Tree Genetics & Genomes (2017) 13: 57
DOI 10.1007/s11295-017-1140-1

ORIGINAL ARTICLE

Association genetics of growth and adaptive traits in loblolly pine (*Pinus taeda* L.) using whole-exome-discovered polymorphisms

Mengmeng Lu[1,2] · **Konstantin V. Krutovsky**[1,2,3,4,5] · **C. Dana Nelson**[6,7] · **Jason B. West**[1,2] · **Nathalie A. Reilly**[8] · **Carol A. Loopstra**[1,2]

Received: 18 November 2016 / Revised: 23 March 2017 / Accepted: 10 April 2017 / Published online: 26 April 2017

Abstract In the USA, forest genetics research began over 100 years ago and loblolly pine breeding programs were established in the 1950s. However, the genetics underlying complex traits of loblolly pine remains to be discovered. To address this, adaptive and growth traits were measured and analyzed in a clonally tested loblolly pine (*Pinus taeda* L.) population. Over 2.8 million single nucleotide polymorphism (SNP) markers detected from exome sequencing were used to test for single-locus associations, SNP-SNP interactions, and correlation of individual heterozygosity with phenotypic traits. A total of 36 SNP-trait associations were found for specific leaf area (5 SNPs), branch angle (2), crown width (3), stem diameter (4), total height (9), carbon isotope discrimination (4), nitrogen concentration (2), and pitch canker resistance traits (7). Eleven SNP-SNP interactions were found to be associated with branch angle (1 SNP-SNP interaction), crown width (2), total height (2), carbon isotope discrimination (2), nitrogen concentration (1), and pitch canker resistance (3). Non-additive effects imposed by dominance and epistasis account for a large fraction of the genetic variance for the quantitative traits. Genes that contain the identified SNPs have a wide spectrum of functions. Individual heterozygosity positively correlated with water use efficiency and nitrogen concentration. In conclusion, multiple effects identified in this study influence the performance of loblolly pines, provide resources for understanding the genetic control of complex traits, and have potential value for assisting breeding through marker-assisted selection and genomic selection.

Keywords Association mapping · Epistasis · Exome heterozygosity · Loblolly pine · SNP

Communicated by P. Ingvarsson

Electronic supplementary material The online version of this article (doi:10.1007/s11295-017-1140-1) contains supplementary material, which is available to authorized users.

✉ Carol A. Loopstra
c-loopstra@tamu.edu

[1] Department of Ecosystem Science and Management, Texas A&M University, 2138 TAMU, College Station, TX 77843-2138, USA

[2] Molecular and Environmental Plant Sciences Program, Texas A&M University, 2474 TAMU, College Station, TX 77843-2474, USA

[3] Department of Forest Genetics and Forest Tree Breeding, Georg-August-University of Göttingen, Büsgenweg 2, 37077 Göttingen, Germany

[4] N. I. Vavilov Institute of General Genetics, Russian Academy of Sciences, 3 Gubkina Str., Moscow 119333, Russia

[5] Genome Research and Education Center, Siberian Federal University, 50a/2 Akademgorodok, Krasnoyarsk 660036, Russia

[6] USDA Forest Service, Southern Research Station, Southern Institute of Forest Genetics, 23332 Success Road, Saucier, MS 39574, USA

[7] Forest Health Research and Education Center, University of Kentucky, 730 Rose Street, Lexington, KY 40546, USA

[8] Department of Biology and Marine Biology, University of North Carolina Wilmington, 601 S. College Road, Dobo Hall 102, Wilmington, NC 28403-5915, USA

Introduction

Loblolly-shortleaf pine forests comprise nearly one fourth or 55 million acres of all southern forests in the USA (Smith et al. 2009). Loblolly pine's (*Pinus taeda* L.) amenability to plantation management, high wood/fiber yields, and fast growth make it one of the most economically important forest species in the world. Lumber and pulpwood are the primary products. Loblolly pine is also a potential renewable feedstock for alternative energy and fuel (Nelson et al. 2013) as well as a promising tool in efforts to mitigate warming and long-lasting climate changes caused by greenhouse gas emissions (Millar et al. 2007; Frederick et al. 2008; Bolte et al. 2009).

Advanced loblolly pine breeding practices have been implemented over the past 60 years, creating favorable production economics (McKeand et al. 2006). Appropriate breeding strategies rely on an understanding of valuable traits including crown structural characteristics, growth, water use efficiency (WUE), and disease resistance. Crown structural characteristics such as branch angle, leaf area, and crown width affect interception of radiation and competition with other trees (Emhart et al. 2007). Larger trees tend to display flatter branch angles and have wider crowns and thus expose more leaf area (Lambeth and Hubert 1997; Emhart et al. 2007). The ability to sustain wood yield and quality under adverse conditions such as drought and diseases is also a key consideration for loblolly pine breeding. Carbon isotope discrimination ($\Delta^{13}C$) has long been used to reflect long-term WUE in forest trees due to the feasibility to screen a large number of individuals over a short period of time (Aitken et al. 1995; Baltunis et al. 2008). Plants with higher WUE discriminate less against ^{13}C when they are exposed to the same fluctuations in environmental conditions (Farquhar et al. 1989; Aitken et al. 1995; Cregg and Zhang 2000; Baltunis et al. 2008; Cumbie et al. 2011). Although fusiform rust, caused by *Cronartium quercuum* f.sp. *fusiforme*, is the most serious disease of loblolly pine (Amerson et al. 2015), we were unable to obtain rust resistance data for this set of clones. However, we were able to obtain resistance data for these clones to pitch canker, caused by *Fusarium circinatum*. Pitch canker is also a serious disease of loblolly pine and has become increasingly more problematic over the last a couple of decades. It endangers loblolly pine via the resinous lesions on stems and branches that lead to high seedling mortality and slower growth rates (Kayihan et al. 2005; Quesada et al. 2010).

Association mapping is widely used to dissect traits of interest, helping to identify genes that control them. Association studies require a mapping or an association population representing a wide spectrum of phenotypic variation, phenotypic measurements, and abundant molecular markers scored on the mapping population. Environmental conditions drive forest phenotypic adaptation and geographic distribution through natural selection. Within the loblolly pine species, trees from different provenances have considerable diversity in phenotypic performance (Schmidtling 2001). To grasp as much phenotypic variation as possible for our association study, we selected the loblolly pine clonally propagated population from the Allele Discovery of Economic Pine Traits 2 (ADEPT2) project. It was specifically designed for association mapping to represent a range-wide collection from regions with different environmental conditions (see "Plant material" in "Materials and methods" for details). As mentioned above, an array of adaptive and growth traits, namely, specific leaf area, branch angle, crown width, stem diameter, height, $\Delta^{13}C$, nitrogen concentration, and disease resistance, was measured in this population.

The abundance of genetic variation in the abovementioned traits and their value for adaptation or growth suggest they are potential subjects for artificial selection by tree breeders. With advances in molecular and genomic methods, the selection process may be accelerated in forest trees breeding by marker-assisted selection and genomic selection (Thavamanikumar et al. 2013). However, these methods have not been widely tested or applied in the breeding of loblolly pines or other conifers. One reason is that the large and complex genome of loblolly pine poses challenges for gene discovery (Neale et al. 2014; Wegrzyn et al. 2014; Westbrook et al. 2015), association studies and genomic selection (Resende et al. 2012; Isik 2014). The number and identity of the genes controlling productivity and adaptive traits are largely unknown (Gonzalez-Martinez et al. 2006). A second reason is that high-density SNP genotyping assays/platforms remain to be developed for loblolly pine, although 7K Illumina Infinium SNP genotyping array is now available (Eckert et al. 2009b). Due to these problems and a long generation time, only traditional family-based breeding methods are currently applied to these forest trees.

Association analysis is an efficient method to dissect complex traits using large numbers of widely distributed SNPs in the loblolly pine genome. Successful association mapping has been implemented within diverse loblolly pine populations, producing an array of SNP markers and genes that are apparently connected with various traits. For example, using nearly 4000 SNPs derived from expressed sequence tags (ESTs) based PCR amplicons of 18 megagametophytes (Eckert et al. 2009a), 10 SNPs were detected to be associated with pitch canker disease resistance (Quesada et al. 2010), 7 SNPs with $\Delta^{13}C$, 1 SNP with height, 6 SNPs with foliar nitrogen concentration (Cumbie et al. 2011), 28 SNPs with metabolites (Eckert et al. 2012), and 80 SNPs with expression of 33 xylem development genes (Palle et al. 2013). Additionally, 101 associations with 27 gene expression phenotypes (Seeve 2010), and numerous SNPs with height, diameter at breast height, volume, fusiform rust resistance, wood specific gravity, and stem forking index (Chhatre et al. 2013) were identified.

Loblolly pine occurs naturally on both sides of the Mississippi River, and there are differences observed between eastern and western sources. This presumably can be dated back to the Pleistocene geologic era. During the last glaciation, loblolly pine supposedly retreated to two refugia, one in southeast Texas and/or northeast Mexico and the other in south Florida and the Caribbean. Isolation prevented widespread pollen exchange between populations and hence resulted in some of the differences observed now (Wells et al. 1991; Schmidtling 2001; Eckert et al. 2010). Thus, population structure will be an issue in a range-wide association study such as ours and will need to be properly accounted for in the association analysis.

Problems and challenges have emerged with the application of association mapping to loblolly pines and other conifer species. First, for most quantitative and complex traits, large numbers of alleles explain only a small portion of the genetically heritable variation (Quesada et al. 2010; Cumbie et al. 2011). The missing heritability might be due to the small number of markers and rare variants that were usually excluded from the genotyping chips (Manolio et al. 2009). Second, though most efforts were put on the discovery of additive marker associations, non-additive effects imposed by dominance and epistasis play important roles in determining the genetic variation and need to be further studied (Eckert et al. 2009a; Cumbie et al. 2011). Third, the gene-based SNP discovery focused primarily on coding sequence (CDS) regions. However, regulatory elements in non-coding regions tend to have more polymorphisms with small effects associated with quantitative traits (Flint and Mackay 2009).

