ORIGINAL ARTICLE

Association genetics of carbon isotope discrimination, height and foliar nitrogen in a natural population of *Pinus taeda* L

WP Cumbie[1], A Eckert[2,3], J Wegrzyn[4], R Whetten[1], D Neale[4,5] and B Goldfarb[1]

[1]*Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC, USA;* [2]*Section of Evolution and Ecology, University of California at Davis, Davis, CA, USA;* [3]*Center for Population Biology, University of California at Davis, Davis, CA, USA;* [4]*Department of Plant Sciences, University of California at Davis, Davis, CA, USA and* [5]*Institute of Forest Genetics, USDA Forest Service, Davis, CA, USA*

Loblolly pine, *Pinus taeda* L., is one of the most widely planted, commercially and ecologically important tree species in North America. We took an association genetics approach, using an unimproved population of 380 clonally replicated unrelated trees, to test 3938 single nucleotide polymorphisms (SNPs) in as many genes for association with phenotypic variation in carbon isotope discrimination, foliar nitrogen concentration and total tree height after two growing seasons. Best linear unbiased prediction (BLUP) was used with a spatial adjustment to remove environmental variation from phenotypic data derived from a common garden experiment. After correction for multiple testing, a total of 14 SNPs were associated with the traits of carbon isotope discrimination ($n = 7$), height ($n = 1$) and foliar nitrogen concentration ($n = 6$) using 380 clones. Tails of the population phenotypic distribution were compared for allele frequency differences, revealing 10 SNPs with allele frequency in at least one tail significantly different from the overall population. Eight associated SNPs were in sequences similar to known genes, such as an AP2 transcription factor related to carbon isotope discrimination and glutamate decarboxylase associated with foliar nitrogen concentration, and others were from unknown genes without homologs in *Arabidopsis*.
Heredity (2011) **107**, 105–114; doi:10.1038/hdy.2010.168; published online 19 January 2011

Introduction

Understanding the relationship between genotype and phenotype is essential for the improvement of complex traits in economically important plant species. Improvements in genomic technology and knowledge gained from research in model organisms are creating opportunities for large-scale genomic research in commercially important species. Association genetics studies in humans have demonstrated the potential to discover DNA sequence variants that are correlated with disease phenotypes (Smith and Newton-Cheh, 2009). Both candidate gene and genome-wide approaches are potentially successful methods for the discovery of variants that are either causal, or are linked to, the causal variant for disease and quantitative traits (Hirschhorn and Daly, 2005). Association studies for herbaceous plants have been successful in identifying polymorphisms related to phenotypic variation in adaptive traits in *Arabidopsis* (Chan *et al.*, 2010), as well as economically important

traits in maize (Buckler *et al.*, 2009), sugar beet (Stich *et al.*, 2008b) and wheat (Jing *et al.*, 2007).

Forest trees have recently been used in several association studies (González-Martínez *et al.*, 2007, 2008; Ingvarsson *et al.*, 2008; Eckert *et al.*, 2009). Conifers are well-suited for use in association genetics studies because of their large random-mating populations, nucleotide diversity, rapid decay of linkage disequilibrium and haploid tissue obtainable from seeds (Neale and Savolainen, 2004). Economic and adaptive traits have been explored in conifers including Douglas-fir (*Pseudotsuga menziesii*) and loblolly pine, and in angiosperms such as *Populus* (Neale and Ingvarsson, 2008). Results of association studies for wood properties and carbon isotope discrimination have been published in loblolly pine revealing potential associations using a candidate gene approach (González-Martínez *et al.*, 2007, 2008).

The limitation of available water is a critical factor affecting plant growth and survival. The regulation and control of water use by plants is a complex system, and is difficult to measure in a high-throughput assay suitable for a population-scale study. Carbon isotope discrimination between atmosphere and plant organic material (Δ) has been used in a wide range of plant species to assess water use efficiency. Carbon isotopic composition of plant organic material has been established as an indirect but integrated measure of the ratio of photosynthetic

activity to stomatal conductance (Farquhar *et al*., 1989). The response by plants to limitations in water has implications for both an understanding of adaptive variation in natural populations, as well as economic importance for increasing production or yield in economically important plant species. Carbon isotope discrimination has also been used in crop species in an attempt to improve water use efficiency (Condon *et al*., 2004; Rebetzke *et al*., 2006; Anyia *et al*., 2007).

Genetic variation in carbon isotopes (Δ or δ) has been reported for several species of forest trees. Foliar carbon isotope discrimination was under moderate genetic control in *Picea mariana* and was strongly correlated with growth, making it a candidate for indirect selection to improve growth (Johnsen *et al*., 1999). Studies in *Araucaria cunninghamii* (Prasolova *et al*., 2000), *Pinus elliotii x Pinus caribea* hybrids (Prasolova *et al*., 2005) and *Pinus pinaster* (Brendel *et al*., 2002) report variable levels of inheritance for carbon isotope discrimination in foliage and wood samples where individual tree h^2 levels ranged from 0.07 to 0.72. Carbon isotope discrimination in loblolly pine seems to be under a lower level of genetic control when compared with *Picea*, *Araucaria* and *Pinus pinaster*. Recent work by Baltunis *et al*. (2008) reported a low individual tree heritability ($h^2 = 0.09$) from a clonally replicated, controlled mating design, field-grown experiment in loblolly pine. However, the population-wide genetic variation in carbon isotope discrimination has not been well documented in loblolly pine. Carbon isotope discrimination has also been the target trait in genetic marker-based studies in other forest trees. Brendel *et al*. (2002) reported four QTL explaining ~25% of the phenotypic variance for carbon isotope discrimination in *Pinus pinaster*. Previous association testing in loblolly pine found SNPs from four candidate genes (*dhn-1, sod-chl, wrky-like* and *lp5-like*) potentially associated with carbon isotope discrimination (González-Martínez *et al*., 2008), in which candidate sequences were selected based on putative functions in drought response. However, to date, a large number of potential loci have not been tested in a large population for water use efficiency in loblolly pine.

Previous analyses of carbon isotope discrimination in loblolly pine suggested that isotope ratios may be related to either water use efficiency or photosynthetic capacity (González-Martínez *et al*., 2007; Baltunis *et al*., 2008), thus traits related to growth should be measured to distinguish between SNPs influencing carbon isotope discrimination or growth traits. Prasolova *et al*. (2000) identified relationships in *Araucauria cunninghamii* between δ¹³C and tree growth and nitrogen concentration in the crown. The photosynthetic process in leaves uses a large proportion of foliar nitrogen (Evans, 1989), and foliar nitrogen concentration has been related to photosynthesis in loblolly pine (Springer *et al*., 2005), thus looking at foliar nitrogen content and tree growth with Δ¹³C is important to understand whether the isotopic variation relates to growth.

The objectives of this study were to estimate the variation in carbon isotope discrimination, as well as height and foliar nitrogen concentration in a population of unrelated *P. taeda* trees; and to use an association genetics approach to test 3938 SNPs for associations with the previously mentioned traits, estimate effects of associated SNPs, and compare their allele frequencies.

