

# Genetic Architecture of Transcript-Level Variation in Differentiating Xylem of a *Eucalyptus* Hybrid

Matias Kirst,<sup>\*,†,1</sup> Christopher J. Basten,<sup>‡</sup> Alexander A. Myburg,<sup>§</sup>  
Zhao-Bang Zeng<sup>‡</sup> and Ronald R. Sederoff<sup>\*</sup>

<sup>\*</sup>Forest Biotechnology Group, North Carolina State University, Raleigh, North Carolina 27695, <sup>†</sup>Functional Genomics and Genetics Graduate Program, North Carolina State University, Raleigh, North Carolina 27695, <sup>‡</sup>Bioinformatics Research Center, North Carolina State University, Raleigh, North Carolina 27695 and <sup>§</sup>Department of Genetics, Forestry and Agricultural Biotechnology Institute, University of Pretoria, Pretoria 0002, South Africa

Manuscript received December 2, 2004  
Accepted for publication January 11, 2005

## ABSTRACT

Species diversity may have evolved by differential regulation of a similar set of genes. To analyze and compare the genetic architecture of transcript regulation in different genetic backgrounds of *Eucalyptus*, microarrays were used to examine variation in mRNA abundance in the differentiating xylem of a *E. grandis* pseudobackcross population [*E. grandis* × F<sub>1</sub> hybrid (*E. grandis* × *E. globulus*)]. Least-squares mean estimates of transcript levels were generated for 2608 genes in 91 interspecific backcross progeny. The quantitative measurements of variation in transcript abundance for specific genes were mapped as expression QTL (eQTL) in two single-tree genetic linkage maps (F<sub>1</sub> hybrid paternal and *E. grandis* maternal). EQTL were identified for 1067 genes in the two maps, of which 811 were located in the F<sub>1</sub> hybrid paternal map, and 451 in the *E. grandis* maternal map. EQTL for 195 genes mapped to both parental maps, the majority of which localized to nonhomologous linkage groups, suggesting *trans*-regulation by different loci in the two genetic backgrounds. For 821 genes, a single eQTL that explained up to 70% of the transcript-level variation was identified. Hotspots with colocalized eQTL were identified in both maps and typically contained genes associated with specific metabolic and regulatory pathways, suggesting coordinated genetic regulation.

COMPARATIVE analysis of genome sequences of yeast, animal, and plant species indicates that the morphological and developmental diversity of eukaryotes arose primarily from the differential regulation of a core set of genes, rather than by extensive acquisition or creation of new genes (KING and WILSON 1975; BALTIMORE 2001; LEVINE and TJIAN 2003). The evolution of gene expression variation may arise from the expansion of gene families regulating transcription differentially or from an increase in complexity of regulatory sequences and protein complexes involved in transcriptional regulation. These mechanisms modify mRNA levels and cell type specificity while maintaining the structure and functionality of proteins. Phenotypic variation at the species level is also created more often by changes in gene expression regulation than through variant forms of proteins. DNA sequence variation is more frequent in regulatory sequences than in coding regions in humans (ROCKMAN and WRAY 2002).

Despite the contribution of gene expression variation to diversity and evolution, variation normally has been studied either at the DNA sequence level or at the phe-

notype level. The evolution of species can now be studied at the transcriptional level using genomic tools that assess genome-wide variation in gene expression (SCHENA *et al.* 1995; VELCULESCU *et al.* 1995). The genetic control and heritability of transcript-level variation have been demonstrated in yeast, mice, and flies (DUMAS *et al.* 2000; KARP *et al.* 2000; BREM *et al.* 2002; WAYNE and MCINTYRE 2002; SCHADT *et al.* 2003; YVERT *et al.* 2003). Transcriptional regulation may also be affected by nongenetic sources of variation (GIBSON 2003), much like a typical quantitative trait. Therefore, transcript levels measured by microarrays may be considered molecular quantitative phenotypes with genetic, environmental, and developmental sources of variation, which contribute to a transcript-level phenotype.

The nature of transcript abundance variation at the individual, population, and species levels has only recently been investigated. OLEKSIK *et al.* (2002) found that 18% of the genes examined showed differences in transcript abundance among 15 individuals from three populations of fish in the genus *Fundulus*. However, interpopulation variation was minor, affecting the transcript levels of only 15 genes. A comparison of transcript variation in humans, chimpanzees, and orangutans identified species-specific patterns of gene transcript levels, particularly in the brain, which may reflect differences

<sup>1</sup>Corresponding author: School of Forest Resources and Conservation, Newins-Ziegler Hall, Room 367, University of Florida, Gainesville, FL 32611.

among these taxonomic groups (ENARD *et al.* 2002). A similar analysis identified transcript abundance patterns that are conserved in humans, flies, worms, and yeast and in others that are specific for different animals (STUART *et al.* 2003).

Microarrays can be used to study the genetic regulation of gene expression on a genome-wide scale by measuring transcript levels in a segregating progeny. Gene-specific transcript levels estimated from microarrays can be individually analyzed as quantitative traits and mapped using established quantitative genetic methods (MACKAY 2001). This strategy has recently been demonstrated in yeast, mice, humans, and maize (BREM *et al.* 2002; SCHADT *et al.* 2003; YVERT *et al.* 2003), providing a genome-wide overview of the genetic architecture of gene transcript regulation and the relative extent of *cis*- and *trans*-acting factors involved in the control of metabolic and regulatory pathways.

The genetic architecture of variation in transcript abundance represents a complex genetic regulatory network. Understanding the number and nature of the interactions involved is one of the major problems of genetics. For instance, how does the genetic architecture of transcript abundance regulation differ among individuals in a species or among taxonomically related species? Are the genomic regions involved in the regulation of orthologous gene transcripts conserved in different genetic backgrounds, or does variation arise from divergence at many different regulatory loci? These questions were addressed here by analyzing the genetics of transcript variation in wood-forming tissues (differentiating xylem) in a wide interspecific cross of an F<sub>1</sub> hybrid of *Eucalyptis grandis* and *E. globulus* ssp. *globulus*, backcrossed to a different *E. grandis* parent (Figure 1).

