### Association Genetics of Wood Physical Traits in the Conifer White Spruce and Relationships With Gene Expression

Jean Beaulieu,*† Trevor Doerksen,*† Brian Boyle,† Sébastien Clément,*, Marie Deslauriers,*, Stéphanie Beauseigle,† Sylvie Blais,† Pier-Luc Poulin,† Patrick Lenz,† Sébastien Caron,† Philippe Rigault,† Paul Bicho,§ Jean Bousquet† and John MacKay†

*Natural Resources Canada, Canadian Wood Fibre Centre, Québec City, Québec G1V 4C7, Canada, †Université Laval, Canada Research Chair in Forest and Environmental Genomics, Centre for Forest Research and Institute for Systems and Integrative Biology, Québec City, Québec G1V 0A6, Canada, ‡Gyde, Québec City, Québec G1T 1Z2, Canada and §FPInnovations, Vancouver, British Columbia V6S 2L9, Canada

Manuscript received December 9, 2010
Accepted for publication February 22, 2011

#### ABSTRACT

Marker-assisted selection holds promise for highly influencing tree breeding, especially for wood traits, by considerably reducing breeding cycles and increasing selection accuracy. In this study, we used a candidate gene approach to test for associations between 944 single-nucleotide polymorphism markers from 549 candidate genes and 25 wood quality traits in white spruce. A mixed-linear model approach, including a weak but nonsignificant population structure, was implemented for each marker–trait combination. Relatedness among individuals was controlled using a kinship matrix estimated either from the known half-sib structure or from the markers. Both additive and dominance effect models were tested. Between 8 and 21 single-nucleotide polymorphisms (SNPs) were found to be significantly associated ($P \leq 0.01$) with each of earlywood, latewood, or total wood traits. After controlling for multiple testing ($Q \leq 0.10$), 13 SNPs were still significant across as many genes belonging to different families, each accounting for between 3 and 5% of the phenotypic variance in 10 wood characters. Transcript accumulation was determined for genes containing SNPs associated with these traits. Significantly different transcript levels ($P \leq 0.05$) were found among the SNP genotypes of a 1-aminocyclopropane-1-carboxylate oxidase, a β-tonoplast intrinsic protein, and a long-chain acyl-CoA synthetase 9. These results should contribute toward the development of efficient marker-assisted selection in an economically important tree species.

---

**ACQUISITION** of a perennial woody growth habit has had an enormous impact on the evolution of plants and of life on land (Groover 2005). Trees have shaped many terrestrial ecosystems and provide wood and fiber used by diverse industries worldwide. The adaptive importance of wood (secondary xylem) is indicated by its remarkable diversity among species, its significant variation within species, and its developmental plasticity. Although many of the genes involved in wood growth and development are known (e.g., Whetten et al. 2001; Aspeborg et al. 2005; Pavy et al. 2008a), the genetic basis underlying the variability of wood structure is only partly understood.

Revealing the genetics of wood and growth traits in forest trees initially proceeded through mapping quantitative trait loci (QTL) (Groover et al. 1994; Bradshaw and Stettler 1995; Grattapaglia et al. 1996; Ploimion et al. 1996; Sewell et al. 2000; Wheeler et al. 2005; Ukrainetz et al. 2008). Major QTL were identified and reported to explain up to a total of ≥25% of the naturally occurring variation depending on the species and the trait (Kirst et al. 2004). One limitation with linkage mapping is its requirement for large families with known relatedness, i.e., full-sib families in plants or extended pedigrees in humans (Myles et al. 2009). Furthermore, QTL do not replicate well across populations of different genetic background and across environments (Mackay 2001; Pelgas et al. 2011; Ritland et al. 2011).

In contrast to QTL mapping, genetic association mapping [or linkage disequilibrium (LD) mapping] can readily be applied to natural or breeding populations to identify marker–trait associations that remain linked following a large number of recombinations over a population’s history. LD mapping has led to research advances in human, animal, and plant genomics and is increasingly being adopted as the method of choice to dissect quantitative traits (Hirschhorn et al. 2002; Cheung et al. 2005; Zhu et al. 2008; Goddard and Hayes 2009; Grattapaglia et al. 2009; Rafalski 2010; Neale and Krémer 2011). By utilizing past recombination events, association mapping increases the
resolution of marker–trait associations. It is advantageous for species with long generation times because it may capture a broader sampling of allelic variation in a single study. Its application in undomesticated species like forest trees also holds potential to enhance and accelerate genetic resource management activities, including gene conservation and marker-assisted selection (MAS) for quantitative traits (Hauussmann et al. 2004).

Two approaches can be taken in association mapping to identify marker–trait associations. Randomly chosen markers may be distributed evenly across the genome to capture associations through LD with the genes controlling the traits. The alternative is to select a subset of candidate genes that are hypothesized to contribute to the quantitative variation on the basis of prior knowledge. Because conifers have immense genomes (Murray 1998) and low LD within gene limits (Neale and Savolainen 2004), genome-wide coverage of randomly selected markers is not practical. On the other hand, the rapid accumulation of cDNA sequences and expression data in conifers (e.g., Pavy et al. 2008a) enable large-scale candidate gene approaches.

Association studies of wood traits in a handful of tree species suggest the approach is promising but results have been species dependent. In Eucalyptus nitens (H. Deane & Maiden) Maiden, allelic variation in a cinnamoyl CoA reductase gene explained a small but significant proportion of the variation in the cellulose microfibril angle (Thumma et al. 2005), and a variant of a cobrak-like gene was linked to cellulose content and pulp yield (Thumma et al. 2009). In Pinus taeda L., four cell wall genes, i.e., Sadenosyl methionine synthase 2, cinnamyl alcohol dehydrogenase, the water-stress inducible protein *lp3-4*, and α-tubulin, explained ~3% of the variation in wood specific gravity, in the proportion of latewood to earlywood and in cellulose content and pulp yield (Thumma et al. 2009).

Association population: The discovery population consisted of 492 30-year-old trees representing 165 open-pollinated families from 40 provenances sampled in Quebec, Canada. For each family, 3 trees were selected to cover as much as possible of the existing range of diameters at breast height (dbh) observed in the family. Needle tissue as well as 12-mm increment cores taken at breast height were collected, stored on ice, and transported to the Canadian Forest Service facilities where they were stored in a freezer at −10°C until further treatment.

Phenotypic data: Pith to bark profiles of wood physical attributes were obtained using the SilviScan technology (Evans and Ilic 2001) at FPInnovations (Vancouver, BC, Canada). SilviScan is designed to rapidly measure wood attributes using a combination of X-ray densitometry, X-ray diffractometry, and image analysis. Measurements were taken from the radial surface of 2 × 7-mm wood chip samples prepared from the 12-mm cores. The thawed samples were acetone extracted and conditioned at 40% relative humidity. Air-dry (20°C) density was measured in 25-μm steps using X-ray densitometry. Microfibril angle (MFA) was measured in 1-mm steps using X-ray diffractometry, and wood modulus of elasticity (MOE) was calculated from the densitometry and diffractometry data. Cell wall thickness as well as radial and tangential cell dimensions (~1.5 μm) were obtained using optical microscopy and image analysis. From the raw data, early-, late-, and total wood averages were weighted by calculated annual ring area for each of the attributes (Table 1).

Wood traits (supporting information, File S1) were submitted to principal component analysis (PCA) to create a small number of orthogonal composite wood characters that could explain a large proportion of the variation observed in the raw characters (Legendre and Legendre 1998). PCA was carried out using PROC PRINCOMP (SAS Institute 2008) with standardized data. Three components were found to be significant (Frontier 1976), and altogether they explained 83% of the total variation. The first principal component was mainly influenced by wood stiffness, density, and cell wall...
and by 35 cycles of a 30-sec denaturation step at 94°C for 4 min at 94°C. PCR reactions were performed for each primer, and 0.04 unit Platinum Taq DNA polymerase (Invitrogen, Carlsbad, CA). PCR reactions were performed in volumes of 30 μL containing 10–20 ng template DNA, 20 mM Tris-HCL (pH 8.4), 1.5–2.5 mM MgCl2, 220 μM each dNTP, 0.25 μM each primer, and 0.04 unit Platinum Taq DNA polymerase (Invitrogen, Carlsbad, CA). PCR reactions were performed on a Peltier Thermal Cycler PTC200 (DNA Engine, DYAD; MJ Research, Waltham, MA) with an initial denaturation of 4 min at 94°C, followed by 12 cycles of touchdown PCR (a 30-sec denaturation step at 94°C, an annealing step with temperature decreasing from 60°C to 62°C, and a 3-min extension step at 72°C) and by 35 cycles of a 30-sec denaturation step at 94°C, then a 30-sec annealing step at 58°C, and a 3-min extension at 72°C, with a final extension of 10 min at 72°C.

Single-locus products were sequenced for both DNA strands using BigDye Terminators v3.0 cycle sequencing reaction kits (Applied Biosystems, Foster City, CA) and an automated ABI Prism 3700 Genetic Sequencer (Applied Biosystems). DNA sequence fragments were analyzed with SEQUENcher software (GENEVAL, Saint Albans, VT). SNPs were detected from the polymorphic positions indicated by double peaks or frame-shifts in sequence chromatograms of the DNA pool (Pelgas et al. 2004). For each SNP detected, haploid DNA sequences from individual megagametophyte tissue were used to eliminate any heterozygous nucleotide position due to paralogy (Pelgas et al. 2006; Pavv et al. 2008b). A total of 809 of the SNPs discovered using genomic sequences were used to construct part of a 1536-SNP GoldenGate genotyping array (Illiminia, San Diego).

