

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

Jean Beaulieu,<sup>\*,†,1</sup> Trevor Doerksen,<sup>\*,†</sup> Brian Boyle,<sup>†</sup> Sébastien Clément,<sup>\*</sup> Marie Deslauriers,<sup>\*</sup> Stéphanie Beauseigle,<sup>†</sup> Sylvie Blais,<sup>†</sup> Pier-Luc Poulin,<sup>†</sup> Patrick Lenz,<sup>†</sup> Sébastien Caron,<sup>†</sup> Philippe Rigault,<sup>‡</sup> Paul Bicho,<sup>§</sup> Jean Bousquet<sup>†</sup> and John MacKay<sup>†</sup>

<sup>\*</sup>Natural Resources Canada, Canadian Wood Fibre Centre, Québec City, Québec G1V 4C7, Canada, <sup>†</sup>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, <sup>‡</sup>Gydlle, Québec City, Québec G1T 1Z2, Canada and <sup>§</sup>FPIInnovations, 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  $\beta$ -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.

**A** CQUISITION 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; PLOMION *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  $\geq 25\%$  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 KREMER 2011). By utilizing past recombination events, association mapping increases the

Supporting information is available online at <http://www.genetics.org/cgi/content/full/genetics.110.125781/DC1>.

<sup>1</sup>Corresponding author: Natural Resources Canada, Canadian Wood Fibre Centre, 1055 du P.E.P.S., P.O. Box 10380, Stn. Sainte-Foy, Québec City, Québec G1V 4C7, Canada. E-mail: jean.beaulieu@NRC.can.gc.ca

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 (HAUSSMANN *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 cobra-like gene was linked to cellulose content and pulp yield (THUMMA *et al.* 2009). In *Pinus taeda* L., four cell wall genes, *i.e.*, S-adenosyl methionine synthase 2, cinnamyl alcohol dehydrogenase, the water-stress inducible protein *lp3-l*, and  $\alpha$ -tubulin, explained ~3% of the variation in wood specific gravity, in the proportion of latewood to earlywood and in cellulose microfibril angle (GONZÁLEZ-MARTÍNEZ *et al.* 2007). In *P. radiata* L., 10 single-nucleotide polymorphisms (SNPs) located on nine genes were associated with several traits (DILLON *et al.* 2010) but none of the genes overlapped with previous studies. This observation and the small number of genes tested to date suggest that these associations likely represent a small fraction of the genes that influence wood trait variation. Larger-scale analyses are needed to resolve this question and have become feasible with the availability of genome sequences for *Populus* and *Eucalyptus* (available at <http://www.phytosome.net/>), as well as large-scale cDNA resources in conifers such as *Picea* and *Pinus* (MACKAY and DEAN 2011).

The objective of the present study was to test for associations between SNP markers and wood quality and radial growth traits in white spruce [*Picea glauca* (Moench) Voss]. White spruce is widely distributed in North America (NIENSTÄEDT and TEICH 1972) and colonizes a wide range of sites covering most of Canada and a part of the United States (NIENSTÄEDT and

ZASADA 1990). It exhibits high genetic variability for neutral markers (JARAMILLO-CORREA *et al.* 2001; BOUSQUET *et al.* 2007; NAMROUD *et al.* 2008), for growth and phenology (LI *et al.* 1993, 1997), and for wood traits (CORRIVEAU *et al.* 1991; BEAULIEU 2003; LENZ *et al.* 2010). As in many other tree species, spruce breeding aimed at wood quality has seen limited application because wood traits are time consuming and expensive to assess and change over the course of tree development. It is expected that DNA markers that explain a significant portion of the variation in wood traits would help to develop more rapid and effective breeding methods. Our specific objectives were (1) to identify single-marker associations with 25 wood traits using SNPs from several hundred candidate genes, (2) to determine the proportion of the variation in wood traits explained by these associations, (3) to test whether RNA transcript accumulation varies between the different genotypes for significant associations, and (4) to confirm the potential of a candidate gene approach for characterizing the genetic basis of wood traits.

## MATERIALS AND METHODS

**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  $-10^{\circ}$  until further treatment.

**Phenotypic data:** Pith to bark profiles of wood physical attributes were obtained using the SilviScan technology (EVANS and ILIC 2001) at FPIInnovations (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  $2 \times 7$ -mm wood flitch samples prepared from the 12-mm cores. The thawed samples were acetone extracted and conditioned at 40% relative humidity. Air-dry ( $20^{\circ}$ ) density was measured in 25- $\mu$ m 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 ( $\sim 1.5 \mu$ m) 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

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 (PAVY *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 (PAVY *et al.* 2008a) and three such studies of secondary xylem carried out in pine species (PLOMION *et al.* 2000; PAVY *et al.* 2005b; PAIVA 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 (PAVY *et al.* 2008b). Primer pairs were typically designed upon either 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; PAVY *et al.* 2008b). PCR amplifications were performed in volumes of 30  $\mu$ l containing 10–20 ng template DNA, 20 mM Tris-HCL (pH 8.4), 1.5–2.5 mM MgCl<sub>2</sub>, 220  $\mu$ M each dNTP, 0.25  $\mu$ M 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°, followed by 12 cycles of touchdown PCR (a 30-sec denaturation step at 94°, an annealing step with temperature decreasing from 66° to 60°, and a 3-min extension step at 72°) and by 35 cycles of a 30-sec denaturation step at 94°, then

a 30-sec annealing step at 58°, and a 3-min extension at 72°, with a final extension of 10 min at 72°.

Single-locus products were sequenced for both DNA strands using BigDye Terminator v3.0 cycle sequencing reaction kits (Applied Biosystems, Foster City, CA) and an automated ABI Prism 3700 Genetic Sequencer (Applied Biosystems). DNA sequence strands were analyzed with SEQMAN software (SWINDELL and PLASTERER 1997). 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; PAVY *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 (Illumina, 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 (PAVY *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 (MARTH *et al.* 1999), with a *p*-prior of 0.01 and a probability that a position represents one SNP ( $P_{\text{SNP}} \geq 0.99$ ) (PAVY *et al.* 2006). A total of 727 SNPs among the *in silico* SNPs identified were selected 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 (HUERTZ *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/ $\mu$ l) 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 (Illumina), 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; PAVY *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 (PAVY *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 orthologous 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



using the squared allelic correlation coefficient ( $r^2$ ). The decay of LD with distance of base pairs between sites within the same candidate gene was estimated with both a linear model and a nonlinear regression, using the `nls` function in the `nlme` package of the R software v2.12.2 (HILL and WEIR 1988; R DEVELOPMENT CORE TEAM 2010).