To solve the aforementioned problems, a large amount of genetic variation including the rare alleles and regulatory sequences should be included in the association mapping. Target enrichment combined with genotyping by sequencing (GBS) technologies provides opportunities to survey large-scale populations for ample variants including rare alleles in a cost-effective and efficient manner (Neves et al. 2013; Suren et al. 2016). We utilized the NimbleGen SeqCap EZ system (Roche NimbleGen, Inc., Madison, WI) for genome target enrichment and discovered over 2.8 million SNPs using exon-based probes and the DNA of 375 trees from a clonally propagated loblolly pine population (Lu et al. 2016). Phenotypic data collected from this population was associated with the genotyped SNPs to investigate the associations of genetic variation with the traits. Additionally, we examined genetic correlations between the traits, geographical variation within the traits, and the exome-wide individual heterozygosity-trait correlation (HTC). The main reason to study HTC was to see if higher individual heterozygosity is associated with homeostasis and better performance (Babushkina et al. 2016). The main objective of this study was to explore the genetic factors that influence adaptive performance of loblolly pines and to contribute to future breeding efforts.

Materials and methods

Plant material

The loblolly pine population used in this study was originally established for the Allele Discovery of Economic Pine Traits 2 (ADEPT2) project (Cumbie et al. 2011). Maternal parents of the ADEPT2 population were sampled across the natural range of loblolly pine. Trees were grown from open-pollinated seeds from these maternal parents for 1 year and then one to three seedlings from each parent were hedged and established for use in the ADEPT2 project. During the spring of 2010, rooted cuttings of 384 trees from the ADEPT2 population were established at the Harrison Experimental Forest at the Southern Institute of Forest Genetics (30°63′N, 89°06′ W, near Saucier, Mississippi). A randomized incomplete block alpha design was used, with 3 replications of 24 incomplete blocks of size 16 ($r = 3$, $s = 24$, $k = 16$, 4 trees × 4 trees in each block; trees were spaced 3 m by 3 m). These trees were used for phenotyping and collection of foliage for DNA isolation, exome enrichment, and genotyping by sequencing. Prior to field planting, each clone was SSR genotyped using DNA isolated from bulked leaf tissue samples of the three trees per clone. Individual trees were genotyped for clones that showed evidence of multiple genotypes in the bulked DNA. Non-conforming trees were removed from the clone designation and not planted within the experiment.

Phenotyping

Sample collection and measurement of the traits we studied were conducted during the fourth growing season at the Harrison Experimental Forest. Between May 25 and June 24, 2014, the following traits were measured and recorded for each tree: total height was measured using a meter pole; branch angle, represented by the average of three branch angles relative to level at the third major whorl from the top, was measured using a digital level inclinometer; stem diameter at 0.46 m (18 in) high above the ground was measured using a caliper; crown width at the third major whorl from the top was measured using a measuring tape. In addition, total height was measured again at the beginning and part way into the 2015 growing season, and height growth in the first half of the 2015 growing season was calculated as the difference of two measurements in 2015.

South-facing and fully expanded needles from a point half way to the top of each tree were collected for assessment of specific leaf area, $\Delta^{13}C$, and nitrogen concentration. Specific leaf area was calculated as 20 needles' leaf area divided by the dry weight of these 20 needles. The leaf area was measured using a LAI 3000 scanner (Li-Cor, Lincoln, NE). The needles were dried at 65 °C for 72 h.

For $\Delta^{13}C$ and nitrogen concentration analyses, 5 or 6 needles from each dried needle sample were ground into fine, homogeneous powders with a ball mill (MM400, Retsch, Hann, Germany). The samples were weighed in tin capsules and analyzed by EA-IRMS (Delta V, Thermo Scientific, Waltham, MA) in the Stable Isotopes for Biosphere Science Laboratory at Texas A&M University (http://sibs.tamu.edu; College Station, TX). The carbon isotope ratios were reported against VPDB with calibrated laboratory standards. We calculated $\Delta^{13}C$ values using the formula $\frac{\delta a-\delta p}{1+\delta p}$ (Farquhar et al. 1989), where δa and δp represented the isotope composition of air and leaf tissue, respectively (δa was assumed to be −8 ‰). The nitrogen concentration was reported as a mass percentage.

The pitch canker disease resistance data was taken from a published study conducted at the University of Florida (Gainesville, FL) (Quesada et al. 2010). The mean lesion lengths from 4 replications per clone were available for 317 trees used in this study. Therefore, only 317 trees were included for pitch canker resistance analyses.

Phenotypic data analyses

A mixed model analysis (Mclean et al. 1991) was used in order to assess the clonal effects for the measured traits: $y_{ijk} = \mu + r_i + b_k(r_i) + c_j + e_{ijk}$, where y_{ijk} is the phenotypic value for the jth clone in the ith replication and kth block, μ is the population mean, r_i is the fixed variable of replication (i = 1–3), c_j is the random variable of clone (j = 1–384, approx. NID(0, σ^2_C)), $b_k(r_i)$ is the random variable of block nested within replication (k = 1–24, approx. NID(0, $\sigma^2_{b(r)}$)), and e_{ijk} is the error term (approx. NID(0, σ^2_e)). The best linear unbiased prediction (BLUP) estimates for each trait were used as phenotypic values in further analyses. The clonal repeatability was estimated using the formula: $r_{clone} = \frac{\sigma^2_{\text{clone}}}{\sigma^2_{\text{clone}} + \frac{\sigma^2_e}{rep}}$. The standard errors for clonal repeatability estimates were calculated as described by Jayaraman (1999).

Quesada et al. (2010) reported the mean pitch canker lesion length phenotypes using different ramets of the same clones that were genotyped in our study. The dataset contained four replicates with missing data for some clones. To deal with this unbalanced dataset, we used a slightly different mixed model analysis to enable analysis with missing data. The BLUP estimate for mean lesion length induced by pitch canker disease was acquired using the model: $y_{ij} = \mu + r_i + c_j + r_i c_j + e_{ijk}$, where y_{ij} is the phenotypic value for the jth clone in the ith replication, μ is the population mean, r_i is the fixed variable of replication (i = 1–4), c_j is the random variable of clone (j = 1–317, approx. NID(0, σ^2_C)), $r_i c_j$ is the random variable for the interaction of replication by clone (approx. NID(0, σ^2_{rc})), and e_{ij} is the error term (approx. NID(0, σ^2_e)).

A total of 362 trees within this population have known maternal origins. This population was divided into three regions as described by Schmidtling (2001) using maternal origins. The eastern region includes states east of the Mississippi River, the western region includes the states of Arkansas and Louisiana, and the far west region includes the states of Texas and Oklahoma (Fig. 1). An analysis of variance (ANOVA) was applied to compare the BLUP differences of each trait for individuals grouped by their regions of maternal origin. The BLUP estimates were also used to correlate with the climate variables (Supplementary material 2—Table S2) from Eckert et al. (2010). All the statistical analyses were conducted using the JMP Pro 12 statistical software (SAS Institute, Cary, NC).

Genotypic data

Genotypic data were obtained by the authors for 375 trees in this ADEPT2 population (Lu et al. 2016). The NimbleGen SeqCap EZ system (Roche NimbleGen, Inc., Madison, WI) was used to capture and enrich the exome of each tree. The detailed procedures of probe design, raw SNP detection, and genotyping are described in Lu et al. (2016). In this study, the raw SNPs were filtered, accepting only bi-allelic sites with at least 5× sequencing coverage for all of the individuals without missing data and a minor allele frequency (MAF) ≥0.01. A total of 2,822,609 SNPs were retained. Among these SNPs, 1,199,938 (43%) reside in CDS, 36,533 (1%) in five prime untranslated regions (5′ UTR), 70,377 (2%) in three prime untranslated regions (3′ UTR), 516,268 (18%) in introns,

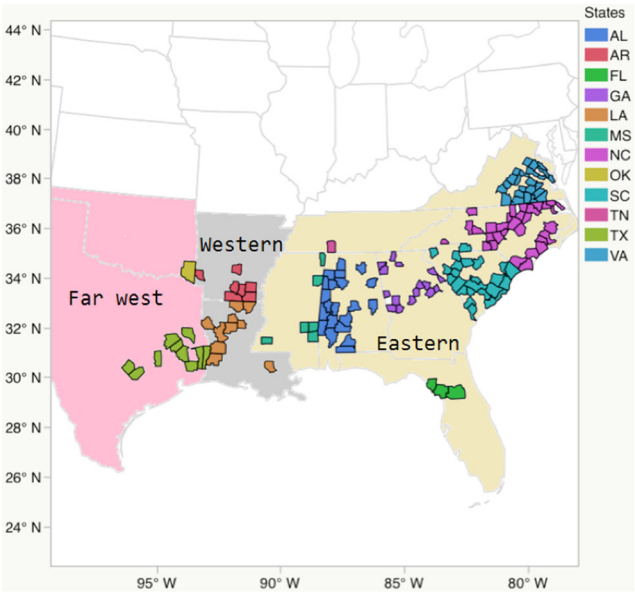

Fig. 1 The states and counties of origin of the maternal parent of the studied loblolly pines. The range was divided into three regions: far west (the states of Texas and Oklahoma, highlighted in *pink*), western (the states of Arkansas and Louisiana, highlighted in *gray*), and eastern (east of the Mississippi River, highlighted in *beige*)

and the remaining SNPs (36%) in unclassified regions, probably unannotated regulatory elements of genes or intergenic regions. A total of 94,478 haplotype blocks were detected for this population using PLINK 1.9 (http://pngu.mgh.harvard.edu/purcell/plink/) (Purcell et al. 2007).