**TABLE 1**

Eight SilviScan wood physical attributes measured for earlywood (EW), latewood (LW), and total wood (TW) as well as the percentage of earlywood in white spruce

| Abbreviation | Trait definition |
|--------------|-----------------|
| DEN          | Density at 8% moisture content |
| MFA          | Microfibril angle     |
| MOE          | Modulus of elasticity |
| CRS          | Tracheid coarseness estimate |
| CWT          | Tracheid cell wall thickness estimate |
| RCD          | Tracheid cell diameter in radial direction |
| TCD          | Tracheid cell diameter in tangential direction |
| ARW          | Average ring width |
| PCT          | Percentage of earlywood |

thickness. The second one was composed of coarseness and cell diameters in both the tangential and the radial directions. Microfibril angle was the trait having the largest influence on the third one.

**Selection of candidate genes:** White spruce candidate gene sequences were identified from >16,500 white spruce consensus sequences that were assembled from 50,000 partially sequenced cDNA clones produced by the Arborea project (Pavv et al. 2005a). A total of 747 candidate genes were identified for SNP discovery and classified on the basis of expression data or information available for each gene. The expression data included an analysis of vascular tissues and needles in *P. glauca* with a custom cDNA array based on 9503 gene clusters (Pavv et al. 2008a) and three such studies of secondary xylem carried out in pine species (Plomin et al. 2000; Pavv et al. 2005b; Pavv 2006). The functional annotations considered white spruce transcription factors, as well as sequences known to be involved in cell wall and lignin synthesis, as well as sequences reported to be potentially involved in wood formation.

**SNP discovery by resequencing genomic DNA:** PCR primers for amplification and resequencing were designed using Primer3 software (Rozen and Skaletsky 2000). Whenever possible, one of the primers was anchored outside of the coding regions to ensure specificity of amplification (Pavv et al. 2008b). Primer pairs were typically designed upon one or two regions per gene.

Genomic DNA was extracted from 100 mg of fresh foliage of each of 25 individuals with the DNeasy Plant Mini Kit (QIAGEN, Mississauga, ON, Canada). DNA concentrations were assessed by GeneSpec spectrophotometer (MiraiBio, Alameda, CA). For the detection of polymorphism, DNA samples were pooled in equal amounts (4 ng of DNA per individual) before polymerase chain reaction (PCR) amplification following Pelgas et al. (2004). Genomic DNA was also extracted from ground haploid megagametophyte tissue to detect paralogs in SNP discovery (Pelgas et al. 2006; Pavv et al. 2008b). PCR amplifications were performed in volumes of 30 μL containing 10–20 ng template DNA, 20 mM Tris-HCL (pH 8.4), 50 ng of DNA template. SNP discovery was performed using BigDye Terminators v3.0 cycle sequencing reaction kits (Applied Biosystems, Foster City, CA) and an automated ABI Prism 3700 Genetic Sequencer (Applied Biosystems). DNA sequence fragments were analyzed with SEQUENcher software (GENEVAL, Saint Albans, VT). SNPs were detected from the polymorphic positions indicated by double peaks or frame-shifts in sequence chromatograms of the DNA pool (Pelgas et al. 2004). For each SNP detected, haploid DNA sequences from individual megagametophyte tissue were used to eliminate any heterozygous nucleotide position due to paralogy (Pelgas et al. 2006; Pavv et al. 2008b). A total of 809 of the SNPs discovered using genomic sequences were used to construct part of a 1536-SNP GoldenGate genotyping array (Illiminia, San Diego).

**In silico SNP discovery in expressed sequence tags:** In silico SNPs were detected using expressed sequence tags (ESTs) obtained from Sanger sequencing of 17 white spruce cDNA libraries (Pavv et al. 2005a, 2006). Essentially, the search for SNPs was made using ~6500 contigs derived from at least two sequenced clones. SNP prediction was made using a proprietary method developed by Gydle (Quebec City, QC, Canada) or PolyBayes version 3.0 (Marty et al. 1999), with a prior of 0.01 and a probability that a position represents one SNP ($P_{SNP}$) $\geq 0.99$ (Pavv et al. 2006a). A total of 727 SNPs among the in silico SNPs were identified and added to the 809 already discovered to complete the panel of 1536 SNPs.

**SNP genotyping:** The minimum distance between any pair of SNPs selected for genotyping was 200 bp, whether considering genomic DNA or cDNA sequences. This distance was based on the information on low LD in Norway spruce, black spruce, and white spruce genes (Hurekhtz et al. 2006; Namroud et al. 2010). About 20% of the SNPs that had been identified by resequencing were retained for genotyping.

Genotyping was conducted at the Genome Quebec Innovation Centre (McGill University, Montreal), using the highly multiplexed Illumina GoldenGate assay (Shen et al. 2005) with 250 ng of template DNA (at a rate of 50 ng/μl) for each of the 492 trees. Positive and negative controls were added to each 96-well sample plate. The intensity data for each SNP were normalized and assigned a cluster position (and resulting genotype) with the BeadStudio software (Illiminia), and a quality score for each genotype was generated. A GenCall score cutoff of 0.25 was used to determine valid SNP genotypes and each retained SNP was required to have a minimum GenTrain score of 0.25 (Fan et al. 2003; Pavv et al. 2008b). GenTrain scores measure the reliability of SNP detection on the basis of the distribution of data points for each genotypic class.

Of the 1536 SNPs arrayed, 1021 (66.5%) were found to be polymorphic with a minimum GenTrain score of 0.25, which is similar to the success rate observed with a previous array in the same species (Pavv et al. 2008b). On the basis of positive controls, the repeatability of the assay for these SNPs was 99.98%. By selecting 472 trees that could be genotyped efficiently for at least 90% of their SNPs, the number of orthogonal valid SNPs was reduced to 969. Moreover, rare SNPs as well as SNPs showing abnormal distributions in this population subset were excluded, such as those with very low (<1%) heterozygosity. Hence, the final data set consisted of 472 trees genotyped for 944 high-quality SNPs representing 549 different genes. Three of these 472 individuals could not be used in the association study because their wood traits could not be assessed (File S2).

**Linkage disequilibrium:** Unphased linkage disequilibrium between pairs of SNPs within each candidate gene was evaluated.
matrix with Nason's kinship coefficients (LOISELLE et al. 1995) to check whether the SNPs found to be significantly associated with wood traits using the half-sib structure of relatedness would still be significant when using Nason's kinship coefficients (File S4). Estimates were obtained using the 944 SNPs for the 469 trees with the software SPADE, version 1.2 (HARVY and VEKEMANS 2002). As suggested by YU et al. (2006), negative kinship values were set to zero and a 469-square matrix was fit using the SAS MIXED procedure (LITTELL et al. 2006; SAS INSTITUTE 2008) in the following “$K$” (per notation in YU et al. 2006) MLM,

$$y = S \alpha + Q v + Z u + e,$$  \hspace{1cm} (1)

where $y$ is a vector of preadjusted wood quality measurements, $\alpha$ is a vector of fixed SNP effects, $u$ is the vector of random breeding values or “background” polygenic effects, $e$ is the vector of random residual effects, and $S$, $Q$, and $Z$ are incidence matrices of 0’s and 1’s relating records ($y$) to fixed ($\alpha$ and $v$) and random ($u$) effects. Expectation of the model is $E(y) = E(\alpha) = 0$ and thus $E(y) = S \alpha + Q v$, where random additive and residual variances are assumed to be $\text{Var}(u) = G = 2K$ and $\text{Var}(e) = R = I$), but are estimated with REML and $K$ and $I$ equal to kinship and identity coefficient matrices, respectively.

Failure to appropriately adjust for multiple testing may produce excessive false positives or overlook true positive signals in association studies when using large numbers of SNP markers. To correct for multiple testing (944 analyses per wood trait), the positive false discovery rate (FDR) method (STOREY AND TibSHIRANI 2003) was used to identify significant SNPs after correcting for multiple testing using the QVALUE software, version 1.0 (STOREY 2002; STOREY et al. 2004). Normality of studentized residuals was checked using the Kolmogorov $D$ and the Shapiro–Wilk $W$ statistics obtained with the “UNIVARIATE” procedure in SAS (SAS INSTITUTE 2008). Phenotypic data were not transformed as they were normally, or approximately normally, distributed.

**Microarray transcript profiling:** Given current limitations and costs of high-throughput genotyping technologies, a candidate gene approach should represent a suitable strategy for association studies with complex traits in species harboring large genomes and little LD. One criterion to consider for selecting genes is the preferential accumulation of transcripts in tissues directly linked with the traits of interest but this had not been directly tested in forest trees. This report aimed to test whether the 549 candidate genes tested in the association study were preferentially expressed in wood tissues and whether the candidate genes harboring significant SNPs were more likely to show a tissue-specific expression pattern than genes not harboring significant SNPs. A transcript profiling experiment compared secondary xylem (Xy), secondary phloem (Ph), and needles (Ne) using a large-scale custom microarray that included the 549 candidate genes.

**Test statistics for association:** Single-marker–based tests were performed for each of the wood traits assessed for earlywood, latewood, and total wood, as well as for the three principal components retained. Moreover, two genetic models were tested for each trait, i.e., the additive (three genotype classes) and dominant (two genotype classes) effects models. A mixed linear model (MLM) was fit for each pair of phenotypic traits and SNP markers (Yu et al. 2006; MALOSETTI et al. 2007). This model accounts for relatedness among individuals through a pairwise kinship matrix estimated from marker genotypes. As kinship coefficients do not always reflect with accuracy the known relatedness among individuals, such as the half-sib structure present in our discovery population, we first built a matrix with known half-sib relationships and used it in a first series of analyses (File S3). We also built a relatedness
BIOCONDUCTOR (http://www.bioconductor.org) in the R statistical environment and following the method described in Pavy et al. (2008). Briefly, data were normalized and the fold changes in spot intensity were estimated by fitting a linear model, using the limma package (release 2.0.7, http://www.bioconductor.org/packages/bioc/1.8/vignettes/lmma/inst/doc/usersguide.pdf). Contrasts between the three types of tissues were estimated for each of the 549 genes, and significance of the differential expression was tested with a moderated t-statistic with P-values adjusted for multiple testing with the method of Benjamini and Hochberg (1995). The candidate genes were classified according to their expression profiles and placed into 1 of 10 tissue preference classes (see Figure 1). A likelihood-ratio test (Sokal and Rohlf 1981) allowed testing for the independence of the tissue preference class distributions between the genes that carried SNPs that were associated with each of the wood traits at \( \alpha \leq 0.05 \) before correction for multiple testing, as well as the genes that had nonsignificant SNPs.

