FORWARD GENOMICS IN *EUCALYPTUS*: FROM PHENOTYPES TO GENES INVOLVED IN WOOD FORMATION

Dario Grattapaglia[1,2]

[1]Research Scientist – EMBRAPA Genetic Resources and Biotechnology, Estação Parque Biológico Final W5 Norte – 70770-970 Brasília DF and [2]Professor, Genomic Sciences and Biotechnology Program, Catholic University of Brasília, SGAN 916 Módulo B Brasília, DF.

ABSTRACT

While the reverse genomics approach to wood quality manipulation aims to determine or alter the phenotype by mutating a given candidate gene through transgenic technology, the forward genomics approach, i.e. going from the analysis of existing phenotypic variation for wood properties to the discovery of the causal genetic variants, is based on exploiting the natural genetic variation that exists in the target species. The genus *Eucalyptus* is outstanding in this respect as it displays a very wide and promptly exploitable intra and interspecific phenoytpic variation for several wood properties. The advent of high throughput genomic technologies in the last ten years has opened new perspectives about the speed, scale and detail with which one can investigate the underlying genetic nature of such variation. The key technologies involved in the forward genomics approach include genetic mapping, QTL (Quantitative Trait Loci) discovery, physical mapping and whole genome sequencing. A number of genes involved in lignin composition have been intensively investigated and manipulated in recent years by transgenic technology demonstrating the great potential that exists with this approach. Recently genes involved in cellulose and hemicelluloses metabolism have been tackled as well. However high impact, industrial level applications of transgenics in eucalypt production forestry are yet to come at the same time that biosafety, regulatory and certification challenges persist. DNA markers have also been successfully used to understand the genetic architecture of complex wood properties traits and operational applications are in routine use to select better trees based on their genotype composition. However the complexity of most industrially relevant traits still represents a technical challenge faced by molecular breeders. The successful application of marker assisted selection in *Eucalyptus* and trees in general has depended heavily on our ability to demonstrate and validate the clear-cut association between a DNA polymorphism and a quantitatively inherited phenotypic trait. In highly heterogeneous eucalypts, while conventional QTL mapping has revealed useful markers that are currently exploited in within-family selection programs, only a more direct linkage disequilibrium mapping approach can uncover population wide applicable gene-trait associations. Selection of candidate genes for direct manipulation by transgenesis or association studies based on its presumed biochemical role is not an easy task even for well defined phenotypes and/or known metabolic pathways. Taking from human genetics, 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 wood quality traits. The contemplation of marker assisted selection or transgenic technology should be done on a case-by-case basis. Expectations should not be overstated until experimental data of realized gains are validated into an industrial forest setting beyond those attained with comparable investments by exploiting the extraordinary genetic variation that exists in the genus *Eucalyptus*.

Keywords: Molecular breeding, QTLs, genomics, *Eucalyptus*, forward genetics, transgenic tree

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 tropics and subtropics on either side of the equator 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 "fiber 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, the theme of this brief presentation. Fifteen years have passed since the first experiments in genetic mapping and molecular breeding of forest trees. Since the outset, 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 knowledge gathered prompted some short term opportunities for the incorporation of genomic analysis in 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 interspecific variation that exists in *Eucalyptus*. The key technologies involved in this approach include genetic mapping, QTL (Quantitative Trait Loci) discovery, physical mapping and genome sequencing (Figure 1). In this presentation, I will 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 molecular breeding in *Eucalyptus*. More in-depth reviews on some of these topics have been or will soon be published (Grattapaglia, 2004, 2007; Kirst et al., 2004a; Myburg et al., 2007; Poke et al., 2005; Shepherd and Jones, 2004).

GENETIC RESOURCES FOR *EUCALYPTUS* GENOMICS

While public genomic resources including a complete genome sequence should become more easily available in the very near future, biological resources and precise phenotyping represent the real limitation of many genomic projects in many instances. 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, *Eucalyptus 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 supply 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 now 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 to produce the same ton of cellulose. While only 2.98 cubic meters of *E. globulus* wood are required per ton of pulp, 3.89 cubic meters are needed from *E. grandis*. *E. globulus* displays a very adequate wood density in the range of 550 kg/m3, the longest fibre length and the largest content of holocellulose and pentosans than any other intensively planted *Eucalyptus* species (Sanchez, 2002). 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 hybridization 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 (Teotônio de Assis pers. comm.). 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 realized in the mill. Several trials have been established in recent years and elite hybrid clones of *E. globulus* with *E. grandis* and *E. urophylla* with 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 interspecific families involving different *Eucalyptus* species were generated and are currently being used for QTL mapping, gene expression and proteomics work. By establishing a rich resource of genetic variation resulting from hybridization we hope to contribute to uncovering the genetic causes that make the *E. globulus* wood so different from the wood of *E. grandis*.