*E. grandis* and *E. globulus* belong to the same taxonomic subgenus, Symphyomyrtus, but they are members of different sections (Latoangulatae and Maidenaria, respectively) that diverged in the late Miocene-Pliocene (5–10 MYA) (PRYOR and JOHNSON 1971; ELDRIDGE *et al.* 1993; LADIGES *et al.* 2003). *E. grandis* occurs naturally along the subtropical regions of Australia's east coast and *E. globulus* is found in the temperate climates of Tasmania and the extreme south of the Australian continent (ELDRIDGE *et al.* 1993). The two species have contrasting wood properties and growth. *E. globulus* has high wood density and relatively slow growth. *E. grandis* has superior growth and lower wood density. These phenotypic differences may be due to fixed alleles within the species. Eucalyptus hybrids have been created and propagated widely because of the unique properties of specific hybrid genotypes. Crosses between *E. grandis* and *E. globulus* have resulted in progeny populations with wide genetic and phenotypic segregation. The hybrids are particularly suited to describe the genetic architecture of quantitative variation in wood quality and growth traits (MYBURG *et al.* 2003). Therefore, such a hybrid

provides an ideal pedigree to study the genetic architecture of transcript abundance variation.

This report extends a previous study (MYBURG *et al.* 2003) in which two separate single-tree genetic maps were generated; a paternal map described the allelic segregation of the F<sub>1</sub> hybrid (tree BBT01058/H1A1) and a maternal map described the *E. grandis* backcross parent (tree 678.2.1). QTL analysis of progeny from the F<sub>1</sub> hybrid reflects the effects associated with the presence of the *E. globulus* parent (genotype not described) or the *E. grandis* (tree G50) alleles in backcross individuals (Figure 1), revealing species differentiation in the regulation of transcript abundance in wood-forming tissues. QTL reported for the *E. grandis* (678.2.1) backcross parent reflect differences in genetic regulation in the pure species. Synteny between the F<sub>1</sub> hybrid maternal and the *E. grandis* paternal maps allows the comparison of the genetic architecture of transcript abundance variation of specific genes in different genetic backgrounds. The transcript-level variation measured in the 91 segregating progeny was mapped as expression QTL (eQTL) in the two paternal maps to identify the *cis*- or *trans*-acting loci that regulate the transcript-level variation. This approach identified eQTL for 1067 genes and suggests that the genetic control of transcript levels is modulated by variation at different regulatory loci in different genetic backgrounds.

## MATERIALS AND METHODS

**Linkage maps:** Transcript-level QTL were mapped onto two single-tree genetic maps [F<sub>1</sub> hybrid (genotype H1A1) and *E. grandis* (genotype 678.2.1)] generated previously by genotyping 331 individuals from the *E. grandis* backcross population with 803 polymorphic AFLP fragments (MYBURG *et al.* 2003). Framework markers were selected to obtain an average spacing of ~10 cM and comprise 96 AFLP fragments mapped to 12 major linkage groups in the *E. grandis* (678.2.1) maternal map and 122 fragments mapped to 11 linkage groups in the F<sub>1</sub> hybrid (H1A1) paternal map. The two maps were aligned and synteny was established among linkage groups on the basis of dominant intercross markers, *i.e.*, markers inherited as heterozygotes from both the F<sub>1</sub> hybrid (H1A1) and the *E. grandis* (678.2.1) parents, segregating in a 3:1 ratio in the progeny.

**Plant material and tissue collection:** The 91 individuals used in this study, from the *E. grandis* backcross mapping population (Figure 1), were used to construct the F<sub>1</sub> hybrid [*E. grandis* (genotype G50) × *E. globulus* (unknown parent)] paternal map and *E. grandis* (genotype 678.2.1) maternal map (MYBURG *et al.* 2003). Stump sprouts were collected from the *E. grandis* backcross progeny population when the trees were 6 years old and rooted. The rooted cuttings were planted on a single field plot near Paysandu (Uruguay) by Forestal Oriental S.A. When the trees were 20 months old, differentiating xylem, including cells in the stages of cambial cell division, cell expansion, and secondary-wall formation, was collected from the first two meters of the entire stem circumference of one ramet of each clone. Tissue collection occurred at the peak of the growth season and was restricted to 2 consecutive days to minimize environmental variation.

**RNA preparation, labeling, and hybridization:** The tissue was stored in RNAlater solution (Ambion, Austin, TX) until

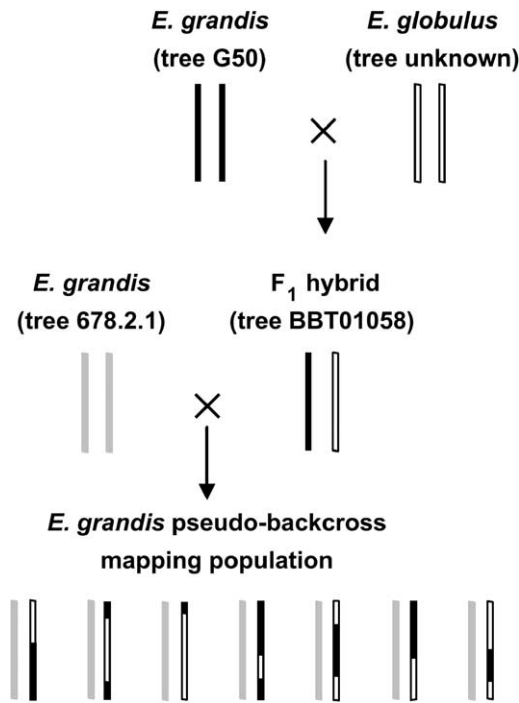


FIGURE 1.—Mating design of the *E. grandis* pseudobackcross mapping population. One homologous chromosome pair is represented for each genotype. Two separate single-tree genetic maps were generated for each backcross parent, *E. globulus* (678.2.1) and the *F*<sub>1</sub> hybrid (BBT01058). *F*<sub>1</sub> hybrid parental alleles can be followed in the *E. grandis* backcross progeny on the basis of testcross markers inherited from either of the two parents of the hybrid [*E. grandis* (solid segments) or *E. globulus* (open segments)].

RNA was extracted (CHANG *et al.* 1993) and purified on RNeasy plant mini kit (QIAGEN, Chatsworth, CA) columns. Labeling was carried out with Cy3 and Cy5 dyes according to the aminoallyl method (HEGDE *et al.* 2000) after first-strand cDNA synthesis with SuperScript II (Life Technologies). After hybridization (20 hr at 42°) and high-stringency washes (HEGDE *et al.* 2000), the slides were scanned with a ScanArray 4000 microarray analysis system scanner (Packard Bioscience). Images were processed using QuantArray software (Packard Bioscience). The raw intensity microarray data were deposited in the Gene Expression Omnibus database under the accession nos. GPL348 (platform), GSM7637–GSM7727 (sample), and GSE502 (series).