Transcript analysis of SNP genotypes: RNA transcript levels were determined for genes underlying the SNPs significantly associated with wood traits to test whether RNA transcript accumulation varied between the different genotypic classes of the SNPs. Because of the large number of determinations, only genes with significant SNPs after correction for multiple testing were targeted (see RESULTS). Transcript levels were determined by RT-qPCR with gene-specific primers. The latter could not be designed for either of the Tubulin genes because of the high level of sequence similarity among family members. For each SNP genotypic class, 8–25 trees were individually sampled by removing three small plugs of \( \sim 1 \text{ cm}^2 \) each of actively growing secondary xylem tissue at 1.3 m from the ground; these represented the entire annual growth produced near the end of the earlywood phase. Tissue handling and RNA extractions were performed as described for white spruce (Bedon et al. 2007); complementary DNA synthesis was prepared from 500 ng of total RNA (Boyle et al. 2009). PCR mixtures were assembled using LightCycler 480 SYBR Green I Master (Roche, Basel, Switzerland) with an epMotion 5075 pipetting robot (Eppendorf, Hamburg, Germany) and amplifications used a LightCycler 480 (Roche) with conditions and melting curve analyses as described (Boyle et al. 2009). The number of RNA transcript molecules was determined using the LRE methodology (Rutledge and Stewart 2008) adapted for Excel (Boyle et al. 2009). Normalization was performed using GeNorm (Vandesompele et al. 2002) and the reference genes ef1a (BT102965), cdc2 (BT106071), and rpl3A (BT115036). The measured number of transcript molecules was transformed to a log scale, and differential expression across three or two genotypic classes was tested by ANOVA or \( t \)-tests, respectively.

RESULTS AND DISCUSSION

Linkage disequilibrium: The pattern of the squared allelic correlation coefficient (\( r^2 \)) with base pair distance within candidate genes (Figure 2) illustrated the rapid LD decay in our white spruce population, with \( r^2 \) values dropping to \(<0.25\) within \(<100\) bp. These results are congruent with those of Namroud et al. (2010) from complete or nearly complete gene sequences in the same species.

Population structure: A weak but nonsignificant population structure with the presence of two subpopulations was detected using the model-based clustering method, as the patterns obtained were as expected for structured populations (Pritchard et al. 2007); i.e., the log probability of \( Pr(X|K) \) values reached a maximum for \( K = 2 \) (Figures 3 and 4). The presence of a weak population structure or even its absence in white spruce at the regional level was expected on the basis of results of studies using various classes of neutral genetic
markers (Alden and Loopstra 1987; Furnier and Stine 1995; Jaramillo-Correa et al. 2001; Namroud et al. 2008, 2010). Despite the nonsignificance of this weak population structure, it was nevertheless considered in the mixed model analyses to minimize the possibility of any false positive association due to the weak population structure signal.

**Genetic association:** Results of single-marker associations with each of the 25 wood traits as well as with each of the significant principal components are presented in File S5 and File S6. For each of the three principal components, we found between 11 and 15 significant associations (P ≤ 0.01) with SNPs when using the additive effects model and between 10 and 14 SNPs for the dominant effects model. However, after correction for multiple testing with a significance level of Q ≤ 0.10, none of these associations were still significant. For the individual wood traits, a total of 23,600 (944 SNPs × 25 traits) and 22,400 (896 SNPs × 25 traits) association tests were performed for the additive and dominant effects models, respectively. Globally, 6.4% and 5.8% of these tests were significant at a level of P ≤ 0.05, respectively, and 405 (43%) SNPs and 425 (47%) SNPs were found associated with at least one of the wood traits. These proportions are of the same order of magnitude as those found for cold hardiness-related traits in Douglas fir [Pseudotsuga menziesii var. menziesii (Mirb.) Franco], for which 7.8% of the association tests performed were found to be significant at this nominal threshold level.

After correction for multiple testing, the number of significant associations was reduced to 25, with 8, 4, and 3 significant associations for early-, late-, and total wood traits, respectively, with an additive effects model, and with 4, 4, and 2 associations, respectively, with a dominant effects model (Table 2). Some SNPs were found to be significantly associated with more than one trait whereas others were positive with both the additive and the dominant effects models. Thus, globally, 13 different SNPs (discovery rate of 1.4%) harbored by as many genes (2.4% of the sampled gene set) were found to be significantly associated with 10 different characters. Among these 13 SNPs, 5 represented synonymous substitutions, 3 were nonsynonymous, and others were located in untranslated regions. Silent SNPs...
| Trait | SNP | Gene ID | Functional annotation | Population structure | Kinship structure | Half-sib structure |
|-------|-----|---------|-----------------------|----------------------|------------------|-------------------|
|       |     |         |                       | F  | P  | F  | P  | Q   | R²d | No. homo1 | hetero | homo2 | No. homo1 | hetero | homo2 | GenBank accession |
| Earlywood ARW | PGGD1-1094 | GQ03806_H09 | LACS9 | 0.00 | 0.9990 | 17.54 | <0.0001 | 16.87 | 5.2 × 10⁻³ | 0.0487 | 0.035 | 123 | 546 | 2.50 | 2.77 | BT116774 |
|        | PGGD1-1396 | GQ02829_R04 | Glycoside hydroxylase family 28 | 0.12 | 0.7263 | 16.25 | <0.0001 | 17.88 | 4.0 × 10⁻³ | 0.0570 | 0.034 | 39 | 430 | 1.86 | 1.77 | BT105847 |
|        | PGGD1-1096 | GQ04010_H15 | Pectinesterase | 0.29 | 0.5875 | 12.20 | <0.0001 | 11.36 | 0.0008 | 0.0932 | 0.026 | 324 | 113 | 1.79 | 1.75 | BT118812 |
|        | PGGD1-1282 | GQ09113_N22 | MB14 | 0.01 | 0.9423 | 9.05 | 0.0001 | 8.83 | 0.0002 | 0.0437 | 0.058 | 64 | 232 | 1.72 | 1.78 | BT107883 |
|        | PGGD1-0511 | GQ0224_E23 | Galactosyltransferase | 2.85 | 0.0930 | 8.08 | 0.0004 | 8.29 | 0.0003 | 0.0561 | 0.037 | 40 | 178 | 247 | 81.0 | 82.4 | BT102859 |
| Latewood ARW | PGGD1-0863 | GQ09005_C12 | Tubulin, TUA2 | 2.20 | 0.1389 | 7.91 | 0.0004 | 8.17 | 0.0005 | 0.0875 | 0.059 | 4 | 55 | 406 | 0.86 | 0.62 | BT105658 |
|        | PGGD1-1313 | GQ07221_001 | Receptor-like kinase | 2.09 | 0.1492 | 9.23 | 0.0001 | 9.34 | 0.0001 | 0.0437 | 0.045 | 1 | 49 | 419 | 1.19 | 0.60 | BT109754 |
|        | PGGD1-0560 | GQ03010_F13 | ACC oxidase | 2.84 | 0.0928 | 10.01 | <0.0001 | 9.01 | 0.0001 | 0.0437 | 0.043 | 1 | 49 | 419 | 1.19 | 0.60 | BT109754 |
|        | PGGD1-1035 | GQ0172_O22 | b-Expansin | 1.59 | 0.2076 | 10.52 | <0.0001 | 10.55 | 5.1 × 10⁻³ | 0.0279 | 0.050 | 1 | 151 | 308 | 0.62 | 0.82 | BT106780 |
| Total wood ARW | PGGD1-1094 | GQ03806_H09 | Acyl lipid metabolism | 0.15 | 0.6965 | 17.79 | <0.0001 | 16.67 | 5.2 × 10⁻³ | 0.0487 | 0.035 | 123 | 146 | 3.08 | 3.38 | BT116774 |
|        | PGGD1-1396 | GQ02829_R04 | Glycoside hydroxylase family 28 | 0.38 | 0.5865 | 15.58 | 4.3 × 10⁻³ | 17.04 | 4.3 × 10⁻³ | 0.0590 | 0.035 | 39 | 430 | 1.86 | 1.77 | BT105847 |
|        | PGGD1-1096 | GQ04010_H15 | Pectinesterase | 0.60 | 0.4379 | 13.25 | 0.0005 | 12.76 | 0.0004 | 0.0914 | 0.029 | 324 | 113 | 1.95 | 1.90 | BT118812 |