Association analyses

Association analyses for the individual SNPs and phenotypic traits were conducted using TASSEL 5.0 (Bradbury et al. 2007). To avoid false positive associations, association analysis requires taking into account confounding factors, such as neutral population structure that is caused by selectively neutral random factors. Selectively neutral markers are required to infer neutral population structure, but the 2.8 million SNPs discovered in this study were mostly within exons and gene-adjacent regions and could be potentially under selection. Therefore, for estimating covariates to adjust for neutral population structure, we used putatively selectively neutral simple sequence repeat (SSR) markers that were previously genotyped in 249 out of the 375 genotyped trees in this population (Eckert et al. 2010). We identified these 249 trees as an individual population, named the structure (*str*) population. Population structure within this population was mainly due to the Mississippi River discontinuity (Lu et al. 2016). We identified the 307 trees from east of the Mississippi River as an individual population named the *east* population. There are 55 trees in our association mapping population with known origins from west of the Mississippi River named the *west* population, but the size was too small to be used as a separate population for any association analyses. Therefore, three populations, *total* (N = 375), *east* (N = 307), and *str* (N = 249), were used to perform association analyses. For the *total* and *east* populations, the simple general linear model (GLM) method (*S* model) and the mixed linear model (MLM) method incorporating a kinship matrix (*K* model) were applied, but without taking into account neutral population structure due to the lack of putative selectively neutral markers to infer it. For the *str* population, in addition to the *S* and *K* models, the GLM incorporating the covariate to adjust for neutral population structure (*Q* model) and the MLM incorporating both the kinship matrix and neutral population structure covariate (*QK* model) were applied. Population structure covariate was estimated using the software STRUCTURE (Pritchard et al. 2000; Hubisz et al. 2009) and 23 SSR markers as described by Eckert et al. (2010). A kinship matrix for each population was estimated by TASSEL 5.0 (Bradbury et al. 2007) using the SNP markers. The kinship relatedness is low in this population with an average range between −0.03 and 0.10 (excluding the self-relatedness). Quantile-quantile (Q-Q) plots were generated for observed against expected $-\log_{10}$ (P) to examine the model fitness, where observed P values were obtained from association mapping and expected P values from the assumption that no association occurred between marker and trait. Significance of associations between loci and traits were determined by the P values. A corrected Bonferroni threshold 0.05/94,478 = 5.29E−7, where 94,478 was the number of haplotype blocks, was applied to screen for significant loci.

Individual exome-wide HTC analyses

The observed individual exome-wide heterozygosity values were calculated for each of 375 trees within this population using the 2,822,609 genotyped SNPs and the "het" function in the software VCFtools (Danecek et al. 2011). The raw SNPs were filtered by accepting only bi-allelic sites with at least 5× sequencing coverage for all of the individuals without missing data and MAF ≥0.01. Therefore, the missing data in the SNP set were excluded for all analyses including heterozygosity detection. BLUP estimates for each trait were used to correlate with the heterozygosity (HTCs). Pearson correlation coefficients (r) were used to evaluate the HTCs. Within this population, 362 trees have known maternal origins and could be separated accordingly to the two populations on the west and east sides of the Mississippi River, respectively. Since distinct geographical and genetic structure patterns along the Mississippi River exist within this population, r was calculated separately in the *total* (N = 375), *east* (N = 307), and *west* (N = 55) populations.

SNP interaction analyses

The epistatic SNP-SNP interaction test was implemented using PLINK 1.9 (http://pngu.mgh.harvard.edu/purcell/plink/) (Purcell et al. 2007). A Bonferroni threshold of 0.05/3,176,320,469,273 = 1.57E−14, where 3,176,320,469,273 was the number of all SNP pairs, was used to correct the multiple comparisons.

Annotation of genes that contained SNPs associated with traits

The information for genes that contained SNPs associated with traits was obtained from loblolly pine Gene Annotation v3.0 (http://dendrome.ucdavis.edu/ftp/Genome_Data/genome/pinerefseq/Pita/v1.01/Pita_Annotation_v3.0/) (Wegrzyn et al. 2014). The loblolly pine reference genome assembly and annotation are under active improvement. The regulatory sequences such as promoters, enhancers and silencers have not been identified yet. SNPs within 5000 bp downstream or upstream of a gene were considered to be within a putative regulatory sequence of the gene. If a SNP was located in a region without annotation, the flanking sequence 700 bp upstream and downstream of the SNP was used as a query to do a Blastx search against the entire

National Center for Biotechnology Information (NCBI) non-redundant (nr) protein database (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

Clonal variances were obtained from the models used to obtain the BLUPs. The percentage of clonal variance explained by the effects of identified SNPs or SNP-SNP interactions was estimated by comparing the model incorporating the SNPs or SNP-SNP interactions as random effects with the reduced model without SNP effects. The percentage of phenotypic variance accounted for by the effects of identified SNPs or SNP-SNP interactions was determined using a similar approach. The detailed formulas were described in Quesada et al. (2010). To obtain the additive and dominance effects for the SNPs detected by association analyses, the loci in Hardy-Weinberg Equilibrium (HWE) with all three genotype classes present were treated as fixed effects and tested in linear regressions.

For the identified SNPs located in CDS, the effect of SNP substitution on the amino acid was investigated by aligning the sequences with the SNPs and the corresponding transcripts from the loblolly pine Gene Annotation v3.0 (http://dendrome.ucdavis.edu/ftp/Genome_Data/genome/pinerefseq/Pita/v1.01/Pita_Annotation_v3.0/) (Wegrzyn et al. 2014) using the Clustal Omega software (http://www.ebi.ac.uk/Tools/msa/clustalo/) (Sievers et al. 2011). The software ExPASy Translate (http://web.expasy.org/translate/) (Gasteiger et al. 2003) was utilized to translate the DNA sequences to amino acid sequences.

Results

Clonal repeatability, genetic correlations, and geographical variation

Clonally replicated loblolly pines from the ADEPT2 population were measured for traits of interest during their fourth growing season at the Harrison Experimental Forest (Supplementary material 1—Table S1). At the time of phenotyping, 78% of the trees planted in 2010 had survived. For each trait measured, we calculated the clonal repeatability within this population (Table 1). Clonal repeatability calculations indicated that 31% of the variation in specific leaf area could be attributed to genetic effects. For growth and architecture traits, 60% of the variation in branch angle, 62% in crown width, 56–62% in total height, 54% in stem diameter, and 11% in 2015 height growth could be attributed to genetic effects. Due to the low clonal repeatability observed, height growth in 2015 (i.e., difference between two height measurements in the 2015 growing season) was excluded from further analyses. Clonal repeatability values for $\Delta^{13}C$ (85%) and nitrogen concentration (76%) were higher than for other traits.

Bivariate Pearson correlation coefficients of clone means (i.e., BLUP estimates) were calculated to investigate the genetic correlations between the traits (Table 2). Strong positive correlations were observed between total height and crown width, diameter and crown width, and diameter and total height. $\Delta^{13}C$ was correlated with branch angle, crown width, total height, and nitrogen concentration. A small negative correlation existed between crown width and branch angle. Nitrogen concentration was positively correlated with specific leaf area, crown width and total height.

We compared BLUP estimates of individuals grouped by their regions of maternal origin (Fig. 1). Significant differences ($P < 0.05$) were identified for crown width ($P = 0.002$); stem diameter ($P = 0.005$); nitrogen concentration ($P = 0.044$); and $\Delta^{13}C$ ($P = 0.004$) between eastern, western, and far west regions (Fig. 2). For the other traits, no significant differences were found among different regions. Trees from the eastern region showed greater crown width, stem diameter, and nitrogen concentration than those from the western and far west regions, indicating genotypes from east of the Mississippi River tend to have a higher early growth rate (through 4 years in the field) than those from west. The lower $\Delta^{13}C$ values of eastern trees indicate higher WUE than those of trees from western and far west regions.

The BLUP estimates were correlated with climate variables including minimum and maximum temperature, monthly average precipitation, and 19 bioclimatic variables (Supplementary material 2—Fig. S1). Height and stem diameter showed positive correlations with the temperature of the coldest months and negative correlations with the annual temperature range and temperature seasonality. Plus, they showed negative correlations with the precipitation of the coldest and driest months and positive correlations with the precipitation of the warmest months. The nitrogen concentration was negatively correlated with temperature except it was positively correlated with the temperature seasonality and annual temperature range. The nitrogen concentration was generally negatively correlated with the precipitation, especially with the precipitation of the coldest quarter. No significant correlation was observed between the climate variables and other traits.

