Copyright © 2007 by the Genetics Society of America
DOI: 10.1534/genetics.106.061127

# Association Genetics in *Pinus taeda* L. I. Wood Property Traits

**Santiago C. González-Martínez,*,† Nicholas C. Wheeler,‡,[1] Elhan Ersoz,* C. Dana Nelson§ and David B. Neale*,**,[2]**

**Department of Plant Sciences, University of California, Davis, California 95616, †Department of Forest Systems and Resources, Forest Research Institute, CIFOR-INIA, 28040 Madrid, Spain, ‡Weyerhaeuser, Weyerhaeuser Technical Center, Tacoma, Washington 98477, §Southern Institute of Forest Genetics, USDA Forest Service, Saucier, Mississippi 39574 and **Institute of Forest Genetics, Pacific Southwest Research Station, USDA Forest Service, Davis, California 95616*

Manuscript received May 22, 2006
Accepted for publication October 26, 2006

ABSTRACT

Genetic association is a powerful method for dissecting complex adaptive traits due to (i) fine-scale mapping resulting from historical recombination, (ii) wide coverage of phenotypic and genotypic variation within a single experiment, and (iii) the simultaneous discovery of loci and alleles. In this article, genetic association among single nucleotide polymorphisms (58 SNPs) from 20 wood- and drought-related candidate genes and an array of wood property traits with evolutionary and commercial importance, namely, earlywood and latewood specific gravity, percentage of latewood, earlywood microfibril angle, and wood chemistry (lignin and cellulose content), was tested using mixed linear models (MLMs) that account for relatedness among individuals by using a pairwise kinship matrix. Population structure, a common systematic bias in association studies, was assessed using 22 nuclear microsatellites. Different phenotype: genotype associations were found, some of them confirming previous evidence from collocation of QTL and genes in linkage maps (for example, *4cl* and percentage of latewood) and two that involve nonsynonymous polymorphisms (*cad* SNP M28 with earlywood specific gravity and *4cl* SNP M7 with percentage of latewood). The strongest genetic association found in this study was between allelic variation in α-*tubulin*, a gene involved in the formation of cortical microtubules, and earlywood microfibril angle. Intragenic LD decays rapidly in conifers; thus SNPs showing genetic association are likely to be located in close proximity to the causative polymorphisms. This first multigene association genetic study in forest trees has shown the feasibility of candidate gene strategies for dissecting complex adaptive traits, provided that genes belonging to key pathways and appropriate statistical tools are used. This approach is of particular utility in species such as conifers, where genomewide strategies are limited by their large genomes.

A major goal of molecular population and evolutionary genetic studies is to identify the causal genes of natural variation in traits that affect fitness and result in evolutionary change through natural selection and adaptation to local environments. Similarly, plant breeders seek to identify causative polymorphisms for important agronomic traits, thus providing a powerful resource for genetic improvement of plant crops through direct allele selection (Haussmann *et al.* 2004) and/or biotechnology (Boerjan 2005). In forest trees, adaptive traits often have both evolutionary and agronomic relevance, a perfect example of which is an array of wood property traits that combine fundamental roles in the evolution of land plants, plant growth, and resistance to biotic and abiotic environmental stress in nature with practical importance in lumber and pulp production (Peter and Neale 2004). The primary goal of this study was to identify allelic effects of genes controlling wood property phenotypes using a population genomics approach (association genetic or linkage disequilibrium mapping) to genetic dissection.

Molecular approaches to genetically dissecting wood properties and other adaptive traits (*e.g.*, growth rhythm, cold hardiness) in forest trees have historically focused on quantitative trait locus (QTL) mapping (Groover *et al.* 1994; Frewen *et al.* 2000; Sewell *et al.* 2000, 2002; Jermstad *et al.* 2001a,b, 2003; Brown *et al.* 2003; Wheeler *et al.* 2005; Casasoli *et al.* 2006). QTL mapping has provided a comprehensive understanding of the number of QTL, the size of effects of individual QTL and the approximate location of QTL in a number of tree genomes. However, QTL studies are primarily relevant within the pedigree(s) being evaluated, severely limiting their utility to make broad evolutionary inferences. Furthermore, given the large genetic-to-physical distance in most conifers (~3000 kb/cM), identification of specific genes responsible for phenotypic variation in these species is unlikely via QTL mapping.

Association genetics—a population-level survey that takes advantage of historical recombination to identify

[1]*Present address:* Molecular Tree Breeding Services, Centralia, WA 98531.

[2]*Corresponding author:* Department of Plant Sciences, University of California, 1 Shields Ave., Davis, CA 95616.
E-mail: dbneale@ucdavis.edu

trait–marker relationships on the basis of linkage disequilibrium (Cardon and Bell 2001; Flint-Garcia *et al.* 2003)—has become a favored genetic approach for dissecting quantitative traits in many organisms (Thornsberry *et al.* 2001; Flint-Garcia *et al.* 2003; Neale and Savolainen 2004; González-Martínez *et al.* 2006a). Loblolly pine (*Pinus taeda* L.), a conifer distributed in large, out-crossing, natural populations, possesses virtually all of the genetic properties deemed of value for association genetic studies. For instance, the species (i) possesses substantial levels of nucleotide diversity and has low linkage disequilibrium (Brown *et al.* 2004; González-Martínez *et al.* 2006b), (ii) can be easily propagated to create large detection and verification populations, maintainable over many years (and multiple sites), and (iii) retains virtually all of its natural genetic variability, even in relatively small breeding populations two to three generations removed from wild stands.

The essence of the association genetic approach is the identification of statistical associations between variation in relevant phenotypic traits and allelic polymorphism in known genes. Wood properties of evolutionary and practical significance (Zobel and van Buijtenen 1989; Zobel and Jett 1995) chosen for study here are (i) wood specific gravity or density, (ii) microfibril angle, (iii) percentage of latewood, and (iv) chemical composition of the cell wall (*i.e.*, the relative proportions of polysaccharides, such as α-cellulose and hemi-cellulose and lignins). Identifying the relevant genes for study is more complex. Due to the large genome size (30 billion bp) and rapid decay of linkage disequilibrium in pine, a candidate gene approach was deemed appropriate.

In forest trees, a number of genes involved in the biosynthesis of polysaccharides, lignins, and cell-wall proteins have been identified via classical biochemical analysis and gene or protein expression profiling (see reviews in Whetten *et al.* 1998; Plomion *et al.* 2001; Peter and Neale 2004; Boerjan 2005). Several of these candidate genes for wood formation have been confirmed by forward or reverse genetic mutant analyses in model species (see, for example, Goujon *et al.* 2003 for xylem lignification in Arabidopsis) or by the study of natural mutants (see Ralph *et al.* 1997 and Gill *et al.* 2003 for *cad* gene). In addition, on the basis of the pattern of DNA sequence polymorphism, Pot *et al.* (2005) suggested a functional role in wood formation for three genes: *ppl* (a glycine-rich protein), *cesA3* (a cellulose synthase), and *korrigan* (a gene involved in cellulose–hemi-cellulose assembly; Pot *et al.* 2006), and González-Martínez *et al.* (2006b) found that *ccoaomt-1*, a gene involved in lignin biosynthesis, might be a target of balancing selection in natural populations of *P. taeda*.

QTL studies have often shown that phenotypic effects of individual genes are small [for example, most QTL explained <5% of the phenotypic variance of wood property traits in *P. taeda* (Brown *et al.* 2003)]. Furthermore, complex interactions among genes may mask single-allele effects. Thus, it is generally difficult to demonstrate that a particular allelic variant is indeed causally related to a phenotype (Weigel and Nordborg 2005). Association genetics simultaneously allows for the detection of alleles with moderate effects on phenotypes and the study of epistatic interactions among loci (Tabor *et al.* 2002; Hirschhorn and Daly 2005; Szalma *et al.* 2005). Successful association studies require large population sizes because variants that contribute to complex traits are likely to have only modest phenotypic effects (Hirschhorn and Daly 2005). Studies of statistical power for association using coalescence simulations of single random-mating populations with mutation, genetic drift, and recombination showed that ~500 individuals are necessary for detecting causative polymorphisms of small effect and that greater power is achieved more by increasing the sample size than by increasing the number of polymorphisms (Long and Langley 1999).

