption of the technology for wood quality manipulation have been raised, including: (1) what is the relative magnitude of the attainable gain and cost/benefit relationship by manipulating lignification or cellulose genes when compared to the directed exploitation of the genetic variation in *Eucalyptus* by hybridisation and intensive selection? (2) What are the specific biosafety and intellectual property issues relevant to transgenic eucalypts, and the time and investment necessary to solve them, to actually be able to plant transgenic trees on a large scale? (3) What is the speed by which breeding programs generate new and better clones for several adaptability traits (growth, pest resistance, clonability etc.) as compared to the time needed to do the biosafety job for every new transgenic clone? (4) What is the lifespan of a patent in the local regulation as compared to the time needed to effectively make returns on the patent from the planted forest before the patent goes into public domain? (5) What are the market issues that the company has to consider in adopting transgenics both in relation to public perception and forest certification processes? All these and other issues will have to be carefully considered without overlooking that, just as occurred in annual crops such as soybean, maize and cotton, the use of transgenics could become a major technology divide and represent the necessary condition for a forest-based industry to continue to be competitive in the world scenario.

Conclusions and perspectives

The successful application of molecular breeding in *Eucalyptus* will depend heavily on first validating the association between a DNA polymorphism and a quantitatively inherited phenotypic trait. In highly heterogeneous eucalypts, only a more direct linkage disequilibrium mapping approach can uncover population-wide applicable marker–trait associations. Following the successful path taken in human genetics, the forward genomics approach based on colocalisation of candidate genes and QTLs for relevant traits together with integrative expression-QTL mapping could be a powerful strategy to find such associations. At the moment, there are two possibilities for circumventing the dilemma of choosing candidate genes correctly. The first is microarray-based genotyping with ultra-dense arrays of short (25 nt) oligonucleotides coupled to methods to reduce genome complexity of DNA samples to be hybridised either by cDNA synthesis or methyl filtration (e.g. West *et al*., 2006) that would allow sufficient throughput for association genetic analysis of thousands of genes at a time. Such an array format could later turn out to be a useful instrument for MAS once validated marker–trait associations have been established. The second would be to have access to a whole genome sequence so that candidate genes in a fine mapping interval delimited by markers flanking a QTL with

centimorgan resolution could be mined, reannotated and then analysed in association mapping experiments.

A major advance in this respect was very recently accomplished with the announcement, on 8 June 2007, that the proposal to sequence the *Eucalyptus grandis* genome, submitted by the International *Eucalyptus* Genome Network (EUCAGEN) (http://www.ieugc.up.ac.za; Myburg, 2004), was selected as a target eukaryotic genome for the 2008 fiscal year, among 120 proposals presented to the 2007 Community Sequencing Program of the Joint Genome Institute of the US Department of Energy (http://www.jgi.doe.gov/News/news_6_8_07.html). As published in the press release, 'The biomass production and carbon sequestration capacities of eucalyptus trees match DOE's and the nation's interests in alternative energy production and global carbon cycling. The consortium of eucalyptus draws upon the expertise from dozens of institutions and hundreds of researchers worldwide.'

This public collaborative effort will contribute greatly to the advancement of *Eucalyptus* genetics, genomics and molecular breeding by bringing together existing private databases and genomic resources and thereby expanding the value of such genome sequences. Key contributions will come from a number of countries, including the USA with a large EST collection to be donated by Arborgen, Inc. (M Hinchee pers. comm.), and the GENOLYPTUS network in Brazil will make available its collection of ESTs, a high-density microsatellite map and will collaborate with the Arizona Genomics Institute in the construction of a physical map for the sequenced genome. Furthermore, prospects exist that a low-coverage draft genome of *E. camaldulensis*, currently being sequenced at the Kazusa DNA Research Institute in Japan (T Hibino pers. comm.) will also be made public through the EUCAGEN initiative.

As this genome project advances and new and more powerful analytical tools become accessible, the true challenge to dissecting the complexity of economically and adaptation important traits in *Eucalyptus*, and identifying key genes to be manipulated by molecular breeding or transgenic technologies, will depend to a large extent on our ability to phenotype trees accurately, analyse the overwhelming amount of genomic data available and translate this into relevant information for breeding. The actual use of genomic information in molecular breeding should be considered on a case-by-case basis. Expectations should not be overemphasised until experimental data of realised gains are validated into an industrial forest setting beyond those attained with comparable investments in conventional breeding by exploiting the extraordinary genetic variation that exists in the genus *Eucalyptus*.

Acknowledgements — I am very thankful to the Brazilian Ministry of Science and Technology (MCT) for the continued competitive grant support for research through FINEP (PADCT, Fundo Verde Amarelo); the Brazilian National Research Council-CNPq for the research fellowship, the participating institutions both public and private, in the GENOLYPTUS project; and especially all the undergraduate and graduate students, research collaborators and breeders for the continued discussions and scientific inputs.