Individual exome-wide HTCs

To test whether gene-based individual exome-wide heterozygosity affects adaptive traits, we used exome-based SNPs to calculate individual multi-locus heterozygosity and correlated the heterozygosity values with the BLUP estimates of each trait within the clonally tested populations. Two significant HTCs were detected in the *total* population and one HTC in the *east* population (Table 3). No significant correlations were detected in the *west* population. Heterozygosity was negatively correlated with $\Delta^{13}C$ ($r = -0.173$ in the *total* population and -0.137 in the *east* population). Since a smaller $\Delta^{13}C$ value indicates higher WUE, the negative

Table 1 Phenotypic data summary

Trait	Mean	Standard deviation	Maximum	Median	Minimum	Number of trees measured	Clonal repeatability (standard error)
SLA (cm^2/mg)	28.88	5.19	43.38	29.06	14.11	920	0.31 (0.10)
BA (°)	35.36	8.75	62.50	35.93	9.30	921	0.60 (0.18)
CW (m)	1.40	0.34	3.60	1.39	0.43	922	0.62 (0.19)
DIA (m)	0.05	0.02	0.10	0.05	0.01	922	0.54 (0.17)
2014H (m)	3.25	0.73	5.44	3.30	0.85	922	0.56 (0.17)
2015HB (m)	3.99	0.93	6.78	4.04	1.07	918	0.61 (0.19)
2015HA (m)	4.66	0.97	7.44	4.70	1.52	908	0.62 (0.19)
2015HG (m)	0.67	0.23	2.41	0.65	0.07	908	0.11 (0.04)
Δ^{13}C, permil, VPDB	24.28	0.56	25.81	24.30	22.21	920	0.85 (0.24)
N (%)	0.93	0.09	1.28	0.94	0.58	920	0.76 (0.22)

SLA specific leaf area, *BA* branch angle, *CW* crown width, *DIA* stem diameter, *2014H* total height in 2014, *2015HB* total height at the beginning of the 2015 growing season, *2015HA* total height part way into the 2015 growing season, *2015HG* height growth in the first half of the 2015 growing season, Δ^{13} *C* carbon isotope discrimination, *N* nitrogen concentration

relationship between the heterozygosity and the Δ^{13}C may indicate that higher heterozygosity is positively associated with WUE. A positive correlation between the heterozygosity and the nitrogen concentration ($r = 0.124$) was also observed.

Marker-trait association analyses

Association analyses were performed using *S* and *K* models for the *total* and *east* populations, and all *S*, *K*, *Q*, and *QK* models were used for the *str* population. The Q-Q plots indicated that better fits could be observed with different models. For the traits of specific leaf area, branch angle, stem diameter, total height in 2014, total height at the beginning, and part way into the 2015 growing season and Δ^{13}C, the *S* model was the best fit for the *east* population, but the *Q* model was better for the *str* population, and they were selected for further analyses. For nitrogen concentration, the *K* model was the best fit for the *east* population and the *Q* model for the *str* population. Similarly, for crown width, the *K* model for the *east* population and the *QK* model for the *str* population were selected.

Table 2 Pearson correlation coefficients of clone BLUP between traits

	SLA	BA	CW	2014H	DIA	Δ^{13}C	N
SLA		NS	NS	NS	NS	NS	0.140
BA			−0.107	NS	NS	0.154	NS
CW				0.757	0.769	−0.189	0.123
2014H					0.897	−0.138	0.113
DIA						NS	NS
Δ^{13}C							−0.251
N							

SLA specific leaf area, *BA* branch angle, *CW* crown width, *DIA* stem diameter, *2014H* total height in 2014, Δ^{13} *C* carbon isotope discrimination, *N* nitrogen concentration, *NS* non-significant at $P < 0.05$

For pitch canker resistance, the *K* model for the *total* and *east* populations and the *K* and *QK* models for the *str* population were selected (Supplementary material 2—Table S3). The purpose of selecting different models using Q-Q plots is to select the models that correct the best for the systematic bias. Models that showed the least distorted Q-Q plots were selected. The associations with *P* values that passed the Bonferroni threshold were kept for further analyses.

Associations were identified with specific leaf area (5 SNPs), branch angle (2), crown width (3), stem diameter (4), total height in 2014 (4), total height at the beginning of the 2015 growing season (8), total height part way into the 2015 growing season (7), Δ^{13}C (4), nitrogen concentration (2), and pitch canker resistance (7) (Table 4). When the height-related SNPs were combined, 9 different SNPs were associated with total height. Two loci were detected for both total height and stem diameter; therefore, a total of 34 different SNPs were identified as associated.

The VCFtools software (Danecek et al. 2011) was used to calculate the minor allele frequencies (MAFs) and to test for HWE (Table 4). Of the 34 associated SNPs, 4 (12%) departed from HWE. Fourteen SNPs (41%) had MAF greater than 0.05 with a range from 0.07 to 0.32. The MAFs of other SNPs were between 0.01 and 0.04. The effects of these SNPs were relatively small, as r^2 ranged between 0.08 and 0.14. The percentages of clonal and phenotypic variance explained by the SNPs were estimated by comparing the full models (including significant SNPs as random variables in the model) and reduced models (without the significant SNPs in the model). The percentage of clonal and phenotypic variance explained by each locus ranged from 5 to 27% and from 2 to 6%, respectively (Table 4). If the individual SNPs for each trait are applied in the models altogether, they can explain 4–18% of the phenotypic variance (Supplementary material 2—Table S4). As the phenotypic data used in this study were collected from only one location, the percentage of each SNP composing the phenotypic variance might be overestimated.

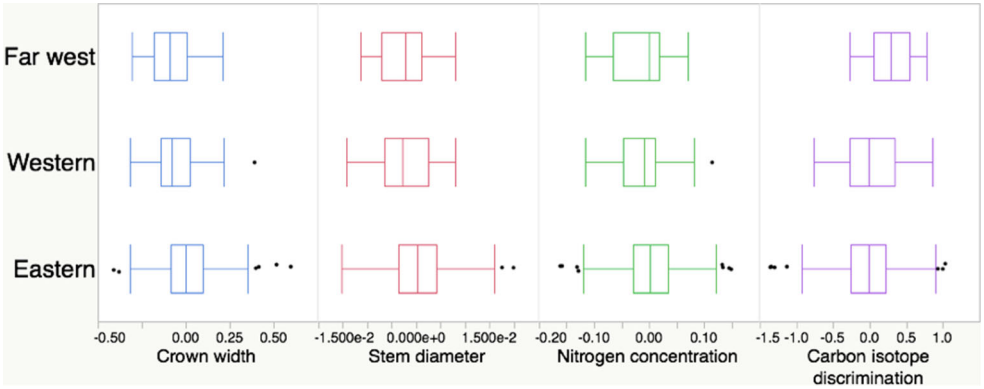

Fig. 2 BLUP estimate distribution for the traits with significant differences ($P < 0.05$) among far-west, western, and eastern regions

To analyze the additive and dominance effects of the associated SNPs, only the six SNP loci in HWE with all three genotype classes represented were tested using linear regressions (Table 5). Additive and dominance effects were significant for all the tested SNP loci at $P < 0.05$, but the estimate for additive effect of locus tscaffold8954_1999_T_G for specific leaf area was not precise as the standard error was greater than half the value of the estimate. Among the six tested loci, the ratio of dominance to additive effects showed the dominance effect was larger in magnitude than the additive effect at four loci. Among them, locus scaffold894573_84188_C_T had the largest dominance effect resulting in an increase in stem diameter. At the remaining loci, dominance effects showed a similar magnitude to additive effects.

Epistasis analyses

Eleven SNP-SNP interactions were identified in associations with branch angle (1 SNP-SNP interaction), crown width (2), total height in 2014 (2), $\Delta^{13}C$ (2), nitrogen concentration (1), and pitch canker resistance (3) (Table 6). None of the epistatic loci were detected in the marker-trait association test. All the identified SNPs were in HWE. The loci scaffold901027_91326_G_C and tscaffold6745_463233_G_A were found to be involved in two identified interactions. Sixteen SNPs (80%) had MAF greater than 0.05.

Each SNP-SNP interaction accounted for 15–30% of the clonal variance and 8–10% of the phenotypic variance for the associated trait (Table 6). The combined SNP interaction effects for each trait accounted for 9–16% of the phenotypic variance (Supplementary material 2—Table S4). The addition of these epistatic loci to the associated loci increased the explained proportion of phenotypic variance by 2–14% (Fig. 3). All the identified interactions were between SNPs located on different scaffolds. The most significant SNP-SNP interaction was observed between the scaffold892137_41285_G_A and scaffold461440_154634_T_A loci. Their interaction was related to branch angle. The genotype combination of these two loci showed distinct phenotypic differences (Fig. 4).

Annotation of genes that contained SNPs associated with traits

The loci with the identified SNPs were annotated using loblolly pine Gene Annotation 3.0 (http://dendrome.ucdavis.edu/ftp/Genome_Data/genome/pinerefseq/Pita/v1.01/Pita_Annotation_v3.0) (Wegrzyn et al. 2014). Of the 34 SNPs identified by association, 19 resided in CDS, 2 in UTRs, 3 in introns, 4 in putative regulatory sequences, and 6 in unannotated (unclassified) regions (Table 4). Of the 20 SNPs identified in the epistasis analysis, 8 resided in CDS, 7 in introns, 1 in a UTR, 1 in a putative regulatory sequence, and 3 in unannotated regions (Table 6).

For the identified SNPs located in tentative genes, the function annotations are presented in Tables 4 and 6. For the identified SNPs located in regions without annotation, the flanking sequences around the SNPs were used for a Blastx search to identify the functions of genes that contained SNPs associated with traits.