**Population structure:** Associations between markers and traits are susceptible to errors and biases caused by covariance between the marker and polygenic effects (KENNEDY *et al.* 1992). Although population structure is rare in outbred forest species (NEALE and SAVOLAINEN 2004), it is considered the most serious systematic bias producing false-positive associations (MARCHINI *et al.* 2004). When population structure is present, bias occurs because subgroups of relatives tend to share more markers and gene alleles genome-wide (BRESEGHELLO and SORRELLS 2006) than a pair of individuals drawn at random from the population.

White spruce has previously been shown to harbor little or no population genetic structure for neutral markers in the region sampled (JARAMILLO-CORREA *et al.* 2001; NAMROUD *et al.* 2008, 2010). The presence of population structure was further checked for our data set using the model-based clustering algorithm implemented in the software STRUCTURE, version 2.2 (PRITCHARD *et al.* 2000, 2007). The entire population of 469 trees was analyzed, assuming that the 944 SNP loci were unlinked. Contrary to previous studies (see, for instance, GONZÁLEZ-MARTÍNEZ *et al.* 2007), the discovery population was not genotyped using another class of neutral genetic markers (*e.g.*, microsatellites) to test for population structure, as the large number of available SNPs was considered sufficient. The correlated allele frequency model was used as it is likely the most appropriate as white spruce populations are almost always expected to be connected by gene flow. This model implies that allele frequencies in each putative cluster are likely to be similar (FALUSH *et al.* 2003). Preliminary tests showed that STRUCTURE was detecting substructures partially confounded with the half-sib structure. Furthermore, as  $K$  values were increasing, log probability of data  $\Pr(X|K)$  (Equation 12 in PRITCHARD *et al.* 2000) never reached a maximum value, which is needed to determine the uppermost level of structure. To avoid the detection of substructure confounded with half-sib structure, we randomly selected one individual in each of the 165 open-pollinated families making up the discovery population and analyzed this subset with STRUCTURE to determine the uppermost hierarchical level of structure in our discovery population. Two other subsets containing one tree per open-pollinated family were also composed, using the remaining trees. Then, 20 replicate runs of STRUCTURE analyses at the uppermost level of structure found with the first subset were performed for each of the three subsets, and the software CLUMPP (JAKOBSSON and ROSENBERG 2007) was used to rank the clusters and estimate  $Q$  coefficients' averages over replicate runs.

**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

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 SPAGeDi, version 1.2 (HARDY 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,

$$y = S\alpha + Qv + Zu + e, \quad (1)$$

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 error 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  $E(u) = E(e) = 0$  and thus  $E(y) = S\alpha + Qv$ , where random additive and residual variances are assumed to be  $\text{Var}(u) = G = 2K\sigma_a^2$  and  $\text{Var}(e) = R = I\sigma_e^2$ , 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.

Tissue samples were obtained from 3-year-old *in vitro* clonally propagated white spruce seedlings of one genotype (see PAVY *et al.* 2008a for details). All the tissues were frozen in liquid nitrogen. Total RNA was isolated following the method of CHANG *et al.* (1993) with modifications described in PAVY *et al.* (2008a). The custom microarray was composed of a total of 32,640 oligonucleotides, 25,094 of which represented unique *P. glauca* gene sequences. A description of the sequence data is available at <http://www.arborea.ca>. Microarray hybridizations used five independent samples of each tissue type; two samples labeled with different dyes were hybridized to each slide (dye swaps) and image analyses proceeded following PAVY *et al.* (2008a). Data analysis was carried out using

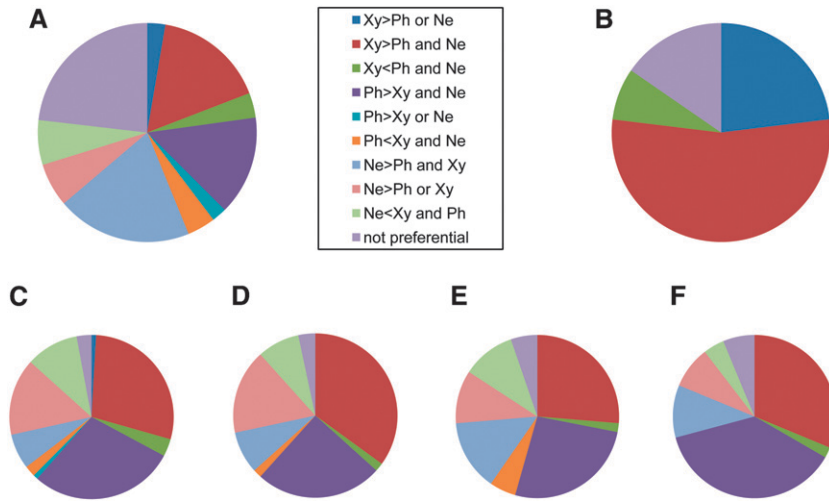


FIGURE 1.—Distribution of genes into tissue-preferential transcript accumulation classes. (A–F) Each chart represents a different set of gene sequences (Xy, xylem; Ph, phloem; Ne, needles). (A) Whole-transcriptome microarray (25,094 genes). (B) Genes containing a SNP significantly associated with wood property traits (13 genes). (C) Genes containing a SNP submitted to association analysis (549 genes). (D–F) Genes containing a SNP significantly associated with wood traits at  $P \leq 0.05$ : (D) MFA (60 genes), (E) tracheid cell diameter in radial direction (RCD) (57 genes), and (F) DEN (48 genes).