**Microarray design:** The cDNA microarray comprised 2608 cDNAs selected from a set of 14,000 ESTs from *E. grandis* (cDNA libraries from differentiating xylem, juvenile and adult leaf, petiole, and root) and from *E. tereticornis* (two cDNA libraries from flower), as described previously (KIRST *et al.* 2004). After PCR amplification and purification in multiscreen 96-well filtration plates (Millipore, Bedford, MA), clones were screened for quality in agarose gels and printed in duplicate on aminosilane-coated glass slides (Corning, Corning, NY) using an Affymetrix 417 spotter. Sequences of the cDNAs used in the microarray were deposited in GenBank (CB967505–CB968059, CD667988–CD670002, CD670004, CD670097, CD670101–CD670112, CD670114–CD670137).

**Experimental design and statistical analysis:** A loop design (CHURCHILL 2002) was used to maximize the number of sampled meioses, while technically replicating each sample (Cy3 and Cy5). The loop design allows for the evaluation of twice

as many genotypes (meiotic events) for a given number of slides as does a reference design. Therefore, while each individual from the segregating population was technically replicated using Cy3 and Cy5 labeling, the loci tested for presence of an eQTL were biologically replicated in each individual carrying the alternative parental alleles for that locus in the population. Raw signal intensity values were transformed ( $\log_2$ ) and normalized using the mixed analysis of variance (ANOVA) model,  $y_{ijk} = \mu + A_i + D_j + P_k + (A \times D)_{ij} + (A \times P)_{ik} + (D \times P)_{jk} + (A \times D \times P)_{ijk} + \epsilon_{ijk}$ , as suggested previously (JIN *et al.* 2001; WOLFINGER *et al.* 2001). This model accounts for systematic (experiment-wide) sources of variation associated with array ( $A_i$ , d.f. = 90, random effect), dye ( $D_j$ , d.f. = 1, fixed effect), and pin effects ( $P_k$ , d.f. = 3, fixed effect), and interactions. Residuals, representing normalized values, were used to estimate individual tree genotype effects for each tree and each cDNA in the mixed ANOVA model  $r_{ilm} = \mu + A_i + N(A)_{li} + T_m + \epsilon_{ilm}$ . In this model,  $T_m$  (d.f. = 90, fixed effect) represents the effect of the individual tree, or genotype, on the transcript level of every gene, from which least-squares means estimates were calculated. Array ( $A_i$ , d.f. = 90, random effect) was included in the model to control for spot effect (JIN *et al.* 2001; WOLFINGER *et al.* 2001) and  $N(A)_{li}$  (d.f. = 91, random effect) accounts for the spot replication within slides. Significant differences in transcript abundance in the segregating population were inferred from the type 3 tests of fixed effects for genotype, carried out on a gene-by-gene basis in the mixed ANOVA model.

**eQTL detection:** Least-squares means estimates of transcript levels, calculated for each individual and cDNA, and the framework marker data of the *F*<sub>1</sub> hybrid paternal and *E. grandis* maternal maps (MYBURG *et al.* 2003) were used for separate, genome-wide QTL detection scans using QTL Cartographer (BASTEN *et al.* 2003). Composite interval mapping (ZENG 1993, 1994), *i.e.*, model 6 of the Zmapqtl module of QTL Cartographer, was used for QTL detection. Likelihood-ratio (LR) profiles [ $-2 \ln(L_0/L_1)$ ], representing the ratio of the likelihood of the null hypothesis ( $L_0$ , no QTL in the marker interval) to the alternative hypothesis ( $L_1$ , presence of a QTL in the marker interval) were generated for each gene at every 2-cM interval of the parental map. A window size of 10 cM was used, and cofactors were chosen by stepwise regression (SRmapqtl). Epistatic interactions were evaluated using multiple interval mapping (KAO *et al.* 1999) in a model with a maximum number of three QTL. Empirical LR thresholds were determined by randomly permuting the trait values among marker genotypes and recording the highest LR peak produced in each random data set using the same QTL detection parameters described above (CHURCHILL and DOERGE 1994; DOERGE and CHURCHILL 1996). The empirical threshold for experiment-wise type I error rates of 0.01, 0.05, and 0.1 was determined by recording the 100th-, 500th-, and 1000th-ranked LR of 10,000 random permutations for 35 cDNAs. The most conservative LR threshold found was used for all the genes.

## RESULTS

**eQTL detection:** Transcript abundance variation for 2608 cDNAs was assessed in 91 genotypes from a pseudo-backcross population (Figure 1). Each individual from the progeny was technically replicated using two dyes, while the parental alleles were biologically replicated in approximately half of the segregating population. Of the 2608 elements represented in the cDNA microarray, 1373 (53%) identified a differential transcript abun-

dance in the backcross progeny after a Bonferroni correction for 2608 tests (experiment-wise  $\alpha = 0.05$ ), reflecting the wide segregation of the mapping population. Every gene (cDNA) was evaluated for eQTL, even when there was no significant difference in expression among the backcross progeny, following the suggestions of BREM *et al.* (2002) and SCHADT *et al.* (2003). Threshold log-likelihood ratios of 11, 13, and 16 were used to obtain an experiment-wise  $\alpha$  of 10, 5, and 1%, respectively, on the basis of permutation tests (CHURCHILL and DOERGE 1994; DOERGE and CHURCHILL 1996). Support intervals were evaluated previously for a set of lignin-related genes and typically extend over 20–40 cM (KIRST *et al.* 2004). Results reported here use a LR threshold of 11, unless stated otherwise, and are fully described in supplementary Tables 1 and 2 at <http://genetics.org/supplemental/>.