Materials and methods

Plant material

A total of 425 ($n = 425$) unrelated genotypes of loblolly pine were selected from the North Carolina State University Cooperative Tree Improvement Program and the Western Gulf Forest Tree Improvement Program at the Texas Forest Service (henceforth referred to as the NCSU + WG population). Trees were grown from seed from trees selected from natural stands during the first generation of improvement to represent the natural range of loblolly pine (Figure 1). A small number of plantation selection seedlots, which had known-seed sources were used to cover geographical areas from which natural stand selections were not available. All seeds were open-pollinated, except for a few control-pollinated seedlots. Control-pollinated seedlots came from 2nd- or 3rd-cycle mating, but care was taken to ensure no relatedness to other entries in the population. Trees were grown from seed for 1 year and then hedged for stem cutting production using established methods for loblolly pine (Lebude *et al*., 2004). In the spring of 2006, rooted cuttings of each genotype were planted in a raised nursery bed comprised of a loamy sand from the coastal plain of North Carolina with a soil texture composed of 85% sand, 12.2% silt and 2.8% clay (Gocke, 2006). The bed was 1.5×40 m with a soil depth of 30 cm and a randomized complete block design was used, with two replicates planted per clone. Trees were planted at a 22-cm spacing in the nursery bed.

The trees were grown for two growing seasons and foliage was collected for carbon isotope discrimination after the end of the second growing season in December 2007. Annual rainfall during 2007 was 69% of 30-year normal rainfall amounts (64 cm compared with the normal of 92 cm), so foliage produced during that growing season developed under drought conditions (http://www.nc-climate.ncsu.edu). The entire second flush of foliage from 2007 was harvested from each tree. Harvested tissue was dried at 65 °C for 48 h and isotope analysis was performed at the COIL Cornell stable isotope facility (http://www.cobsil.com). Samples were pulverized using a Spex certiprep freezer/mill and analyzed using a Thermo Delta V Advantage Isotope Ratio Mass Spectrometer (Thermo Fisher Scientific, Waltham, MA, USA). The relative abundance of ¹³C to ¹²C was determined using 3 mg samples of pine needle tissue. The carbon isotope ratio δ¹³C was reported against a standard of Pee Dee Belemnite (Craig, 1954) and then converted to Δ, the carbon isotope discrimination value using

$$\Delta = \frac{\delta_a - \delta_p}{1 + \delta_p} \tag{1}$$

where δ_a is the atmospheric isotope composition (assumed to be −8‰) and δ_p is the leaf tissue isotope composition. Foliar nitrogen was estimated from mass spectroscopy (as a mass percentage) at COIL using the same foliage samples used for isotopic analysis. Total tree height (cm) was measured after the second growing season from the soil level to the top of the terminal bud.

Analysis of phenotypic data
A two-stage approach was used for association testing. Phenotypic data were analyzed using a mixed model
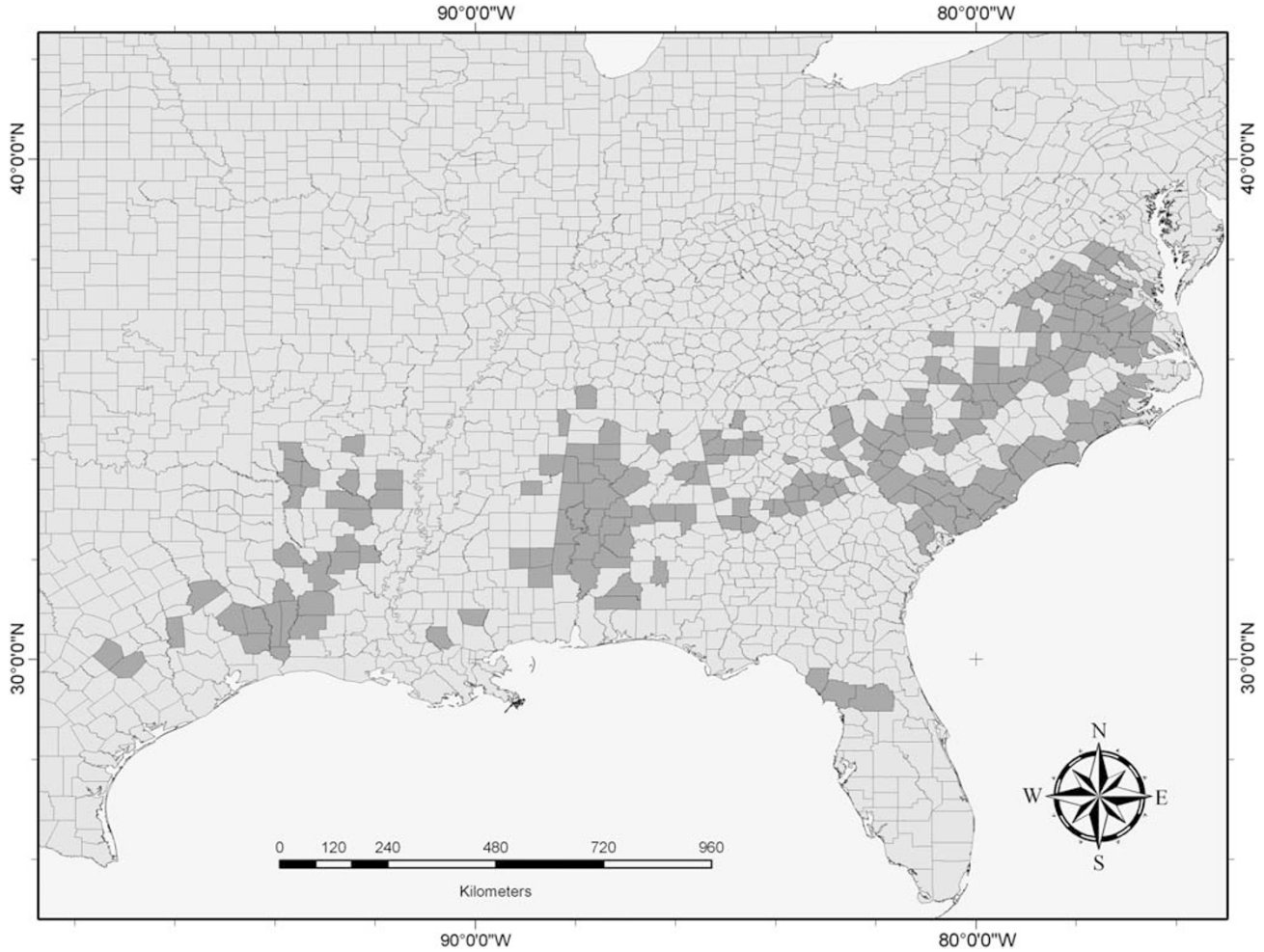
Figure 1 Counties of origin for genotypes selected for the NCSU association population. Shaded counties represent at least one genotype.