BIOCONDUCTOR (<http://www.bioconductor.org>) in the R statistical environment and following the method described in PAVY *et al.* (2008a). 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/limma/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 in-

dividually sampled by removing three small plugs of  $\sim 1$  cm<sup>2</sup> 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  $P_r(X|K)$  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

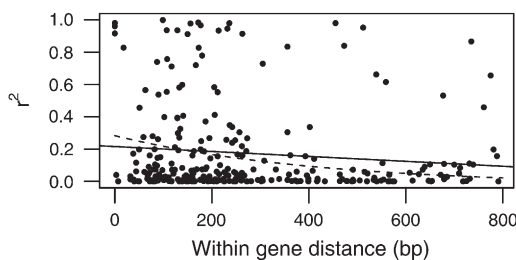


FIGURE 2.—Patterns of intralocus linkage disequilibrium (LD) for white spruce estimated using the squared allelic correlation coefficient between each pair of SNPs within candidate genes.

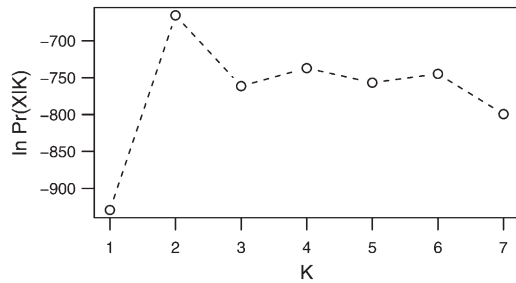


FIGURE 3.—Plot of the mean of  $L(K)$  over 20 runs for each of  $K$  clusters ranging from 1 to 7 as a function of the number of clusters. A constant of 15,500 was removed on the  $y$ -axis.

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 ( $P \leq 0.01$ ) 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  $Q \leq 0.10$ , none of these associations were still significant. For the individual wood traits, a total of 23,600 (944 SNPs  $\times$  25 traits) and 22,400 (896 SNPs  $\times$  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  $P \leq 0.05$ , 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

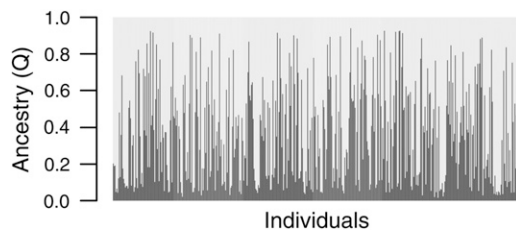


FIGURE 4.—Genetic clustering of the 469 white spruce individuals of the discovery population using the program STRUCTURE. Vertical bars represent individuals, and shaded and solid areas represent proportional membership of each individual in each of the two clusters. Individuals are ranked by latitude and longitude.

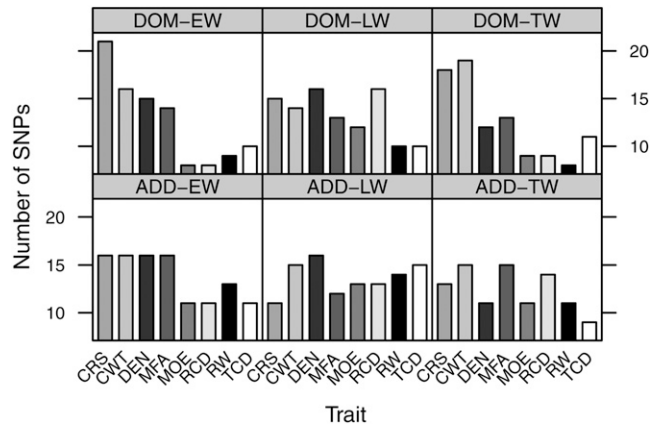


FIGURE 5.—Number of significant ( $P \leq 0.01$ ) SNP associations before correction for multiple testing for eight early- (EW), late- (LW), and total wood (TW) traits (see Table 1) in white spruce. Associations were tested using an additive (ADD) or dominant (DOM) effects mixed linear model.

(ECKERT *et al.* 2009). They are also close to the 8.3% of significant associations (106 of 1276 tests) found by GONZÁLEZ-MARTÍNEZ *et al.* (2007) for wood traits in *P. taeda*.

The number of significant associations was slightly reduced when using a nominal threshold of  $P \leq 0.01$ . Hence, for each of the wood traits assessed, we found 8–19 significant SNPs for earlywood, 11–18 SNPs for late-wood, and 7–18 SNPs for total wood, depending on the genetic effects model (Figure 5); this represents an overall 1.5% of significant association tests and the numbers of SNPs associated with at least one wood trait were 136 (14%) and 148 (17%), for the additive and dominant effects models, respectively. The phenotypic variation explained by each of these significant SNPs varied between 1.5% and 5.4%. The total variation in wood traits that could be explained by considering all these SNPs simultaneously if we were assuming an additive effects model and the absence of interaction between significant SNPs could sum up to 60%. However, the exact percentage could be obtained only when considering simultaneously all these SNPs in a MLM analysis.

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

**TABLE 2**  
**Significant associations between gene SNPs and wood traits in white spruce after correction for multiple testing**