(continued)
| Trait | SNP  | Gene ID | Functional annotation | Population structure | Kinship structure | Half-sib structure |
|-------|------|---------|-----------------------|----------------------|------------------|-------------------|
|       |      |         |                       | F  | P   | F  | P   | F  | P   | Q  | R² | No. homo1 | No. homo2 | Lsmean hetero | Lsmean homo2 | GenBank accession |
|       |      |         |                       |    |     |    |     |    |     |    |    |      |        |                |                |                      |
| Earlywood |     |         |                       |    |     |    |     |    |     |    |    |      |        |                |                |                      |
| CWT   | PGWD-1282 | GQ0113_N22 | MYB4 | 0.00 | 0.9859 | 0.92 | 0.3390 | 17.48 | <0.0001 | 17.01 | 4.2 x 10⁻³ | 0.0374 | 0.036 | 64   | 999 | 255.0 | 64   | 999 | 255.0 | 64   | 999 | 255.0 | BT107883 |
| CRS   | PGWD-1282 | GQ0113_N22 | MYB4 | 0.92 | 0.3390 | 17.28 | <0.0001 | 16.07 | 6.2 x 10⁻³ | 0.0378 | 0.037 | 64   | 999 | 255.0 | 64   | 999 | 255.0 | 64   | 999 | 255.0 | BT107883 |
| FCT   | PGWD-0354 | GQ0100_P08 | Tubulin, TUB3 | 0.59 | 0.4441 | 0.99 | 0.3215 | 14.48 | 0.0002 | 14.46 | 0.0002 | 0.0234 | 0.023 | 41   | 427 | 85.8 | 81.5 | BT107883 |
|       | PGWD-1035 | GQ0172_O22 | β-Expansin | 0.99 | 0.3215 | 14.48 | 0.0002 | 14.46 | 0.0002 | 0.0234 | 0.023 | 41   | 427 | 85.8 | 81.5 | BT107883 |
| Latewood |     |         |                       |    |     |    |     |    |     |    |    |      |        |                |                |                      |
| ARW   | PGWD-0863 | GQ0109_C12 | Tubulin TUA2 | 2.16 | 0.1427 | 14.01 | 0.0002 | 14.18 | 0.0002 | 0.0355 | 0.035 | 4    | 464  | 0.86  | 0.59  | BT106568 |
|       | PGWD-0354 | GQ0100_P08 | Tubulin TUB3 | 2.01 | 0.1368 | 11.61 | 0.0007 | 11.71 | 0.0007 | 0.0417 | 0.040 | 2    | 466  | 0.96  | 0.60  | BT108845 |
| MFA   | PGWD-0107 | GQ0133_K12 | Glycosyl hydrolase family 10 | 0.96 | 0.3279 | 14.14 | 0.0002 | 13.92 | 0.0002 | 0.0464 | 0.037 | 3    | 466  | 27.5  | 19.0  | BT101933 |
|       | PGWD-1070 | GQ0298_P24 | β-TIP | 2.93 | 0.0875 | 15.54 | <0.0001 | 16.10 | 7.5 x 10⁻³ | 0.0656 | 0.040 | 326  | 138  | 25.0  | 25.6  | BT101219 |
| RCD   | Total wood |         |                       |    |     |    |     |    |     |    |    |      |        |                |                |                      |
| CWT   | PGWD-1282 | GQ0113_N22 | MYB4 | 0.05 | 0.8584 | 14.16 | 0.0002 | 15.76 | 0.0002 | 0.0871 | 0.050 | 64   | 999  | 1.88  | 1.95  | BT107883 |
| CRS   | PGWD-1282 | GQ0113_N22 | MYB4 | 1.05 | 0.3054 | 15.41 | <0.0001 | 15.09 | 0.0001 | 0.0742 | 0.035 | 64   | 999  | 269.9 | 279.4 | BT107883 |

*Sec Table 1 for definition of abbreviations.  
*aNason’s kinship coefficients (SPAGeDi) are used to take account of the relatedness of the trees making up the discovery population.  
*cHalf-sib coefficients are used to take account of the known relatedness of the trees making up the discovery population.  
*dR² is the coefficient of determination estimated by GLM analysis.
**TABLE 3**

Tissue and organ preferential transcript accumulation of gene sequences carrying the gene SNPs significantly associated with wood traits

| Gene sequence ID   | Functional annotation     | Trait | Needle vs. xylem | Phloem vs. xylem | Needle vs. phloem | Tissue profile |
|--------------------|----------------------------|-------|------------------|------------------|-------------------|---------------|
|                    |                            |       | FD    | P       | FD    | P       | FD    | P       |                     |
| GQ0172_O22        | β-Expansin                 | PCT   | -6.90 | 1.88E-16 | -8.82 | 2.15E-17 | 1.28  | 0.056  | Xy > Ne and Ph     |
| GQ03005_C12      | Tubulin A2                 | ARW   | -4.96 | 4.96E-11 | -2.20 | 4.71E-05 | -2.26 | 2.37E-05 | Xy > Ne and Ph     |
| GQ03006_P17      | Unknown                    | ARW   | -4.15 | 1.02E-13 | -5.56 | 1.98E-15 | 1.34  | 0.016  | Xy > Ne and Ph     |
| GQ04010_P08      | Tubulin 3B                 | PCT, ARW | -2.22 | 1.57E-11 | -1.88 | 7.11E-09 | -1.18 | 0.047  | Xy > Ne and Ph     |
| GQ03010_F13      | ACC oxidase                | PCT, ARW | -1.81 | 1.94E-11 | -1.23 | 0.002  | -1.47 | 2.25E-07 | Xy > Ne and Ph     |
| GQ03113_N22      | MYB4                       | CWT, CRS | -1.67 | 7.69E-09 | -2.03 | 1.09E-11 | 1.21  | 0.007  | Xy > Ne and Ph     |
| GQ0224_E23       | Galactosyltransferase, XT2 | PCT   | -1.13 | 0.002  | -1.11 | 0.009  | -1.02 | 0.594  | Xy > Ne and Ph     |
| GQ02908_P24      | B-TIP                      | RCD   | -2.20 | 2.32E-11 | -1.10 | NS      | -2.00 | 5.62E-10 | Xy = Ph > Ne       |
| GQ02829_F04      | Glycoside hydrolase family 28 | CWT | -3.62 | 1.95E-14 | 1.45  | 0.001  | -5.25 | 4.77E-17 | Ph > Xy and Ne     |
| GQ04010_I15      | Pectinesterase             | CWT   | -1.52 | 0.001  | 3.84  | 3.81E-14 | -5.85 | 2.89E-17 | Ph > Xy > Ne       |
| GQ03211_O01      | Receptor-like kinase       | ARW   | 1.23  | 0.023  | 3.50  | 2.86E-15 | -2.84 | 1.81E-13 | Ph > Xy and Ne     |
| GQ0333_K12       | Glycosyl hydrolase family 10 | MFA | 1.01  | NS      | -1.03 | NS      | 1.04  | NS      | NA                 |
| GQ03806_H09      | LACS9                      | ARW   | 1.02  | NS      | -1.06 | NS      | 1.08  | NS      | NA                 |

*See Table 1 for abbreviated trait names.
*Tissues were from 3-year-old greenhouse seedlings. Xylem, entire wood or secondary xylem (without bark); phloem, secondary phloem and phelloderm; needle, young foliage produced on year's growth. P, probability; NS, not significant.
*FD, fold difference. The sign of the fold difference (+ or −) indicates preferential accumulation in the first tissue/organ (+) or the second one (−).
*Abbreviations: Xy, xylem; Ne, needles; Ph, phloem; NA, not applicable.
TABLE 4
Transcript accumulation comparing genotypic classes for selected gene SNPs significantly associated with wood traits

| Gene Sequence ID | Functional annotation | Traita | No. of trees testedb | RNA transcript moleculesc | Significance (Pa) |
|------------------|------------------------|--------|----------------------|---------------------------|-----------------|
| GQ0172_O22      | β-Expansin             | PCT    | Homo1 11             | Homo1 539                | 0.245           |
|                  |                        |        | Homo2 16             | Homo2 686                | 0.384           |
| GQ03010_F13     | ACC oxidase            | PCT, ARW| 25 12     | Homo1 87                | 0.030*          |
|                  |                        |        | Homo2 122            | Homo2 NA                 | NA              |
| GQ03113_N22     | MYB4                   | CWT, CRS| 17 11     | Homo1 1488              | 0.158           |
|                  |                        |        | Homo2 1467           | Homo2 1476               | 0.684           |
| GQ02829_F04     | β-TIP                  | RCD    | 10 8 8               | Homo1 2458              | 0.075           |
|                  |                        |        | Homo2 1718           | Homo2 2769               | 0.026*          |
| GQ03806_H09     | Glycoside hydrolase    | CWT    | 8 9                  | Homo1 133               | 0.201           |
| family 28        |                        |        | Homo2 144            | Homo2 NA                 | NA              |

aSee Table 1 for abbreviated trait names.
bTranscript levels were assessed in small plugs of differentiating xylem removed from the tree at 1.3 m from the ground.
cRNA transcripts are the average number of RNA molecules per 1 µg of total RNA, after normalization (see MATERIALS AND METHODS).
dStatistical significance was determined using a one-way analysis of variance. P, probability. The dominance model compares homo1 and hetero to homo2. NA, not applicable. *, significant at P ≤ 0.05.

The number of significant SNPs is consistent with previous reports of association studies in forest trees. For example, González-Martínez et al. (2007) reported 7 significant associations with wood traits in *P. taeda* at a FDR of Q ≤ 0.10. Recently, using the same statistical threshold, Dillon et al. (2010) reported 10 SNPs significant with a variety of wood traits in *P. radiata*. Thumma et al. (2005) reported 2 SNPs associated with microfibril angle in *E. nitens*. For climate-related traits such as cold hardiness and phenological characters, there seems to be a trend toward larger numbers of significant associations. Besides Ingvarsson et al. (2008) who found 2 SNPs associated with bud set in *Populus tremula L.*, Eckert et al. (2009) reported 30 significant associations related to adaptation in *Pseudotsuga menziesii* var. *menziesii*, whereas Holliday et al. (2010) detected 35 significant SNPs for bud set and frost hardiness in *P. sitchensis* (Bong.) Carr.