We investigated the effect of SNP substitution on amino acid sequence by aligning and comparing the amino acid sequences translated from the sequences containing the SNPs

Table 3 Significant SNP heterozygosity-trait correlations

Population (number of trees)	Trait	*r*	*P*
Total ($N = 375$)	Carbon isotope discrimination	−0.173	0.001
	Nitrogen concentration	0.124	0.017
East ($N = 307$)	Carbon isotope discrimination	−0.137	0.017

r Pearson correlation coefficient

Table 4 SNPs significantly associated with measured traits

Trait	SNP[c]	r^2	*P* value	MAF	PCV[a]	PPV[b]	Location	Gene	Annotation
SLA	scaffold585302_58207_A_G	0.10	2.13E−08	0.02	26.89	2.69	Intron	PITA_000065619	Auxin-responsive protein IAA13
SLA	scaffold901438_87380_A_G	0.08	2.24E−07	0.02	22.73	2.28	CDS (I to V)	PITA_000046166	Predicted, importin subunit alpha-like
SLA	tscaffold7553_225423_C_T	0.09	1.94E−07	0.07	27.27	2.73	CDS (S to L)	PITA_000013301	Predicted, E3 ubiquitin-protein ligase UPL3
SLA	tscaffold9126_24299_G_A	0.09	9.28E−08	0.03	27.27	2.73	CDS (E to K)	PITA_000025957	Leucine-rich repeat
SLA	tscaffold8954_1999_T_G	0.09	5.28E−07	0.08	26.14	2.62	P5′RS	PITA_000030967	Putative copper-transporting ATPase HMA5-like isoform X1
BA	scaffold744759_88802_A_G	0.10	4.21E−07	0.08	7.59	2.47	CDS (synonymous)	PITA_000043049	Putative disease resistance protein
BA	scaffold886562_42033_C_T	0.09	1.73E−07	0.02	11.72	3.81	CDS (Q to stop)	PITA_000057448	RNA editing factor OTP81
CW	C32326948_8323_G_C	0.12	1.19E−07	0.02	8.05	2.79	Unclassified	Unclassified	Kinesin heavy chain[e]
CW	C32326948_24090_G_C	0.12	1.19E−07	0.02	8.05	2.79	3′UTR[d]	PITA_000078936	Putative ATP-dependent RNA helicase
CW	scaffold900647.1_1652_T_C	0.12	1.19E−07	0.01	12.25	4.24	CDS (P to L)	PITA_000085365	Sec-independent protein translocase protein TatA
DIA	scaffold4572_62040_G_C	0.10	3.67E−07	0.32	13.56	3.72	P5′RS	PITA_000055333	Probable disease resistance protein At4g33300
DIA	scaffold869612_10751_C_T	0.08	2.92E−07	0.04	10.45	2.88	CDS (S to L)	PITA_000078887	Putative RNA-binding protein
DIA	scaffold894573_84188_C_T	0.09	2.78E−07	0.04	22.57	6.20	CDS (T to I)	PITA_000057281	Voltage-dependent anion channel
DIA	tscaffold5105_261867_C_T	0.08	5.03E−07	0.02	18.44	5.06	CDS (R to stop)	PITA_000008001	Probable receptor-like serine/threonine-protein kinase At5g57670 isoform X2
TH	C32139602_5870_A_G	0.08	2.60E−07	0.03	11.54	3.45	CDS (K to R)	PITA_000087565	Predicted, eukaryotic translation initiation factor 5-like
TH	C32495018_13991_T_C	0.11	7.64E−08	0.21	7.69	2.30	P5′RS	PITA_000063143	Glycoside hydrolase family
TH	scaffold4572_62040_G_C	0.12	1.81E−08	0.32	12.50	3.70	P5′RS	PITA_000055333	Probable disease resistance protein At4g33300
TH	scaffold802276_312_C_A	0.09	5.15E−07	0.09	18.75	5.56	Unclassified	Unclassified	Bifunctional pinoresinol-lariciresinol reductase[e]
TH	scaffold887377_18672_C_T	0.12	8.75E−08	0.26	6.25	1.85	CDS (A to V)	PITA_000087672	Unknown
TH	scaffold898261_152482_C_T	0.08	2.50E−07	0.02	18.75	5.56	5′UTR[d]	PITA_000028701	Rab family GTPase; Small GTPase superfamily, Ras
TH	tscaffold5847_116012_T_C	0.08	3.96E−07	0.01	15.38	4.60	CDS (W to R)	PITA_000020895	Predicted, lysosomal beta glucosidase
TH	tscaffold5105_261867_C_T	0.08	2.24E−07	0.02	15.38	4.60	CDS (R to stop)	PITA_000008001	Probable receptor-like serine/threonine-protein kinase At5g57670 isoform X2
TH	tscaffold7910_321796_C_T	0.11	1.77E−07	0.13	7.69	2.30	Intron	PITA_000014254	*Copia*-type polyprotein, putative
$\Delta^{13}C$	scaffold17165_86468_G_A	0.09	1.71E−07	0.01	10.00	6.25	Unclassified	Unclassified	Polyadenylate binding protein[e]
$\Delta^{13}C$	scaffold75093.3_123328_G_A	0.09	1.79E−07	0.01	10.00	6.25	CDS (synonymous)	PITA_000037598	Predicted, histone-lysine N-methyltransferase
$\Delta^{13}C$	scaffold109938_8086_T_C	0.09	1.78E−07	0.01	5.28	3.13	Unclassified	Unclassified	Organic cation/carnitine transporter[e]
$\Delta^{13}C$	tscaffold4948_129087_G_A	0.10	3.99E−07	0.01	5.00	3.13	CDS (synonymous)	PITAhm_003076	Predicted, small heat shock protein, chloroplastic
N	C31890840_9044_C_T	0.10	1.76E−07	0.14	9.12	4.20	Unclassified	Unclassified	F-box/kelch-repeat protein At3g06240-like[e]
N	C32008620_13926_T_C	0.11	4.70E−07	0.12	8.95	4.12	CDS (synonymous)	PITA_000091703	C3H4 type zinc finger protein
PC	C32571366_88829_G_A	0.14	3.67E−07	0.19	8.82	3.42	CDS (synonymous)	PITA_000019796	Predicted, Leucine-rich repeat
PC	scaffold394991_238149_C_A	0.12	3.98E−07	0.09	7.44	2.89	Unclassified	Unclassified	Probable quinone oxidoreductase[e]
PC	scaffold490830_154983_T_A	0.14	1.31E−08	0.01	14.84	5.77	CDS (I to K)	PITA_000030834	Predicted, probable peptide/nitrate transporter At3g53960-like
PC	scaffold771268_7109_G_A	0.13	2.01E−07	0.08	13.14	5.10	P3′RS	PITA_000095986	Predicted, probable aminotransferase TAT2-like
PC	tscaffold758_520984_C_T	0.12	2.98E−07	0.17	7.14	2.85	CDS (A to V)	PITA_000009116	Unknown
PC	tscaffold1272_36798_C_T	0.11	3.31E−07	0.01	8.47	3.29	Intron	PITAhm_002502	Unknown
PC	tscaffold5506_305141_G_A	0.11	2.07E−07	0.01	11.49	4.46	CDS (synonymous)	PITA_000020926	Predicted, dehydration-responsive protein RD22-like isoform 2

SLA specific leaf area, *BA* branch angle, *CW* crown width, *DIA* stem diameter, *TH* total height, $\Delta^{13}C$ carbon isotope discrimination, *N* nitrogen concentration, *PC* pitch canker resistance, *MAF* minor allele frequency, *CDS* coding sequences, and it is followed by the substitution type, *P5′RS* putative 5′ regulatory sequence, *P3′RS* putative 3′ regulatory sequence

[a] The percentage of clonal variance accounted for by each SNP

[b] The percentage of phenotypic variance accounted for by each SNP

[c] SNPs were named using scaffold names with the SNP position number in the nucleotide sequence followed by the major and minor SNP alleles

[d] Untranslated sequences

[e] SNPs that were not located on annotated sequences, instead, the flanking sequences around SNPs were used to query against the NCBI GenBank non-redundant protein database using Blastx

and the corresponding transcripts from loblolly pine Gene Annotation v3.0 http://dendrome.ucdavis.edu/ftp/Genome_Data/genome/pinerefseq/Pita/v1.01/Pita_Annotation_v3.0/ (Wegrzyn et al. 2014) (Tables 4 and 6). Of the 27 SNPs found in CDS, 19 (70%) caused non-synonymous substitutions. Two non-synonymous substitutions generated stop codons and caused premature truncation of the coding sequences, while the others resulted in amino acid replacements. One non-synonymous substitution resulting in premature truncation occurs on the SNP locus associated with branch angle. It is located in a gene that encodes an RNA-editing factor. Another one occurs on the locus associated with stem diameter and total height. It is located in a gene encoding a receptor-like serine/threonine-protein kinase.

Discussion

Clonal repeatability and genetic correlation

The clonal repeatability was high for all the measured traits except for height growth for the first half of the 2015 growing season. However, with the exception of total height, the traits in this study were only measured on one population in 1 year in one location, and therefore, the clonal repeatability estimates were up-biased. Nonetheless, previous studies have shown these traits to be moderately heritable in loblolly pines. Emhart et al. (2007) determined the broad-sense heritabilities of crown radius, leaf area, and branch angle in loblolly pine to be 0.20, 0.25, and 0.26, respectively, when estimated from a combined analysis across families. Baltunis et al. (2008) reported that the across-site estimate of broad-sense heritability for $\Delta^{13}C$ was 0.19, and Emhart (2005) reported similar estimates 0.23 and 0.17 based on two separate years of sampling. For nitrogen concentration, Cumbie et al. (2011) reported that the broad-sense clone mean heritability was 0.42. Pitch canker resistance in loblolly pines is also a heritable and complex trait with a continuous distribution across clones (Kayihan et al. 2005). Quesada et al. (2010) estimated 30–40% of the disease variation could be attributed to genetic effects. The clonal repeatability and the considerable phenotypic variation suggested the population used in this study was suitable for association mapping. The lower clonal repeatability for height growth during the first part of the 2015 growing season agrees with previous conclusions that annual growth increments have low heritability (White et al. 2007; Shmulsky and Jones 2011) and is more affected by environmental effects, such as availability of light, water, and nutrition rather than the genetic components. Therefore, height growth in 2015 was excluded from further analyses.