Trait <sup>a</sup>	SNP	Gene ID	Functional annotation	Population structure			Kinship structure <sup>b</sup>			Half-sib structure <sup>c</sup>						
				F	P		F	P		F	P	Q	R <sup>2d</sup>	No. homo1	No. hetero	Lnmean homo1
Earlywood	ARW	PGWD1-1094	GQ03806_H09	0.00	0.9990	17.54	<0.0001	16.87	5.2 × 10 <sup>-5</sup>	0.0487	0.035	123	346	2.50	2.77	BT116774
	CWT	PGWD1-1396	GQ02829_F04	0.12	0.7263	16.25	<0.0001	17.88	4.0 × 10 <sup>-5</sup>	0.0370	0.034	39	430	1.86	1.77	BT105847
PCT	PGWD1-1096	GQ04010_I15	Pectinesterase	0.29	0.5875	12.20	<0.0001	11.36	0.0008	0.0932	0.026	324	179	1.75	1.79	BT118812
	PGWD1-1282	GQ03113_N22	MYB4	0.01	0.9423	9.05	0.0001	8.83	0.0002	0.0437	0.038	64	232	1.78	1.79	BT107883
	PGWD1-0511	GQ0224_E23	Galactosyl-transferase	2.83	0.0930	8.08	0.0004	8.29	0.0003	0.0561	0.037	40	178	82.4	80.9	BT103289
Latewood	PGWD1-1313	GQ03211_001	Receptor-like kinase	1.00	0.3167	12.03	<0.0001	12.61	3.6 × 10 <sup>-6</sup>	0.0018	0.054	1	49	81.6	81.5	BT109734
	PGWD1-0560	GQ03010_F13	ACC oxidase	0.98	0.3239	12.38	<0.0001	12.51	4.3 × 10 <sup>-6</sup>	0.0018	0.054	1	151	81.5	81.5	BT106780
	PGWD1-1035	GQ0172_022	β-Expansin	1.66	0.1981	7.25	0.0008	7.21	0.0008	0.0841	0.032	41	209	83.8	81.3	CO47855.2
ARW	PGWD1-0863	GQ03005_G12	Tubulin, TUA2	2.20	0.1389	7.91	0.0004	8.17	0.0003	0.0875	0.039	4	55	0.62	0.59	BT106568
	PGWD1-1313	GQ03211_001	Receptor-like kinase	2.09	0.1492	9.23	0.0001	9.34	0.0001	0.0437	0.045	1	49	0.60	0.60	BT109734
	PGWD1-0581	GQ03006_P17	Unknown	2.84	0.0928	10.01	<0.0001	9.01	0.0001	0.0437	0.043	1	26	0.57	0.60	BT106529
Total wood	ARW	PGWD1-0560	GQ03010_F13	1.59	0.2076	10.52	<0.0001	10.55	3.1 × 10 <sup>-4</sup>	0.0279	0.050	1	151	0.82	0.82	BT106780
	CWT	PGWD1-1094	GQ03806_H09	0.15	0.6965	17.79	<0.0001	16.67	5.2 × 10 <sup>-5</sup>	0.0487	0.035	123	146	3.08	3.38	BT116774
CWT	PGWD1-1396	GQ02829_F04	Glycoside hydrolase	0.38	0.5385	15.58	4.3 × 10 <sup>-5</sup>	17.04	4.3 × 10 <sup>-5</sup>	0.0390	0.033	39	430	1.86	1.77	BT105847
	PGWD1-1096	GQ04010_I15	Pectinesterase	0.60	0.4379	13.25	0.0003	12.76	0.0004	0.0914	0.029	324	113	1.90	1.95	BT118812

(continued)



**TABLE 2**  
(Continued)

Trait <sup>a</sup>	SNP	Gene ID	Functional annotation	Population structure		Kinship structure <sup>b</sup>		Half-sib structure <sup>c</sup>								
				F	P	F	P	F	P	Q	R <sup>2d</sup>	No. homo1 and hetero	No. homo2	Lsmean hetero	Lsmean homo2	GenBank accession
Earlywood	CWT	PGWDJ-1282	GQ03113_N22	0.00	0.9839	17.48	<0.0001	17.01	4.2 × 10 <sup>-5</sup>	0.0374	0.036	64	399	1.72	1.79	BT107883
	CRS	PGWDJ-1282	GQ03113_N22	0.92	0.3390	16.28	<0.0001	16.07	6.2 × 10 <sup>-5</sup>	0.0478	0.037	64	399	255.0	268.9	BT107883
	PCT	PGWDJ-0354	GQ04010_P08	0.59	0.4441	13.34	0.0003	14.12	0.0002	0.0411	0.032	2	466	70.9	81.5	BT107883
		PGWDJ-1035	GQ0172_O22	0.99	0.3215	14.48	0.0002	14.46	0.0002	0.0274	0.032	41	427	83.8	81.3	CO47855.2
Latewood	ARW	PGWDJ-0863	GQ03005_G12	2.16	0.1427	14.01	0.0002	14.18	0.0002	0.0395	0.035	4	464	0.86	0.59	BT106568
	CWT	PGWDJ-0354	GQ04010_P08	2.01	0.1568	11.61	0.0007	11.71	0.0007	0.0947	0.030	2	466	0.96	0.60	BT107883
	MFA	PGWDJ-0107	GQ0133_K12	0.96	0.3279	14.14	0.0002	13.92	0.0002	0.0864	0.031	3	466	27.5	19.0	BT101933
Total wood	RCD	PGWDJ-1070	GQ02908_P24	2.93	0.0875	15.54	<0.0001	16.10	7.5 × 10 <sup>-5</sup>	0.0656	0.040	328	138	23.0	23.6	BT101219
	CWT	PGWDJ-1282	GQ03113_N22	0.03	0.8584	14.16	0.0002	13.76	0.0002	0.0871	0.030	64	399	1.88	1.95	BT107883
	CRS	PGWDJ-1282	GQ03113_N22	1.05	0.3054	15.41	<0.0001	15.09	0.0001	0.0742	0.035	64	399	265.9	279.4	BT107883

Dominance effects model

<sup>a</sup> See Table 1 for definition of abbreviations.

<sup>b</sup> Nason's kinship coefficients (SPAGeDi) are used to take account of the relatedness of the trees making up the discovery population.

<sup>c</sup> Half-sib coefficients are used to take account of the known relatedness of the trees making up the discovery population.

<sup>d</sup> R<sup>2</sup> 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 <sup>a</sup>	Needle vs. xylem <sup>b</sup>		Phloem vs. xylem		Needle vs. phloem		Tissue profile <sup>d</sup>
			FD <sup>c</sup>	P	FD <sup>c</sup>	P	FD <sup>c</sup>	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_F23	Galactosyl-transferase, XT2	PCT	-1.13	0.002	-1.11	0.009	-1.02	0.594	Xy > Ne and Ph
GQ02908_P24	β-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_H15	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
GQ0133_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

<sup>a</sup>See Table 1 for abbreviated trait names.

<sup>b</sup>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.

<sup>c</sup>FD, fold difference. The sign of the fold difference (+ or -) indicates preferential accumulation in the first tissue/organ (+) or the second one (-).