**Number and distribution of eQTL:** A total of 1655 eQTL were detected in both maps for 1067 genes (41%). Of these, 608 had been identified with differentially abundant transcripts in the progeny. EQTL were identified for 811 genes (31%) in the F<sub>1</sub> hybrid paternal map (maximum LR = 103) and for 451 genes (17%) in the *E. grandis* maternal genetic map (maximum LR = 90), using the least-stringent criterium for QTL identification in this study (LR > 11). For 195 genes, eQTL were detected in both maps. At the highest stringency (LR > 16), the number of eQTL identified decreased to 193 (7.4%) and 120 (3.3%) for the F<sub>1</sub> hybrid paternal and *E. grandis* maternal maps, respectively. The maximum LR detected for each of the 2608 genes has a skewed distribution (Figure 2) with a median and upper quartile of 9.3 and 11.8 for the F<sub>1</sub> hybrid paternal and 7.8 and 10.0 for the *E. grandis* maternal mapping set (Figure 2A). As might be expected on the basis of previous QTL studies of this pedigree (MYBURG 2001), eQTL were more readily identified in the highly heterogeneous genetic background of the F<sub>1</sub> hybrid, relative to the pure species (*E. grandis*).

**Genetic architecture:** A total of 821 genes displayed a simple genetic architecture for transcript-level variation, with a single eQTL being detected in 73% of the cases where they could be identified in the F<sub>1</sub> hybrid paternal and in 81% of the cases in the *E. grandis* maternal map (Figure 2B). For both maps, more complex patterns were detected in 246 genes. For example, four eQTL were detected for a putative serine carboxypeptidase cDNA sequence (CD668599) in the *E. grandis* map. Three cDNAs displayed five eQTL in the hybrid map, for transcripts encoding an unknown protein (CD668546), a putative protein (CB967788), and a translation elongation factor eEF-1  $\alpha$ -chain (CB967966).

**Additive effects, direction, and proportion of the phenotypic variation explained:** The marker data in the F<sub>1</sub> hybrid paternal map were recoded so that the genotype at each marker was associated with the presence of the *E. globulus* (unknown parent) or the absence of *E. grandis*

(G50) alleles (MYBURG 2001). Therefore, the direction of QTL effects indicates the effect on transcript abundance associated with the substitution of the *E. grandis* QTL allele (negative sign) with the *E. globulus* (positive sign) QTL allele in the *E. grandis* backcross population. Estimates of transcript-level additive effects reported by QTL Cartographer for the F<sub>1</sub> hybrid paternal map ranged from a 1.74-fold change for a phosphatase (CD668452) (negative effect) to a 2.5-fold change for a gene encoding a low-temperature and salt-responsive protein LTI6B (CD669389) (positive effect) (Figure 2C). The eQTL for LTI6B explains 70% of the transcript abundance variation in the progeny, the largest proportion explained for any gene. Analysis of the direction of QTL effects in the *E. grandis* maternal map is arbitrary for each linkage group and, therefore, no inferences can be made about the grandparent origin of the QTL alleles. For the *E. grandis* maternal map, the maximum additive effect was 2.4-fold change for a putative protein (CD668860). The eQTL explained 68% of the phenotypic variation in transcript-level variation measured in the progeny.

**Homology of eQTL detected in the parental maps:** eQTL were identified in both parental maps for 195 genes (LR > 11). Synteny between the linkage groups from the two parental maps was inferred from intercross AFLP markers shared between the parents of the F<sub>1</sub> hybrid and the *E. grandis* (678.2.1) trees (MYBURG *et al.* 2003) and was used to evaluate the genomic locations of eQTL in both maps. From the 195 genes for which eQTL could be identified in the two parental maps, only 13 had significant eQTL localized in homologous linkage groups and did overlap to a certain extent. This suggests that most eQTL are *trans*-acting. If we raised the minimum LR of 13, the number of genes with eQTL in both maps was 62, of which only 6 had eQTL in homologous linkage groups. Variation in recombination rates throughout the genome of both parental trees does not allow a definitive comparison of eQTL genomic locations, but suggests that for this subset of colocalized eQTL, transcript abundance regulation is associated with the same genomic regions, although they could be regulated by different closely linked loci.

**Epistatic interaction and proportion of the phenotypic variation explained:** Transcription is initiated by the interaction of an RNA polymerase with a gene's promoter, normally enhanced by the cooperation of various transcription factors that interact epistatically. Epistasis, evaluated using multiple interval mapping (KAO *et al.* 1999) and identified for 310 genes in the F<sub>1</sub> hybrid paternal map and for 285 genes in the *E. grandis* maternal map, explained up to 62% (F<sub>1</sub> hybrid) and 69% (*E. grandis*) of the transcript-level variation in the backcross population (supplemental Tables 3 and 4 at <http://genetics.org/supplemental/>). For 19 genes in the F<sub>1</sub> hybrid paternal map and 35 in the *E. grandis* maternal map, the epistatic interactions explained a