Strong positive correlations between crown width, height, and stem diameter along with a weak negative correlation between crown width and branch angle indicated that bigger trees tended to have wider crowns and flatter branches, which is in agreement with previous progeny tests of loblolly pines measuring 9- or 10-year-old trees on four sites (Lambeth and Hubert 1997). The wider crown and flatter branches enabled the trees to capture light better and to be more competitive than other trees, thus accumulating more biomass. The genetic correlations of crown width with other growth traits along with its medium to high heritability as reported in the previous studies suggested crown width could be a key component of productivity, and selection for crown width could favor growth traits (Lambeth and Hubert 1997; Emhart et al. 2007). Genes affecting crown structure in loblolly pine have been rarely explored in the past. Considering heritability of crown structure and growth traits, molecular markers associated with these traits could be valuable in marker-assisted selection (MAS) (Nelson and Johnsen 2008).

Due to a changing climate, forest trees with better adaptive characteristics such as superior photosynthetic and water use abilities will be needed in the future (IPCC 2014). In this study, slight positive correlations existed between nitrogen concentration and specific leaf area, crown width and total height, suggesting a higher nitrogen concentration may have increased the leaf area and tree size through a promoted photosynthetic ability since nitrogen is indispensable for rubisco, a key enzyme of the Calvin cycle (Bloomfield et al. 2014). It should be noted that the correlations shown here for nitrogen concentration as well as $\Delta^{13}C$ were subtle, possibly because the trees contained in this population were originally from a broad geographic range, and these different genotypes may display distinct photosynthetic and water use strategies depending upon environments (Flanagan and Johnsen 1995).

Stable $\Delta^{13}C$ in plants reflects the balance between photosynthetic ability and stomatal conductance. It has long been used as a measure for WUE in forest trees as less discrimination is associated with higher WUE (Aitken et al. 1995; Baltunis et al. 2008). The trees with flatter branch angles, wider crowns, greater heights and higher nitrogen concentrations tended to have lower $\Delta^{13}C$, suggesting the fast-growing trees with higher light capture and photosynthetic ability have a better WUE. It is worth noting that these trees may also exhibit greater stomatal closure under high vapor pressure deficit (VPD) or limiting soil moisture since they are deploying large canopies and have greater nitrogen concentration. These canopies are perhaps more productive over some time period, but would be (perhaps) more vulnerable to hydraulic failure.

The empirical relationship between $\Delta^{13}C$ and growth can be negative, positive, or uncorrelated depending on the specific environment or species (Orians and Solbrig 1977; Aitken et al. 1995; Flanagan and Johnsen 1995; Li et al. 2013). It can be hypothetically explained that high WUE could be at the expense of growth (Orians and Solbrig 1977). However, neither our study, which was conducted in seasons with normal

Table 5 Additive and dominance effects for the SNP loci detected by association

SNP[a]	Trait	Additive effect[b]	Standard error of the additive effect	Dominance effect[b]	Standard error of the dominance effect	Dominance/additive
scaffold894573_84188_C_T	DIA	0.006	0.001	0.027	0.007	4.500
scaffold887377_18672_C_T	2015HA	0.093	0.032	0.11	0.039	1.183
scaffold887377_18672_C_T	2015HB	0.094	0.034	0.114	0.041	1.213
scaffold802276_312_C_A	2014H	−0.185	0.032	−0.364	0.105	1.968
C32008620_13926_T_C	N	0.016	0.006	−0.035	0.02	−2.188
tscaffold8954_1999_T_G	SLA	−0.11	0.105	−0.256	0.108	2.327
C32571366_88829_G_A	PC	0.132	0.031	0.158	0.036	1.197

DIA stem diameter, *2015HA* total height part way into the 2015 growing season, *2015HB* total height at the beginning of the 2015 growing season, *2014H* total height in 2014, *N* nitrogen concentration, *SLA* specific leaf area, *PC* pitch canker resistance

[a] SNPs were named using scaffold names with the SNP position number in the nucleotide sequence followed by the major and minor SNP alleles

[b] Additive and dominance effects were significantly different from zero based on ANOVA ($P < 0.05$)

precipitation, nor the study conducted by Cumbie et al. (2011) during seasons with drought supported this hypothesis for loblolly pine. It should be noted that in both studies, relatively young trees were measured. Our trees were measured in their fourth growing season and those in the Cumbie et al. (2011) study in their second growing season, indicating the possibility that young trees may use a different WUE strategy (Aitken et al. 1995).

Geographical variation and climate adaptation

Environmental heterogeneity and gradients drive the adaptation of forest trees, thus creating geographical variation within the natural range. Through adaptive and selectively neutral processes, loblolly pine developed geographic differences between populations east and west of the Mississippi River due to the 100,000-year refugia isolation. Reports have shown that loblolly pines from west of the Mississippi River are slower growing but more resistant to aridity and crowding (Schmidtling 2001). Our study demonstrated that genotypes with eastern origins tended to grow faster and have a better WUE than western and far west genotypes. The growth rate difference was consistent with previous results (Schmidtling 2001), but the drought tolerance within this population is difficult to judge since relatively young trees were measured and were under normal precipitation conditions during the growing seasons of sampling. Additionally, the trees in our population were grown from open-pollinated seeds collected from local seed orchards, and only the maternal origins can be determined, leaving the paternal origins uncertain. Therefore, the measured trees may have different drought tolerance phenotypes from the originally selected maternal parents.

The climate within the natural loblolly pine range is humid, warm-temperate with long, hot summers and mild winters (Baker and Langdon 1990). Our results supported that the coldness in winter is a limiting factor for loblolly pine growth since there were positive correlations between the temperature of the coldest months and the height and stem diameter. Trees with origins with a smaller temperature range and seasonality tend to grow bigger. The loblolly pine growth is negatively correlated with the precipitation of the coldest and driest months and positively correlated with the precipitation of the warmest months, probably because the growth season of loblolly pine is during the warm months. The negative correlations of nitrogen concentration with temperature and precipitation may be due to the higher nitrogen consumption under a warm and wet temperature, but the temperature seasonality and annual temperature range do not facilitate the nitrogen consumption.

Non-additive effects

Most robustly associated SNPs detected in loblolly pine association studies accounted for only a fraction of the total genetic variance in a trait (González-Martínez et al. 2007; Quesada et al. 2010). Similarly, in our study, only 5–27% of the clonal variance and 2–6% of the phenotypic variance could be explained by the associated SNPs. This agrees with the hypothesis that most quantitative traits are affected by many genes with small effects (Flint and Mackay 2009) as well as previous evidence in loblolly pine (Emhart et al. 2007; Quesada et al. 2010; Cumbie et al. 2011). It is also possible that genes with major effects remain undetected. However, in this study, we analyzed 2,822,609 SNPs in or near 48,391 high-quality tentative genes, so it seems unlikely that we would have missed the genes with major effects for every trait if they exist. Dissecting these quantitative traits and revealing their genetic control remain challenging. To better examine these traits, we extended our investigation beyond additive to non-additive effects, namely, dominance and epistasis. Though only a few SNP-SNP interactions were determined to be associated, they generally contributed more to the clonal and phenotypic

Table 6 Significant SNP-SNP interactions detected for traits

Trait	SNP1 (MAF)[a]	Location	Candidate gene/annotation	SNP2 (MAF)[a]	Location	Gene/annotation	PCV[b]	PPV[c]	*p* value
BA	scaffold892137_41285_G_A (0.09)	CDS (synonymous)	PITA_000033000/predicted, kinesin-like protein KIF18B	scaffold461440_154634_T_A (0.14)	intron	PITA_000021259/unknown	27.97	9.09	7.74E−17
CW	scaffold845670_729_G_A (0.09)	Unclassified	Unclassified/bark storage protein A-like[e]	C32549530_117801_G_T (0.04)	P3′RS	PITA_000046419/predicted, wound-induced protein 1-like	28.52	9.87	1.45E−14
CW	scaffold892370.2_38187_G_A (0.16)	CDS (synonymous)	PITA_000068396/Disease resistance-like protein	scaffold893009.2_42867_G_C (0.19)	CDS (A to P)	PITA_000068998/unknown	23.01	7.97	1.41E−14
2014H	C32560776_5010_G_A (0.09)	CDS (D to G)	PITA_000038233/TIR/NBS/LRR disease resistance protein	tscaffold6745_463233_G_A (0.12)	intron	PITA_000009136/ADP-ribosylation factor	29.53	8.62	9.93E−16
2014H	C32560776_5095_A_G (0.09)	Intron	PITA_000038233/TIR/NBS/LRR disease resistance protein	tscaffold6745_463233_G_A (0.12)	intron	PITA_000009136/ADP-ribosylation factor	27.21	7.94	6.02E−15
$\Delta^{13}C$	tscaffold488_11868_C_T (0.03)	CDS (V to A)	PITA_000007619/indole-3-acetic acid-amido synthetase GH3.6 (GH3 auxin-responsive promoter)	scaffold894878_57816_T_C (0.48)	intron	PITA_000057030/beta-1,3-n-acetylglucosaminyltransferase radical fringe	15.53	9.81	9.22E−15
$\Delta^{13}C$	scaffold117762.1_74566_C_G (0.04)	CDS (L to V)	PITA_000056667/predicted, cytochrome c oxidase subunit 6b-1-like isoform X2	tscaffold2954_107287_G_A (0.10)	intron	PITA_000004221/predicted, probable phosphoinositide phosphatase SAC9	16.67	10.53	1.52E−14
N	scaffold90489_7681_G_A (0.08)	Unclassified	Unclassified/leucine-rich repeat receptor-like serine/threonine-protein kinase BAM2[e]	scaffold807725_59239_A_T (0.28)	CDS (N to Y)	PITA_000043336/predicted, uncharacterized zinc finger protein At4g06634 isoform X2	20.95	9.65	3.69E−15
PC	C31644976_3483_C_T (0.03)	Unclassified	Unclassified/putative leucine-rich repeat receptor-like serine/threonine-protein kinase[e]	scaffold898450_46698_G_T (0.21)	intron	PITA_000069402/pentatricopeptide repeat-containing protein At1g12300, mitochondrial-like	22.81	8.86	2.69E−15
PC	scaffold901027_91326_G_C (0.05)	CDS (R to T)	PITA_000045299/transcription factor	tscaffold2366_165430_C_T (0.14)	3′UTR[d]	PITA_000012460/unknown	23.19	9.01	7.75E−15
PC	scaffold901027_91326_G_C (0.05)	CDS (R to T)	PITA_000045299/transcription factor	tscaffold362_1145764_A_T (0.14)	Intron	PITAhm_003320/mitochondrial processing peptidase subunit beta; GBF interacting protein 1-like	22.60	8.78	4.57E−15