<sup>d</sup>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	Trait <sup>a</sup>	No. of trees tested <sup>b</sup>			RNA transcript molecules <sup>c</sup>			Significance ( <i>P</i> ) <sup>d</sup>	
			Homo1	Hetero	Homo2	Homo1	Hetero	Homo2	Additive	Dominance
<i>GQ0172_O22</i>	β-Expansin	PCT	11	16	16	539	686	750	0.245	0.384
<i>GQ03010_F13</i>	ACC oxidase	PCT, ARW		25	12		87	122	0.030*	NA
<i>GQ03113_N22</i>	MYB4	CWT, CRS	17	22	17	1488	1467	1476	0.158	0.684
<i>GQ02908_P24</i>	β-TIP	RCD	10	8	8	2458	1718	2769	0.075	0.026* <sup>e</sup>
<i>GQ02829_F04</i>	Glycoside hydrolase family 28	CWT		8	9		133	144	0.201	NA
<i>GQ03806_H09</i>	LACS9	ARW		8	14		744	1230	0.015*	NA

<sup>a</sup>See Table 1 for abbreviated trait names.

<sup>b</sup>Transcript levels were assessed in small plugs of differentiating xylem removed from the tree at 1.3 m from the ground.

<sup>c</sup>RNA transcripts are the average number of RNA molecules per 1 μg of total RNA, after normalization (see MATERIALS AND METHODS).

<sup>d</sup>Statistical 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  $P \leq 0.05$ .

<sup>e</sup>Statistically significant when testing for underdominance.

should not be considered *a priori* as potential false positives because they could affect transcript levels and codon usage (e.g., KIMCHI-SARFATI *et al.* 2007; CHAMARY and HURST 2009). While any of these SNPs could be quantitative trait nucleotides (QTNs), they could also be linked to adjacent causal but untested SNPs.

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  $Q \leq 0.10$ . 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; GONZÁLEZ-MARTÍNEZ *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

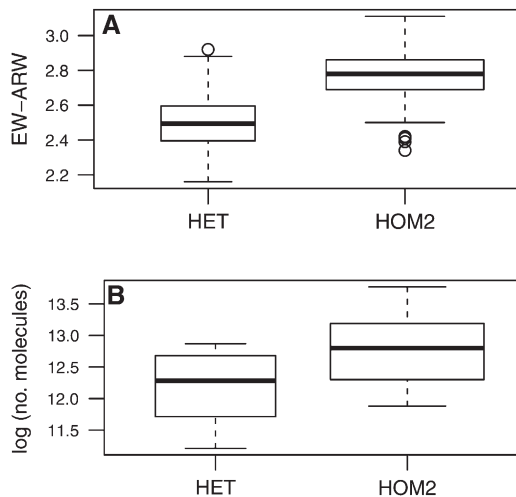


FIGURE 6.—Relationship between SNP genotypic classes and transcript accumulation. (A) Estimated genotypic effects of a significant SNP (*PGWDI-1094*) for earlywood average ring width (and TW-ARW, see Table 2) in the additive effects model. (B) Transcript levels of the corresponding gene determined by 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\_022*), 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 4133 (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 asso-

ciation, 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*,  $\beta$ -*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,

and phenotypic variation in a forest tree population. These authors identified one SNP in a COBRA-like gene of *E. nitens* (EniCOBL4A) that was associated with allelic expression imbalance and with the amount of 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 the synthesis of phospholipids and triacylglycerols in vegetative and seed tissues (PONGDONTRI and HILLS 2001; SHOCKEY *et al.* 2002). They play essential roles 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 (SHOCKEY *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 *LACS9* 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 (*PWD1-1094*) 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 family members were represented on the genotyping chip and all but one of them accumulated transcripts preferentially in phloem (four) or xylem (two). The SNP (*PGWD1-1396*) is a synonymous substitution (TGC–TGT) coding for a cysteine. It was represented by only one class of homozygous individuals in the discovery population. On average, the homozygotes had thinner cell walls than the heterozygotes (1.77  $\mu\text{m}$  *vs.* 1.86  $\mu\text{m}$ ); the transcript levels did not differ between genotypes.

*Glycoside hydrolase family 10*: The glycoside hydrolase protein family 10 (GH10), formerly known as cellulase family F, comprises several enzymes with known activities.

They play important roles in primary cell wall metabolism, particularly in the reassembly and degradation process (GIRKE *et al.* 2004) and in secondary cell wall metabolism (ICHINOSE *et al.* 2010). The *Populus* genome encodes seven GH10 xylanases, but only *PttXyn10A* had detectable ESTs in the wood-forming tissue (GEISLER-LEE *et al.* 2006) and was highly upregulated in different stages of wood formation (ASPEBORG *et al.* 2005; TAKAHASHI-SCHMIDT 2008). Only one family member was included in our genotyping chip. The SNP (*PGWD1-0107*) consisted of a nonsynonymous substitution (CCA–GCA) coding for proline and alanine that was found to be associated with latewood MFA in a dominance effects model.

*Pectinesterase*: Pectin methylesterases (PME) are known to catalyze the demethylesterification of cell wall polygalacturonans, one of the major polysaccharides in the primary wall and middle lamella. In dicotyledonous plants, these ubiquitous cell wall enzymes are involved in important developmental processes including cellular adhesion and stem elongation (MICHELI 2001). It has been suggested that the prevalence of neutral PMEs in active cells would allow cell wall expansion (GUGLIELMINO *et al.* 1996). Pectinesterase transcripts were among the differentially regulated and most strongly expressed genes in tension wood of *Populus* (ANDERSSON-GUNNERAS *et al.* 2006) and in the juvenile cambial region of *E. grandis* (GALLO DE CARVALHO *et al.* 2008). A pectin methyl esterase from wood-forming tissues of hybrid aspen (*P. tremula x tremuloides* Michx.) was found to act as a negative regulator of both symplastic and intrusive growth of developing wood cells (SIEDLECKA *et al.* 2008). Fourteen different pectinesterase sequences were present on our 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 (TGC–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. (GOICOECHEA *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 (GCA–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  $\mu\text{g}/\text{m}$ ) and for earlywood (269  $\mu\text{g}/\text{m}$ ) than the alternate homozygous type (266  $\mu\text{g}/\text{m}$  and 255  $\mu\text{g}/\text{m}$ , 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-xylosyltransferase 2)*: The hemicellulose xyloglucan (XyG) is believed to play an important role in cell wall structure and function (ECKARDT 2008). A number of enzymes participate in XyG biosynthesis, including  $\beta$ -(1/4)-glucan synthase,  $\alpha$ -fucosyltransferases,  $\beta$ -galactosyltransferases, and  $\alpha$ -xylosyltransferases (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 (CAVALIER *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 (SHIU and BLECKER 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 (SCHRADER *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  $\text{CO}_2$  activity. The closest ACC oxidase homolog in *Arabidopsis* was regulated during *in vitro* xylem vessel element differentiation (KUBO *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* Dougl. 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).