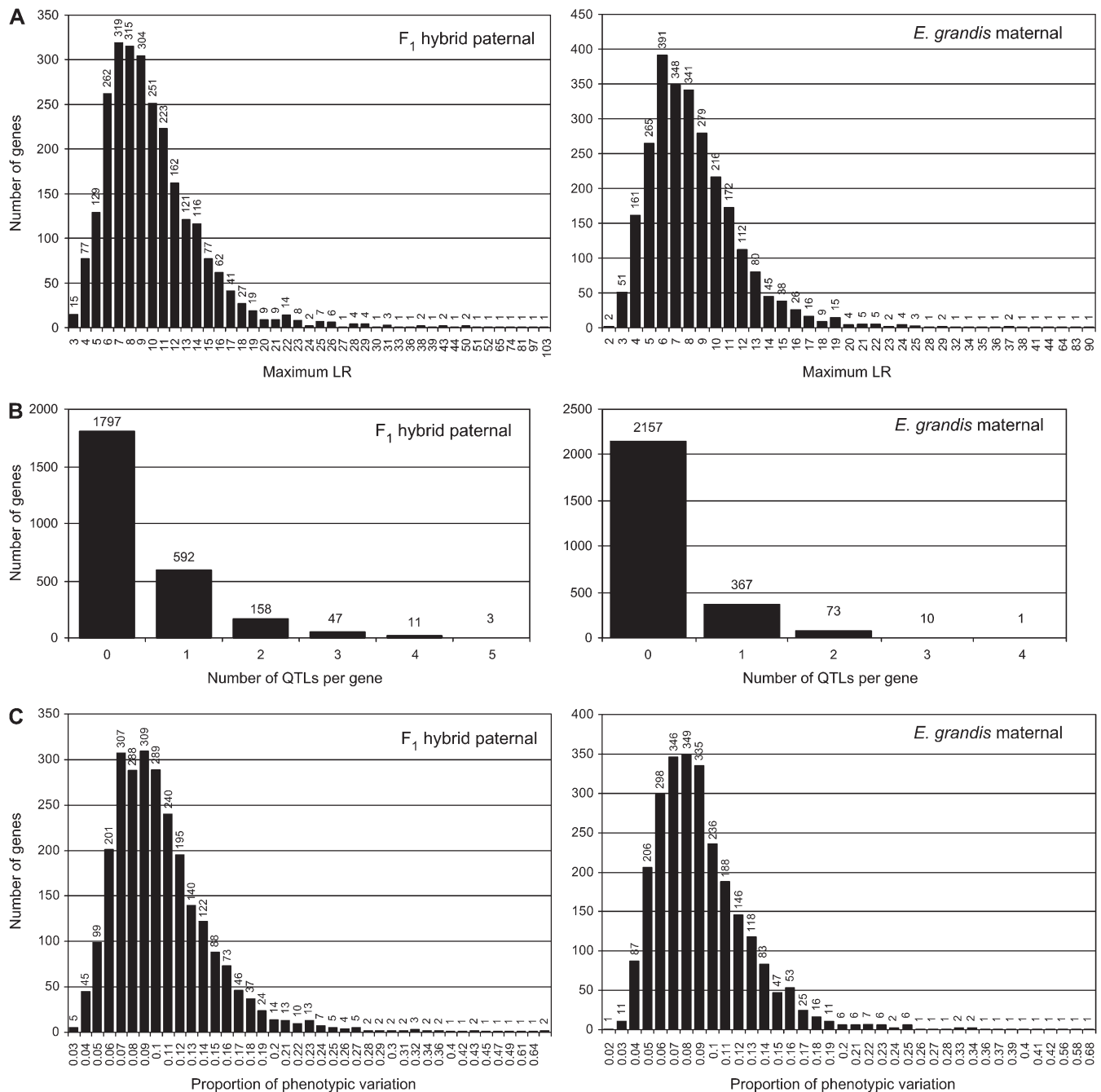


FIGURE 2.—Frequency distribution of eQTL. (A) Distribution of the maximum-likelihood ratios detected for each gene in the F<sub>1</sub> hybrid and *E. grandis* maps. (B) Number of eQTL detected per gene. (C) Proportion of the phenotypic variation explained by the most significant eQTL identified for each gene.

higher proportion of the transcript-level variation than the estimated additive effect of the interacting eQTL.

**Genomic distribution of eQTL:** The total length of the *E. grandis* and the F<sub>1</sub> hybrid (*E. grandis* × *E. globulus*) maps were estimated to be 1335 and 1448 cM, respectively (MYBURG 2001). If eQTL were evenly distributed, they would be detected on average every 2–3 cM in both maps. Instead, eQTL were clustered in certain genomic regions of both the F<sub>1</sub> hybrid paternal map and the *E. grandis* map (Figure 3). Although significant segrega-

tion distortion was observed in this mapping population previously (MYBURG *et al.* 2003), these regions did not colocalize with an eQTL hotspot. Hotspots containing eQTL for 10 or more genes were identified in 18 genomic regions of the F<sub>1</sub> hybrid paternal map, with the exception of LG2, LG3, and LG11. A major cluster containing eight hotspots was identified on LG8, and the largest hotspot, with eQTL for 116 genes, was located on LG9. eQTL hotspots were identified at a lower frequency in

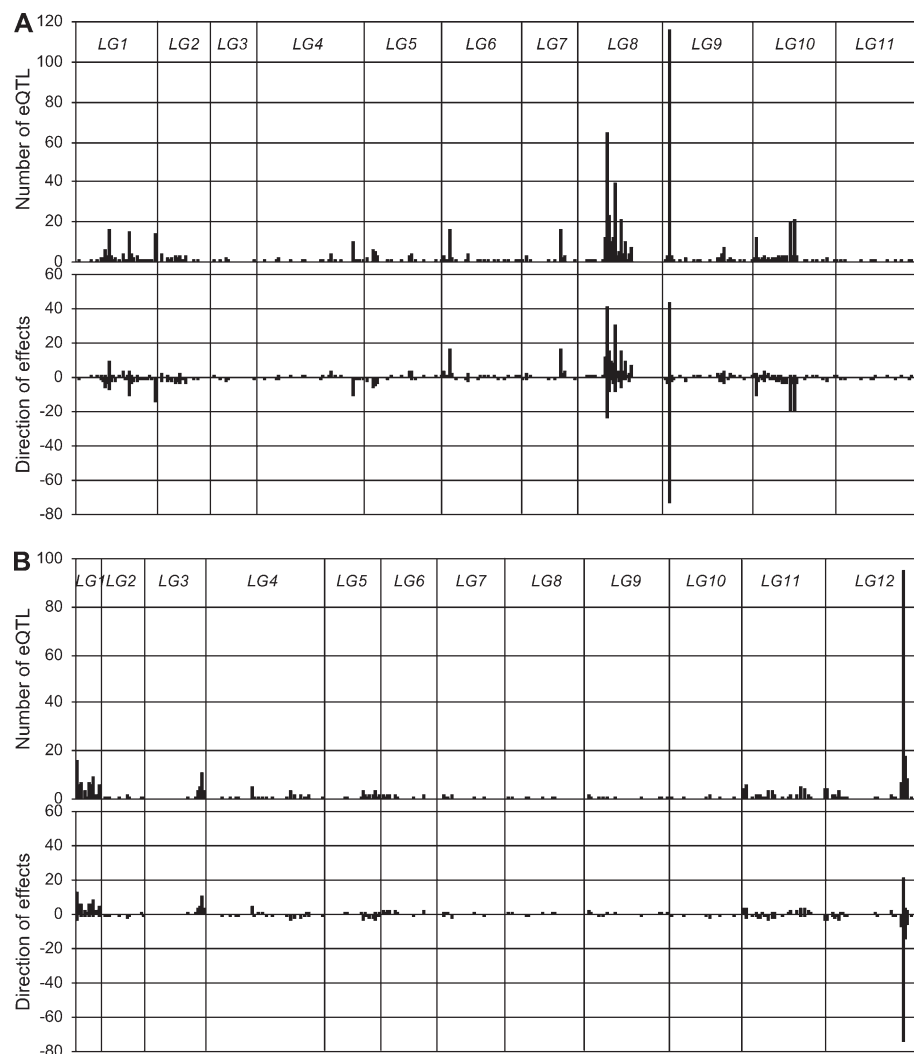


FIGURE 3.—Number and direction of eQTL identified in every 2 cM of the F<sub>1</sub> hybrid (A) and *E. grandis* (B) parental maps. The location of the highest LR was recorded for each gene for which a LR > 11 was detected.

the *E. grandis* maternal map, where only five genomic regions located in LG1, LG3, and LG12 contained QTL for transcript levels of 10 genes or more.