The phenotypic variation in wood traits captured by SNPs deemed significant by FDR in our study varied between 2.6 and 5.4% (Table 2). Small SNP effects are consistent with other association studies in which between 1.5 and 6.5% of the total phenotypic variation was accounted for by SNPs (Thumma et al. 2005; Gonzalez-Martinez et al. 2007; Ingvarsson et al. 2008; Dillon et al. 2010). These small SNP effects are in accordance with polygenic quantitative models of wood traits (Zobel and Jett 1995; Brown et al. 2003; Pot et al. 2006).

When analyzed simultaneously and without taking marker interactions into consideration, the cumulative effect of multiple significant SNPs for the same trait could explain a higher proportion of the total phenotypic variation in some cases (e.g., 9.7%, 8.7%, and 11.1% for average ring width in latewood, cell wall thickness in earlywood, and percentage of earlywood, respectively, with the additive effects model; Table 2). For some of the significant SNPs in the present discovery population, one of the homozygous genotype classes had very few observations (<1%) in both the additive and the dominant effects models (Table 2). In these cases, the corresponding genotypic effects could be biased and partly drive the significance of SNP loci under the dominant effects model. However, when examining phenotypic data, we observed that these individuals

---

**Figure 6.**—Relationship between SNP genotypic classes and transcript accumulation. (A) Estimated genotypic effects of a significant SNP (PGWD1-1094) for earlywood average ring width (and TW-ARW, see Table 2) in the additive effects model. (B) Transcript levels of the corresponding gene determined by qPCR.
possessing rare homozygous genotypes did not necessarily have the most extreme phenotypes. Thus, the effects of these loci may in fact be real. To confirm this, it would require (1) a larger discovery population to capture a greater frequency of low-frequency genotypes and/or (2) an analytical method that accounts for infrequent observations resulting in less bias.

Genes carrying significant SNPs belonged to 11 known gene families and 1 family of unknown function. These gene families are listed in Table 2. The fact that significant associations were found with both growth (e.g., average ring width, percentage of earlywood) and wood quality (e.g., tracheid coarseness, microfibril angle) traits is encouraging, as we are usually interested in selecting for both traits simultaneously. In this study, we could not find any SNP significantly associated with wood density after correction for multiple testing. Wood density is determined by several wood characteristics such as cell size and wall thickness, the ratio of earlywood to latewood, and other factors (Zobel and Van Buijtenen 1989). This is likely the main reason why no significant associations were observed between wood density and SNPs once a correction for multiple testing was performed. This trend might also be specific to the present study because significant associations with wood density were reported for P. taeda (González-Martínez et al. 2007).

Tissue profiling and genotypic transcript accumulation: Transcript accumulation levels were determined by using a large-scale custom oligonucleotide microarray that was composed of 25,094 probes, each estimated to represent a different white spruce gene (see MATERIALS AND METHODS). Total RNA hybridizations against these probes were for comparisons between secondary xylem, secondary phloem, and young needles (Figure 1). Xylem-preferential RNA accumulation compared to that in needles or phloem was found for 10 of the 13 genes harboring SNPs significantly associated with one or more of the wood traits, and the majority of genes (7 of 13) were xylem preferential compared to both needles and phloem (Table 3 and Figure 1B). Transcripts of the expansin sequence (GQ0172_O22), a tubulin (GQ03005_C12), and a sequence of unknown function (GQ03006_P17) were the most xylem-preferential sequences. Two of the genes had low transcript levels in all tissues, and only the receptor kinase-like transcripts accumulated to lower levels in xylem compared to either needles or phloem.

These transcript accumulation profiles were compared to microarray data for the 549 genes tested for association with wood traits and for the transcriptome represented by the microarray (Figure 1). A total of 4138 (16% of the total) of the sequences on the microarray accumulated preferentially in xylem compared to both phloem and needles, and 672 (3%) sequences were xylem preferential in reference to phloem or needles (Figure 1A). Of the 549 genes tested for association, 29% accumulated preferentially in xylem compared to both phloem and needles (Figure 1C), as was expected considering the criteria for selecting the candidate genes. Many of these 549 genes were also preferential to phloem (29%) and only 7% were preferential to needles (compared to 20% in the whole transcriptome). Considering the genes and SNPs putatively associated with wood traits before correction for multiple testing, the distributions of transcript profiles varied slightly from trait to trait but they were not significantly different from the overall set of 549 candidate genes tested (not shown). For example, the proportion of xylem-preferential transcripts varied from 36% for cellulose MFA (Figure 1D) to 24% for tracheid cell diameter (TCD) (Figure 1E), while density (DEN) was slightly enriched for phloem preferential transcripts (Figure 1F).

The association study reported here considered many more candidate genes than previous investigations related to conifer wood properties, which targeted small sets of genes known to be expressed in secondary xylem tissues (e.g., González-Martínez et al. 2007; Dillon et al. 2010). Our analysis selected xylem-expressed transcripts and included sequences that were not preferentially expressed in xylem. For example, the 24 MYB transcription factor sequences and the 13 glycosyl hydrolase protein sequences gave diverse expression profiles (not shown). The observation that most of the SNPs significantly associated with wood property traits were in genes with xylem-preferential transcript accumulation indicates that expression data are relevant for selecting candidate genes. Interestingly, a few genes with low transcript levels in xylem were also associated with phenotypic variation in wood traits, but one of these genes was found to be strongly expressed in differentiating xylem of mature trees (Table 4), whereas tissue preferential patterns were determined in whole xylem of young trees (Figure 1).

Transcript levels could be compared among the different genotypic classes for six of the significantly associated SNPs (after FDR correction for multiple testing) by using RT-qPCR with gene-specific primers as described in Boyle et al. (2009). The assays used differentiating secondary xylem from the 30-year-old trees of the association population. Transcript accumulation varied significantly between genotypic classes of the SNPs for the LACS9, b-TIP, and ACC oxidase genes (Table 4), which represents 50% of the SNPs analyzed. The MYB and the expansin transcripts did not differ significantly between genotypes, and in determinations of the glycoside hydrolase family 28 sequence transcripts were too variable among trees to draw any conclusions. These differences in transcript accumulation among genotypic classes are discussed below with their putative function and relationship with variation in phenotypic trait. To our knowledge, only Thumma et al. (2009) previously reported a link between SNPs, transcript levels,
2001; S in development via fatty acid metabolism, including in Arabidopsis were shown to activate 14 fatty acids (S the synthesis of phospholipids and triacylglycerols in The LACS enzymes provide the acyl-CoA pools used in cellulose deposited in the wood.

Potential role of genes containing significant association SNPs: Long-chain acyl-CoA synthetase 9 (LACS9): The LACS enzymes provide the acyl-CoA pools used in development via fatty acid metabolism, including the provision of energy. The nine LACS characterized in Arabidopsis were shown to activate 14–20 carbon fatty acids (Stockey et al. 2002) and are expressed in all tissues. Extended analyses of the LACS gene family have begun to reveal the roles of individual members in the peroxisome (Fulda et al. 1997) and chloroplast (Schnurr et al. 2002). Interestingly, the LACS transcript levels were low in the young trees used for the tissue profiling (not shown), but were among the highest in the secondary xylem of the mature trees (Table 4).

Individuals of only one of the two homozygous classes were present in the discovery population and, on average, they showed larger earlywood and total wood average ring width than heterozygotes (Table 2 and Figure 6a). The SNP (PGWD1-1094) was located in the 3′-UTR, which could indicate a role in the regulation of transcription, RNA stability, or post-transcriptional RNA processing. The transcript levels differed significantly between genotypes (P = 0.02) (Table 4). The estimated allelic effects of the SNP (PWND1-1094) on the wood trait (Figure 6a) corresponded well with estimates of transcript levels (Figure 6b); i.e., heterozygotes had both lower average width of the latewood ring (ARW) and lower transcript levels than the homozygotes.

Glycoside hydrolase family 28 [polygalacturonase (pectinase) family protein] (GQ02829_F04): The glycoside hydrolase protein family 28 (GH28) is a large family of plant proteins that comprises enzymes with several known activities. Pectin-modifying enzymes have well-established roles in the primary cell wall, in particular during cellular growth and in biotic interactions; however, data are lacking about wood formation. Seven putative GH28 members were represented on the genotyping chip and all but 2 of them accumulated more transcripts in phloem (7) or xylem (5). The SNP (PGWD1-1096) represents a nonsynonymous substitution (TGCC–GGC) coding for cysteine and glycine that was associated with both earlywood and total wood cell wall thickness.

MYB4 (R2R3-MYB domain DNA-binding protein): The MYB4 gene has been linked to secondary cell wall formation, including lignification in the secondary xylem in both P. taeda and P. glauca. In P. glauca, transcripts levels were highest in the secondary xylem of the main stem and of large roots and were upregulated during the formation of compression wood (Bedon et al. 2007). Closely related MYBs of P. taeda (Patzlaff et al. 2003a,b) and of E. gunnii Hook f. (Gorciocehe et al. 2005) have been shown to activate the transcription of genes encoding lignin synthesis enzymes and to regulate lignin accumulation and cell wall formation when overexpressed in angiosperm model plants. Overexpression of the P. taeda MYB8 gene in transgenic spruce resulted in ectopic secondary cell wall thickening (Bomal et al. 2008). The SNP (PGWD1-1282) represents a synonymous substitution (GACA–GCG) coding for alanine. It was associated with cell wall thickness and fiber coarseness, both of which are related to wood density and radial
and tangential cell diameter. Heterozygous and dominant homozygous trees showed significantly higher fiber coarseness for the whole wood ring (279 µg/m) and for earlywood (269 µg/m) than the alternate homozygous type (266 µg/m and 255 µg/m, respectively, Table 2). A similar trend was observed for cell wall thickness (Table 2). The transcript levels did not differ significantly (P = 0.16) between genotypes, despite the large number of individuals tested (Table 4).