$\beta$ -*Expansin [AtEXPB3 (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.  $\beta$ -Expansins act through cell wall loosening in a variety of developmental functions. They have been mostly described in grasses; 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  $\beta$ -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 sequenced 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*  $\alpha$ -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  $\beta$ -tubulin gene was shown to be preferentially expressed in xylem in *Eucalyptus* (PAUX *et al.* 2004).  $\beta$ -Tubulin (*EgrTUB1*) was differentially regulated in *Eucalyptus* branches and was reported to be likely involved in cell wall biosynthesis (QIU *et al.* 2008). The TUA2 SNP (*PGWD1-0863*) represents a synonymous substitution (GGA–GGC) coding for glycine in a tubulin, whereas the TUB3 SNP (*PGWD1-*

0354) is located in the 3'-UTR (T-C). Both of the SNPs were associated with latewood ARW; the SNP in *TUB3* was also associated with earlywood percentage (PCT) (Table 2).

**$\beta$ -TIP ( $\beta$ -TONOPLAST INTRINSIC PROTEIN):** Membrane intrinsic proteins form a family of 35 genes in the Arabidopsis genome, and 13 of them encode tonoplast intrinsic proteins (TIPs) that localize in the vacuolar membrane (tonoplast) (JOHANSON *et al.* 2001). Most TIPs are classified as probable or confirmed aquaporins. In Arabidopsis, both the  $\beta$ -TIP (also known as TIP3,2) and the highly similar  $\alpha$ -TIP (TIP3,1) are expressed preferentially in germinating seeds, while other classes of TIPs are strongly expressed during seed maturation and dehydration (VANDER WILLIGEN *et al.* 2006). In *P. taeda*, ESTs of a TIP and an aquaporin were overrepresented in a xylem cDNA library made from compression wood compared to normal wood (WHETTEN *et al.* 2001). Compression wood produces tracheids with altered cell shape and thicker and more lignified secondary cell walls. We found that  $\beta$ -TIP accumulated preferentially in xylem and phloem compared to needles. The SNP (*PGWD1-1070*) was located in an intron (T\*-C\*) and was associated with the tracheid diameter in the latewood (dominance model). The homozygote (homo2) had an average diameter of 23.6  $\mu\text{m}$ , whereas the other genotypic class (heterozygotes and homo1) had a slightly lower average diameter of 23  $\mu\text{m}$  (Table 2). The transcript levels also differed significantly between genotypic classes but did not appear to be positively correlated with the wood phenotype. The transition from earlywood to latewood is strongly influenced by water availability, and osmotic stress-related transcripts including aquaporins have been found to accumulate in latewood (EGERTSDOTTER *et al.* 2004). The final latewood morphology appears to involve  $\beta$ -TIP but its functional role remains to be tested.

**Methodological and statistical challenges:** Association studies in conifer populations have many attractive features because discovery populations with weak or nonexistent population structure and high nucleotide diversity can be assembled (NEALE and SAVOLAINEN 2004); however, several challenges remain to be overcome. This study shows that a larger number of less frequent genotypes should be captured by increasing the sample size to better estimate their effects. A larger sample size may also offer the opportunity to include experimental design features in the model (*e.g.*, blocks) so that estimates of genetic effects are adjusted accordingly. Second, additional information on marker covariance structure can be accommodated in the model only if significant LD is present among markers and QTL. In natural conifer populations, LD is expected to sink to very low levels over distances that are spanned of average-sized genes (*e.g.*,  $r^2 \leq 0.2 \sim 100$  bp in our data set), making the discovery of additional SNPs unlikely to yield strong LD across candidate genes. More recently,

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 (CALUS *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 (GONZÁLEZ-MARTÍNEZ *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 Arborea 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. Lamothe, J. Fillon and J.-P. Dionne for ideas and support regarding the bioinformatics aspects. We are indebted to P. Watson of Canfor Pulp Limited Partnership who was involved in the early stages of the project and to the team of A. Montpetit at the Illumina genotyping platform of the McGill University and Genome Quebec Innovation Centre. We thank S. Lebihan, C. Collins, and C. Nelson of the British Columbia Microarray Facility at the Vancouver Prostate Centre for printing the microarray slides. This research was supported by funding to the Arborea project from Genome Canada, Genome Quebec, the Canadian Forest Service of Natural Resources Canada (Canadian Wood Fibre Centre and Genomics Research and Development Initiative), and FPInnovations.