Another major concern in association studies is the high rate of false positives, a likely factor in the inability to replicate associations found in the literature (Redden and Allison 2003 and references therein; see also the review of 603 published gene–disease associations involving 268 genes by Hirschhorn *et al.* 2002). In large studies, a significant number of false positives will arise simply by chance. False positives can also arise from population stratification (see, for instance, Aranzana *et al.* 2005 for Arabidopsis) and/or within-population kinship among individuals, and genetic association models that account for the different levels of relatedness found in natural populations have recently been developed (Pritchard *et al.* 2000; Thornsberry *et al.* 2001; Yu *et al.* 2006).

Although association genetic approaches to dissecting complex traits are relatively recent among crop and other plant species, early results are promising. Genomewide genetic association in Arabidopsis identified previously known flowering-time (*fri*) and pathogen resistance (*rpm1*, *rps2*, and *rps5*) genes (Aranzana *et al.* 2005). Studies based on one or two candidate genes were also successful in associating flowering time with allelic variation in different Arabidopsis (*cry2*, Olsen *et al.* 2004; *fri*, Hagenblad *et al.* 2004 and Shindo *et al.* 2005), *Brassica nigra* (*col1*, Österberg *et al.* 2002), and maize (*dwarf8*, Thornsberry *et al.* 2001) genes. In maize, several associations among candidate genes (previously identified by functional genetics approaches or QTL studies) and important commercial phenotypic traits have been found, including associations between digestibility and the *zmpox3* peroxidase gene (Guillet-Claude *et al.* 2004), between kernel composition and *sh1*, *sh2*, and *bt2* (Wilson *et al.* 2004), and between maysin synthesis and the *a1* gene (Szalma *et al.* 2005). The first published association study in forest trees found an association between two single nucleotide

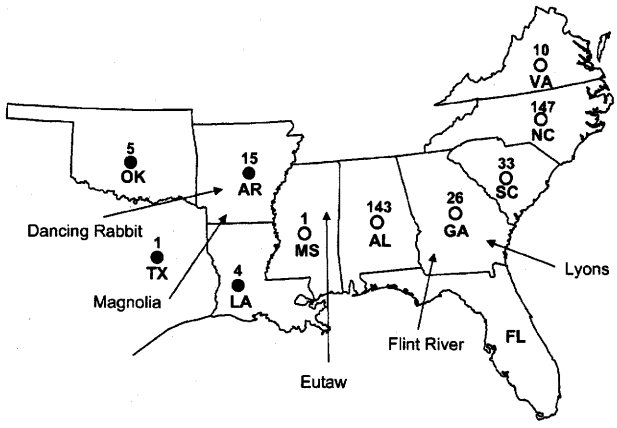


FIGURE 1.—Map showing region of tree origin and locations of sampled clone banks and seed orchards; 34 second-generation selections obtained by controlled crossing of trees from different U. S. states and 16 trees of unknown origin are not shown. Solid dots represent sources that originate west of the Mississippi River.

polymorphism (SNP) markers from the *ccr* gene and microfibril angle in *Eucalyptus nitens*, explaining ~5% of the total phenotypic variation. These results were supported by haplotype analysis in the same population and by screening two full-sib families of *E. nitens* and *E. globulus* (THUMMA *et al.* 2005). The general success of association studies in plants highlights the utility of association mapping in important tree crops, such as pine, Douglas fir, spruce, and poplar, involving a wide range of commercial and fitness traits and well-characterized candidate genes.

For this study, a collection was made of first- and second-generation tree selections (>400) from throughout the natural range of loblolly pine, an ecologically and economically important species of the southeastern United States. Statistical models were used to account for population structure and pairwise kinship, and associations between allelic variation of 58 SNPs and several wood properties were estimated.

## MATERIALS AND METHODS

**Association population:** The association population consisted of trees growing in *ex situ* clone banks and seed orchards containing grafted first- and second-generation parent tree selections (clones) from the Weyerhaeuser loblolly pine improvement program. Clones were originally selected because they were fast growing, had straight stems, and were free of disease; they originated from 10 states in the southeastern United States (Figure 1). Ten separate *ex situ* orchards and clone banks, located at five sites (see supplemental Table S1 at http://www.genetics.org/supplemental/), were sampled. At the time at which clones were chosen for this study, all trees exceeded 15 years of age (16–23) with the exception of 42 clones located in a single clone bank (Atlantic Standard, Lyons, age 9).

Wood and needle tissues were sampled from >480 clones, 73% of which were represented in the collection by two replicates (ramets). Three RFLP markers were used to genotype DNA extracted from needle tissues of clones with two ramets to ensure clonal integrity. Clones with nonmatching genotypes were excluded from the study. Other clones were subsequently excluded due to incomplete phenotypic or genotypic (SNP) data sets. In total, the number of clones used in this genetic association study was 435 and 422 for solid wood and wood chemistry traits, respectively (see supplemental Table S2 at http://www.genetics.org/supplemental/ for clone origin). The majority of the second-generation selections (173 clones) shared one parent with at least one other tree and in rare cases full-sibs were included, but only a few second-generation trees were obtained from first-generation selections included in this study.

**Phenotypic data:** *Specific gravity, percentage of latewood, and microfibril angle:* Multiple radial wood cores (5 mm) were taken from each ramet 24 in. above the graft union (between 4 and 5 ft from the ground). Cores were cropped at the pith and outer edge of ring 15. Wood specific gravity and the volume percentage of latewood were determined using a continuous X-ray densitometry scan. Specific gravity was determined for both earlywood and latewood for each ring (age 3–15) and data were averaged to create composite traits: rings 3–5 (juvenile wood), rings 6–10 (transition wood), rings 11–15 (mature wood), and rings 3–15 (all ages). Similar composite traits were developed for the percentage of latewood. Composite traits were considered more informative than individual ring data because they represented expression over a longer period of time (SEWELL *et al.* 2000, 2002; BROWN *et al.* 2003). All phenotypic measurements were standardized by the mean of the clone bank or seed orchard from which they came.

Microfibril angle was determined for earlywood of rings 3, 5, and 10 using X-ray diffraction techniques described by MEYLAN (1967), EL-OSTA *et al.* (1972), and MEGRAW *et al.* (1998).

*Wood chemistry:* Radial wood cores from each ramet were cropped to include only rings 3–9 and individually ground in a Wiley mill at 20 mesh. For clones with two ramets, ground samples were combined. All samples were subsequently reground at a finer mesh to make a single clonal sample. For each clone, three aliquots (~250 mg, with resampling) were scanned with a near-infrared (NIR) spectra fiber-optic probe (Analytical Spectral Devices) from a FieldSpec FR spectrometer (GARBUTT *et al.* 1992). Spectral data were collected from 350 to 2500 nm and averaged over aliquots. Spectral data were subsequently subjected to *p*rojection to *l*atent *s*tructure (PLS), a multivariate analysis technique used to correlate NIR sample spectra to wet chemistry reference measurements (BURNS and CIURCZAK 1992). Calibration models developed from PLS (see below) were used to predict the chemistry of all population samples. Lignin and cellulose content ranged from 27.8 to 35.1% and from 32.1 to 38.8%, respectively.

The reference chemistry method employed was *h*igh *p*H *a*nion-*e*xchange *c*hromatography with *p*ulsed-*a*mperometric *d*etection (HPAEC–PAD; WALLIS 1996; DAVIS 1998). Specifically, ground wood undergoes a two-stage hydrolysis process in sulfuric acid to convert the principle cellulosic wood polymers to carbohydrate monomers in solution (arabinose, galactose, glucose, xylose, and mannose). The noncarbohydrate portion of the wood is largely retained as an acid-insoluble residue (primarily lignin with some acid-insoluble ash). Following hydrolysis, the solution of wood carbohydrates was separated by HPAEC using a Carbo-Pac PA1 analytical column (Dionex) followed by PAD. For each analyte, the carbohydrate concentration in solution was converted to the equivalent weight percentage of anhydrous polymer on an oven-dry sample basis. PLS models ($N = 28$) correlating NIR spectra with acid-insoluble solids (approximately lignin), glucan, and mannan were developed. Samples subjected to both NIR and reference

chemistry methods suggest that the acid-insoluble solids, glucan, and mannan predictions can be made to within ±2.2, 1.6, and 1.2%, respectively (absolute weight percentage). Cellulose content was estimated from a NIR prediction of the glucan content after subtracting the portion of glucan derived from wood glucomannans (hemi-celluloses). The glucomannan correction follows the method of JANSON (1970; cellulose = glucan − mannan/3.6). Both acid-insoluble solids and carbohydrates were measured in quadruplicate for all calibration samples.