DISCOVERY AND VALIDATION OF QTLs FOR WOOD PROPERTIES TRAITS

Molecular marker maps have been successfully used to detect major effect QTLs (Quantitative Trait Loci) 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, fiber length and fiber 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, QTL validation efforts have started to effectively move from phenotypes to genomic regions controlling traits of interest for marker assisted selection. In the GENOLYPTUS project we have consolidated the construction of multiple genetic maps and the detection of QTLs for several wood properties. We have shown the possibility of performing comparative QTL position analysis between independent experiments. Genetic

maps were constructed for three independent genetically unrelated families. The first involved a cross between two elite Rio Claro natural hybrid trees involving predominantly *E. grandis* and *E. urophylla*. The other two maps were derived form crosses between pure *E. grandis* and *E. urophylla* select trees and the F1 progeny was cloned and planted in replicated trials in five environments throughout Brazil. Map construction used as reference the integrated map involving 234 markers developed by Brondani et al. (2006) and added new markers developed from genomic shotgun as well as EST sequences. QTLs detection was accomplished under different models. In the cross involving hybrid parents, 10 QTLs were detected for parent clone 235, with LODs varying between 2.9 (lignin content) to 4.2 (specific wood consumption). For hybrid parent 221, five QTLs were detected with LODs varying from 2.9 (cellulose yield) to 4.8 (basic wood density). Comparative QTL mapping across the three 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 fiber 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 higher density of microsatellites for QTL validation, directed search for allelic variants at QTLs in multiple pedigrees thus allowing precise determination of target genomic regions for gene discovery, association mapping studies, high resolution mapping and marker assisted selection (Missiaggia et al. 2005).

GENE DISCOVERY

Up until the announcement of the completion of the genome sequence of Poplar (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 (EST's). EST sequencing quickly generates a large index of partial genes for the organism of interest making them available in organized collections of clusterized sequences for further molecular investigation. Partial gene sequences generated have had mutiple 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 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 started about ten years ago when some tens of thousand sequences were generated. With advancements in sequencing technology quickly the size of the databases became on the order of hundreds of thousand of sequences and currently with recent breakthroughs in sequencing technologies databases of over 300.000 short sequences can be generated in a matter of hours.

A number of mostly private ESTs databases for *Eucalyptus* have been built and have represented important sources for genomic experiments. In the GENOLYPTUS project we have completed an initial database of over 120,000 EST sequences derived from twenty 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. 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). With this rich genomic resource in hand, experiments are now being carried out with different approached to test whether phenotypic differences in wood quality traits between *E. grandis* and *E. globulus* for example, could be attributed to specific sequence differences in coding regions of candidate genes. However we should not overlook the fact that sequence differences in regulatory regions of such genes could also play an important role in the definition of phenotypes. The search for those sequence polymorphisms is certainly a more challenging task that will depend, among other factors, in having an adequate genomic resource in the form of a physical map.

PHYSICAL MAPPING

While genetic mapping of *Eucalyptus* has evolved quite rapidly, efforts have been timid in generating physical mapping resources for species of the genus. A complete physical map for *Eucalyptus* will certainly represent a great experimental resource for years to come as it provides a physical organized equivalent of the genome to access genes and regulatory regions for multiple applications both in transgenics and molecular breeding. To this end we have constructed a *Eucalyptus grandis* BAC (Bacterial Artifical Chromosome) library with an initial genome coverage of ~4X with over 70% of the inserts >150 kb long. Using this BAC library we have been isolating and shotgun sequencing a number of candidate genes involved in wood chemical composition. This strategy has allowed us to land on BACs for important genes that code enzymes involved in the lignin or cellulose biosynthetic pathway as well as some transcription factors. For two key genes, 4-CL and CAD, the smaller single BAC (30Kbp for CAD and 120 kbp for 4-CL) as evaluated by PFGE was selected for constructing shotgun libraries with average insert size of 2 kbp. Assembly of 1,052 reads (3.5X) for 4-CL resulted in a 5,477 bp contig covering the whole gene but part of the 5-UTR. For CAD 768 reads (10X) allowed the assembly of a 9,785 bp contig of what