*Galactosyl-transferase XT2 (UDP-xyllosyltransferase 2):* The hemicellulose xyloglucan (XyG) is believed to play an important role in cell wall structure and function (Eckardt et al. 2008). A number of enzymes participate in XyG biosynthesis, including b-(1/4)-glucan synthase, a-fucosyltransferases, b-galactosyltransferases, and α-xyllosyltransferases (Girke et al. 2004). A lack of detectable xyloglucan in the xxt1/xxt2 double mutant in Arabidopsis caused a significant reduction in the stiffness and strength parameters of the mutants (Cavaleri et al. 2008). The xyloglucan acts as a tether in the extracellular matrix and plays a key role in the loosening and tightening of cellulose microfibrils (Hayashi and Kaida 2011). The SNP (PGWD1-0511) represents a synonymous substitution (TCA–TCG) coding for a serine. It was associated with earlywood percentage.

*Receptor-like kinase (ATP binding/kinase/protein serine/threonine kinase) (GQ03211_001):* Receptor-like kinases (RLKs) control a wide range of processes, including development, disease resistance, hormone perception, and self-incompatibility. The Arabidopsis genome contains >600 RLK homologs, representing nearly 2.5% of the annotated protein-coding genes (Shu and Bleecker 2001). In Arabidopsis, the PXY receptor-like kinase plays a role in the maintaining of cell polarity in the meristem and is required for the differentiation of specialized and spatially separated xylem and phloem cells (Fisher and Turner 2007). In poplar, transcripts of three receptor-like kinases accumulate a different stage of secondary xylem differentiation (Schraeder et al. 2004). In *P. glauca*, transcript accumulation was clearly preferential to phloem, suggesting that communication between xylem and phloem may affect wood properties. The SNP (PGWD1-1313) is a synonymous substitution (ACA–ACG) coding for threonine in the exon. It was associated with both the ARW and the proportion of earlywood (PCT), two closely related traits.

*ACC oxidase (1-aminocyclopropane-1-carboxylate oxidase):* This enzyme forms ethylene from its precursor, ACC. In plants, ethylene has signaling functions such as stimulation of fruit ripening and is required for CO₂ activity. The closest ACC oxidase homolog in Arabidopsis was regulated during *in vitro* xylem vessel element differentiation (Kuro et al. 2005). Endogenous ACC was shown to accumulate in the vascular cambium specifically on the lower side of branches and was linked to compression wood differentiation in *P. contorta* Doug. ex Loud. ssp. latifolia (Savidge et al. 1983). Evidence also suggested that ethylene plays a role in the control of xylem differentiation by inducing the activity of enzymes involved in lignification (Miller et al. 1984). In *P. glauca*, ACC oxidase transcripts were moderately preferential to secondary xylem (Table 3). ACC oxidase protein accumulated significantly more in secondary xylem compared to needles and in response to gravitational stress in *P. pinaster* Ait. (Plomion et al. 2000). The SNP (PGWD1-0560) represents a nonsynonymous substitution (GGG–GAG) coding for glycine and glutamine. It was also associated with both latewood average ring width and earlywood percentage. The transcripts varied between the two genotypic classes that could be compared (P = 0.03) (Table 4).

*β-Expansin (AtEXP3 (Arabidopsis thaliana Expansin B3)):* Expansins are a large family of proteins that play major roles in plant growth and development by acting on cell walls. β-Expansins act through cell wall loosening in a variety of developmental functions. They have been mostly described in grapes; no studies on gene expression or function in secondary cell wall development or wood formation have been reported. The SNP was located in an intron (A*–C*) and was associated with the PCT. The homozygous trees had significantly more earlywood (84%) than the two other genotypes (81%). Our data suggested that the transcripts may accumulate to lower levels in the homozygous genotype with higher PCT, but the difference was not statistically significant. The β-expansin gene might be somehow related to the differences in stiffness and cell wall thickness between earlywood and latewood, but this remains untested.

*Tubulins (TUA2 and TUB3):* Several lines of evidence suggest that the xylem predominant TUA and TUB isoforms are specifically associated with cellulose synthesis during secondary cell wall formation (Oakley et al. 2007). Both of the *P. glauca* tubulins were expressed preferentially in xylem and shared homologies with tubulins from *P. taeda* that are highly abundant among ESTs from xylem cDNA libraries (Yang et al. 2004) and SAGE sequencing data (Lorenz and Dean 2002).

The *P. glauca* α-tubulin gene (TUA2) is highly similar to sequences that have been linked to the formation of cortical microtubules in *P. taeda* (González-Martínez et al. 2007), developing xylem-preferential transcript accumulation in *P. radiata* (Cato et al. 2006), ESTs that are unique to xylem libraries in Populus (Pilate et al. 2004), and transcripts levels that correlated with auxin concentration in developing secondary xylem in *P. tremula* x *tremuloides* (Nilsson et al. 2008). A β-tubulin gene was shown to be preferentially expressed in xylem in Eucalyptus (Paux et al. 2004). β-Tubulin (EgrTUB1) was differentially regulated in Eucalyptus branches and was reported to be likely involved in cell wall biosynthesis (Qu et al. 2008). The TUA2 SNP (PGWD1-0863) represents a synonymous substitution (GGA–GAG) coding for glycine in a tubulin, whereas the TUB3 SNP (PGWD1-
LD has been shown to also have substantial gene-to-gene variation, in natural populations of spruces (Namroud et al. 2010). As more SNPs are discovered, LD information within genes may be used to infer haplotypes and predict their effects, which has been found to yield more accurate estimates of genetic effects (Całus et al. 2008). Although LD falls to low levels in natural populations of conifers, smaller populations contain greater potential LD partly as a function of population size (Neale and Savolainen 2004). Thus, if genomic resources are fixed, working with smaller population sizes, such as elite breeding material, would be another way of utilizing LD information in an association analysis. In such situations, genomic selection (Wong and Bernardo 2008; Grattapaglia and Resende 2011) or multiple-gene selection could also become attractive.

Finally, there is a wealth of information on longitudinal phenotypic data that at first appears underutilized by taking overall means or by discretizing the data into cross-sectional “age” traits. From one standpoint, markers found to be significant over developmental ages, year environments, and locations can be interpreted as stable and thus possibly useful for MAS (Gonzalez-Martinez et al. 2007). However, if trait (co)variances change over time, more complex, functional models may better serve both researchers whose interests lie in the prediction of genetic effects (e.g., breeders) and those interested in elucidating gene functions (e.g., geneticists and molecular biologists). In this study, the application of more complex functional models, which require more parameters to be estimated, was limited by the relatively small size of the data set. A forthcoming study is planned in an expanded sampling of our discovery population with individuals genotyped at more SNPs per gene and more candidate genes, taking these considerations into account.

**Conclusion:** Given the large number of SNPs that explained significant amounts of phenotypic variation before corrections for multiple testing, it stands to reason that several of these SNPs could indeed not be false positives and be declared significant if the size of the discovery population was expanded. This is a line of research that we are pursuing actively. This study represents one of the first well-documented cases showing significant differential transcript accumulation among the genotypes of a SNP associated with a trait of interest in a population of conifer trees. It also shows that transcript accumulation profiles are appropriate for selecting candidate genes and finding SNPs associated with wood properties, because most of the significant SNPs were found in genes that had higher transcripts levels in secondary xylem. While the number of significant associations and the proportion of variation explained were small, they were in line with expectations derived from statistical and quantitative genetics theory. These results obtained using large-scale gene sampling and high-throughput genotyping approaches combined
with gene expression represent significant progress toward the identification of genes that control wood traits and the development of successful marker-aided selection in conifer trees. This trend will continue given the current rapid advances in high-throughput sequencing and genotyping.

This study is part of the Arborlea project and all the authors are members of this research project in forest genomics. We thank E. Pouliot, D. Plourde, É. Dussault, P. Labrie, and numerous summer students for technical assistance. We also thank M. Lamotte, J. Fillon and J.-P. Dionne for ideas and support regarding the bioinformatics methods to de

CATO, S., L. McMILLAN, L. DONALDSON, T. RICHARDSON, C. EICH et al., 2006 Wood formation from the base to the crown in *Pinus radiata* gradients of tracheid wall thickness, wood density, radial growth rate and gene expression. Plant Mol. Biol. 60: 565–581.

CAVALIERI, D. M., O. LEROUVEL, L. NEUMETZLER, K. YAMAUCHI, A. REINECKE et al., 2008 Disrupting two *Arabidopsis thaliana* xylosyltransferase genes results in plants deficient in xyloglucan, a major primary cell wall component. Plant Cell 20: 1519–1537.

CHAMARY, J., V. and L. D. HURST, 2009 The price of silent mutations. Sci. Am. 300(6): 46–53.

CHANG, S. J., P. JURJEVIC and J. CAINERY, 1993 Simple and efficient method for isolating RNA from pine trees. Plant Mol. Biol. Rep. 11: 113–116.

CHEUNG, V. G., R. S. SPIELMAN, K. G. EWENS, T. M. WEBER, M. MORLEY et al., 2005 Mapping determinants of human gene expression by regional and genome-wide association. Nature 437: 1363–1369.

CORRIVEAU, A., J. BEAULIEU and G. DAOUT, 1991 Heritability and genetic correlations of wood characters of Upper Ottawa Valley white spruce populations grown in Quebec. For. Chron. 67: 698–705.

DILLON, S. K., M. NOLAN, W. LI, C. BELL, H. XU Wu et al., 2010 Allelic variation in cell wall candidate genes affecting solid wood properties in natural populations and land races of *Pinus radiata*. Genetics 185: 1477–1487.

ECKARDT, N. A. 2008 Role of xyloglucan in primary cell walls. Plant Cell 20: 1421–1422.