**SNP genotyping:** *DNA isolation:* DNA isolation was accomplished first by grinding needle tissue in liquid nitrogen followed by whole-DNA extraction by QIAGEN (Valencia, CA) DNeasy maxi plant DNA extraction kit in a 96-well plate format according to the manufacturer's instructions. A final volume of 200 μl of ~100 ng/μl DNA was obtained. Usage stocks at 6 ng/μl were prepared for downstream PCR applications and used normally with 1:10 dilutions within the PCR reaction mixes (a total of ~5–10 ng/reaction).

*SNP discovery and selection:* SNP discovery was previously conducted by direct sequencing of megagametophyte DNA samples by BROWN *et al.* (2004) and GONZÁLEZ-MARTÍNEZ *et al.* (2006b) in 19 wood- and 18 drought-related candidate genes. From those sets (288 SNPs from wood-related genes and 196 SNPs from drought-related genes), a total of 58 SNPs from 20 candidate genes (1–7/gene), including 8 that collocate with wood property QTL (*c4h-1, 4cl, c3h-1, sams-2, ccoaomt-1, agp-6, agp-like,* and *α-tubulin*; BROWN *et al.* 2003; our unpublished results), were used in this study (see supplemental Table S3 at http://www.genetics.org/supplemental/). Priority for genotyping was given to nonsynonymous substitutions, since they reflect changes at the protein level and thus are putative indicators of functional variation. One to six additional haplotype-tagging SNPs per gene were selected for genotyping to represent most standing haplotypic variation within the candidate genes. About 72% of the markers were common SNPs (minor allele frequency >0.05).

*SNP genotyping:* Genotyping was conducted on a Victor$^2$-Wallac SNP genotyping platform with the Acryloprime Universal Florescence Polarization Terminator Dye Incorporation kit (FP-TDI, Perkin-Elmer, Torrance, CA). Genotyping primers for FP-TDI were designed within the immediate 30 bp upstream or downstream of the SNP to be genotyped observing criteria such as absence of SNPs within the primer binding site and, when possible, >50° melting temperatures (HSU *et al.* 2001). SNPs were scored according to the clustering of genotypic groups (see supplemental Figure S1 at http://www.genetics.org/supplemental/ for an example). Sequences for the genotyping primers, their annealing temperatures, the direction of single nucleotide extension reaction, and the alleles at the SNP loci are listed elsewhere (supplemental Table S3 at http://www.genetics.org/supplemental/).

**Population structure:** Population stratification is the most serious systematic bias producing false-positive associations (MARCHINI *et al.* 2004; HIRSCHHORN and DALY 2005). To test for population structure, all 435 selections were genotyped with 22 nuclear microsatellites (nuSSRs) exhibiting high levels of polymorphism (expected heterozygosity 0.242–0.944; see formulas in NEI 1978). A first test of population structure was done using the 252 first-generation selections collected from undisturbed stands in the 1950s and a model-based clustering algorithm (STRUCTURE software; PRITCHARD *et al.* 2000; ROSENBERG *et al.* 2002). Models with a putative number of clusters ($K$ parameter) from one to eight, noncorrelated allele frequencies, and both burn-in, to minimize the effect of the starting configuration, and run-length periods of $10^6$ were run. In addition, a standard regression analysis between pairwise genetic and geographical distances was performed.

Second, we delimited five major geographical regions on the basis of seed transfer recommendations (SCHMIDTLING 2001) and climate data (Spatial Climate Analysis Service, Oregon State University, http://www.ocs.oregonstate.edu/prism/) and performed an analysis of molecular variance (AMOVA) including 27–34 trees/geographical region. AMOVA was computed following WEIR and COCKERHAM (1984). A permutation test (10,000 permutations) was used to test for significant population genetic structure among regions. To obtain a more balanced data set, the AMOVA analyses included both first- and second-generation selections obtained from crosses of trees belonging to the same region (see supplemental Table S2 at http/www.genetics.org/supplemental/).

**Test statistics for association:** Phenotypic trait variables were normally distributed or closely approximated a normal distribution, and thus it was not necessary to apply data transformations. Variables for the same trait (composite measures for different ring-age groupings; see *Phenotypic data*) were highly correlated (Pearson correlation coefficient of ~0.50–0.95). To create an overall composite measure for each trait, principal component analysis (PCA) was performed and the principal component with the highest eigenvalue was retained. First principal components explained 85.28% for earlywood specific gravity, 81.62% for latewood specific gravity, 80.42% for percentage of latewood, 73.46% for earlywood microfibril angle, and 97.45% for lignin/cellulose content total variation. Subsequent tests of association were done on the basis of both the original variables (to test for differences among juvenile, transition, and mature wood) and these synthetic principal components.

Single-marker-based tests were preferred to haplotype-based tests to avoid uncertainty in haplotype determination from diploid SNP data sets. In addition, single-marker-based association analyses have either similar or greater power than haplotype-based tests, as shown by simulation studies (LONG and LANGLEY 1999). A *m*ixed *l*inear *m*odel (MLM) was fitted for each single marker and trait (see YU *et al.* 2006 for details). This approach takes into account relatedness among individuals by using a pairwise kinship matrix as covariate in a mixed model and was deemed appropriate to address relatedness caused by second-generation tree selections. SPAGeDi version 1.2 software (HARDY and VEKEMANS 2002) was used to estimate Nason's kinship coefficient (LOISELLE *et al.* 1995) using 22 nuSSRs, and a kinship matrix was built for second-generation pairs of trees. Kinship of unrelated first-generation trees and negative kinship values were set to zero following YU *et al.* (2006). To avoid potential biases caused by genetic differentiation between origins west and east of the Mississippi Valley (*i.e.*, population structure), a factor with two levels indicating tree origin was included in the models. MLMs were run using Tassel version 1.9.4 (release March 2006). Corrections for multiple testing were performed using the positive false discovery rate (FDR) method (STOREY 2002; STOREY and TIBSHIRANI 2003; see also http://faculty.washington.edu/~jstorey/qvalue/).

## RESULTS

**Population structure:** The model-based clustering analyses showed a pattern typical of unstructured populations (PRITCHARD and WEN 2004): namely, plateaus in the estimate of the log-likelihood of the data were not reached; the proportion of the sample assigned to each population was roughly symmetric (~$1/K$ in each population; in our case, proportions were 30–36% for $K = 3$

**TABLE 1**

**Significant associations after correction for multiple testing using the positive FDR method (*Q*-values)**

| Trait | Wood-age type | Gene | SNP | *N* | Marker effect | | | FDR *Q*-value |
|---|---|---|---|---|---|---|---|---|
| | | | | | *F* | *P* | $R^2$ | |
| ewsg | Transition | *sams-2* | M44 | 403 | 6.7595 | 0.0013 | 0.0327 | 0.0630 |
| | All ages | *cad* | M28 | 409 | 7.7480 | 0.0005 | 0.0347 | *0.0228* |
| | PCA | *cad* | M28 | 366 | 6.5945 | 0.0015 | 0.0351 | 0.0742 |
| lw | Transition | *lp3-1* | Q5 | 431 | 7.9007 | 0.0004 | 0.0357 | *0.0248* |
| ewmfa | Transition | *α-tubulin* | M10 | 374 | 8.3766 | 0.0040 | 0.0221 | *0.0062* |
| | PCA | *α-tubulin* | M10 | 370 | 13.5080 | 0.0003 | 0.0355 | *0.0078* |

Mixed linear models included a factor identifying clones from west or east of the Mississippi Valley, a potential cause of population structure. ewsg, earlywood specific gravity; lw, percentage of latewood; ewmfa, earlywood microfibril angle. Only SNP per-trait associations with FDR *Q*-values <0.10 are shown. FDR *Q*-values ≤0.05 are in italics.

and 18–21% for $K = 5$); and most individuals were admixed (see supplemental Figure S2A at http://www.genetics.org/supplemental/). Moreover, correlation between pairwise genetic and geographical distances, as shown by a standard regression analysis, was extremely low ($R^2 = 0.0047$; supplemental Figure S2B at http://www.genetics.org/supplemental/).

Genetic differentiation among major geographical regions, as estimated by AMOVA, was low ($F_{ST} = 0.0083$) but significant ($P < 0.00$). However, when populations from west of the Mississippi Valley were removed from the analysis, the differentiation estimate was lowered to 0.0001 and was not significant ($P = 0.92$). To account for this population structure, a factor with two levels indicating tree origin (west or east of the Mississippi Valley) was included in all genetic association models.