with the standard form:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e} \qquad (2)$$

where $\mathbf{y}$ is the vector of response variables (observations); $\mathbf{X}$ is the design matrix relating individual observations to the fixed effects in the model and $\mathbf{b}$ is the vector of fixed effect factors that includes the overall mean, site and replication within site effects; $\mathbf{Z}$ is the incidence matrix relating observations to random effects and $\mathbf{a}$ is the vector of random effects that includes clone effects and interactions with blocks; and $\mathbf{e}$ is the vector of random residual terms. The expectation of fixed terms is $E(\mathbf{y}) = \mathbf{Xb}$. The random terms are assumed to have zero means and variances $\text{var}(\mathbf{a}) = \mathbf{G}$, $\text{var}(\mathbf{e}) = \mathbf{R}$. The variance-covariance matrix of observations (vector y) is $\text{Var}(\mathbf{y}) = \mathbf{ZGZ}^T + \mathbf{R}$ where $\mathbf{G}$ and $\mathbf{R}$ are covariance matrices corresponding to $\mathbf{a}$ and $\mathbf{e}$, respectively. The $\mathbf{G}$ matrix accounts for the genetic effects, while $\mathbf{e}$ accounts for random residual effects. The $\mathbf{G}$ is a block diagonal matrix defined as $\mathbf{I}\sigma_C^2$ for the variance among clones and $\mathbf{I}\sigma_{CB}^2$ for the variance of the clone by block interaction, where I is an identity matrix of dimension $n_i \times n_i$ (n_i = number of levels of the ith term). The $\mathbf{R} = \mathbf{I}\sigma_e^2$ is a diagonal matrix with the residual error variances (σ_e^2) in the diagonal and zero covariances (null submatrix) in the off diagonals when errors are independent; I is the identity matrix of dimension equal to the number

of observations. In this experiment a spatial residual structure was implemented which divides $\mathbf{e}$ into spatially dependent (ξ) and spatially independent (η) residuals (Dutkowski *et al.*, 2002). For spatially dependent residuals a covariance structure was specified using a first-order autoregressive process in rows and columns:

$$\mathbf{R} = \sigma_\xi^2 [\text{AR1}(\rho_{\text{col}}) \otimes \text{AR1}(\rho_{\text{row}})] + \sigma_\eta^2 \mathbf{I}, \qquad (3)$$

where σ_ξ^2 is the spatial residual variance, σ_η^2 is the independent residual variance, $\mathbf{I}$ is an identity matrix equal to the number of observations, and the error term is included in the individual residual variance. $\text{AR1}(\rho)$ is a first order autoregressive correlation matrix with the form:

$$\text{AR1}(\rho) = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^n \\ \rho & 1 & \rho & \cdots & \\ \rho^2 & \rho & 1 & \cdots & \\ \vdots & \vdots & \vdots & \ddots & \\ \rho^n & \cdots & \cdots & \cdots & 1 \end{bmatrix} \qquad (4)$$

Broad-sense clone mean heritability was estimated using the formula:

$$H_c^2 = \frac{\sigma_C^2}{\sigma_C^2 + \frac{\sigma_{CB}^2}{b} + \frac{\sigma_E^2}{bt}}, \qquad (5)$$

where σ_C^2 is the variance among clones, σ_{CB}^2 is the variance due to the clone by block interaction, σ_E^2 is the residual

variance after spatial residuals are removed, b is the number of blocks (2) and t is the trees per plot (1). Standard errors of the heritability estimates were calculated using the Taylor series expansion (Gilmour *et al*., 2006). The best linear unbiased prediction (BLUP) values for clonal genotypes were used as phenotypes in the association analysis. All statistical analysis was performed using the ASReml 2.0 statistical software package (VSN International Ltd., Hertfordshire, UK) (Gilmour *et al*., 2006).

Genotypic data

Genotypes for single nucleotide polymorphisms (SNPs) were obtained using the Illumina Infinium assay (Illumina, San Diego, CA, USA). The discovery and genotyping of the SNPs has been previously described (Eckert *et al*., 2009). A description of the development of these SNPs is available online at http://dendrome.ucdavis.edu/adept2/. Briefly, SNPs were detected and genotyped for 7508 resequenced amplicons generated from all available unique EST contigs representing all pine ESTs known to date using an Infinium genotyping chip. From the resequenced amplicons, roughly 22,000 SNPs were discovered and 7216 were selected for genotyping. SNPs were chosen based on quality of the SNP call (for example, PolyPhred), coverage across unique amplicons and spacing within those amplicons (SNPs were chosen to be far apart). This resulted in most amplicons being represented by one SNP. SNP genotypes were selected for association analysis using the BeadStudio ver. 3.1.3.0 software (Illumina), based on quality, reliability of genotype calls, and polymorphism which produced 3938 SNP loci out of the 7216 that were informative in this population. Predicted gene function was not a criterion for the choice of SNPs for inclusion among the SNPs assayed, nor for inclusion of the resulting genotypic data in the association analysis. Genotypic data from 3938 SNP loci were available for 380 of the 425 clones measured in this experiment.

Association analyses

Association testing was performed in TASSEL using both a general linear model (GLM) and a mixed model (MLM) approach (Bradbury *et al*., 2007). Population structure covariates were estimated from a set of 23 nuclear microsatellite markers using STRUCTURE with a cluster number of five (Eckert *et al*., 2010). Marker-based kinship was estimated using a function in the EMMA (efficient mixed model analysis) package (Kang *et al*., 2008) in the R programming environment (R Development Core Team, 2010). We used the positive false discovery rate approach to adjust P-values for multiple testing (Storey, 2003), using the qvalue package in R with a false discovery rate of 0.05.

We chose to compare the use of population and marker-based kinship to determine whether these factors enhanced analysis of the NCSU association population. Yu *et al*. (2006) demonstrated the value of accounting for population structure and relatedness through the incorporation of genomic control and marker-based kinship in mixed model association testing. We compared observed p-values from association testing against a uniform distribution of expected P-values using the mean of the squared differences of potential models for association testing (Stich *et al*., 2008a). We compared four models to examine the distribution of P-values from association tests: a GLM with no structure or kinship effects (GLM), a GLM with covariates to account for population structure (Q), a mixed model with a marker-based kinship component (K) and a mixed model that incorporated both population structure and marker-based kinship estimates (QK).

SNP effects and genotypic frequencies

For each associated SNP, we estimated the additive (a) and dominance (d) effects using ASReml 2.0. SNPs were treated as fixed effects to test for significance and to generate best linear unbiased estimates for genotype classes of each SNP. Additionally, variance estimates of significant SNPs with three genotypic classes were generated by adding a SNP effect into the mixed model analyzing phenotypic data, where the SNP effect was treated as a random effect with two degrees of freedom. To further evaluate the effect of potential associations with phenotypes, we compared the genotypic frequencies of the tails of the population phenotypic distribution with the frequencies observed in the entire population. The tails of each phenotypic distribution were truncated at >1.5 s.d. above and below the population mean for each trait. Genotypic frequencies and tests for Hardy-Weinberg Equilibrium were performed in the Allele procedure in SAS (SAS, 1989).

To obtain annotations for SNPs, flanking sequences of the corresponding EST contig were obtained from the Dendrome database (http://dendrome.ucdavis.edu/treegenes) for all SNPs associated with a trait after multiple testing correction. We performed a BLASTx query against the NCBI non-redundant protein database (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Gene and site annotations for the strongest hits (lowest e-value) for each sequence are reported.