was found to be CAD2 (Brommonschenkel et al. 2005). With the full genomic sequence in hands it is now possible to identify regulatory regions and carry out a detailed analysis of polymorphism in a set of individuals by resequencing specific upstream regions in an association mapping approach. It is in our immediate plan to build a complete physical map for *E. grandis*, by fluorescent fingerprinting as a contribution to an international *Eucalyptus* Genome Network recently established. This physical genomic resource from *E. grandis* will facilitate the identification of specific genomic regions in *E. globulus*. It should be faster, for example, to clone the full homolog gene from *E. globulus* and thus compare in detail potential regulatory regions responsible for differential patterns of gene expression and resulting phenotypic variation.

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 analyzed 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 biossynthesis 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 co-localization 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 applied in *Eucalyptus* 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 effect were discovered highly correlated with volume growth as well (Kirst et al., 2004b). 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 probes representing all the 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 interspecifically between *E. grandis* and *E. globulus*. The preliminary analyses showed that the number of genes differentially expressed both within and between species is unexpectedly small but with some interesting genes emerging. 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 (MAS)

Eucalyptus breeding programs vary broadly according to several aspects including 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 by hybridization and substantial amounts of non-additive genetic variation can be captured by vegetative propagation, favorable conditions for MAS in forest trees. Favorable alleles at QTL segregating within-families could be efficiently tagged with microsatellite markers in linkage equilibrium with the actual functional polymorphisms and used for marker assisted within-family selection for superior individuals. 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. Recently this general MAS scheme has been adopted by some companies. 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 which provide significant added value to the final product such as branching habit (for solid wood), wood chemical traits, or allow clonal deployment such as adventitious rooting or somatic embryogenesis response. 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.

ASSOCIATION MAPPING

With the rapid advancement of genome projects generating a large amount of sequence information and single nucleotide polymorphism (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 linkage disequilibrium (LD) (non random association between alleles at linked loci) between detectable sequence polymorphisms SNPs and QTN (quantitative trait nucleotide) that ultimately determine the patterns of phenotypic variation.

Some laboratories have started association mapping work in *Eucalyptus* (Thumma et al. 2005) for wood traits by sampling trees in the wild or from breeding programs that display contrasting phenotypes for wood quality traits. The challenge however is considerable, as LD in *Eucalyptus* decays very rapidly, in general within 100 to 1000 bp as seen in the few *Eucalyptus* genes analyzed to date (Thumma et al. 2005; Kirst et al. 2005b; 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 at this moment the only alternative left is a candidate gene approach..

From the operational point of view, the candidate gene approach has the advantage that once a major effect gene is determined and validated, marker assisted selection could then be practiced directly on the gene and therefore would not rely on the need for strong association (linkage disequilibrium) between the marker allele and the favorable 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 maximize 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 co-segregation 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 pathogens 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, might radically modify the ways that 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 hybridization 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 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 co-localization 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 (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 analyzed in association mapping experiments. A draft genome of *E. camaldulensis* is currently being sequenced in Japan, and the possibility exists that a fully public 4X draft of the *E. grandis* genome will be sequenced by JGI (Joint Genome Institute) of the U.S. Department of Energy within the next years, following a proposal we recently submitted as an international group of *Eucalyptus* geneticists who formed the *Eucalyptus* Genome Network (EUCAGEN) (www.ieugc.up.ac.za). Such public collaborative efforts should 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. As a full draft sequence of the *Eucalyptus* genome becomes available, the true challenge to dissecting the complexity of economically-important traits in *Eucalyptus* will depend to a large extent on our ability to phenotype trees accurately, analyze the overwhelming amount of genomic data available and translate this into truly useful molecular tools for breeding. Marker-assisted selection or transgenics should be considered on a case-by-case basis, without overstating the gains to be expected until hard experimental data are accumulated on the actual gains made from its application in an industrial forest setting beyond those which can be attained by comparable investment in conventional breeding.

ACKNOWLEDGMENT

I am very thankful to the Brazilian Ministry of Science and Technology (MCT) for the continued competitive grant support for research through FINEP; 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

1.Grattapaglia, D. Integrating genomics into *Eucalyptus* breeding. **Genetics and Molecular Research 3**: 369-379 (2004).