**Genetic association:** *Specific gravity in earlywood and latewood:* Significant ($P < 0.05$) associations were found for 4 (juvenile wood) to 10 (synthetic PCAs) SNPs in earlywood and 1 (mature wood) to 7 (transition wood) SNPs in latewood specific gravity (see supplemental Tables S4A and S4B at http://www.genetics.org/supplemental/). Several genes, namely, *cad, sams-2, comt-2, dhn-2, lp3-3, ccr-1, α-tubulin, c3h-1,* and *c4h-1* for earlywood and *c3h-1, c4h1,* and *c4h-2* for latewood, had SNPs that gave consistent genetic associations with specific gravity in different wood-age types. However, correction for multiple testing using the FDR method resulted in only two significant associations for earlywood, involving *cad* and *sams-2* genes, and none for latewood (Table 1).

*Cinnamyl alcohol dehydrogenase* (*cad*) SNP M28 showed significant *Q*-values obtained by FDR for earlywood specific gravity (ewsg) in pooled rings 3–15. Uncorrected *P*-values for association of SNP M28 and ewsg were ≤0.004 in single variables representing juvenile (rings 3–5), transition (rings 6–10), and mature (rings 11–15) wood, but were not significant after correction for multiple testing. Heterozygous trees for this SNP showed higher average earlywood specific gravity than both homozygous types (0.0032 in AT *vs.* −0.0056 and −0.0023 in AA and TT, respectively, for rings 3–15 referred to the overall average; see also Figure 2), possibly indicating overdominance.

*S-adenosyl methionine synthetase 2* (*sams-2*) SNP M44 showed genetic association to earlywood specific gravity in juvenile and transition wood (see supplemental Table S4A at http://www.genetics.org/supplemental/), with *Q*-values obtained by FDR nearly significant for transition wood (FDR *Q*-value = 0.063). In this case, the mode of action seems additive with the G allele conferring a higher earlywood specific gravity (−0.0050 in AA, −0.0027 in AG, and 0.0026 in GG for rings 3–15 referred to the overall average; see also Figure 2 for transition wood).

Combining the putative significant associations of the *cad* and *sams-2* genes in a general lineal model, these genes explained ~7% (in mature wood) to 10% (all age) of the total phenotypic variance and ~14–20% of the total genetic variance (considering $h^2 \sim 0.50$; Zobel and Jett 1995) for earlywood specific gravity in loblolly pine.

*Percentage of latewood:* Eight SNPs showed genetic association ($P < 0.05$) with the percentage of latewood (see supplemental Table S4C at http://www.genetics.org/supplemental/). However, only the genetic association between SNP Q5 from the *water-stress inducible protein 1* (*lp3-1*) gene and the percentage of latewood in transition wood (rings 6–10) remained significant after correction for multiple testing using FDR (Table 1). This association was caused by two ramets of the same clone sampled from different clonal banks (Taxa46 and Taxa47 in supplemental Table S2 at http://www.genetics.org/supplemental/) with a very high percentage of transition and mature latewood (~17 and ~12% more latewood percentage than the average, respectively; see also Figure 2) and homozygous for the G allele; the only other tree with a GG genotype showed a moderately higher than average percentage of transition and mature latewood (Taxa398: ~5 and ~1% more latewood percentage than the average, respectively).

Uncorrected tests repeatedly showed (across wood-age types) genetic associations between *4-coumarate CoA ligase* (*4cl*) SNPs M3 and M7 and percentage of latewood

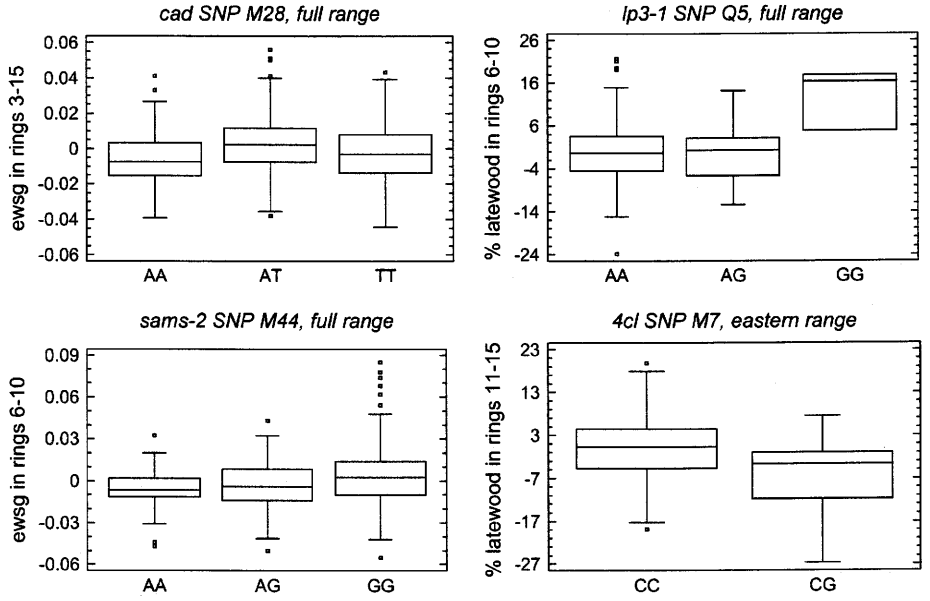


FIGURE 2.—Genotypic effects (box plots) of SNPs that showed significant genetic association (after correction for multiple testing) with earlywood specific gravity (*cad* SNP M28 and *sams-2* SNP M44) and percentage of latewood (*lp3-1* SNP Q5 and *4cl* SNP M7 in the east of the Mississippi Valley range).

(see supplemental Table S4C at http://www.genetics.org/supplemental/), but none of these associations were significant after multiple testing corrections. However, a significant interaction between marker and population structure effects was found (data not shown) and, once the 42 clones from west of the Mississippi Valley were removed, genetic associations for *4cl* SNP M7 and both mature wood and pooled rings 3–15 were significant after correction for multiple testing (FDR *Q*-values of 0.015 and 0.040, respectively). In the east of the Mississippi Valley range, there was a heterozygous (CG) genotypic effect of 4.57 reduction in latewood percentage (−4.4000 in CG and 0.1675 in GG for rings 3–15 referred to the overall average; see also Figure 2) that was not present in the west of the Mississippi Valley range.

*Earlywood microfibril angle:* Fourteen SNPs showed significant genetic association ($P < 0.05$) with earlywood microfibril angle (see supplemental Table S4D at http://www.genetics.org/supplemental/). The strongest genetic association found for this trait involves a silent mutation in intron I of the α-*tubulin* gene (SNP M10). In addition to α-*tubulin* SNP M10, uncorrected tests showed genetic association with earlywood microfibril angle in more than one wood-age type for *dhn-1* SNP Q1 (juvenile and mature wood) and *comt-2* SNP M38 (transition and mature wood). Another candidate gene possibly associated with earlywood microfibril angle might be *cinnamoyl CoA reductase 1* (*ccr-1*); a significant association was found between SNPs M46 and M48 in this gene and earlywood microfibril angle in rings 3 and 5, but *P*-values for these associations (range of 0.0098–0.0330) were not significant after correction for multiple testing using FDR.

With respect to α-*tubulin* SNP M10, mixed models (MLM) showed strong genetic association with earlywood microfibril angle for ring 5 (FDR *Q*-value = 0.0062), representing transition wood, and synthetic PCAs (FDR *Q*-value = 0.0078). Uncorrected *P*-values for other types of wood were 0.0203 and 0.0040 for rings 3 and 10, respectively (see supplemental Table S4D at http://www.genetics.org/supplemental/). Minor allele frequency (allele G) for this SNP in the association population was 0.024 (20/417). Given Hardy–Weinberg proportions, GG genotypes are expected to be rare (frequency <0.001) and were not found in the association population (see also Figure 3A). Heterozygous (AG) genotypic effects increased earlywood microfibril angle 4.59°, considering all samples (4.4000 in AG and −0.1886 in GG for ring 5 referred to the overall average) and 5.76°, considering only the eastern range (5.7000 in AG and −0.0648 in GG for ring 5 referred to the overall average). α-*tubulin* SNP M10 explained ~2–4% of the total phenotypic variance in earlywood microfibril angle (*i.e.*, ~4–8% of the total genetic variance, considering $h^2 \sim 0.50$; ZOBEL and JETT 1995).