References

Brondani RPV, Williams ER, Brondani C and Grattapaglia D (2006) A microsatellite-based consensus linkage map for *Eucalyptus* species and a novel set of 230 microsatellite markers for the genus. *BMC Plant Biology* 6: 20

de Assis TF (2000) Production and use of *Eucalyptus* hybrids for industrial purposes. In: Dungey HS, Dieters MJ and Nikles DG (eds) *Hybrid breeding and genetics of forest trees: proceedings of the QFRI/CRC-SPF symposium, 9–14 April 2000, Noosa, Queensland, Australia*. Department of Primary Industries, Brisbane. pp 63–74

Dekkers JC (2004) Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *Journal of Animal Science* 82: 313–328

Dekkers JC and Hospital F (2002) The use of molecular genetics in the improvement of agricultural populations. *Nature Reviews Genetics* 3: 22–32

Faria DA, Alves TPM, Pereira RW and Grattapaglia D (2006) Frequência de SNPs e extensão do desequilíbrio de ligação ao longo dos genes CCR e CAD em *E. grandis*, *E. globulus* e *E. urophylla*. Abstracts of the 52nd Brazilian Genetics Congress, Foz do Iguaçu, Brazil, 3–6 September 2006, abstract GP251

González-Martínez SC, Wheeler NC, Ersoz E, Nelson CD and Neale DB (2007) Association genetics in *Pinus taeda* L. I. Wood property traits. *Genetics* 175: 399–409

Grattapaglia D (2007) Marker assisted selection in *Eucalyptus*. In: Guimarães E (ed) *Marker-assisted selection (MAS) in crops, livestock, forestry and fish: current status and the way forward*. Food and Agricultural Organization of the United Nations, Rome. pp 251–282

Kirst M, Myburg AA, De Leon JP, Kirst ME, Scott J and 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

Kirst M, Basten CJ, Myburg AA, Zeng ZB and Sederoff RR (2005a) Genetic architecture of transcript-level variation in differentiating xylem of a *Eucalyptus* hybrid. *Genetics* 169: 2295–2303

Kirst M, Marques CM and Sederoff R (2005b) Nucleotide diversity and linkage disequilibrium in three *Eucalyptus globulus* genes. IUFRO Tree Biotechnology 2005, 7–11 November 2005, Pretoria, South Africa, abstract 5.28. Available at http://www.eyevisual.co.za/biotree/viewAbstract.asp [accessed 29 July 2008]

Meuwissen TH, Hayes BJ and Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819–1829

Missiaggia AA, Mamani EM, Novaes E, Pappas MCR, Padua JG, Brondani RPV, Coelho ASG and Grattapaglia D (2005) Microsatellite based QTL mapping and validation across multiple pedigrees of *Eucalyptus*. IUFRO Tree Biotechnology 2005, 7–11 November 2005, Pretoria, South Africa, abstract s5.34. Available at http://www.eyevisual.co.za/biotree/viewAbstract.asp [accessed 29 July 2008]

Myburg AA (2004) The International *Eucalyptus* Genome Consortium (IEuGC): opportunities and resources for collaborative genome research in *Eucalyptus*. In: Li B and McKeand S (eds) *Forest genetics and tree breeding in the age of genomics: progress and future. Proceedings of the IUFRO Joint Conference of Division 2, 1–5 November 2004, Charleston, South Carolina, USA*. pp 154–155

Myburg AA, Potts BM, Marques C, Kirst M, Gion JM, Grattapaglia D and Grima-Pettenati J (2007) *Eucalyptus*. In:

Kole C (ed) *Genome mapping and molecular breeding in plants, vol. 7: Forest trees*. Springer, Heidelberg. pp 115–160

Neale DB and Savolainen O (2004) Association genetics of complex traits in conifers. *Trends in Plant Science* 9: 325–330

Novaes E, Drost D, Farmerie B and Kirst M (2007) Rapid, high-throughput gene discovery in *Eucalyptus* by massive parallel pyrosequencing. IUFRO Tree Biotechnology 2007, 3–8 June 2007, Azores, Portugal, abstract SIII. Available at http://www.itqb.unl.pt/iufro2007/SciProg.html [accessed 29 July 2008]

O'Malley D, Sederoff R and Grattapaglia D (1994) Methods for within family selection in woody perennials using genetic markers. United States Patent and Trademark Office. Patent no. 6 054 634 (www.uspto.gov)

Pasquali G, Bastolla FM, Kirch RP, Brondani RPV, Coelho ASG, Grattapaglia G, Brommonschenkel SH, Pappas GJ Jr, Pereira GAG and Cascardo JCM (2005) Sequencing of the *Eucalyptus* transcriptome in the Genolyptus project. IUFRO Tree Biotechnology 2005, 7–11 November 2005, Pretoria, South Africa, abstract S1.21. Available at http://www.eyevisual.co.za/biotree/viewAbstract.asp [accessed 29 July 2008]

Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE and Tanksley SD (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* 335: 721–726

Poke F S, Vaillancourt RE, Potts BM and Reid JB (2005) Genomic research in *Eucalyptus*. *Genetica* 125: 79–101

Schimleck LR, Michell AJ and Vinden P (1996) Eucalypt wood classification by NIR spectroscopy and principal components analysis. *Appita Journal* 49: 319–324

Strauss SH, Lande R and Namkoong G (1992) Obstacles to molecular-marker-aided selection in forest trees. *Canadian Journal of Forest Research* 22: 1050–1061

Tanksley SD (1993) Mapping polygenes. *Annual Review of Genetics* 27: 205–233

Thumma BR, Nolan MF, Evans R and Moran GF (2005) Polymorphisms in cinnamoyl coa reductase (CCR) are associated with variation in microfibril angle in *Eucalyptus* spp. *Genetics* 171: 1257–1265

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A *et al.* (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313: 1596–1604

West MA, van Leeuwen H, Kozik A, Kliebenstein DJ, Doerge RW, St Clair DA and Michelmore RW (2006) High-density haplotyping with microarray-based expression and single feature polymorphism markers in *Arabidopsis*. *Genome Research* 16: 787–795