2.Grattapaglia, D. Marker assisted selection in *Eucalyptus*. In: E. Guimarães (ed.) **Marker-assisted selection (MAS) in crops, livestock, forestry and fish: Current status and the way forward**. FAO - Food and Agricultural Organization, Rome, in press (2007).

3.Kirst, M.; Myburg, A.; Sederoff, R. Genetic mapping in forest trees: markers, linkage analysis and genomics. **Genet Eng (NY) 26**: 105-141 (2004a)

4.Myburg, A. A.; Potts, B. M.; Marques, C.; Kirst, M.; Gion, J. M.; Grattapaglia, D.; Grima-Pettenati,

J. *Eucalyptus*. In: C. Kole, C (ed.) **Genome Mapping & Molecular Breeding in Plants Volume 7: Forest Trees**. Springer, Heidelberg, Berlin, New York, Tokyo, in press (2007).

5.Poke, F. S.; Vaillancourt, R. E.; Potts, B. M.; Reid, J. B. Genomic research in *Eucalyptus*. **Genetica 125**: 79-101 (2005)..

6.Shepherd, M.; Jones, M. Molecular markers in tree improvement: Characterisation and use in *Eucalyptus*. In: **Biotechnology in Agriculture and Forestry, Vol 55: Molecular Marker Systems in Plant Breeding and Crop Improvement** (H. Lörz and G. Wenzel, eds.), (2004) pp. 399-409. Springer-Verlag, Heidelberg.

7.Sanchez, J.R. Panel de tecnologia, pasta y papel. In: **Socioeconomia, patologia, tecnologia y sostenibilidade del eucalipto**. Universidad de Vigo (ed.) pp. 187-192. Ed. Graficas Diumaró (2002).

8.Brondani, R. P. V.; Williams, E. R.; Brondani, C.; Grattapaglia, D.. 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. doi:10.1186/1471-2229-6-20 (2006).

9.Missiaggia, A. A.; Mamani, E. M.; Novaes, E.; Pappas, M. C. R.; Padua, J. G.; Brondani, R. P. V.; Coelho, A. S. G.; Grattapaglia, D. Microsatellite based QTL mapping and validation across multiple pedigrees of *Eucalyptus*. **Abstracts of the IUFRO Tree Biotechnology 2005**, Pretoria, South Africa, November 7-11, (2005) abstract s5.34 http://www.eyevisual.co.za/biotree/viewAbstract.asp.

10.Tuskan, G. A.; Difazio, S.; Jansson, S.; Bohlmann, J.; Grigoriev, I. et al. The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). **Science 313**:1596-1604 (2006)

11.Pasquali, G.; Bastolla, F. M.; Kirch, R. P.; Brondani, R. P. V.; Coelho, A. S. G.; Grattapaglia, G.; Brommonschenkel, S. H.; Pappas, G. J. Jr.; Pereira, G. A. G.; Cascardo J. C. M. Sequencing of the *Eucalyptus* transcriptome in the Genolyptus project. **Abstracts of the IUFRO Tree Biotechnology 2005**, Pretoria, South Africa, November 7-11, (2005) abstract S1.21 http://www.eyevisual.co.za/biotree/viewAbstract.asp.

12.Brommonschenkel, S. H.; Brondani, R. P. V.; Bucelli, R. F.; Lourenço, R. T.; Novaes, E.; Bandeira, L. F.; Pappas, G. J. Jr; Santos, S. N., Póvoa, A.; , Pappas, M. C. R.; Coelho, A. S. G.; Grattapaglia, D. A BAC library of *Eucalyptus grandis*: characterization, fingerprinting, bac-end sequencing and shotgun assembly of lignification genes. **Abstracts of the IUFRO Tree Biotechnology 2005**, Pretoria, South Africa, November 7-11, (2005) abstract S3.08 http://www.eyevisual.co.za/biotree/viewAbstract.asp

13.Kirst, M.; Myburg, A. A.; De Leon, J. P.; Kirst, M. E.; Scott, J.; Sederoff, R.. Coordinated genetic regulation of growth and lignin revealed by quantitative trait locus analysis of cDNA microarray data in an interspecific backcross of *Eucalyptus*. **Plant Physiol 135**: 2368-2378 (2004b).

14.Kirst, M.; Basten, C. J.; Myburg, A. A.; Zeng, Z. B.; Sederoff, R. R.. Genetic architecture of transcript-level variation in differentiating xylem of a *Eucalyptus* hybrid. **Genetics 169**: 2295-2303 (2005a).