*Lignin and cellulose content:* Wood chemistry traits (*i.e.*, lignin and cellulose content) were highly (and inversely) correlated (~95% of correlation, judging by the Pearson correlation coefficient), showing similar trends in the association analyses. Five SNPs (*4cl*, *c3h-1*, *cesA3*, *ccr-1*, and *mt-like* genes) showed significant association ($P < 0.05$) with lignin and cellulose content (see supplemental Table S4E at http://www.genetics.org/supplemental/), but none remained significant after corrections for multiple testing.

## DISCUSSION

A candidate-gene-based association mapping strategy has been proposed as a promising approach to dissect complex traits in forest trees (NEALE and SAVOLAINEN

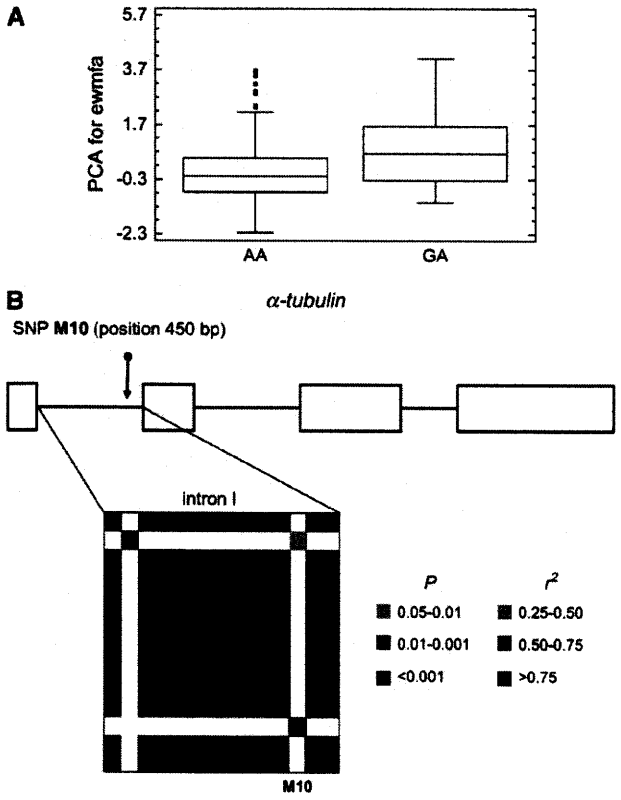


FIGURE 3.—(A) Genotypic effects (box plot) of *α-tubulin* SNP M10 on earlywood microfibril angle (ewmfa). (B) LD plot (only informative sites) for intron I, where SNP M10 is located, based on polymorphism data from BROWN *et al.* (2004). Exons are represented by boxes. SNP M10 position is referred to the beginning of the gene using AY832609 accession as reference (KRUTOVSKY and NEALE 2005).

2004; GONZÁLEZ-MARTÍNEZ *et al.* 2006a). The application of such a strategy in loblolly pine identified several SNPs from lignification and other wood- and drought-related genes that showed genetic association with an array of wood property traits and establishes the feasibility of a candidate gene approach in species with large genomes (30 billion bp in pine) and low linkage disequilibrium. This study also showed that the number of false positives due to a moderate level of relatedness in the association population (introduced in this case by second-generation selections) was low. This was illustrated by the similar results obtained in standard general linear models and MLMs that included a pairwise kinship matrix as a covariate (see supplemental Table S4 at http://www.genetics.org/supplemental/). One possible explanation for this result is that even when most second-generation selections had two to three half-sibs in the association population, the average pairwise kinship for single trees was 0.0131, much less than the average expectation for either half-sibs (0.1250) or full-sibs (0.2500).

Previous studies in loblolly pine showed several candidate genes to collocate with wood physical and chemical traits in linkage maps (BROWN *et al.* 2003; our unpublished results). This study was able to confirm some of the previously suggested causative relationships between QTL and collocated candidate genes. Namely, *4cl* collocated with a latewood percentage QTL verified by repeated detection and significant genetic association was found between *4cl* SNP M7 and this same trait. Two other genes (*sams-2* and *α-tubulin*) had significant genetic associations and mapped to regions where several QTL for wood property traits were detected. Finally, *cesA3* collocated with a tentative QTL for cell-wall chemistry, although the evidence for genetic association between *cesA3* and wood chemistry traits in this article was weak.

Although more QTL than significant genetic associations were found for wood property traits in loblolly pine, the amount of variation explained per QTL/SNP was similar. Indeed, single-SNPs explained <5% of the phenotypic variance for all traits measured, indicating genetic control by many loci with relatively small individual effects, a finding repeatedly reported in conifer QTL studies involving wood property traits (*e.g.*, BROWN *et al.* 2003; JERMSTAD *et al.* 2003; POT *et al.* 2006). The only other association study in forest trees, in Eucalyptus, found two SNPs, each of them explaining 4.6% of the total variance (THUMMA *et al.* 2005). However, for traits with multiple significant associations, the cumulative amount of phenotypic variance explained is appreciable (*e.g.*, ~20% in earlywood specific gravity, when considering uncorrected *P*-values), and the amount of additive genetic variance explained (in this case, ~40%, given $h^2 \sim 0.50$; ZOBEL and JETT 1995) makes molecular marker-assisted selection in tree breeding attractive. Still, the current lack of validation across different association populations and field-testing environments and the high rate of false positives typically seen in association studies (HIRSCHHORN *et al.* 2002; ARANZANA *et al.* 2005), suggest that these estimates should be considered with caution.

Some genetic associations were found repeatedly in different wood-age types and across different association models and have significant *Q*-values after correction for multiple testing using FDR. Thus, these are likely associations and not false positives. Because LD appears to decay rapidly in conifers studied to date (BROWN *et al.* 2004; RAFALSKI and MORGANTE 2004; GONZÁLEZ-MARTÍNEZ *et al.* 2006b), SNPs showing genetic association with wood property traits are likely located in close proximity to the causative polymorphisms or are the quantitative trait nucleotides themselves. A nonsynonymous substitution in exon I of *cad* (SNP M28) was in strong association with earlywood specific gravity. DNA sequence data showed full linkage of this SNP with another nonsynonymous substitution in exon IV (Figure 4A; our unpublished results), and partial linkage with the cad null allele causing lignin modifications in wild trees (see RALPH *et al.* 1997; GILL *et al.* 2003). *cad* is a key enzyme in the lignin biosynthesis

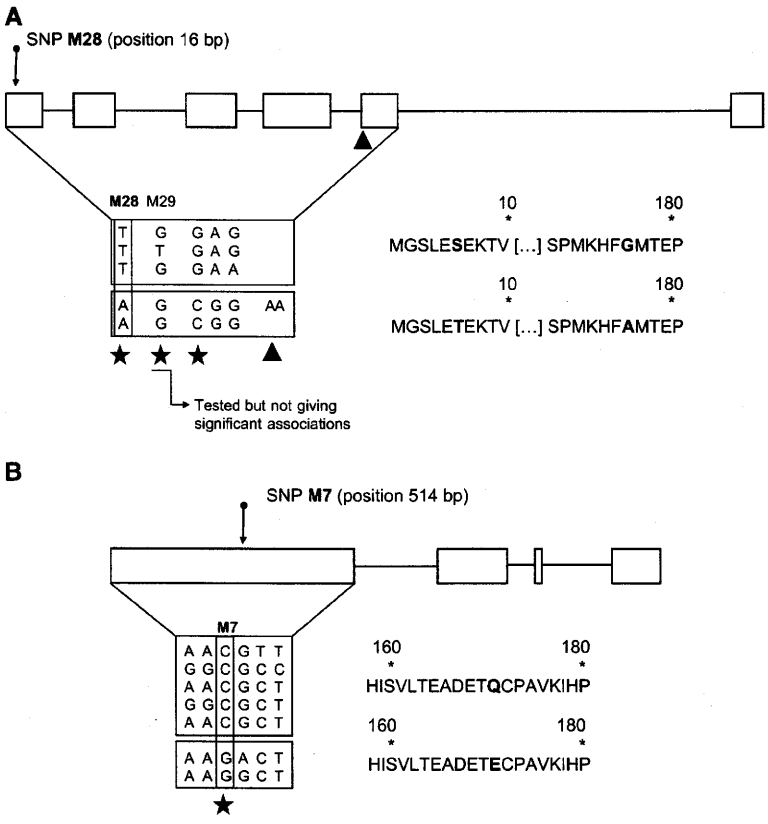


FIGURE 4.—Schematic of candidate genes with nonsynonymous mutations that showed genetic association with wood property traits. Exons are represented by boxes. Solid stars indicate nonsynonymous substitutions. Allelic variation at the DNA sequence (elaborated from BROWN *et al.* 2004) and protein levels is also shown. (A) *cad*: SNP M28 position is referred to the beginning of the gene using Z37991 and Z37992 accessions as references (MACKAY *et al.* 1995); the 2-base indel (GILL *et al.* 2003) that causes the *cad*-null allele described by RALPH *et al.* (1997) is indicated by a solid triangle. (B) *4cl*: SNP M7 position is referred to the beginning of the gene using U39405 accession as reference (ZHANG and CHIANG 1997).