ECKERT, A. J., A. D. BOWER, J. L. WERGRYN, B. PANDE, K. J. JERMSTAD et al., 2009 Association genetics of coastal Douglas fir (*Pseudotsuga menziesii* var. *menziesii*). In: *P. menziesii* Pine. E. Cold-hardiness related traits. Genetics 182: 1299–1302.

EGERTSDOTTER, U., L. VAN ZYL, J. MACKAY, R. WHETTEN, G. PETER et al., 2004 Gene expression profiling during formation of earlywood and latewood in loblolly pine: expression profiles of 350 genes. Plant Biol. 6: 654–663.

EVANS, R., and J. ILIC, 2001 Rapid prediction of wood stiffness from microfibril angle and density. For. Prod. J. 51: 53–57.

FAULH, D., M. STEPHENS and J. K. FRITCHARD, 2003 Inference of population structure using multifocus genotype data: linked loci and correlated allele frequencies. Genetics 164: 1567–1587.

FAN, J.-B., A. OLIPIAN, R. SHEN, B. G. KERMANI, F. GARCIA et al., 2003 Highly parallel SNP genotyping. Cold Spring Harbor Symp. Quant. Biol. 68: 69–78.

FISHER, K., and S. TURNER, 2007 PXY, a receptor-like kinase essential for maintaining polarity during plant vascular-tissue development. Curr. Biol. 17: 1061–1066.

FROST, S., 1976 Study of the decrease of eigenvalues in a principal component analysis comparison with the broken stick model. J. Exp. Mar. Biol. Ecol. 25: 67–75 (in French).

FULDA, M., E. HEINZ and F. P. WOLTER, 1997 *Brassica napus* cDNAs encoding fatty acyl-CoA synthetase. Plant Mol. Biol. 33: 911–922.

FURNER, G. R., and M. STINE, 1995 Interpopulation differentiation of nuclear and chloroplast loci in white spruce. Can. J. For. Res. 25: 736–742.

GALLO DE CARVALHO, M. C. D. C., D. G. G. CALDAS, R. T. CARNEIRO, D. H. MOON, G. R. SALVATIERRA et al., 2008 SAGE transcript profiling of the juvenile cambial region of *Eucalyptus grandis*. Treec Physiol. 28: 903–919.

GIESLER-LEE, J., M. GEISLER, P. M. COUNTINHO, B. SAGERMAN, N. NISHIKURO et al., 2006 Poplar carbohydrate-active enzymes. Gene identification and expression analyses. Plant Physiol. 140: 946–992.

GIRKE, T., J. LAURICHA, H. TRAN, K. KEEGSTRA and N. RAHMKEL, 2004 The cell wall navigator database. A systems-based approach to organism-unrestricted mining of protein families involved in cell wall metabolism. Plant Physiol. 136: 3003–3008.

GODBOUT, M. E., and B. HAYES, 2009 Mapping genes for complex traits in domestic animals and their use in breeding programs. Nat. Rev. Genet. 10: 381–391.

GOICHECHA, M., E. LACOMBE, S. LEGAY, S. MIHALJEVIC, P. RECH et al., 2005 *EgMYB2*, a new transcriptional activator from *Eucalyptus* xylem, regulates secondary cell wall formation and lignin biosynthesis. Plant J. 43: 553–567.
in an interspecific backcross of *Eucalyptus*. Plant Physiol. 135: 2368–2378.

Kuro, M., M. Udagawa, N. Nishikubo, G. Horiguchi, M. Yamaguchi et al., 2005 Transcription switches for protoxym and metaxym vessel formation. Genes Dev. 19: 1855–1860.

Legendre, P., and L. Legendre, 1998 *Numerical Ecology*, Second English Edition. Elsevier, New York.

Lenz, P., A. Cloutier, J. MacKay and J. Beaulieu, 2010 Genetic control of wood properties in *Picea glauca*—an analysis of trends with cambial age. Can. J. For. Res. 40: 703–715.

Li, P., J. Beaulieu, A. Corriveau and J. Bousquet, 1993 Genetic variation in juvenile growth and phenology in a white spruce provenance—progeny test. Silvae Genet. 42: 52–60.

Li, P., J. Beaulieu and J. Bousquet, 1997 Genetic structure and patterns of genetic variation among populations in eastern white spruce (*Picea glauca*). Can. J. For. Res. 27: 189–198.

Littell, R. C., G. A. Milliken, W. W. Stroup, R. D. Wolfinger and O. Schabenberger, 2006 *SAS for Mixed Models*, Ed. 2. SAS Institute, Cary, NC.

Loudelle, B. A., V. L. Sork, J. Nason and C. Graham, 1995 Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). Am. J. Bot. 82: 1420–1425.

Lorenz, W. W., and J. F. D. Dean, 2002 *SAGE* profiling and demonstration of differential gene expression along the axial developmental gradient of lignifying xylem in loblolly pine (*Pinus taeda*). Tree Physiol. 22: 301–310.

MacKay, T. F. C., 2001 Genetic architecture of quantitative traits. Annu. Rev. Genet. 35: 303–339.

MacKay, J., and J. Dean, 2011 Transcriptomics, in *Genetics, Genomics and Breeding of Conifers*, edited by C. Plomion, J. Bousquet and C. Kole. CRC Press, New York (in press).

Malosetti, M., C. G. van der Linden, B. Vosman and F. van Eeuwijk, 2007 A mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato. Genetics 175: 879–889.

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.

Marth, G. T., I. Korf, M. D. Yandell, R. T. Yeh, Z. Gu et al., 1999 A general approach to single-nucleotide polymorphism discovery. Nat. Genet. 23: 452–456.

Michie, F., 2001 Pectin methyltransferases: cell wall enzymes with important roles in plant physiology. Trends Plant Sci. 6: 414–419.

Miller, A. R., W. L. Pengelly and L. W. Roberts, 1984 Introduction of xylem differentiation in *Lactuca* by ethylene. Plant Physiol. 75: 1165–1168.

Murray, B. G., 1998 Nuclear DNA amounts in gymnosperms. Annu. Rev. Ecol. Syst. 29: 5–15.

Myres, S., J. Peiffer, P. J. Brown, E. S. Ersöz, Z. Zhang et al., 2009 Association mapping: critical considerations shift from genotyping to experimental design. Plant Cell 21: 2194–2202.

Namroud, M.-C., J. Beaulieu, N. Juge, J. Laroche and J. Bousquet, 2008 Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. Mol. Ecol. 17: 3599–3613.

Namroud, M.-C., C. Guillet-Claude, J. MacKay, N. Isabel and J. Bousquet, 2010 Molecular evolution of regulatory genes in spruces from different species and continent: heterogeneous patterns of linkage disequilibrium and selection but correlated recent demographic changes. J. Mol. Evol. 70: 371–386.

Neale, D. B., and A. Kremer, 2011 Forest tree genomics: growing resources and applications. Nat. Rev. Genet. 12: 111–122.

Neale, D. B., and O. Savolainen, 2004 Association genetics of complex traits in conifers. Trends Plant Sci. 9: 925–303.

Nienstaedt, H., and A. H. Teich, 1972 *The Genetics of White Spruce* (Research Paper WO-15). USDA Forest Service, Washington, DC.

Nienstaedt, H., and J. C. Zasada, 1990 *Picea glauca* (Moench) Voss white spruce, pp. 209–226 in *Conifers*. Agriculture Handbook No. 654 (Silvics of North America, Vol. 1) coordinated by R. M. Burns and B. H. Honkala. USDA Forest Service, Washington, DC.

Nilsson, J., A. Karlberg, H. Antti, M. Lopez-Vernaza, E. Mellerowicz et al., 2008 Dissecting the molecular basis of the regulation of wood formation by auxin in hybrid aspen. Plant Cell 20: 843–855.
Oakley, R. V., Y.-S. Wang, W. Ramakrishna, S. A. Harding and C.-J. Tsai, 2007 Differential expansion and expression of a- and b-tubulin gene families in *Populus*. Plant Physiol. 145: 961–973.

Paiva, J., 2006 Phenoypic and molecular plasticity of wood forming tissues in maritime pine (*Pinus pinaster* Ait.). Ph.D. Thesis, Bordeaux University, Cestas, France and New University of Lisbon, Oeiras, Portugal.

Patzlaff, A., S. McInnis, A. Courtenay, C. Surman, L. J. Newman et al., 2005a Characterisation of a pine MYB that regulates lignification. Plant J. 36: 743–754.

Patzlaff, A., L. J. Newman, C. Dubos, R. W. Whetten, C. Smith et al., 2005b Characterisation of PtMYB1, an R2R3-MYB from pine xylem. Plant Mol. Biol. 55: 207–208.

Paux, E., M. Tamashiyuki, N. Ladjouche, P. Sivadon and J. Grima-Pettenati, 2004 Identification of genes preferentially expressed during wood formation in *Eucalyptus*. Plant Mol. Biol. 55: 263–280.

Pavy, N., C. Paule, L. Parsons, J. A. Crow, M.-J. Morency et al., 2005a Generation, annotation, analysis and database integration of 16,500 white spruce EST clusters. BMC Genomics 6: 144.

Pavy, N., J. Laroche, J. Bousquet and J. Mackay, 2005b Large-scale statistical analysis of secondary xylem ESTs in pine. Plant Mol. Biol. 57: 203–224.

Pavy, N., L. S. Parsons, C. Paule, J. Mackay and J. Bousquet, 2006 Automated SNP detection from a large collection of white spruce expressed sequences: contributing factors and approaches for the categorization of SNPs. BMC Genomics 7: 174.

Pavy, N., B. Boyle, C. Nelson, C. Paule, I. Giguerre et al., 2008a Identification of conserved core xylem gene sets: cofiber cDNA microarray development, transcript profiling and computational analyses. New Phytol. 180: 766–786.