15.Thumma, B. R., M. F. Nolan, R. Evans, and G. F. Moran. 2005. Polymorphisms in cinnamoyl coa reductase (ccr) are associated with variation in microfibril angle in *Eucalyptus* spp. **Genetics 171**: 1257-1265.

16.Kirst, M.; Marques, C. M.; Sederoff, R. Nucleotide diversity and linkage disequilibrium in three *Eucalyptus globulus* genes. **Abstracts of the IUFRO Tree Biotechnology 2005**, Pretoria, South Africa, November 7-11, (2005b) abstract 5.28. http://www.eyevisual.co.za/biotree/viewAbstract.asp

17.Faria, D. A.; Alves, T. P. M.; Pereira, R. W.; Grattapaglia, D. 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 52º Brazilian Genetics Congress**, Foz do Iguaçu, abstract GP251 (2006).

18.West, M.A.; van Leeuwen, H.; Kozik, A.; Kliebenstein, D.J.; Doerge, R.W.; St Clair, D.A.; Michelmore, R.W. High-density haplotyping with microarray-based expression and single feature polymorphism markers in arabidopsis. **Genome Research 16**: 787-795 (2006).

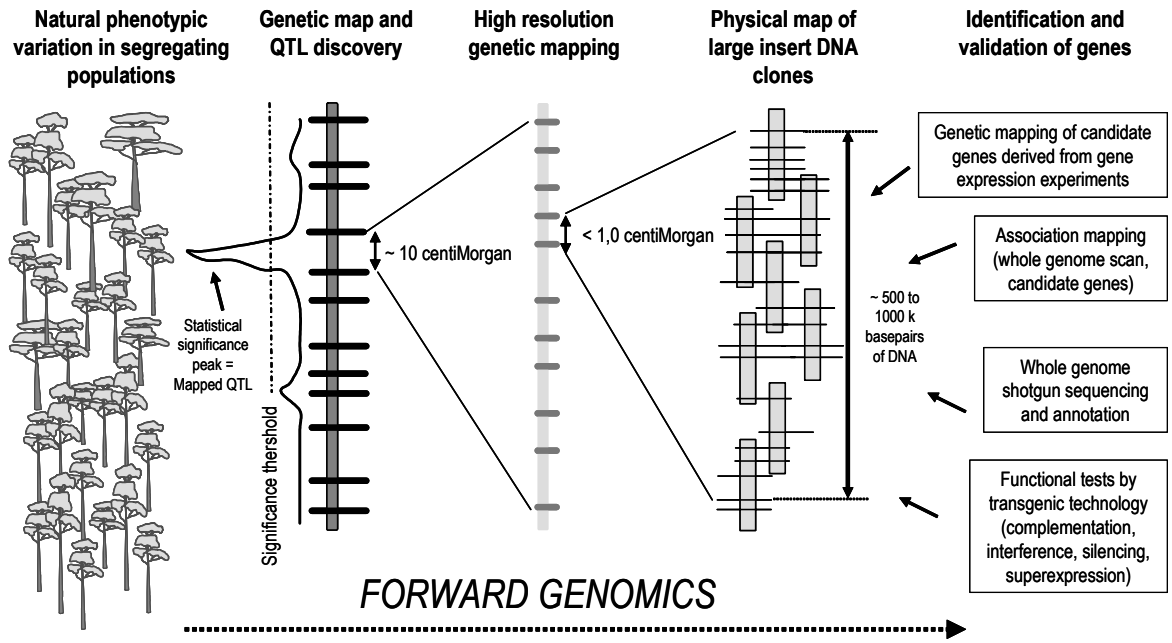

Figure 1. Schematic representation of the forward genomics approach. Genetic mapping and QTL discovery is carried out in segregating populations specifically generated to display large amounts of natural phenotypic variation for wood properties, growth and disease resistance. Major effects QTLs are detected with a resolution of 10 centiMorgans, defining target genomic segments involved in the control of the measured traits. By high resolution mapping, narrower genomic windows of < 1 cM are defined that can be anchored to physical maps of large insert clones such as BAC (Bacterial Artifical Chromosomes) covering ~500 a 1000 kbp of DNA. At this point several reverse genomics approaches (boxes) can be used to identify, test and validate specific genes that underlie the target QTL.