#### LITERATURE CITED

- ALDEN, J., and C. LOOPSTRA, 1987 Genetic diversity and population structure of *Picea glauca* on an altitudinal gradient in interior Alaska. *Can. J. For. Res.* **17**: 1519–1526.
- ANDERSSON-GUNNERAS, S., E. J. MELLEROWICZ, J. LOVE, B. SEGERMAN, Y. OHMIYA *et al.*, 2006 Biosynthesis of cellulose-enriched tension wood in *Populus*: global analysis of transcripts and metabolites identifies biochemical and developmental regulators in secondary wall biosynthesis. *Plant J.* **45**: 144–165.
- ASPEBORG, H., J. SCHRADER, P. M. COUNTINHO, M. STAM, A. KALLAS *et al.*, 2005 Carbohydrate-active enzymes involved in the secondary cell wall biogenesis in hybrid aspen. *Plant Physiol.* **137**: 983–997.
- BEAULIEU, J., 2003 Genetic variation in tracheid length and relationships with growth and wood traits in eastern white spruce (*Picea glauca*). *Wood Fib. Sci.* **35**: 609–616.
- BEDON, F., J. GRIMA-PETTENATI and J. MACKAY, 2007 Conifer R2R3-MYB transcription factors: sequence analyses and gene expression in wood-forming tissues of white spruce (*Picea glauca*). *BMC Plant Biol.* **7**: 17.
- BENJAMINI, Y., and Y. HOCHBERG, 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**: 289–300.
- BOMAL, C., F. BEDON, S. CARON, S. D. MANSFIELD, C. LEVASSEUR *et al.*, 2008 Involvement of *Pinus taeda* MYB1 and MYB8 in phenylpropanoid metabolism and secondary cell wall biogenesis: a comparative in *planta* analysis. *J. Exp. Bot.* **59**: 3925–3939.
- BOYLE, B., N. DALLAIRE and J. MACKAY, 2009 Evaluation of the impact of single nucleotide polymorphisms and primer mismatches on quantitative PCR. *BMC Biotechnol.* **9**: 75.
- BOUSQUET, J., N. ISABEL, B. PELGAS, J. COTTRELL, D. RUNGIS *et al.*, 2007 Spruce, pp. 93–114 in *Forest Trees* (Genome Mapping and Molecular Breeding in Plants, Vol. 7), edited by C. KOLE. Springer-Verlag, Berlin.
- BRADSHAW, H. D., and R. F. STETTNER, 1995 Molecular genetics of growth and development in *populus*. 4. Mapping QTLs with large effects on growth, form, and phenology traits in a forest tree. *Genetics* **139**: 963–973.
- BRESEGHELLO, F., and M. E. SORRELLS, 2006 Association analysis as a strategy for improvement of quantitative traits in plants. *Crop Sci.* **46**: 1323–1330.
- 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.
- CALUS, M. P. L., T. H. E. MEUWISSEN, A. P. W. DE ROOS and R. F. VEERKAMP, 2008 Accuracy of genomic selection using different methods to define haplotypes. *Genetics* **178**: 553–561.
- CATO, S., L. McMILLAN, L. DONALDSON, T. RICHARDSON, C. ECHT *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.
- CAVALIER, D. M., O. LEROUXEL, 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. PURYEAR and J. CAIRNEY, 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**: 1365–1369.
- CORRIVEAU, A., J. BEAULIEU and G. DAoust, 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. X. 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. WEGRZYN, B. PANDE, K. D. JERMSTAD *et al.*, 2009 Association genetics of coastal Douglas fir (*Pseudotsuga menziesii* var. *menziesii*, Pinaceae). I. Cold-hardiness related traits. *Genetics* **182**: 1289–1302.
- 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.
- FALUSH, D., M. STEPHENS and J. K. PRITCHARD, 2003 Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**: 1567–1587.
- FAN, J.-B., A. OLIPHANT, 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.
- FRONTIER, 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.
- FURNIER, 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*. *Tree Physiol.* **28**: 905–919.
- GEISLER-LEE, J., M. GEISLER, P. M. COUNTINHO, B. SEGERMAN, N. NISHIKUBO *et al.*, 2006 Poplar carbohydrate-active enzymes. Gene identification and expression analyses. *Plant Physiol.* **140**: 946–962.
- GIRKE, T., J. LAURICHA, H. TRAN, K. KEEGSTRA and N. RAIKHEL, 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.
- GODDARD, M. E., and B. J. HAYES, 2009 Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat. Rev. Genet.* **10**: 381–391.
- GOICOECHEA, 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.