Results

Phenotypic variation

Carbon isotope discrimination, height and %N displayed significant variation among clones. Use of the spatial model allowed additional environmental variation to be removed, which enhanced heritability estimates and removed environmental bias from the clonal BLUP values. Figure 2 displays a variogram of the residuals from the spatial model indicating spatial trends in the nursery bed that would not have been accounted for without the spatial model in the two-replicate design. Broad-sense clone mean heritability estimates (H_c^2) explained half of the variation in $\Delta^{13}C$ (Table 1), whereas $\sim 40\%$ of the variation in second-year height and %N were explained by variation among clones. Bivariate analyses (Table 2) revealed that $\Delta^{13}C$ was negatively genetically correlated with height ($r_{g(xy)} = -0.37$) and %N ($r_{g(xy)} = -0.42$), whereas height and %N were positively correlated ($r_{g(xy)} = 0.32$). Phenotypic correlations taking into account both genetic and environmental effects were weak with the exception of $\Delta^{13}C$ and %N ($r = -0.55$). Clonal BLUPs for carbon isotope discrimination ranged from 20.7 to 23.4 with a population mean of 22.11. Height ranged from 116 to 240 cm with a mean of 194 cm and %N ranged from 1.15 to 1.52 with a mean of 1.40. BLUPs
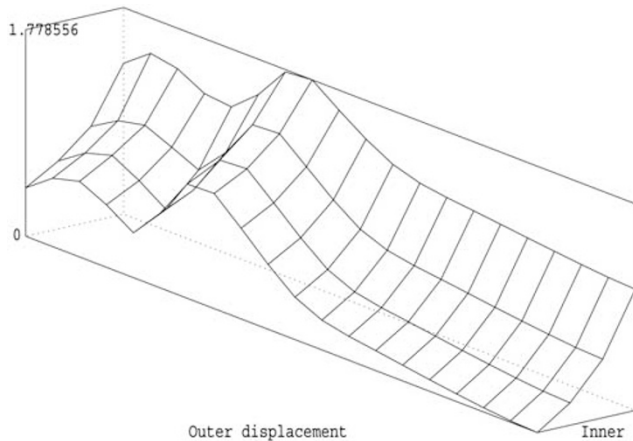
Figure 2 Variogram plot of residuals for $\Delta^{13}C$. The three dimensional plot reflects the similarity of trees because of environmental trends in a row and column spatial configuration.

Table 1 Phenotypic mean $_{Pop}$ and s.e.$_{Pop}$, population minimum (Min), population maximum (Max), σ_c^2, σ_ε^2 and H_c^2 and H_c^2 s.e. for $\delta^{13}C$, height and %nitrogen

Trait	$\Delta^{13}C$	Height (cm)	Nitrogen (%)
Mean $_{Pop}$	22.11	194	1.41
s.e.$_{Pop}$	0.05	14	0.08
Min	20.7	116	1.15
Max	23.4	240	1.52
σ_ε^2	0.25	1440	0.0064
σ_c^2	0.37	2787	0.0128
H_c^2	0.50	0.43	0.42
H_c^2 s.e.	0.05	0.05	0.06

Abbreviations: σ_c^2, clonal variance; H_c^2, clone mean heritability; σ_ε^2, error variance; Mean $_{Pop}$, phenotypic mean; s.e.$_{Pop}$, standard error; H_c^2 s.e.

Table 2 Phenotypic (below-diagonal) and Clonal (above-diagonal) correlations (and standard errors) among $\Delta^{13}C$, height and %N

	$\Delta^{13}C$	Height	%N
$\Delta^{13}C$	—	−0.37 (0.22)*	−0.42 (0.21)*
Height	0.02 (0.04)	—	0.32 (0.16)*
%N	−0.55 (0.03)*	0.04 (0.04)	—

*Indicates significance at $P < 0.05$.

were used as the phenotype in subsequent association testing.

Annotations for associated SNPs
Association analyses revealed SNPs located in regions of sequences of previously described function and also in unknown sequences (Table 3). BLASTx queries using the flanking sequences around the associated SNPs found similar proteins in the Genbank database for five out of 14 SNPs. Putative orthologs were a mitochondrial protein (0_8304_02_414), a heme activated DNA-binding protein (CL599Contig1_07_109) and an AP2 domain transcription factor (0_3648_01_357) for $\Delta^{13}C$ -associated SNPs. For %N-associated SNPs, these analyses revealed a receptor protein kinase-like protein (2_7865_01_156), and glutamate decarboxylase (0_17195_01_417) (Table 3).

SNP-trait associations
Significant associations between BLUP values and one or more SNP loci were found for all traits. Before multiple testing correction, the number of associations significant at the $P < 0.05$ level ranged from 144 (K) to 193 (GLM) for $\Delta^{13}C$, 133 (K and QK) to 172 (GLM) for height, and 135 (QK) to 176 (GLM) for %N, across the four different models tested. After correction for multiple testing, we found four to seven SNPs associated with $\Delta^{13}C$, one SNP associated with height, and five to six SNPs associated with %N, depending on the model used for testing. For each trait, the model with the highest number of significant associations following multiple testing correction was reported: QK for $\Delta^{13}C$, GLM for height, and GLM for %N. The four models used in this analysis identified similar numbers of SNPs as associated with all traits at the four levels of P-value thresholds (see Supplementary Information).

The associations in this study were largely caused by SNPs segregating rare alleles (MAF 2–35%) with small effects (r^2: 4–9%) on $\Delta^{13}C$, height and %N (Table 4). All significantly associated SNPs failed to depart from Hardy Weinberg Equilibrium except for SNP 2_1501_01_109. Additive effects were significantly different from zero for six loci associated with $\Delta^{13}C$ and five loci associated with %N. Dominance effects were significant for the locus associated with height, four of the loci associated with $\Delta^{13}C$, and five of the loci associated with %N (Table 4).

SNP effects
When individual SNPs were used in a mixed model as an independent random effect with clones, no SNP variance estimates were significantly different from zero using a chi-square test of −2logliklihood model values (not shown). Marker r^2 values for significantly associated SNPs ranged from 0.04 to 0.08 indicating that individual SNP loci account for small amounts of the variation among clones (Table 4). Using a model that included the effect of clones and individual SNPs as random effects, individual SNP effects accounted for small amounts of the phenotypic variance ranging from less than 0.001% up to 3.2% in $\Delta^{13}C$ and as high as 7% for %N.

Additive and dominance estimates were estimated for all SNPs observed to be in Hardy Weinberg Equilibrium with all three genotype classes present. Additive effects were significant for six SNPs associated with $\Delta^{13}C$, but the estimates of SNP effects were not precise as standard errors were greater than half the value of the estimate (Table 4). The ratio of dominance to additive effects (d/a) showed dominance effects similar in magnitude to additive effects ranging from −0.79 to 1.30 for $\Delta^{13}C$ (Table 4). Dominance effects for %N were similar in range (−0.75 to −1.04) with the exception of locus CL1074Contig1_03_101 where dominance was only 1% of the additive effect. The dominance effect for tree height at locus 0_14415_01_190 was the largest dominance effect observed at more than three times the additive effect, which resulted in a reduction in height (Table 4).