pathway that catalyzes the final step in the synthesis of monolignols by converting cinnamaldehydes to cinnamyl alcohols. Downregulation of *cad* does not reduce lignin content in wood but notably affects lignin structure and composition (MACKAY *et al.* 2001; PILATE *et al.* 2002; VAN FRANKENHUYZEN and BEARDMORE 2004). Earlywood specific gravity is also affected by allelic variation in *sams-2*, a gene that is thought to be involved in methyl transfer during biosynthesis of monolignols during wood formation. This gene is also an intermediate in the synthesis of ethylene and induced under water-deficit conditions (CHANG *et al.* 1996).

Another reliable genetic association, albeit restricted to the east of the Mississippi Valley loblolly pine range, was found between *4cl* SNP M7 and percentage latewood. The SNP M7 is the only nonsynonymous mutation found in *4cl* exon I, resulting in the substitution of a glutamine by a glutamic acid at the protein level (Figure 4B). *4-coumarate CoA ligase* genes in plants are involved in several biosynthetic pathways, including the biosynthesis of flavonoids and monolignols (CUKOVIC *et al.* 2001; PETER and NEALE 2004). Transcript levels for this gene (and less notably for other lignin biosynthetic genes) have been found to be higher in latewood than in earlywood (EGERSTDOTTER *et al.* 2004). Reducing *4cl* expression in transgenic poplar resulted in a reduction of the amount of lignin in wood (up to 45%; HU *et al.* 1999; LI *et al.* 2003), although it was compensated by an increase in cellulose and did not affect cellular or whole-plant structural integrity (HU *et al.* 1999).

The strongest genetic association found in this study was between allelic variation in α-*tubulin* SNP M10 and earlywood microfibril angle. This SNP was also marginally associated with earlywood specific gravity variation (see supplemental Table S4A at http://www.genetics.org/supplemental/). The two traits were correlated ($r \sim 0.3$–$0.4$). The SNP M10 is found within the intron I of the gene and it is not in LD with any other polymorphism found in this region, otherwise in tight linkage (Figure 3B). Tubulins are often suggested as candidate genes for S2-layer microfibril angle in wood because the orientation of newly deposited cellulose microfibrils is thought to coalign with cortical microtubules (the alignment hypothesis; BASKIN 2001; BASKIN *et al.* 2004), a major component of the cell cytoskeleton that is made of heterodimers of α- and β-tubulins (NICK 2000; PILATE *et al.* 2004; YANG and LOOPSTRA 2005 and references therein). Tubulins belong to multigene families in plants. The gene member tested for association in this study was identical to the contig 8045 assembled from loblolly pine xylem EST libraries (available at http://biodata.ccgb.umn.edu/nsfpine/contig_dir20/), which did not show differences in expression between early and latewood or between south Arkansas and south Louisiana seed sources (YANG and LOOPSTRA 2005).

A significant genetic association between a polymorphism in a *ccr* gene and microfibril angle has been recently reported in Eucalyptus (THUMMA *et al.* 2005). Moreover, there is evidence of *ccr* variation in gene expression affecting lignin content and composition (see review in PETER and NEALE 2004), and suppressing *ccr* activity is considered a promising approach to reducing lignin content for pulping in poplar (BOERJAN *et al.* 1999). Genetic association between SNPs M46 and M48 in *ccr-1* with microfibril angle and lignin content in loblolly pine, although not statistically significant after correction for multiple testing using FDR, might reflect the effect of allelic variation of this key enzyme on wood property traits. However, our results on loblolly pine are not conclusive and cannot confirm the significant genetic association found in Eucalyptus.

Virtually all fitness and economic traits in forest trees appear to be under the control of many genes with small-to-modest effect as evidenced by decades of empirically based quantitative genetic studies (NAMKOONG 1979) and, more recently, dozens of QTL studies (see, for instance, BROWN *et al.* 2003; JERMSTAD *et al.* 2003; POT *et al.* 2006). This study reinforces the quantitative model and, perhaps more importantly, provides keen insight into the genetic basis of natural variation by revealing estimates of allelic affects on specific phenotypic traits. By evaluating relatively large natural populations in association studies, the effect of allelic substitutions against a diverse genetic background can be accurately estimated. This has both evolutionary and tree breeding implications. It is notable that QTL and association genetic studies provide similar estimates of the magnitude of effect for individual genetic factors. This reinforces the value of (i) QTL studies for revealing genomewide estimates of the number of genetic factors contributing to a trait and (ii) collocation studies to identify candidate genes. It also provides confidence in the power and accuracy of all these estimation procedures.

A relatively modest, but highly targeted, array of genes belonging to key pathways was evaluated in this first multigene association genetic study of forest trees. For several genes, the full-length coding region was used for SNP discovery but for others only an ~500- to 1200-bp fragment was available (see BROWN *et al.* 2004; GONZÁLEZ-MARTÍNEZ *et al.* 2006b for details). Given the rapid decay of within-gene linkage disequilibrium in conifers, it is probable that some genetic associations involving partially screened genes were not detected in this study; thus extension of the SNP discovery to the full-length DNA sequence, including promoter regions, is desirable. Ultimately it will be necessary to conduct association studies with virtually all the genes in the genome to have a complete understanding of the genetic architecture of a trait. This no longer seems the daunting challenge it may have been only a few years ago. Rapidly developing, cost-effective technologies for SNP discovery and large-scale genotyping make genomic-scale studies feasible. Furthermore, efforts to locate all the genes in some tree species are moving forward rapidly. For tree species such as loblolly pine, Monterey pine (*Pinus radiata* D. Don), and the genus Eucalyptus, deep EST-sequencing projects, in some cases coupled with BAC libraries, are providing excellent representation of the expressed genome, and in poplar (*Populus* spp.), a nearly complete genome sequence has been produced (TUSKAN *et al.* 2006). Efforts are currently underway in loblolly pine to discover SNPs in 10,000 genes, both structural and regulatory, to associate allelic variation with an array of adaptive traits, and to develop validation studies to confirm already detected associations.

The authors thank Gary F. Peter, Glenn Howe, and two anonymous reviewers for helpful commentary on the manuscript. Thanks are extended to Brian Penttila, Jennifer Roers, Garth R. Brown, Geoffrey P. Gill, Robert J. Kuntz, Julie Beal, Mark H. Wright, and Jill Wegrzyn for technical assistance. Santiago C. González-Martínez was supported by the Ramón y Cajal fellowship RC02-2941 (Ministerio de Educación y Ciencia, Spain). We are indebted to Weyerhaeuser for plant material, technical assistance, and general support over several years. This research was funded by the Initiative for Future Agriculture and Food Systems (U. S. Department of Agriculture).

## LITERATURE CITED

ARANZANA, M. J., S. KIM, K. ZHAO, E. BAKKER, M. HORTON *et al.*, 2005 Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. PLoS Genet. **1:** e60.

BASKIN, T. I., 2001 On the alignment of cellulose microfibrils by cortical microtubules: a review and a model. Protoplasma **215:** 150–171.

BASKIN, T. I., G. T. S. BEEMSTER, J. E. JUDY-MARCH and F. MARGA, 2004 Disorganization of cortical microtubules stimulates tangential expansion and reduces the uniformity of cellulose microfibril alignment among cells in the root of *Arabidopsis*. Plant Physiol. **135:** 2279–2290.

BOERJAN, W., 2005 Biotechnology and the domestication of forest trees. Curr. Opin. Biotechnol. **16:** 159–166.

BOERJAN, W., H. MEYERMANS, C. CHEN, J. CHRISTENSEN, J.-C. LEPLÉ *et al.*, 1999 Improved wood quality for the pulp and paper industry by genetic engineering of lignin biosynthesis. International Poplar Symposium II, September 13–17, 1999, Orleans, France.

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.