**Direction of effects in eQTL hotspot:** The general direction of effects in eQTL hotspots was evaluated in the F<sub>1</sub> hybrid paternal map. Hotspots containing eQTL with effects predominantly in one direction suggest a genetic locus that affects a large number of genes in the same way and may be associated with differences in wood or growth characteristics between *E. grandis* and *E. globulus*. The proportions of genes with positive and negative eQTL effects at each hotspot were contrasted using a chi-square test. Nine hotspots had a significantly higher proportion of eQTL with either positive or negative effect, at a level of 0.01 (Figure 3), of which five were mostly of positive and four of negative effect. Six of the eight eQTL hotspots clustered in LG8 had effects in both directions. Hotspots could be due to gene-rich regions or to single or a few genes regulating transcript levels of a large number of genes (such as general transcription regulators). Many hotspots included a number of genes coding for enzymes of metabolic pathways, sug-

gesting that some loci regulate the pathway as a whole (KIRST *et al.* 2004).

## DISCUSSION

The quantitative control of expression of 2608 genes was analyzed in an *E. grandis* backcross progeny, generated by mating an F<sub>1</sub> hybrid of *E. grandis* and *E. globulus*, to an unrelated *E. grandis* tree. Transcript levels were estimated for each of 91 genotypes with a technical replication using two dyes (Cy3 and Cy5), while the presence of a QTL was evaluated by partition of the progeny set into two populations, carrying alternative parental alleles for each locus. Therefore, when testing for the presence of a QTL, each allele was biologically replicated in ~45 individuals from the segregating population. A small progeny population was used in this study (91 individuals), implying that some eQTL may have been missed or that their effect was overestimated (BEAVIS 1997). The high environmental variation in tree plantations may also lower the power of detection of eQTL. Despite these limitations, it was possible to identify eQTL for

>40% of the genes represented in the cDNA microarray in both genetic maps. Distribution of eQTL was not random, as would have been expected by chance. Instead, many eQTL were clustered in specific genomic regions. The high genetic diversity and the contrasting phenotypic characteristics of wood (differentiating xylem) from *E. globulus* and *E. grandis* (MYBURG 2001) translated into wide segregation of transcript abundance in the xylem of the progeny, and significant differences were identified for more than half of the genes in the cDNA microarray.

For 195 genes, detecting QTL for transcript-level variation in two genetic maps identified genetic loci that regulate quantitative variation in mRNA abundance in both the F<sub>1</sub> hybrid (*E. grandis* × *E. globulus*) and the *E. grandis* genetic backgrounds. Because synteny had been previously established (MYBURG *et al.* 2003), it was possible to evaluate whether homologous genomic regions were involved in regulation of expression of the same genes. EQTL for only 13 of 195 genes could be identified in both parental homologs. The proportion of genes with eQTL in homologous linkage groups was similar when a higher likelihood ratio (LR > 13) was considered as a threshold for eQTL detection. Lack of conservation of the genetic architecture of transcript abundance regulation in different genetic backgrounds suggests that many different genetic loci could be involved in modulation of transcription of these genes and that there is a complex and variable network of gene expression control.

Mapping of eQTL can indicate *cis*-acting or *trans*-acting regulation. The variation in transcript level can be determined by variation among alleles, that is, some property of the functional gene itself (*cis*-acting). Alternatively, variation may occur in a factor contributing to transcription rate, specificity, or stability of the transcript (*trans*-acting). In most cases, these alternatives may be distinguished by genetic or physical mapping of the transcribed gene. In the case of *cis*-regulation, the genomic location of the eQTL will colocalize with the gene. If variation is *trans*-regulated, the site controlling the variation may be located elsewhere. The strength of inference in distinguishing these alternatives lies in concluding that the locations of the gene and the eQTL are significantly different. The location of eQTL is rarely precise, and the precision of genetic map or physical location, although typically greater, may also have a significant error of estimation. In such experiments, on the basis of the limits of an eQTL interval, two eQTL may appear to overlap, although this could be due to effects of different genes. However, mapping of eQTL to nonhomologous chromosomes in different genetic backgrounds suggests that variation in the regulation of transcript abundance is predominantly in *trans*-acting loci. The complete genome sequence of Eucalyptus is not available and therefore only limited inferences can be made about whether the eQTL identified correspond

to the physical location of the gene (*cis*-regulation) or to the genetic loci of its *trans*-regulator. Some of the genes represented in the cDNA set have been genetically mapped in the *E. grandis* backcross population and examples of *cis*- and *trans*-regulation have been identified (Figure 4). However, if most of the transcript-level regulation occurred in *cis*, a high level of homology among the genetic location of eQTL would have been expected. In yeast, only 25% of the genes for which eQTL were identified colocalized to the position of the eQTL (BREM *et al.* 2002).

The developmental, morphological, and behavioral complexity of higher, multicellular eukaryotes is believed to have arisen as a result of more elaborate regulation of gene expression through a larger set of transcription regulators and more sophisticated DNA and protein regulatory complexes (LEVINE and TJIAN 2003). Therefore, a large proportion of genes in higher plants may display complex patterns of *trans*-regulation. Furthermore, we were able to detect only those regulatory loci that diverged between the two parental species (heterozygous in the F<sub>1</sub> hybrid) or those that are heterozygous in the *E. grandis* backcross parent.

Interactions among multiple segregating genetic loci also appear to play an important role in the control of variation in transcript abundance in this experimental population, because epistasis was identified for 310 genes in the F<sub>1</sub> hybrid paternal map and for 285 genes in the *E. grandis* maternal map. For many genes, a higher proportion of transcript-level variation was explained by epistatic interactions rather than by the estimated additive effects of the interacting loci. In eukaryotes, upstream transcription factors often bind as homo- or heterodimers. Interaction among multiple genetic loci, representing different transcription regulators, may be important for genetic control of gene expression as many of these regulators act in combination to modulate transcription. Numerous other opportunities occur for the modification of mRNA levels before and after initiation of transcription and many could involve multiple loci interacting epistatically rather than additively.