Pavy, N., B. Pelgas, S. Beauségle, S. Blais, F. Gagnon et al., 2008b Enhancing genetic mapping of complex genomes through the design of highly-multiplexed SNP arrays: application to the large and unsequenced genomes of white spruce and black spruce. BMC Genomics 9: 21.

Pelgas, B., N. Isabel and J. Bousquet, 2004 Efficient screening for expressed sequence tag polymorphisms (ESTPs) by DNA pool sequencing and denaturing gradient gel electrophoresis (DGGE) in spruces. Mol. Breed. 13: 263–279.

Pelgas, B., S. Beauségle, V. Achere, S. Jeandroux, J. Bousquet et al., 2006 Comparative genome mapping among *Picea glauca*, *P. mariana* and *P. abies*, and correspondence with other Pinaceae. Theor. Appl. Genet. 113: 1371–1393.

Pelgas, B., J. Bousquet, P. G. Meirmans, K. Ritland and N. Isabel, 2011 QTL mapping in white spruce: gene maps and genomic regions underlying adaptive traits across pedigrees, years and environments. BMC Genomics 12: 114.

Pilate, G., A. Déjardin, F. Laurans and J.-C. Leple, 2004 Tension wood as a model for functional genomics of wood formation. New Phytol. 164: 63–72.

Plomin, C., C. E. Durell and D. M. O’Malley, 1996 Genetic dissection of height in maritime pine seedlings raised under accelerated growth conditions. Theor. Appl. Genet. 93: 849–858.

Plomin, C., C. Pionneau, J. Brach, P. Costa and H. Balléres, 2000 Compression wood-responsive proteins in developing xylem of maritime pine (*Pinus pinaster* Ait.). Plant Physiol. 123: 959–970.

Pongdontri, P., and M. Hills, 2001 Characterization of a novel plant acyl-CoA synthetase that is expressed in lipogenic tissues of *Brassica napus* L. Plant Mol. Biol. 47: 717–726.

Pot, D., J.-C. Rodrigues, P. Rozenberg, G. Chantre, J. Tibrift et al., 2006 QTLs and candidate genes for wood properties in maritime pine (*Pinus pinaster* Ait.). Tree Genet. Genomes 2: 10–24.

Pritchard, J., X. Wen and D. Falush, 2007 *Documentation for Structure Software: Version 2.2*. Department of Human Genetics, University of Chicago, Chicago.

Pritchard, J. K., M. Stephens and P. Donnelly, 2000 Inference of population structure using multilocus genotype data. Genetics 155: 945–959.

Qu, D., I. W. Wilson, S. Gan, R. Washusen, G. F. Moran et al., 2008 Gene expression in *Eucalyptus* branch wood with marked variation in cellulose microfibril orientation and lacking G-layers. New Phytol. 179: 94–101.

R Development Core Team, 2010 *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna.

Rafalski, A., 2010 Association genetics in crop improvement. Curr. Opin. Plant Biol. 13: 174–180.

Ritland, K., K. Krutovsky, R. McInnis, B. Pelgas, N. Isabel et al., 2011 Genetic mapping in conifers, in *Genetics, Genomics and Breeding of Conifers*, edited by C. Plomin, J. Bousquet and C. Kole. CRC Press, New York (in press).

Rozen, S., and H. Skalesky, 2000 Primer3 on the WWW for general users and for biologist programmers, pp. 363–366 in *Bioinformatics Methods and Protocols: Methods in Molecular Biology*, edited by S. Misener and S. A. Krawetz. Humana Press, Totowa, NJ.

Rutledge, R. G., and D. Stewart, 2008 A kinetic-based sigmoidal model for the polymerase chain reaction and its application to high-capacity absolute quantitative real-time PCR. BMC Biotechnol. 8: 47.

SAS Institute, 2008 *The SAS System for Windows*, Version 9.1. SAS Institute, Cary, NC.

Savidge, R. A., G. M. C. Mutumba, J. K. Heath and P. F. Wareing, 1983 Gas chromatography-mass spectroscopy identification of 1-aminocyclopentane-1-carboxylic acid in compression wood vascular cambium of *Pinus contorta* Doug. Plant Physiol. 71: 434–436.

Schnurr, J. A., J. M. Shockey, G.-J. de Boer and J. A. Browse, 2002 Fatty acid export from the chloroplast. Molecular characterization of a major plastidial acyl-coenzyme A synthetase from *Arabidopsis*. Plant Physiol. 129: 1701–1709.

Schraden, J., J. Nilsson, E. Mellerowicz, A. Berglund, P. Nilsson et al., 2004 A high-resolution transcript profile across the wood-forming meristem of poplar identifies potential regulators of cambial stem cell identity. Plant Cell 16: 2278–2292.

Sewell, M. M., D. L. Basso, R. A. Magraw, 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.

Shen, R., L.-B. Fan, D. Campbell, W. Chang, J. Chen et al., 2005 High-throughput SNP genotyping on universal bead arrays. Mutat. Res. 573: 70–82.

Shiu, S.-H., and A. B. Bleecker, 2001 Plant receptor-like kinase gene family: diversity, function, and signaling. Sci. STKE. 2001 (113): re29.

Shockey, J. M., M. S. Fulda and J. A. Browse, 2002 Arabidopsis contains nine long-chain acyl-coenzyme A synthetase genes that participate in fatty acid and glycerolipid metabolism. Plant Physiol. 129: 1710–1722.

Siedlecka, A., S. Wiklund, M.-A. Pérone, F. Michell, J. Lennénska et al., 2008 FecA methyl esterase inhibits intrusive and symplastic cell elongation in developing wood cells of *Populus*. Plant Physiol. 146: 554–565.

Sokal, R. R., and F. J. Rohlf, 1981 *Biometry*. Ed. 2. W. H. Freeman, San Francisco.

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 genomewide studies. Proc. Natl. Acad. Sci. USA 100: 9440–9445.

Storey, J. D., J. E. Taylor and D. Siegmund, 2004 Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. J. R. Stat. Soc. Ser. B 66: 187–205.

Swindell, S. R., and T. N. Plasterer, 1997 *SEQMAN: contig assembly*, pp. 77–89 in *Sequence Data Analysis Guidebook*, edited by S. R. Swindell. Humana Press, Totowa, NJ.

Takahashi-Schmidt, J., 2008 Functional studies of selected extra-cellular carbohydrate-active hydrolases in wood formation. Ph.D. Thesis, Faculty of Forest Sciences, Department of Forest Genetics and Plant Physiology, Umeå University, Umeå, Sweden.

Thumma, B. R., M. F. Nolan, R. Evans and G. F. Moran, 2001 Polymorphisms in *Cinnamyl CoA Reductase* (CCR) are associated with variation in microfibril angle in *Eucalyptus* spp. Genetics 171: 1257–1265.

Thumma, B. R., B. A. Matheson, D. Zhang, C. Meeske, R. Meeder et al., 2009 Identification of cis-acting regulatory polymorphisms in a eukalypt *COBR*1-like gene affecting cellulose content. Genetics 183: 1153–1164.
Ukrainetz, N. K., K. Ritland and S. D. Mansfield, 2008  Identification of quantitative trait loci for wood quality and growth across eight full-sib coastal Douglas-fir families. Tree Genet. Genomes 4: 159–170.

Vander Willigen, C., O. Postaire, C. Tournaire-Roux, Y. Boursiac and C. Maurel, 2006  Expression and inhibition of aquaporins in germinating Arabidopsis seeds. Plant Cell Physiol. 47: 1241–1250.

Vandesompele, J., K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, et al., 2002  Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 3: research0034.1–0034.11.

Wheeler, N. C., K. Jermstad, K. 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: 143–156.

Whetten, R., Y. H. Sun, Y. Zhang and R. Sederoff, 2001  Functional genomics and cell wall biosynthesis in loblolly pine. Plant Mol. Biol. 47: 275–291.

Wong, C. K., and R. Bernardo, 2008  Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. Theor. Appl. Genet. 116: 815–824.

Yang, S.H., L. van Zyl, E.-G. No and C. A. Loopstra, 2004  Microarray analysis of genes preferentially expressed in differentiating xylem of loblolly pine (Pinus taeda). Plant Sci. 166: 1185–1195.

Yu, J., G. Pressoir, W. H. Briggs, I. V. 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.

Zhu, C., M. Gore, E. S. Buckler and J. Yu, 2008  Status and prospects of association mapping in plants. Plant Genome 1: 5–20.

Zobel, B. J., and J. P. van Buijtenen, 1989  Wood Variation. Its Causes and Control. Springer-Verlag, New York.

Communicating editor: M. Kirst
Association Genetics of Wood Physical Traits in the Conifer White Spruce and Relationships With Gene Expression

Jean Beaulieu, Trevor Doerksen, Brian Boyle, Sébastien Clément, Marie Deslauriers, Stéphanie Beauseigle, Sylvie Blais, Pier-Luc Poulin, Patrick Lenz, Sébastien Caron, Philippe Rigault, Paul Bicho, Jean Bousquet and John MacKay

Copyright © 2011 by the Genetics Society of America
DOI: 10.1534/genetics.110.125781
Files S1-S6 are available for download as Excel or Text files at http://www.genetics.org/cgi/content/full/genetics.111.125781/DC1.

File S1: Genotypes of the white spruce discovery population for 944 SNP loci.

File S2: Overall total wood, earlywood and latewood averages of each tree making up the white spruce discovery population for 25 wood traits.

File S3: Matrix of kinship coefficients built up using the known half-sib structure of trees making up the discovery population.

File S4: Matrix of Nason's kinship coefficients between each pair of trees making up the discovery population.

File S5: Results of the association studies for each of the 25 wood traits using mixed linear and additive effects models.

File S6: Results of the association studies for each of the 25 wood traits using mixed linear and dominance effects models.