BROWN, G. R., G. P. GILL, R. J. KUNTZ, C. H. LANGLEY and D. B. NEALE, 2004 Nucleotide variation and linkage disequilibrium in loblolly pine. Proc. Natl. Acad. Sci. USA **101:** 15255–15260.

BURNS, D. A., and E. W. CIURCZAK (Editors), 1992 *Handbook of Near-Infrared Analysis.* Marcel Dekker, New York.

CARDON, L. R., and J. I. BELL, 2001 Association study designs for complex diseases. Nat. Rev. Genet. **2:** 91–99.

CASASOLI, M., J. DERORY, C. MORERA-DUTREY, O. BRENDEL, I. PORTH *et al.*, 2006 Comparison of quantitative trait loci for adaptive traits between oak and chestnut based on an expressed sequence tag consensus map. Genetics **172:** 533–546.

CHANG, S., J. D. PURYEAR, M. A. D. L. DIAS, E. A. FUNKHOUSER, R. J. NEWTON *et al.*, 1996 Gene expression under water deficit in loblolly pine (*Pinus taeda*): isolation and characterization of cDNA clones. Physiol. Plantarum **97:** 139–148.

CUKOVIC, D., J. EHLTING, J. A. VANZIFFLE and C. J. DOUGLAS, 2001 Structure and evolution of 4-coumarate: coenzyme A ligase (4CL) gene families. Biol. Chem. **382:** 645–654.

DAVIS, M., 1998 A rapid modified method for compositional carbohydrate analysis of lignocellulosics by high pH anion-exchange chromatography with pulsed amperometric detection (HPAEC/PAD). J. Wood Chem. Technol. **18:** 235–252.

EGERSTDOTTER, U., L. VAN ZYL, J. MACKAY, G. F. PETER, M. KIRST *et al.*, 2004 Gene expression during formation of earlywood and latewood in loblolly pine: expression profiles of 350 genes. Plant Biol. **6:** 654–663.

EL-OSTA, M. L. M., R. W. WELLWOOD and R. G. BUTTERS, 1972 An improved X-ray technique for measuring microfibril angle of coniferous wood. Wood Sci. **5:** 113–117.

FLINT-GARCÍA, S. A., J. M. THORNSBERRY and E. S. BUCKLER, IV, 2003 Structure of linkage disequilibrium in plants. Annu. Rev. Plant Biol. **54:** 357–374.

FREWEN, B. E., T. H. CHEN, G. T. HOWE, J. DAVIS, A. ROHDE *et al.*, 2000 Quantitative trait loci and candidate gene mapping of bud set and bud flush in Populus. Genetics **154:** 837–845.

GARBUTT, D. C. F., M. J. DONKIN and J. H. MEYER, 1992 Near infrared reflectance analysis of cellulose and lignin in wood. Pap. S. Afr. **12:** 45–48.

GILL, G. P., G. R. BROWN and D. B. NEALE, 2003 A sequence mutation in cinnamyl alcohol dehydrogenase gene associated with altered lignification in loblolly pine. Plant Biotechnol. J. **1:** 253–258.

GONZÁLEZ-MARTÍNEZ, S. C., K. V. KRUTOVSKY and D. B. NEALE, 2006a Forest tree population genomics and adaptive evolution. New Phytol. **170:** 227–238.

GONZÁLEZ-MARTÍNEZ, S. C., E. ERSOZ, G. R. BROWN, N. C. WHEELER and D. B. NEALE, 2006b DNA sequence variation and selection of tag single-nucleotide at candidate genes for drought-stress response in *Pinus taeda* L. Genetics **172:** 1915–1926.

GOUJON, T., R. SIBOUT, A. EUDES, J. MACKAY and L. JOUANIN, 2003 Genes involved in the biosynthesis of lignin precursors in *Arabidopsis thaliana*. Plant Physiol. Biochem. **41:** 677–687.

GROOVER, A., M. DEVEY, J. LEE, R. MEGRAW, T MITCHELL-OLDS *et al.*, 1994 Identification of quantitative trait loci influencing wood specific gravity in an outbred pedigree of loblolly pine. Genetics **138:** 1293–1300.

GUILLET-CLAUDE, C., C. BIROLLEAU-TOUCHARD, D. MANICACCI, P. M. ROGOWSKY, J. RIGAU *et al.*, 2004 Nucleotide diversity of the *ZmPox3* maize peroxidase gene: relationships between a MITE insertion in exon 2 and variation in forage maize digestibility. BMC Genet. **5:** 19.

HAGENBLAD, J., C. TANG, J. MOLITOR, J. WERNER, K. ZHAO *et al.*, 2004 Haplotype structure and phenotypic associations in the chromosomal regions surrounding two *Arabidopsis thaliana* flowering time loci. Genetics **168:** 1627–1638.

HARDY, O. J., and X. VEKEMANS, 2002 SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol. Ecol. Notes **2:** 618–620.

HAUSSMANN, B. I. G., H. K. PARZIES, T. PRESTERI, Z. SUŠIĆ and T. MIEDANER, 2004 Plant genetic resources in crop improvement. Plant Genet. Resour. **2:** 3–21.

HIRSCHHORN, J. N., and M. J. DALY, 2005 Genome-wide association studies for common diseases and complex traits. Nat. Rev. Genet. **6:** 95–108.

HIRSCHHORN, J. N., K. LOHMUELLER, E. BYRNE and K. HIRSCHHORN, 2002 A comprehensive review of genetic association studies. Genet. Med. **4:** 45–61.

HU, W. J., S. A. HARDING, J. LUNG, J. L. POPKO, J. RALPH *et al.*, 1999 Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. Nat. Biotechnol. **17:** 808–812.

HSU, T. M., X. CHEN, S. DUAN, R. D. MILLER and P. Y. KWOK, 2001 Universal SNP genotyping assay with fluorescence polarization detection. Biotechniques **31:** 560, 562, 564–568.

JANSON, J., 1970 Calculation of the polysaccharide composition of wood and pulp. Paperi ja Puu **5:** 323–329.

JERMSTAD, K. D., D. L. BASSONI, K. S. JECH, N. C. WHEELER and D. B. NEALE, 2001a Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir: I. Timing of vegetative bud flush. Theor. Appl. Genet. **102:** 1142–1151.

JERMSTAD, K. D., D. L. BASSONI, N. C. WHEELER, T. S. ANEKONDA, S. N. AITKEN *et al.*, 2001b Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir: II. Spring and fall cold-hardiness. Theor. Appl. Genet. **102:** 1152–1158.

JERMSTAD, K. D., D. L. BASSONI, K. S. JECH, G. A. RITCHIE, N. C. WHEELER *et al.*, 2003 Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas fir. III. Quantitative trait loci-by-environment interactions. Genetics **165:** 1489–1506.

KRUTOVSKY, K. V., and D. B. NEALE, 2005 Nucleotide diversity and linkage disequilibrium in cold-hardiness- and wood quality-related candidate genes in Douglas fir. Genetics **171:** 2029–2041.

LI, L., Y. ZHOU, X. CHENG, J. SUN, J. M. MARITA *et al.*, 2003 Combinatorial modification of multiple lignin traits in trees through multigene cotransformation. Proc. Natl. Acad. Sci. USA **100:** 4939–4944.

LOISELLE, B. A., V. L. SORK, J. NASON and C. GRAHAM, 1995 Spatial genetic structure of a tropical understorey shrub, *Psychotria officinalis* (Rubiaceae). Am. J. Bot. **82:** 1420–1425.

LONG, A. D., and C. H. LANGLEY, 1999 The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. Genome Res. **9:** 720–731.

MACKAY, J. J., W LIU, R. WHETTEN, R. R. SEDEROFF and D. M. O'MALLEY, 1995 Genetic analysis of cinnamyl alcohol dehydrogenase in loblolly pine: single gene inheritance, molecular characterization and evolution. Mol. Gen. Genet. **247:** 537–545.

MACKAY, J. J., D. M. O'MALLEY, T. PRESNELL, F. L. BOOKER, M. M. CAMPBELL *et al.*, 2001 Inheritance, gene expression, and lignin characterization in a mutant pine deficient in cinnamyl alcohol dehydrogenase. Proc. Natl. Acad. Sci. USA **94:** 8255–8260.

MARCHINI, J., L. R. CARDON, M. S. PHILLIPS and P. DONNELLY, 2004 The effects of human population structure on large genetic association studies. Nat. Genet. **36:** 512–517.