EQTL identified in this study were in many cases clustered in specific genomic regions, or eQTL hotspots, as observed previously in yeast and mice (BREM *et al.* 2002; SCHADT *et al.* 2003; YVERT *et al.* 2003). eQTL hotspots may represent gene-rich regions or the location of a transcription regulator that controls a large set of genes. For certain metabolic or regulatory pathways, mapping of transcript-level variation may identify the genetic loci that control the flux through the pathway. Functional annotation of the genes represented on the cDNA microarray identified functional relationships among genes represented by eQTL in several eQTL hotspots, as observed in yeast (BREM *et al.* 2002). A large set of genes involved in lignin biosynthesis had colocalized eQTL on LG4 and LG9 (KIRST *et al.* 2004). The ability to identify QTL clusters or eQTL hotspots

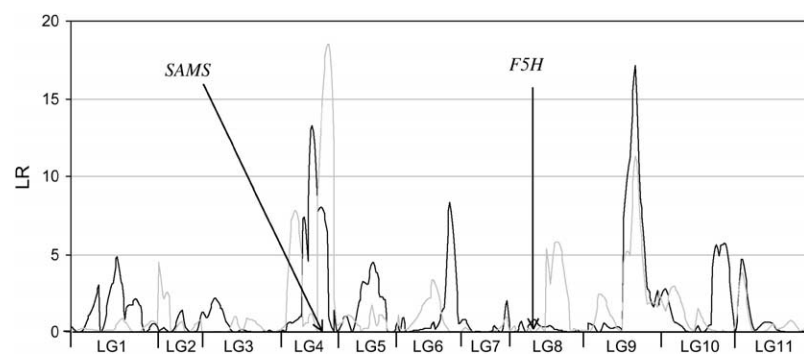


FIGURE 4.—*Cis* and *trans*-acting regulation of gene expression. LR profiles generated by composite interval mapping in the  $F_1$  hybrid paternal map. The arrows indicate the genetic location of the *S*-adenosyl-methionine synthase gene (*SAMS*, CB967747) and the coniferaldehyde 5-hydroxylase (*F5H*, CD669804) gene. The eQTL profile of *SAMS* (LR profile, shaded line) indicates *cis*-regulation of this gene, while the eQTL profile of *F5H* (LR, solid line) suggests *trans*-regulation by multiple loci.

related to specific metabolic functions may depend on the stringency of the genetic regulation of genes in specific pathways. The phenylpropanoid pathway and lignin biosynthesis may require stringent control, as some pathway intermediates are toxic to the cell, requiring an immediate response by other pathways to either process or secrete these compounds. Other metabolic and regulatory pathways may not be under such mechanisms of control and the genetic regulation of a network of genes may not be distinguishable in experiments with microarrays. In many cases, major effects of regulation may not be at the level of transcription.

Finally, the high incidence of eQTL hotspots in some locations may be partially biased, because the genes included here were selected to include specific functional categories (related to wood formation). A similar analysis with a microarray comprising cDNAs representing a large sample of functional genes of *Eucalyptus* could identify hotspots in other genomic regions.

This study provided a genome-wide map of the genetic loci that regulate transcript-level variation for genes involved in biosynthesis of the secondary wall in the differentiating xylem of *Eucalyptus*. Analysis of the genetic architecture of gene regulation involved in wood formation is also a powerful tool for identification of candidate genes for quantitative traits of commercial and biological interest, when combined with phenotypic data (SCHADT *et al.* 2003). Colocalization of eQTL and QTL for traditional quantitative traits such as wood density and growth suggests candidate genes for further analysis. The concept is similar to the approach of mapping genes of metabolic or regulatory pathways that are potential candidates for regulation of quantitative variation in complex traits. This strategy in some cases has been able to identify genes that colocalize with QTL in forestry, but QTL typically span large genomic regions (20–30 cM) and many genes are tested (GION *et al.* 2000; BROWN *et al.* 2003). Identification of colocalized gene expression (eQTL) and trait QTL may be more informative, because only genes transcribed in the tissue of interest and differentially expressed between parental strains are identified. The identification of colocalized gene expression and phenotypic QTL can be supported by the analysis of correlation between tran-

script level and phenotype for positional candidates (genes whose eQTL overlaps the phenotypic QTL). Transcript levels measured by microarrays could account for genetic sources of variation associated with dominance and epistasis, as well as for nongenetic sources of variation, such as developmental and environmental variation. Therefore, an association between transcript-level variation and phenotypic variation in quantitative traits may be higher than that between genetic variation (genetic markers) and phenotype in QTL analysis. This higher association may be useful in distinguishing a specific gene from other positional candidates that control a quantitative trait.

We thank J. Vogel and S. V. Tingey (DuPont) for kindly providing database access and biological materials (cDNAs) used for the microarrays; J. G. De León (Forestal Oriental S.A.) for providing the plant material; G. Gibson and R. Wolfinger for assistance in establishing the appropriate ANOVA model for the analysis of the microarray data and for comments about the manuscript; B. Sosinski and L. He [Genome Research Laboratory-North Carolina State University (NCSU)] for establishing the microarray facility; C. Stasolla, L. van Zyl, D. Craig, and G. Passador-Gurgel for technical assistance and Rebecca Doerge for useful comments about the manuscript. This work was supported in part by the National Science Foundation (grant no. DBI 9975806), by the NCSU Forest Biotechnology Industry Consortium, and by a fellowship from the NCSU genomics program to M.K.

## LITERATURE CITED

- BALTIMORE, D., 2001 Our genome unveiled. *Nature* **409**: 814–816.
- BASTEN, C. J., B. S. WEIR and Z.-B. ZENG, 2003 *QTL Cartographer, Version 1.17: A Reference Manual and Tutorial for QTL Mapping*. Department of Statistics, North Carolina State University, Raleigh, NC.
- BEAVIS, W. D., 1997 QTL analysis: power, precision and accuracy, pp. 145–162 in *Molecular Dissection of Complex Traits*, edited by A. H. PATERNON. CRC Press, Boca Raton, FL.
- BREM, R. B., G. YVERT, R. CLINTON and L. KRUGLYAK, 2002 Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**: 752–755.
- BROWN, G. R., D. L. BASSONI, G. P. GILL, J. R. FONTANA, N. C. WHEELER *et al.*, 2003 Identification of quantitative trait loci influencing wood property traits in loblolly pine (*Pinus taeda* L.). III. QTL verification and candidate gene mapping. *Genetics* **164**: 1537–1546.
- CHANG, S., J. PURYEAR and J. CAIRNEY, 1993 A simple and efficient method for isolating RNA from pine trees. *Plant Mol. Biol. Rep.* **11**: 117–121.
- CHURCHILL, G. A., 2002 Fundamentals of experimental design for cDNA microarrays. *Nat. Genet.* **32**: 490–495.