BA branch angle, *CW* crown width, *2014H* total height in 2014, *$\Delta^{13}C$* carbon isotope discrimination, *N* nitrogen concentration, *PC* pitch canker resistance, *CDS* coding sequences, and it is followed by the substitution type, *P3'RS* putative 3′ regulatory sequence

[a] SNPs were named using scaffold names with the SNP position number in the nucleotide sequence followed by the major and minor SNP alleles

[b] The percentage of clonal variance accounted for by each SNP

[c] The percentage of phenotypic variance accounted for by each SNP

[d] Untranslated sequences

[e] SNPs that were not located on annotated sequences; instead, the flanking sequences around SNPs were used to query against the NCBI GenBank non-redundant protein database using Blastx

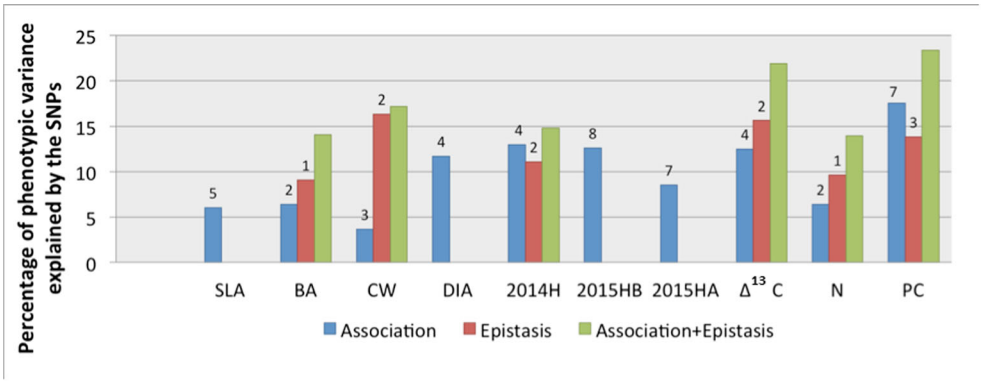

Fig. 3 Percentage of phenotypic variance for each trait contributed by the SNPs detected by association and epistasis. The numbers of the identified SNPs (association) or SNP-SNP interactions (epistasis) are presented above the *bars*. *SLA* specific leaf area, *BA* branch angle, *CW* crown width, *DIA* stem diameter, *2014H* total height in 2014, *2015HB* and *2015HA* total height at the beginning and part way into the 2015 growing season, respectively, $\Delta^{13}C$ carbon isotope discrimination, *N* nitrogen concentration, *PC* pitch canker resistance

variance than the additive loci (Fig. 3). None of the epistatic loci were discovered in association studies, indicating the additive and epistasis effects may determine the traits using independent networks (Zhang et al. 2015). The dominance effects are similar or larger in magnitude than additive effects. Although the additive effects are the main focus for loblolly pine breeding, the results from this research as well as previous studies indicated that both the dominance and epistasis effects should be considered in MAS of loblolly pine, since they might play important roles in capturing desirable traits (Eckert et al. 2009a; Cumbie et al. 2011). Additionally, only SNPs were investigated in this study; however, other kinds of polymorphisms such as indels, copy number variations, transposable elements, and stable epigenetic modifications could also explain the phenotypic variation.

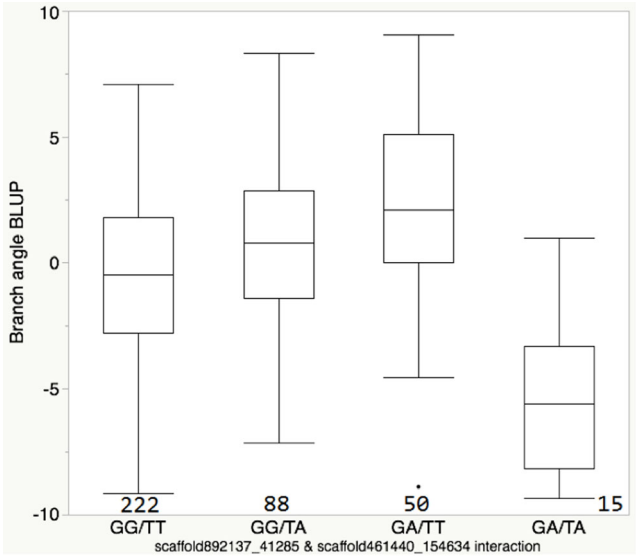

Fig. 4 Phenotypic differences between genotype combinations of the SNPs scaffold892137_41285_G_A and scaffold461440_154634_T_A. The interaction of these two loci is in association with branch angle. Numbers of individuals with the genotype combinations are at the bottom of the *bars*. The y-axis represents the BLUP estimates of branch angle for the individuals within the population

Non-coding and rare variants

Non-coding and rare variants were also used to address quantitative trait dissection problems. In this study, we used exome sequencing to identify over 2.8 million SNPs when filtering conditions were relaxed to include SNPs with MAF greater than 0.01, assuring a broad spectrum of SNPs containing rare alleles to be investigated. Moreover, in exome sequencing, the capture often extends to non-target regions, and therefore, variants adjacent to CDS and UTRs, including introns and putative regulatory elements, were also identified. Among the loci detected by the association analyses, more than half had MAF smaller than 0.05 (Table 4) and 44% resided in non-CDS (Table 4). These low-frequency and non-CDS variants are important resources for exploring the quantitative traits of loblolly pine. As the reference assembly and gene annotation for the loblolly pine genome are under active improvement, the variant locations and annotations might be revised in the future.

Application of the identified variants

MAS utilizing the identified alleles and those that will be discovered in the future using the SNPs identified in this project may facilitate loblolly pine breeding. For instance, within this population, tree 276B contained the most desired alleles at 18 additive and 2 epistatic loci, which were associated with the traits of total height, stem diameter, $\Delta^{13}C$, nitrogen concentration, and pitch canker resistance. This corresponded to its superior performance indicated by its above median values of measurement for growth traits. The height-related loci, which were detected multiple times using measurements at different times, and two alleles detected in association with both stem diameter and total height could be regarded with

higher validity than the other loci. However, before these loci can be applied for breeding, they need to be verified using more samples at different ages and with replications in different locations.

Heterozygosity-trait correlations

It is hypothesized that individual heterozygosity may correlate with individual fitness and superior trait performance due to dominance or overdominance (heterosis) (Charlesworth and Willis 2009; Ruiz-Lopez et al. 2012; Rodríguez-Quilón et al. 2015). Correlating individual heterozygosity measured using genetic markers with individual fitness-related or adaptation-associated traits (heterozygosity-trait correlations or HTCs) could test this. Forest trees have been used to test for HTCs. Ledig et al. (1983) reported a significant correlation between mean annual basal area increment and heterozygosity in pitch pine suggesting that growth may be positively associated with isozyme heterozygosity. Using a maritime pine (*P. pinaster* Ait.) population, no significant correlation was found between survival and genome-wide heterozygosity (Rodríguez-Quilón et al. 2015); nonetheless, the authors suggested the heterozygosity of specific genes was of great importance to increase fitness. In another study focusing on a Siberian larch (*Larix sibirica* Ledeb.) population, no relationship was found between individual heterozygosity and radial growth (Babushkina et al. 2016), but the authors pointed that relationships could be rather complex depending on the tree age, and more markers and samples are needed to address it.

In this study, individual heterozygosity was found to be associated with $\Delta^{13}C$ and nitrogen concentration within a clonally tested loblolly pine population. Different correlation results detected using different populations may be due to population size, and as the size gets smaller, significance goes down. To verify the effects of heterozygosity on the traits, a population including more samples should be tested for the correlation. Additionally, it would be interesting to calculate and compare the HTCs using individual heterozygosity based on supposedly neutral markers and loci under selection in the loblolly pine genome separately.

Putative functions of genes that contained SNPs associated with traits

In this study, exome-derived probes were used for sequence capture; hence, the SNPs were identified mostly in gene spaces or very close to them. Since linkage disequilibrium decays rapidly within this population (Lu et al. 2016), the identified SNPs are likely to be within or close to the genes controlling the phenotypic traits. Previous studies have used nearly 4000 SNPs to identify loci associated with height, $\Delta^{13}C$, nitrogen concentration, pitch canker disease resistance, and other important traits (Quesada et al. 2010; Cumbie et al. 2011). To validate the SNPs identified in this study, we mapped those sequences with previously identified SNPs to loblolly pine reference assembly v1.01 (http://dendrome.ucdavis.edu/ftp/Genome_Data/genome/pinerefseq/Pita/v1.01) using the software GMAP (Wu and Watanabe 2005). However, none of them could be mapped to the genes or flanking sequences that contained SNPs associated with traits reported in this study. Nonetheless, genes reported in this study provide valuable clues to understand the genetic architecture of complex traits.