MEGRAW, R. A., G. LEAF and D. BREMMER, 1998 Longitudinal shrinkage and microfibril angle in loblolly pine, pp. 27–61 in *Microfibril Angle in Wood*, edited by B. A. BUTTERFIELD. University of Canterbury Press, Christchurch, New Zealand.

MEYLAN, B. A., 1967 Measurement of microfibril angle by X-ray diffraction. For. Prod. J. **17:** 51–58.

NAMKOONG, G., 1979 *Introduction to Quantitative Genetics in Forestry*. USDA Forest Service Technical Bulletin 1588, Washington, DC.

NEALE, D. B., and O. SAVOLAINEN, 2004 Association genetics of complex traits in conifers. Trends Plant Sci. **9:** 325–330.

NEI, M., 1978 Estimation of average heterozygosity and genetic distance for a small number of individuals. Genetics **89:** 583–590.

NICK, P., 2000 *Plant Microtubules*. Springer-Verlag, Berlin.

OLSEN, K. M., S. S. HALLDORSDOTTIR, J. R. STINCHCOMBE, C. WEINIG, J. SCHMITT *et al.*, 2004 Linkage disequilibrium mapping of Arabidopsis *CRY2* flowering time alleles. Genetics **167:** 1361–1369.

ÖSTERBERG, M. K., O. SHAVORSKAYA, M. LASCOUX and U. LAGERCRANTZ, 2002 Naturally occurring indel variation in the *Brassica nigra COL1* gene is associated with variation in flowering time. Genetics **161:** 299–306.

PETER, G., and D. B. NEALE, 2004 Molecular basis for the evolution of xylem lignification. Curr. Opin. Plant Biol. **7:** 737–742.

PILATE, G., E. GUINEY, K. HOLT, M. PETIT-CONIL, C. LAPIERRE *et al.*, 2002 Field and pulping performances of transgenic trees with altered lignification. Nat. Biotechnol. **20:** 607–612.

PILATE, G., A. DÉJARDIN, F. LAURANS and J.-C. LEPLÉ, 2004 Tension wood as a model for functional genomics of wood formation. New Phytol. **164:** 63–72.

PLOMION, C., G. LE-PROVOST and A. STOKES, 2001 Wood formation in trees. Plant Physiol. **127:** 1513–1523.

POT, D., L. MCMILLAN, C. ECHT, G. LE-PROVOST, P. GARNIER-GÉRÉ *et al.*, 2005 Nucleotide variation in genes involved in wood formation in two pine species. New Phytol. **167:** 101–112.

POT, D., J. C. RODRIGUES, P. ROZENBERG, G. CHANTRE, J. TIBBITS *et al.*, 2006 QTLs and candidate genes for wood properties in maritime pine (*Pinus pinaster* Ait.). Tree Genet. Genomes **2:** 10–24.

PRITCHARD, J. K., and W. WEN, 2004 *Documentation for Structure Software, Version 2*. Department of Human Genetics, University of Chicago, Chicago (http://pritch.bsd.uchicago.edu).

PRITCHARD, J. K., M. STEPHENS and P. DONNELLY, 2000 Inference of population structure using multilocus genotype data. Genetics **155:** 945–959.

RAFALSKI, A., and M. MORGANTE, 2004 Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. Trends Genet. **20:** 103–111.

RALPH, J., J. J. MACKAY, R. D. HATFIELD, D. M. O'MALLEY, R. W. WHETTEN *et al.*, 1997 Abnormal lignin in a loblolly pine mutant. Science **277:** 235–239.

REDDEN, D. T., and D. B. ALLISON, 2003 Nonreplication in genetic association studies of obesity and diabetes research. J. Nutr. **133:** 3323–3326.

ROSENBERG, N. A., J. K. PRITCHARD, J. L. WEBER, H. M. CANN, K. K. KIDD *et al.*, 2002 Genetic structure of human populations. Science **298:** 2381–2385.

SCHMIDTLING, R. C., 2001 *Southern Pine Seed Sources.* USDA, GTR SRS-44, Asheville, NC.

SEWELL, M. M., D. L. BASSONI, R. A. MEGRAW, N. C. WHEELER and D. B. NEALE, 2000 Identification of QTLs influencing wood property traits in loblolly pine (*Pinus taeda* L.). I. Physical wood properties. Theor. Appl. Genet. **101:** 1273–1281.

SEWELL, M. M., M. F. DAVIS, G. A. TUSKAN, N. C. WHEELER, C. C. ELAM *et al.*, 2002 Identification of QTLs influencing wood property traits in loblolly pine (*Pinus taeda* L.). II. Chemical wood properties. Theor. Appl. Genet. **104:** 214–222.

SHINDO, C., M. J. ARANZANA, C. LISTER, C. BAXTER, C. NICHOLLS *et al.*, 2005 Role of *FRIGIDA* and *FLOWERING LOCUS C* in determining variation in flowering time of *Arabidopsis.* Plant Physiol. **138:** 1163–1173.

STOREY, J. D., 2002 A direct approach to false discovery rates. J. R. Stat. Soc. Ser. B **64:** 479–498.

STOREY, J. D., and R. TIBSHIRANI, 2003 Statistical significance for genome-wide studies. Proc. Natl. Acad. Sci. USA **100:** 9440–9445.

SZALMA, S. J., E. S. BUCKLER, IV, M. E. SNOOK and M. D. MCMULLEN, 2005 Association analysis of candidate genes for maysin and chlorogenic acid accumulation in maize silks. Theor. Appl. Genet. **110:** 1324–1333.

TABOR, H. K., N. J. RISCH and R. M. MYERS, 2002 Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat. Rev. Genet. **3:** 1–7.

THORNSBERRY, J. M., M. M. GOODMAN, J. DOEBLEY, S. KRESOVICH, D. NIELSEN *et al.*, 2001 *Dwarf8* polymorphisms associate with variation in flowering time. Nat. Genet. **28:** 286–289.

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.

TUSKAN, G. A., S. DIFAZIO, S. JANSSON, J. BOHLMANN, I. GRIGORIEV *et al.*, 2006 The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science **313:** 1596–1604.

VAN FRANKENHUYZEN, K., and T. BEARDMORE, 2004 Current status and environmental impact of transgenic forest trees. Can. J. For. Res. **34:** 1163–1180.

WALLIS, A., 1996 Chemical analysis of polysaccharides in plantation eucalypt woods and pulp. Appita J. **49:** 258–262.

WEIGEL, D., and M. NORDBORG, 2005 Natural variation in *Arabidopsis.* How do we find the causal genes? Plant Physiol. **138:** 567–568.

WEIR, B. S., and C. C. COCKERHAM, 1984 Estimating F-statistics for the analysis of population structure. Evolution **38:** 1358–1370.

WHEELER, N. C., K. D. JERMSTAD, K. V. KRUTOVSKY, S. N. AITKEN, G. T. HOWE *et al.*, 2005 Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-Fir. IV. Cold-hardiness QTL verification and candidate gene mapping. Mol. Breed. **15:** 145–156.

WHETTEN, R., J. J. MACKAY and R. SEDEROFF, 1998 Recent advances in understanding lignin biosynthesis. Annu. Rev. Plant Physiol. Plant Mol. Biol. **49:** 585–609.

WILSON, L. M., S. R. WHITT, A. M. IBÁÑEZ, T. R. ROCHEFORD, M. M. GOODMAN *et al.*, 2004 Dissection of maize kernel composition and starch production by candidate gene association. Plant Cell **16:** 2719–2733.

YANG, S. H., and C. A. LOOPSTRA, 2005 Seasonal variation in gene expression for loblolly pine (*Pinus taeda*) from different geographical regions. Tree Physiol. **25:** 1063–1073.

YU, J., G. PRESSOIR, W. H. BRIGGS, I. VROH BI, M. YAMASAKI *et al.*, 2006 A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. **38:** 203–208.

ZHANG, X.-H., and V. L. CHIANG, 1997 Molecular cloning of *4-Coumarate:Coenzyme A Ligase* in loblollv pine and the roles of this enzyme in the biosynthesis of lignin in compression wood. Plant Physiol. **113:** 65–74.

ZOBEL, B. J., and J. B. JETT, 1995 *Genetics of Wood Production.* Springer-Verlag, New York.

ZOBEL, B. J., and J. P. VAN BUIJTENEN, 1989 *Wood Variation: Its Causes and Control.* Springer, Berlin/Heidelberg, Germany/New York.

Communicating editor: B. J. WALSH