Genotypic frequencies
We tested for differences in genotypic frequency in the tails of the population phenotypic distribution for all loci

Table 3 SNP loci annotations and significance values for second-year Ht, carbon isotope discrimination ($\Delta^{13}C$) and foliar nitrogen level (%N)

SNP Locus	Trait	Annotation	P value	Q value	E-value
0_14415_01_190[NC]	Ht	Unknown	2.38E-05	8.87E-02	—
0_17030_01_94[NC]	$\Delta^{13}C$	Unknown	4.95E-07	1.91E-05	—
0_8304_02_414[NS]	$\Delta^{13}C$	Mitochondrial protein	2.84E-05	4.04E-02	5E-66
0_17543_01_196[SYN]	$\Delta^{13}C$	Unknown	3.98E-05	2.97E-03	—
0_10921_01_353[NC]	$\Delta^{13}C$	Unknown	8.67E-05	7.88E-03	—
2_1501_01_109[NC]	$\Delta^{13}C$	Unknown	1.43E-05	1.97E-02	—
CL599Contig1_07_109[SYN]	$\Delta^{13}C$	Putative heme-activated protein	3.79E-04	2.24E-02	1E-53
0_3648_01_357[NS]	$\Delta^{13}C$	AP2 domain transcription factor	6.26E-04	3.63E-02	1E-20
UMN_6338_01_99[NS]	%N	Predicted protein	2.05E-07	2.11E-04	7E-58
2_7865_01_156[NC]	%N	Receptor protein kinase-like protein	3.53E-07	2.11E-04	9E-29
2_1087_01_86[NC]	%N	Hypothetical protein	1.98E-06	1.08E-03	5E-37
CL1074Contig1_03_101[NC]	%N	Predicted protein	6.17E-06	2.77E-03	5E-54
0_17195_01_417[NC]	%N	Glutamate decarboxylase	1.89E-05	4.58E-03	3E-47
2_4191_01_104[NC]	%N	Unknown	1.55E-04	2.71E-02	—

Abbreviations: Ht, height; NC, non-coding; NS, non-synonymous; SNP, single nucleotide polymorphism; SYN, synonymous[a].
[a]P-values and Q-values are reported for the model which yielded the most significant associations (Q<0.05) for individual traits.

Table 4 Description of significant SNP loci, MAF, r^2, %variance estimates, additive and dominance effects

SNP Locus	Trait	SNP[a]	MAF	r^{2b}	% Phenotypic variance explained by SNP	Additive effect a[c]	Standard error of the additive effect	Dominance effect d[d]	Standard error of the dominance effect	d/a
0_14415_01_190	Ht	[T/A]	0.35	0.06	<0.01	−1.28	0.84	−3.99*	1.19	3.13
0_17030_01_94	$\Delta^{13}C$	[A/G]	0.05	0.08	0.4	0.83*	0.15	−0.79*	0.16	−0.95
0_8304_02_414	$\Delta^{13}C$	[A/G]	0.11	0.06	<0.01	0.34*	0.08	−0.27*	0.09	−0.79
0_17543_01_196	$\Delta^{13}C$	[A/G]	0.08	0.06	<0.01	0.51*	0.11	−0.47	0.12	−0.92
0_10921_01_353	$\Delta^{13}C$	[A/G]	0.06	0.05	0.2	0.47*	0.11	−0.48	0.12	−1.03
2_1501_01_109	$\Delta^{13}C$	[A/G]	0.02	0.05	NE[e]	NE	NE	NE	NE	NE
CL599Contig1_07_109	$\Delta^{13}C$	[G/C]	0.11	0.05	2.7	0.28*	0.06	−0.28*	0.07	−1.01
0_3648_01_357	$\Delta^{13}C$	[G/A]	0.09	0.04	3.8	0.24*	0.07	0.31*	0.08	1.30
UMN_6338_01_99	%N	[A/T]	0.06	0.09	<0.01	−0.10*	0.02	0.11*	0.02	−1.09
2_7865_01_156	%N	[A/G]	0.06	0.08	NA	−0.10*	0.02	0.09*	0.02	−0.95
2_1087_01_86	%N	[G/A]	0.05	0.07	<0.01	−0.06	0.01	0.05*	0.01	−0.80
CL1074Contig1_03_101	%N	[A/G]	0.03	0.07	<0.01	0.04*	0.02	<−0.01	0.02	−0.01
0_17195_01_417	%N	[A/C]	0.04	0.06	7.0	−0.06*	0.01	0.04*	0.01	−0.75
2_4191_01_104	%N	[G/A]	0.08	0.05	<0.01	−0.05*	0.01	0.05*	0.01	−1.04

Abbreviations: MAF, minor allele frequencies; SNP, single nucleotide polymorphism.
[a]SNPs are listed in order of common allele/minor allele for each locus.
[b]*r*-square of the markers from the single marker models regressed on Clonal BLUP values.
[c]Additive effect is significantly different from zero from ANOVA (P<0.05).
[d]Dominance effect is significantly different from zero from ANOVA (P<0.05).
[e]NE (Not Estimated) indicates the SNP locus did not meet criteria of three genotypic classes and/or departed from Hardy Weinberg Equilibrium.

significantly associated with $\Delta^{13}C$, height and %N. For each trait, we compared the extremes of the population, which we considered to be clonal BLUP values greater than 1.5 s.ds. above and below the mean of clonal values. Using the population mean as the expected genotypic frequencies, we observed nine loci with significant frequency changes (P<0.05) from the mean for one of the phenotypic tails (Table 5). For each trait, we observed SNPs with departures from the overall population genotype frequencies (one SNP for height (0_14415_01_190), four SNPs for $\Delta^{13}C$ (0_17030_01_94, 0_10921_01_353, CL599Contig1_07_109 and 0_8304_02_414), and five SNPs for %N (0_17195_01_417, 2_1087_01_86, 2_4191_01_104, 2_7865_01_156 and CL1074Contig1_03_101). The single SNP that survived multiple testing correction for height displayed an increased frequency in the heterozygous class for the tail with lower height (P<0.01), whereas SNPs for $\Delta^{13}C$

and %N reflected changes in genotypes with the minor alleles in one tail.

Discussion

The ability of plants to respond to different levels of available water is variable and complex. Members of the genus *Pinus* use a drought avoidance strategy where under well-watered conditions water use is maximized, but decreases quickly when water is limiting to avoid low water potential (Martinez-Vilalta *et al.*, 2004). Moderate genetic correlations indicate an increase in height growth and nitrogen content in foliage as $\Delta^{13}C$ decreases, suggesting that under moderate water stress during the growing season, faster growth is related to assimilation rather than stomatal control as seen by Brendel *et al.* (2002) in *Pinus pinaster*. Previous analyses of height and $\Delta^{13}C$ revealed similar results suggesting that

Table 5 Genotypic frequencies by SNP and phenotypic class (± 1.5 std dev) with χ^2 tests for significant differences from the population