- CHURCHILL, G. A., and R. W. DOERGE, 1994 Empirical threshold values for quantitative trait mapping. *Genetics* **138**: 963–971.
- DOERGE, R. W., and G. A. CHURCHILL, 1996 Permutation tests for multiple loci affecting a quantitative character. *Genetics* **142**: 285–294.
- DUMAS, P., Y. L. SUN, G. CORBEIL, S. TREMBLAY, Z. PAUSOVA *et al.*, 2000 Mapping of quantitative trait loci (QTL) of differential stress gene expression in rat recombinant inbred strains. *J. Hypertens.* **18**: 545–551.
- ELDRIDGE, K., J. DAVIDSON, C. HARWOOD and G. VAN WYK, 1993 *Eucalypt Domestication and Breeding*. Oxford University Press, Oxford.
- ENARD, W., P. KHAITOVICH, J. KLOSE, S. ZOLLNER, F. HEISSIG *et al.*, 2002 Intra- and interspecific variation in primate gene expression patterns. *Science* **296**: 340–343.
- GIBSON, G., 2003 Population genomics: celebrating individual expression. *Heredity* **90**: 1–2.
- GION, J. M., P. RECH, J. GRIMA-PETTENATI, D. VERHAEGEN and C. PLOMON, 2000 Mapping candidate genes in *Eucalyptus* with emphasis on lignification genes. *Mol. Breed.* **6**: 441–449.
- HEGDE, P., R. QI, K. ABERNATHY, C. GAY, S. DHARAP *et al.*, 2000 A concise guide to cDNA microarray analysis. *Biotechniques* **29**: 548–550.
- JIN, W., R. M. RILEY, R. D. WOLFINGER, K. P. WHITE, G. PASSADOR-GURGEL *et al.*, 2001 The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nat. Genet.* **29**: 389–395.
- KAO, C. H., Z.-B. ZENG and R. D. TEASDALE, 1999 Multiple interval mapping for quantitative trait loci. *Genetics* **152**: 1203–1216.
- KARP, C. L., A. GRUPE, E. SCHADT, S. L. EWART, M. KEANE-MOORE *et al.*, 2000 Identification of complement factor 5 as a susceptibility locus for experimental allergic asthma. *Nat. Immunol.* **1**: 221–226.
- KING, M. C., and A. C. WILSON, 1975 Evolution at two levels in humans and chimpanzees. *Science* **188**: 107–116.
- KIRST, M., A. A. MYBURG, J. P. G. DE LEÓN, M. E. KIRST, J. SCOTT *et al.*, 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 Physiol.* **135**: 2368–2378.
- LADIGES, P. Y., F. UDOVICIC and G. NELSON, 2003 Australian biogeographical connections and the phylogeny of large genera in the plant family *Myrtaceae*. *J. Biogeogr.* **30**: 989–998.
- LEVINE, M., and R. TJIAN, 2003 Transcription regulation and animal diversity. *Nature* **424**: 147–151.
- MACKAY, T. F. C., 2001 The genetic architecture of quantitative traits. *Annu. Rev. Genet.* **35**: 303–339.
- MYBURG, A. A., 2001 Genetic architecture of hybrid fitness and wood quality traits in a wide interspecific cross of *Eucalyptus* tree species. Ph.D. Thesis, North Carolina State University, Raleigh, NC (<http://www.lib.ncsu.edu/theses/available/etd-20010723-175234/>).
- MYBURG, A. A., A. R. GRIFFIN, R. R. SEDEROFF and R. W. WHETTEN, 2003 Comparative genetic linkage maps of *Eucalyptus grandis*, *Eucalyptus globulus* and their F<sub>1</sub> hybrid based on a double pseudo-backcross mapping approach. *Theor. Appl. Genet.* **107**: 1028–1042.
- OLEKSIK, M. F., G. A. CHURCHILL and D. L. CRAWFORD, 2002 Variation in gene expression within and among natural populations. *Nat. Genet.* **32**: 261–266.
- PRYOR, L. D., and L. A. S. JOHNSON, 1971 *A Classification of the Eucalypts*. Australian National University Press, Canberra, Australia.
- ROCKMAN, M. V., and G. A. WRAY, 2002 Abundant raw material for cis-regulatory evolution in humans. *Mol. Biol. Evol.* **19**: 1991–2004.
- SCHADT, E. E., S. A. MONKS, T. A. DRAKE, A. J. LUSIS, N. CHE *et al.*, 2003 Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422**: 297–302.
- SCHENA, M., D. SHALON, R. W. DAVIS and P. O. BROWN, 1995 Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**: 467–470.
- STUART, J. M., E. SEGAL, D. KOLLER and S. K. KIM, 2003 A gene-coexpression network for global discovery of conserved genetic modules. *Science* **302**: 249–255.
- VELCULESCU, V. E., L. ZHANG, B. VOGELSTEIN and K. W. KINZLER, 1995 Serial analysis of gene-expression. *Science* **270**: 484–487.
- WAYNE, M. L., and L. M. MCINTYRE, 2002 Combining mapping and arraying: an approach to candidate gene identification. *Proc. Natl. Acad. Sci. USA* **99**: 14903–14906.
- WOLFINGER, R. D., G. GIBSON, E. D. WOLFINGER, L. BENNETT, H. HAMADEH *et al.*, 2001 Assessing gene significance from cDNA microarray expression data via mixed models. *J. Comput. Biol.* **8**: 625–637.
- YVERT, G., R. B. BREM, J. WHITTLE, J. M. AKEY, E. FOSS *et al.*, 2003 Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat. Genet.* **35**: 57–64.
- ZENG, Z.-B., 1993 Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proc. Natl. Acad. Sci. USA* **90**: 10972–10976.
- ZENG, Z.-B., 1994 Precision mapping of quantitative trait loci. *Genetics* **136**: 1457–1468.

Communicating editor: S. R. McCouch