- GONZÁLEZ-MARTÍNEZ, S. C., N. C. WHEELER, E. ERSOZ, C. D. NELSON and D. B. NEALE, 2007 Association genetics in *Pinus taeda* L. I. Wood property traits. *Genetics* **175**: 399–409.
- GRATTAPAGLIA, D., and M. V. D. RESENDE, 2011 Genomic selection in forest tree breeding. *Tree Genet. Genomes* **7**: 241–255.
- GRATTAPAGLIA, D., F. L. G. BERTOLUCCI, R. PENCHEL and R. R. SEDEROFF, 1996 Genetic mapping of quantitative trait loci controlling growth and wood quality traits in *Eucalyptus grandis* using a maternal half-sib family and RAPD markers. *Genetics* **144**: 1205–1214.
- GRATTAPAGLIA, D., C. PLOMION, M. KIRST and R. R. SEDEROFF, 2009 Genomics of growth traits in forest trees. *Curr. Opin. Plant Biol.* **12**: 148–156.
- GROOVER, A., 2005 What genes make a tree a tree? *Trends Plant Sci.* **10**: 210–214.
- GROOVER, A., M. DEVEY, T. FIDDLER, J. LEE, R. MEGRAW *et al.*, 1994 Identification of quantitative trait loci influencing wood specific gravity in an outbred pedigree of loblolly pine. *Genetics* **138**: 1293–1300.
- GUGLIELMINO, N., M. LIBERMANN, A. M. CATESSON, A. MARECK, R. PRAT *et al.*, 1996 Pectin methylesterases from poplar cambium and inner bark: localization, properties and seasonal changes. *Planta* **202**: 70–75.
- 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. PRESTERL, Z. SUŠIĆ and T. MIEDANER, 2004 Plant genetic resources in crop improvement. *Plant Genet. Resour.* **2**: 3–21.
- HAYASHI, T., and R. KAIDA, 2011 Functions of xyloglucan in plant cells. *Mol. Plant* **4**: 17–24.
- HILL, W. G., and B. S. WEIR, 1988 Variances and covariances of squared linkage disequilibria in finite populations. *Theor. Popul. Biol.* **22**: 54–78.
- HIRSCHHORN, J. N., K. LOHMEYER, B. EDWARDS and K. HIRSCHHORN, 2002 A comprehensive review of genetic association studies. *Genet. Med.* **4**: 45–61.
- HOLLIDAY, J. A., K. RITLAND and S. N. AITKEN, 2010 Widespread, ecologically relevant genetic markers developed from association mapping of climate-related traits in Sitka spruce (*Picea sitchensis*). *New Phytol.* **188**: 501–514.
- HUERTZ, M., E. DE PAOLI, T. KÄLLMAN, H. LARSSON, I. JURMAN *et al.*, 2006 Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway Spruce [*Picea abies* (L.) Karst]. *Genetics* **174**: 2095–2105.
- ICHINOSE, H., N. NISHIKUBO, T. DEMURA and S. KANAKO, 2010 Characterization of  $\alpha$ -L-arabinofuranosidase related to the secondary cell walls formation in *Arabidopsis thaliana*. *Plant Biotechnol.* **27**: 259–266.
- INGVARSSON, P., M. V. GARCIA, V. LUQUEZ, D. HALL and S. JANSSON, 2008 Nucleotide polymorphism and phenotypic associations within and around the *phytochrome B2* locus in European aspen (*Populus tremula*, Salicaceae). *Genetics* **178**: 2217–2226.
- JAKOBSSON, M., and N. A. ROSENBERG, 2007 CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**: 1801–1806.
- JARAMILLO-CORREA, J. P., J. BEAULIEU and J. BOUSQUET, 2001 Contrasting evolutionary forces driving population structure at expressed sequence tag polymorphisms, allozymes and quantitative traits in white spruce. *Mol. Ecol.* **10**: 2729–2740.
- JOHANSON, U., M. KARLSSON, I. JOHANSSON, S. GUSTAVSSON, S. SJOVALL *et al.*, 2001 The complete set of genes encoding major intrinsic proteins in *Arabidopsis* provides a framework for a new nomenclature for major intrinsic proteins in plants. *Plant Physiol.* **126**: 1358–1369.
- KENNEDY, B. W., M. QUINTON and J. A. VAN ARENDONK, 1992 Estimation of effects of single genes on quantitative traits. *J. Anim. Sci.* **70**: 2000–2012.
- KIMCHI-SARFATI, C., J. M. OH, I.-W. KIM, Z. E. SAUNA, A. M. CALCAGNO *et al.*, 2007 A “silent” polymorphism in the *MDRI* gene changes substrate specificity. *Science* **315**: 525–528.
- KIRST, M., A. A. MYBURG, J. P. G. DE LEÓN, M. E. KIRST, J. SCOTT *et al.*, 2004 Coordinated genetic regulation of growth and lignin revealed by quantitative trait locus analysis of cDNA microarray data in an interspecific backcross of *Eucalyptus*. *Plant Physiol.* **135**: 2368–2378.
- KUBO, M., M. UDAGAWA, N. NISHIKUBO, G. HORIGUCHI, M. YAMAGUCHI *et al.*, 2005 Transcription switches for protoxylem and metaxylem 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.
- LOISELLE, 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 The 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.
- MICHELL, F., 2001 Pectin methylesterases: 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**: 1165–1166.
- MURRAY, B. G., 1998 Nuclear DNA amounts in gymnosperms. *Ann. Bot.* **82**: 3–15.
- MYLES, S., J. PEIFFER, P. J. BROWN, E. S. ERSOZ, 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**: 325–330.
- 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. 204–226 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  $\alpha$ - and  $\beta$ -tubulin gene families in *Populus*. *Plant Physiol.* **145**: 961–973.
- PAIVA, J., 2006 Phenotypic 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.*, 2003a 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.*, 2003b Characterisation of *PtMYB1*, an R2R3-MYB from pine xylem. *Plant Mol. Biol.* **53**: 597–608.
- PAUX, E., M. TAMASLOUKHT, N. LADOUCE, P. SIVADON and J. GRIMAPETTENATI, 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. GIGUÈRE *et al.*, 2008a Identification of conserved core xylem gene sets: conifer cDNA microarray development, transcript profiling and computational analyses. *New Phytol.* **180**: 766–786.
- PAVY, N., B. PELGAS, S. BEAUSEIGLE, 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. BEAUSEIGLE, V. ACHERÉ, S. JEANDROZ, J. BOUSQUET *et al.*, 2006 Comparative genome mapping among *Picea glauca*, *P. mariana* x *P. rubens* and *P. abies*, and correspondance 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**: 145.
- 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., 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.
- PLOMION, C., C. PIONNEAU, J. BRACH, P. COSTA and H. BAILLÈRES, 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. 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., 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.
- QIU, 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, Y. TSUMURA, B. PELGAS, N. ISABEL *et al.*, 2011 Genetic mapping in conifers, in *Genetics, Genomics and Breeding of Conifers*, edited by C. PLOMION, J. BOUSQUET and C. KOLE. CRC Press, New York (in press).
- ROZEN, S., and H. SKALETSKY, 2000 Primer3 on the WWW for general users and for biologist programmers, pp. 365–386 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. HEALD and P. F. WAREING, 1983 Gas chromatography-mass spectroscopy identification of 1-aminocyclopropane-1-carboxylic acid in compression wood vascular cambium of *Pinus contorta* Dougl. *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**: 1700–1709.
- SCHRADER, 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. 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.
- SHEN, R., J.-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): re22.
- 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. PÉRONNE, F. MICHELI, J. LESNIEWSKA *et al.*, 2008 Pectin methyl esterase inhibits intrusive and symplastic cell growth 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. 75–89 in *Sequence Data Analysis Guidebook*, edited by S. R. SWINDELL. Humana Press, Totowa, NJ.
- TAKAHASHI-SCHMIDT, J., 2008 Functional studies of selected extracellular 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, 2005 Polymorphisms in *Cinnamoyl 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. MEDER *et al.*, 2009 Identification of a *cis*-acting regulatory polymorphism in a eucalypt *COBRA*-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. BOURSIAIC 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. D. 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**: 145–156.
- 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. 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-Verlag, New York.

Communicating editor: M. KIRST

# GENETICS

## Supporting Information

<http://www.genetics.org/cgi/content/full/genetics.110.125781/DC1>

### **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****Supporting Files**

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.