Trait	Locus	Class	XX	XY	YY	#obs	$P > \chi^2$
$\Delta^{13}C$	0_17030_01_94	Population	311	33	1	345	
		High	20	0	1	21	0.03
		Low	24	1	0	25	0.61
	0_10921_01_353	Population	308	36	2	346	
		High	18	2	1	21	0.04
		Low	20	5	0	25	0.27
	CL599Contig1_07_109	Population	274	64	6	344	
		High	14	4	3	21	<0.001
		Low	20	5	0	25	0.79
	0_17543_01_196	Population	287	57	8	352	
		High	17	3	1	21	0.73
		Low	21	4	0	25	0.74
	0_3648_01_357	Population	285	54	5	344	
		High	15	5	1	21	0.24
		Low	18	7	0	25	0.21
	0_8304_02_414	Population	276	66	4	346	
		High	15	4	2	21	0.0016
		Low	22	3	0	25	0.56
Height	0_14415_01_190	Population	152	140	46	338	
		High	12	5	1	18	0.17
		Low	8	21	2	31	0.01
%N	UMN_6338_01_99	Population	313	32	1	346	
		High	15	3	0	18	0.54
		Low	17	1	1	19	0.84
	0_17195_01_417	Population	318	25	3	346	
		High	18	0	0	18	0.45
		Low	15	3	1	19	0.03
	2_1087_01_86	Population	311	32	2	345	
		High	18	0	0	18	0.37
		Low	16	2	1	19	0.026
	2_4191_01_104	Population	297	46	3	346	
		High	15	3	0	18	0.85
		Low	16	1	2	19	<0.001
	2_7865_01_156	Population	309	36	1	346	
		High	18	0	0	18	0.34
		Low	17	1	1	19	0.0002
	CL1074Contig1_03_101	Population	328	17	1	346	
		High	12	5	1	18	<0.001
		Low	19	0	0	19	0.59

Abbreviations: SNP, single nucleotide polymorphism; XX, major allele homozygote; XY the heterozygote; YY, minor allele homozygote.

carbon isotope analysis may be used to improve water use efficiency in loblolly pine populations (Baltunis *et al.*, 2008). An understanding of the genes involved in carbon isotope discrimination and foliar nitrogen content are valuable for both the environmental services provided by natural forests managed with low intensity and wood productivity in more intensively managed plantation forests. The underlying genetic variation in water use efficiency may be valuable for modeling forest stand dynamics, as well as for producing more wood products in a sustainable and efficient manner. In the future we will likely be able to associate alleles of genes with more specific traits to dissect complex traits of economic and ecological value (Nelson and Johnsen, 2008). Carbon isotopes have shown promise for use in improving water use efficiency in pine species, but adequate sampling across environments, such as the natural range of loblolly pine, should be taken into account.

Significantly associated SNPs in this experiment were found in both known and unknown gene sequences and in some cases SNPs were in functional genes related to similar traits in other plant species. The sequence flanking SNP 0_3648_01_357 for $\Delta^{13}C$ was similar to an AP2 domain transcription factor. The AP2 domain family of transcription factors has been associated with ABA-sensitive abiotic stress response in *Arabidopsis* (Finkelstein *et al.*, 1998) in seed and leaf tissue. In *Pinus strobus* the over-expression of the ERF/AP2 transcription factor *CaPF1* conferred increased drought and freeze tolerance in young plants (Tang *et al.*, 2007). SNP 0_3648_01_357 results in a non-synonomous codon change, and accounted for 3.8% of the phenotypic variation in this experiment suggesting that a response to stress is linked to $\Delta^{13}C$ variation in this experiment.

Nitrogen is critical to plant growth and photosynthetic activities, and it is often a limiting nutrient for growth in forest plantations (Fox *et al.*, 2007). To date, genes in loblolly pine affecting the use or uptake of nitrogen have not yet been identified. The sequence flanking SNP 0_17195_01_417 is similar to glutamate decarboxylase (*GAD*), which is involved in the production of GABA, nitrogen metabolism and C:N ratio balance in *Arabidopsis* (Bouché and Fromm, 2004). SNP 2_7865_01_156 was identified as a receptor-like protein kinase-like (RLK) protein, a class of proteins which are important for many

plant functions including nitrogen fixation in legumes such as alfalfa, soybean and peas (Morris and Walker, 2003).

Association analyses have been performed in populations of loblolly pine for wood quality traits (González-Martínez *et al.*, 2007) and carbon isotope discrimination (González-Martínez *et al.*, 2008), and cold-tolerance in Douglas-fir (Eckert *et al.*, 2009) using a candidate gene approach. To date, associated SNPs for traits in loblolly pine explained the highest proportion of phenotypic variance in wood quality traits (20% of phenotypic variance) while only 7% of the phenotypic variance was accounted for in this experiment for the Δ^{13}C-associated SNPs. Recent association analyses in conifers have explained small proportions of the phenotypic variation with individual SNPs (González-Martínez *et al.*, 2007, 2008; Eckert *et al.*, 2009), supporting the treatment of these traits as polygenic and quantitative in nature (Falconer, 1989).

Our results complement those of González-Martínez *et al.* (2008) with new SNPs in previously unidentified sequences, as well as those found in known sequences. Associations identified in this study and candidate genes from the work of González-Martínez *et al.* (2008) in carbon isotope discrimination association testing in loblolly pine reveal functional annotations of genes that are involved in abiotic stress response rather than growth functions, which might suggest that Δ^{13}C is related to both stomatal conductance and photosynthetic capacity, but may be an artifact of the candidate genes tested in these experiments. The gene annotations may be somewhat contradictory to the trait relationships observed in this study, but the candidate genes used in this experiment do not cover a majority of the loblolly pine genome, so gene annotations by themselves are not conclusive.

In both studies, the most significant SNPs accounted for small portions of the phenotypic variance supporting a polygenic response to regulate water loss in loblolly pine. González-Martínez *et al.* (2008) tested only a small number of candidate genes (46 SNPs from 41 genes), whereas we tested associations for 3938 SNPs from nearly as many genes. Unfortunately, the candidate gene sequences from González-Martínez *et al.* (2008) were not included in this study, thus validation of previous associations was not possible. In both studies, the SNPs represent a small proportion of the total variation in the loblolly pine genome, which is $\sim$24,000 MB (Ahuja and Neale, 2005; Morse *et al.*, 2009). Thus, there is future opportunity to identify additional genes that influence Δ^{13}C and growth traits.

The population used in the González-Martínez *et al.* experiment consisted of 61 families from 31 parents in a partial-diallel mating design from a narrower geographical range, whereas the population in this experiment consisted of largely unrelated trees sampled from across the native range of loblolly pine. Allele frequencies for associated SNPs in their experiment were higher than observed in this study, which may be a product of the different population sizes or the degree of kinship among the individuals analyzed. Our study examined SNP alleles across 380 unrelated parents, compared with offspring of 31 unrelated individuals in the González-Martínez *et al.* (2008) work. This study was not replicated on multiple sites, but did experience water stress during the growing season, while the study of

González-Martínez *et al.* (2008) was installed on two sites that were not limited in available water. A replicated experiment in a water-limited environment may be necessary to validate the marker-trait associations identified in this experiment.

Improved methods and larger populations are needed to have greater precision in estimates of marker effects. F-tests show significant additive and dominance effects at several SNP loci associated with measured phenotypes, but best linear unbiased estimates of the genotypic effects revealed large standard errors. The estimated magnitudes of allelic effects in this population are likely to be biased upward, because effects are estimated from a truncated distribution (Xu, 2003). In addition, the majority of minor allele frequencies observed for significant SNPs in this population are low (0.05–0.10) with one exception at 0.35, thus the small number of observations with these alleles are more heavily weighted in the estimation of allelic effects and variances. Experiments using a larger population are likely to more accurately estimate the true effect of any associated SNP polymorphism. A spatial model was implemented to remove environmental variability that would not have been removed by the randomized complete block experimental design. The individual SNP models regressed the BLUP of each clone for each trait on individual SNPs. BLUP predictions were based on two replicates of each genotype planted in a relatively small, homogenous nursery bed. *R*-square values and the strength of any relationship between a SNP polymorphism and clonal BLUP values may not be repeatable in a larger, more heterogeneous field trial, which is typical of forest tree progeny trials. Future experiments should incorporate efficient designs to remove environmental variation and improve the precision of phenotypes used in association studies.

In this population the magnitude of the dominance effects was similar to that of additive effects. Excluding two SNPs, 0_14415_01_190, which was associated with height, and CL1074Contig1_03_101, which was associated with %N, the ratio of dominance to additive ranged from 0.75 to 1.3 (Table 5). Eckert *et al.* (2009) reported similar ratios of dominance to additive effects for cold-tolerance-related traits in Douglas-fir, where most dominance effects were 0.8x–1.3x the additive effect. Additive effects have been the focus of tree breeding programs for loblolly pine but the use of non-additive effects may be an opportunity for capturing desirable trait attributes in deployment populations.

The largest dominance effect observed in this study was the effect of SNP 0_14415_01_190 on height. This effect is supported by the increased number of heterozygotes in the low-end tail for height in the population (Table 5). This difference in genotypic frequency was significant ($P < 0.01$), and the proportion of heterozygotes in the low-end tail increases from 0.41 in the entire population to 0.68 in the low-end tail. However, variance in height due to this SNP was very small (< 0.001%) and was not significant, highlighting the challenge to discover variants and estimate the magnitude of their effects in field trials. Eckert *et al.* (2009) reported that 86% of the SNPs tested for associations with cold tolerance in Douglas-fir showed non-additive effects. If such SNPs are validated and account for significant variation, non-additive effects could have an important role in the improvement of growth.

We observed significant changes in genotypic frequency between at least one tail of the population and the overall population for nine SNPs, lending support that the SNP variants are having some impact on the phenotypes observed in this study. Finding significant changes in allele frequency between the tails of the population suggests that a pooled sampling approach may be of value in forest trees. Pooling individuals based on phenotype for SNP discovery and allele frequency estimation has shown promise in effectively discovering polymorphic SNPs and estimating allele frequencies within the pooled phenotypic classes in human blood and disease studies (Craig *et al.*, 2009; Druley *et al.*, 2009). Such an approach could significantly reduce genotyping cost as compared with association studies in large populations; and identify candidate SNPs for future studies based upon allele frequency differences.

Recent characterization of population structure in loblolly pine, which included the NCSU + WG population, identified 24 SNPs as F_{st} outliers and five loci associated with geographical variation for potential evapotranspiration (Eckert *et al.*, 2010). Two of the F_{st} outliers were associated with traits in this analysis: 0_14415_01_190 for height and 2_1087_01_86 for %N. If selection is occurring on these genes in different subpopulations, this may bias the associations and care must be taken in interpreting the result.

Conclusion

We identified 14 new marker-trait associations for $\Delta^{13}C$, height and %N in loblolly pine. Results suggest that $\Delta^{13}C$, height and %N in loblolly pine are influenced by many genes, each having a small effect, in keeping with the 'infinitesimal model' of quantitative genetics (Fisher, 1918). A recent review summarized data from many experiments, and reported that additive genetic models adequately account for variation in complex traits (Hill *et al.*, 2008), but results from this study suggest that there are both additive and non-additive effects on adaptive and economic traits. As breeding programs select for desirable genotypes of loblolly pine, the allele frequency distribution in pine breeding populations are likely to change from that found in the non-domesticated individuals analyzed in this study. As allele frequencies move closer to 0.5, and alleles that were originally rare in the wild population become more abundant through selection, more complex models that account for non-additive effects are likely to be useful (Mackay *et al.*, 2009). The application of analytical tools such as spatial models will help remove the environmental variation from phenotypic data and may help strengthen future association testing in field-grown trees. Traditional breeding methods based on quantitative genetic approaches have been successful in the improvement of loblolly pine populations, but the development of forest tree genomic resources will aid the improvement and understanding of ecological and economically valuable traits in forest trees.

Conflict of interest

The authors declare no conflict of interest.

References

Ahuja MR, Neale DB (2005). Evolution of genome size in conifers. *Silvae gen* **54**: 126–137.

Anyia AO, Slaski JJ, Nyachiro JM, Archambault DJ, Juskiw P (2007). Relationship of Carbon Isotope Discrimination to Water Use Efficiency and Productivity of Barley Under Field and Greenhouse Conditions. *J Agron Crop Sci* **193**: 313–323.

Baltunis BS, Davis JM, Huber DA, Martin TA (2008). Inheritance of foliar stable carbon isotope discrimination and third-year height in Pinus taeda clones on contrasting sites in Florida and Georgia. *Tree gen & genom* **4**: 797–807.

Bouché N, Fromm H (2004). GABA in plants: just a metabolite? *Trends Plant Sci* **9**: 110–115.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**: 2633–2635.

Brendel O, Pot D, Plomion C, Rozenberg P, Guehl JM (2002). Genetic parameters and QTL analysis of delta13C and ring width in maritime pine. *Plant, Cell Environ* **25**: 945–953.

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C *et al.* (2009). The Genetic Architecture of Maize Flowering Time. *Science* **325**: 714–718.

Chan EKF, Rowe HC, Kliebenstein DJ (2010). Understanding the Evolution of Defense Metabolites in Arabidopsis thaliana Using Genome-Wide Association Mapping. *Genetics* **185**: 991–1007.

Condon AG, Farquhar GD, Rebetzke GJ, Richards RA (2004). Breeding for high water-use efficiency. *J Exp Bot* **55**: 2447–2460.

Craig H (1954). Carbon-13 in plants and the relationship between carbon-13 and carbon-14 variations in nature. *J Geology* **62**: 115–149.

Craig JE, Hewitt AW, McMellon AE, Henders AK, Ma L, Wallace L *et al.* (2009). Rapid inexpensive genome-wide association using pooled whole blood. *Genome Res* **19**: 2075–2080.

Druley TE, Vallania FLM, Wegner DJ, Varley KE, Knowles OL, Bonds JA *et al.* (2009). Quantification of rare allelic variants from pooled genomic DNA. *Nat Meth* **6**: 263–265.

Dutkowski GW, Costa E, Silva J, Gilmour AR, Lopez GA (2002). Spatial analysis methods for forest genetic trials. *Can J For Res* **32**: 2201–2214.

Eckert AJ, Bower AD, Wegrzyn JL, Pande B, Jermstad KD, Krutovsky KV *et al.* (2009). Association genetics of coastal Douglas fir (Pseudotsuga menziesii var. menziesii, Pinaceae). I. cold-hardiness related traits. *Genetics* **182**: 1289–1302.

Eckert AJ, van Heerwaarden J, Wegrzyn JL, Nelson CD, Ross-Ibarra J, Gonzalez-Martinez SC *et al.* (2010). Patterns of population structure and environmental associations to aridity across the range of loblolly pine (Pinus taeda L., Pinaceae). *Genetics* **185**: 969–982.

Evans JR (1989). Photosynthesis and nitrogen relationships in leaves of C3 plants. *Oecologia* **78**: 9–19.

Falconer DS (1989). *Introduction to Quantitative Genetics*, Vol. 3rd edn. Burnt Mill, Harlow, Essex, England: Longman, Scientific & Technical; New York: Wiley.

Farquhar GD, Ehleringer JR, Hubick KT (1989). Carbon isotope discrimination and photosynthesis. *Annu Rev Plant Physiol Plant Mol Biol* **40**: 503–537.

Finkelstein RR, Wang ML, Lynch TJ, Rao S, Goodman HM (1998). The arabidopsis abscisic acid response locus ABI4 encodes an APETALA 2 domain protein. *Plant Cell* **10**: 1043–1054.

Fisher RA (1918). The correlation between relatives under the supposition of Mendelian inheritance. *Trans R Soc Edin* **52**: 399–433.

Fox TR, Allen HL, Albaugh TJ, Rubilar R, Carlson CA (2007). Tree nutrition and forest fertilization of pine plantations in the southern United States. *South J Appl For* **31**: 5–11.

Gilmour AR, Gogel BJ, Cullis BR, Thompson R (2006). *ASReml User Guide Release 2*. VSN International Ltd: Hemel Hempstead, UK. 320 pp.

Gocke MH (2006). *Production system influences the survival and morphology of rooted stem cuttings of loblolly pine (Pinus taeda L.) and sweetgum (Liquidambar styraciflua L.)*. Master of Science thesis, North Carolina State University: Raleigh.

González-Martínez SC, Huber D, Ersoz E, Davis JM, Neale DB (2008). Association genetics in Pinus taeda L. II. Carbon isotope discrimination. *Heredity* **101**: 19–26.

González-Martínez SC, Wheeler NC, Ersoz E, Nelson CD, Neale DB (2007). Association Genetics in Pinus taeda L. I. Wood Property Traits. *Genetics* **175**: 399–409.

Hill WG, Goddard ME, Visscher PM (2008). Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet* **4**: e1000008.

Hirschhorn JN, Daly MJ (2005). Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* **6**: 95–108.

Ingvarsson PK, Garcia MV, Luquez V, Hall D, Jansson S (2008). Nucleotide polymorphism and phenotypic associations within and around the phytochrome B2 locus in European aspen (Populus tremula, Salicaceae). *Genetics* **178**: 2217–2226.

Jing H-C, Kornyukhin D, Kanyuka K, Orford S, Zlatska A, Mitrofanova OP *et al.* (2007). Identification of variation in adaptively important traits and genome-wide analysis of trait marker associations in Triticum monococcum. *J Exp Bot* **58**: 3749–3764.

Johnsen KH, Flanagan LB, Huber DA, Major JE (1999). Genetic variation in growth, carbon isotope discrimination, and foliar N concentration in Picea mariana: analyses from a half-diallel mating design using field-grown trees. *Can J For Res* **29**: 1727–1735.

Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ *et al.* (2008). Efficient control of population structure in model organism association mapping. *Genetics* **178**: 1709–1723.

Lebude AV, Goldfarb B, Blazich FA, Wise FC, Frampton J (2004). Mist, substrate water potential and cutting water potential influence rooting of stem cutting of loblolly pine. *Tree Physiol* **24**: 823–831.

Mackay TFC, Stone EA, Ayroles JF (2009). The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet* **10**: 565–577.

Martinez-Vilalta J, Sala A, Pinol J (2004). The hydraulic architecture of Pinaceae—a review. *Plant Ecol* **171**: 3–13.

Morris ER, Walker JC (2003). Receptor-like protein kinases: the keys to response. *Curr Opin Plant Biol* **6**: 339–342.

Morse AM, Peterson DG, Islam-Faridi MN, Smith KE, Magbanua Z, Garcia SA *et al.* (2009). Evolution of Genome Size and Complexity in *Pinus*. *PLoS One* **4**: e4332.

Neale DB, Ingvarsson PrK (2008). Population, quantitative and comparative genomics of adaptation in forest trees. *Curr Opin Plant Biol* **11**: 149–155.

Neale DB, Savolainen O (2004). Association genetics of complex traits in conifers. *Trends Plant Sci* **9**: 325–330.

Nelson CD, Johnsen KH (2008). Genomic and physiological approaches to advancing forest tree improvement. *Tree Physiol* **28**: 1135–1143.

Prasolova MV, Xu ZH, Farquhar GD, Saffigna PG, Dieters MJ (2000). Variation in branchlet delta13C in relation to branchlet nitrogen concentration and growth in 8-year-old hoop pine families (Araucaria cunninghamii) in subtropical Australia. *Tree Physiol* **15**: 1049–1055.

Prasolova NV, Lundkvist K, Xu ZH (2005). Genetic variation in foliar nutrient concentration in relation to foliar carbon isotope composition and tree growth with clones of the F1 hybrid between slash pine and Caribbean pine. *For Ecol Manage* **210**: 173–191.

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria (ISBN: 3-900051-07-0, http://www.r-project.org/).

Rebetzke GJ, Richards RA, Condon AG, Farquhar GD (2006). Inheritance of carbon isotope discrimination in bread wheat (Triticum aestivum L.). *Euphytica* **150**: 97–106.

SAS (1989). *SAS/STAT User's Guide, Version 6* 4th edn. SAS Institute Inc.: Cary, NC.

Smith JG, Newton-Cheh C (2009). Genome-wide association study in humans. *Methods Mol Biol (Totowa, NJ, U S)* **573**: 231–258.

Springer CJ, Delucia EH, Thomas RB (2005). Relationships between net photosynthesis and foliar nitrogen concentrations in a loblolly pine forest ecosystem grown in elevated atmospheric carbon dioxide. *Tree Physiol* **25**: 385–394.

Stich B, Mohring J, Piepho H-P, Heckenberger M, Buckler ES, Melchinger AE (2008a). Comparison of Mixed-Model Approaches for Association Mapping. *Genetics* **178**: 1745–1754.

Stich B, Piepho H-P, Schulz B, Melchinger A (2008b). Multi-trait association mapping in sugar beet (Beta vulgaris L.). *Theor Appl Genet* **117**: 947–954.

Storey JD (2003). The positive false discovery rate: A Bayesian interpretation and the q-value. *Ann Stat* **31**: 2013–2035.

Tang W, Newton R, Li C, Charles T (2007). Enhanced stress tolerance in transgenic pine expressing the pepper CaPF1 gene is associated with the polyamine biosynthesis. *Plant Cell Rep* **26**: 115–124.

Xu S (2003). Theoretical basis of the Beavis effect. *Genetics* **165**: 2259–2268.

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF *et al.* (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* **38**: 203–208.

Supplementary Information accompanies the paper on Heredity website (http://www.nature.com/hdy)