A SNP residing in an RNA editing factor gene implies that RNA-edited proteins possibly incorporate into polypeptide complexes (Brennicke et al. 1999) to influence branch angle. The association of a bark storage protein gene suggests nitrogen resorption is involved in the development of crown width (Zhu and Coleman 2001). Genes encoding an auxin-responsive protein and a copper-transporting ATPase were found to be associated with specific leaf area. Discovery of these genes suggests that auxin-regulation participates in the process of leaf meristem growth and determines the leaf size and shape (Zgurski et al. 2005). Copper is an essential element for leaf growth since copper deficiency defects in plants induce a general reduced growth rate, chlorosis, and curling of leaf margins (Puig 2014).

One interesting gene that contained SNPs associated with total height encodes bifunctional pinoresinol-lariciresinol reductase. This gene is involved in lignan biosynthesis (Renouard et al. 2014). In trees, lignans are synthesized and deposited in significant amounts in the heartwood region, probably preventing heart rot caused by fungi (Suzuki and Umezawa 2007).

Two interesting genes that contained SNPs associated with $\Delta^{13}C$ encode histone-lysine N-methyltransferase and a small heat shock protein (sHSP). $\Delta^{13}C$ reflects the water use efficiency of plants, which is regulated by stomatal responses to changes in VPD. The phytohormone abscisic acid (ABA) was found to be the means by which angiosperm stomata respond to natural changes in VPD (McAdam et al. 2016). As reported by Zheng et al. (2012), expression of a histone methyltransferase is regulated by ABA. Sun et al. (2016) reported sHSPs may function as a protein chaperone to modulate ABA biosynthesis and ABA signaling. It is possible that these two genes are part of the ABA biosynthesis and signaling pathway, hence impacting the stomatal responses and water use in loblolly pines.

One gene that contained SNPs associated with nitrogen concentration encodes a F-box/kelch-repeat protein. Its association with nitrogen concentration suggests the regulation of F-box/kelch-repeat protein may be under the control of nitrogen concentration. The LRR RLK BAM2 gene was also found to be associated with nitrogen concentration. BAM2 has been recognized for its role in the development of vascular strands within leaves and a correlated control of leaf shape,

size, and symmetry. A similar receptor-like protein kinase-like protein (RLK) was reported by Cumbie et al. (2011) as a candidate gene for regulation of nitrogen concentration.

For pitch canker resistance, two SNPs located on resistance-related genes were identified, one is on a leucine-rich repeat gene, and the other is on a dehydration-responsive protein gene. The leucine-rich repeat domain and the nucleotide-binding site domain are the major parts that compose the main class of R-genes (Leister and Katagiri 2000). When exposed to disease, plants produce R proteins to detect the presence of pathogen effectors, resulting in activation of multiple signaling pathways and transcription of specific genes that limit pathogen proliferation and disease symptom expression (Arango-Velez et al. 2014). A dehydration-responsive gene often acts to suppress the drought stress response (Yamaguchi-Shinozaki and Shinozaki 1993). One syndrome of pitch canker disease infection is the wilt of tips of girdled branches due to obstructed water flow. The association of a dehydration-responsive protein RD22-like gene suggests that water deficit resistance has also been induced as part of the defense mechanism against pitch canker disease. As pointed out before, pitch canker resistance is a quantitative trait influenced by many genes with relatively small effects (Quesada et al. 2010). In this study, the identified resistance-related SNPs explained a small proportion of clonal and phenotypic variation, but SNP-SNP interaction analyses explained a higher proportion. A transcription factor gene was identified to interact with a gene encoding a mitochondrial processing peptidase subunit beta. Another interaction was also found to be related to mitochondria. A gene encoding a leucine-rich repeat receptor-like serine/threonine-protein kinase interacts with a gene encoding a mitochondrial-like pentatricopeptide repeat-containing protein. The latter gene is one of the major mediators for mitochondrial posttranscriptional regulation (Manna 2015). Both abovementioned interactions imply regulation of mitochondrial gene expression is involved in pitch canker resistance. However, it remains unclear how the mitochondrial gene products function in the resistance pathway. It does offer new ideas to explore genes related to pitch canker resistance.

Conclusions

Within the tested loblolly pine population, genetic correlations between traits were detected as well as geographical variation. More than 2.8 million SNPs identified and genotyped by exome capture and sequencing were applied in this study. Individual heterozygosity as measured by individual SNP loci was found to potentially correlate with $\Delta^{13}C$ and nitrogen concentration. Thirty-four SNPs and 11 SNP-SNP interactions were associated with important traits including crown structure, tree height, stem diameter, water-use efficiency, nitrogen concentration, and pitch canker resistance. For the significant SNP loci that could be analyzed for the source of genetic variance, dominance and epistatic effects were substantial complements to additive effects. These results provide direction for loblolly pine breeding strategies and improved tree deployment through MAS and genomic selection. Candidate genes with a broad spectrum of functions were identified. The functional analyses of the genes that contained identified SNPs will facilitate elucidation of the genetic architecture of the loblolly pine traits.

Acknowledgements We thank the Allele Discovery for Economic Pine Traits 2 (ADEPT 2) project (National Science Foundation Grant DBI-0501763) for developing the population; Dr. John Davis and Dr. Tania Quesada for providing the pitch canker disease resistance data and comments on the manuscript; Dr. Jill Wegrzyn and the PineRefSeq Project (USDA National Institute of Food and Agriculture, Award #2011-67009-30030) for providing the draft loblolly pine reference sequences, exon annotation, and bioinformatics assistance; Dr. Tomasz Koralewski for the exome capture probe design; Dr. Claudio Casola for the discussion on the analysis methods; and Texas A&M Institute for Genome Sciences and Society (TIGSS) for providing the computational resources and system administration support for the TIGSS HPC Cluster. Special gratitude goes to Jeffrey Puryear for the assistance on sample grinding and Dr. Tom Byram, Dr. Fred Raley, and technicians of the Harrison Experimental Forest at Southern Institute of Forest Genetics (Gay Flurry, Chance Parker, Chuck Burdine) for the help on sample collections. This study was funded by the Pine Integrated Network: Education, Mitigation, and Adaptation Project (PINEMAP), a Coordinated Agricultural Project funded by the USDA National Institute of Food and Agriculture, award no. 2011-68002-30185.

Author's contributions ML performed the sample collection and measurement, data analysis, and wrote the manuscript. KVK and CAL conceived and designed the study, coordinated the research, and participated in the drafting of the manuscript. CDN helped with the sampling, interpretation, and manuscript editing. JBW helped with the measurement of the carbon isotope discrimination and nitrogen concentration and interpretation. NAR performed the sample collection and measurement. All authors read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest.

Data archiving statement The Illumina HiSeq short read sequences that were used to detect the SNPs are deposited in the Sequence Read Archive (SRA) (accession number SRP075363; https://www.ncbi.nlm.nih.gov/sra).

References

Aitken SN, Kavanagh KL, Yoder BJ (1995) Genetic variation in seedling water-use efficiency as estimated by carbon isotope ratios and its relationship to sapling growth in Douglas-fir. For Genet 2:199–206

Amerson HV, Nelson CD, Kubisiak TL, Kuhlman EG, Garcia SA (2015) Identification of nine pathotype-specific genes conferring resistance to fusiform rust in loblolly pine (*Pinus taeda* L.) Forests 6:2739–2761. doi:10.3390/f6082739

Arango-Velez A, González LMG, Meents MJ, El Kayal W, Cooke BJ, Linsky J, Lusebrink I, Cooke JE (2014) Influence of water deficit on the molecular responses of *Pinus contorta*× *Pinus banksiana* mature trees to infection by the mountain pine beetle fungal associate, *Grosmannia clavigera*. Tree Physiol 34:1220–1239. doi:10.1093/treephys/tpt101

Babushkina EA, Vaganov EA, Grachev AM, Oreshkova NV, Belokopytova LV, Kostyakova TV, Krutovsky KV (2016) The effect of individual genetic heterozygosity on general homeostasis, heterosis and resilience in Siberian larch (*Larix sibirica* Ledeb.) using dendrochronology and microsatellite loci genotyping. Dendrochronologia 38:26–37

Baker JB, Langdon OG (1990) *Pinus taeda* L., loblolly pine. In: Burns RM, Honkala BH (eds) Silvics of North America, vol 1, Confiers. Agriculture handbook no. 654. U.S. Department of Agriculture, Forest Service, Washington, D.C., p 497–512

Baltunis B, Martin T, Huber D, Davis J (2008) Inheritance of foliar stable carbon isotope discrimination and third-year height in *Pinus taeda* clones on contrasting sites in Florida and Georgia. Tree Genet Genomes 4:797–807. doi:10.1007/s11295-008-0152-2

Bloomfield KJ, Farquhar GD, Lloyd J (2014) Photosynthesis-nitrogen relationships in tropical forest tree species as affected by soil phosphorus availability: a controlled environment study. Funct Plant Biol 41:820–832. doi:10.1071/Fp13278

Bolte A, Ammer C, Lof M, Madsen P, Nabuurs GJ, Schall P, Spathelf P, Rock J (2009) Adaptive forest management in central Europe: climate change impacts, strategies and integrative concept. Scand J Forest Res 24:473–482. doi:10.1080/02827580903418224

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. doi:10.1093/bioinformatics/btm308

Brennicke A, Marchfelder A, Binder S (1999) RNA editing. FEMS Microbiol Rev 23:297–316

Charlesworth D, Willis JH (2009) The genetics of inbreeding depression. Nat Rev Genet 10:783–796. doi:10.1038/nrg2664

Chhatre VE, Byram TD, Neale DB, Wegrzyn JL, Krutovsky KV (2013) Genetic structure and association mapping of adaptive and selective traits in the east Texas loblolly pine (*Pinus taeda* L.) breeding populations. Tree Genet Genomes 9:1161–1178. doi:10.1007/s11295-013-0624-x

Cregg B, Zhang J (2000) Carbon isotope discrimination as a tool to screen for improved drought tolerance. In: 11th Metropolitan Tree Improvement Alliance (METRIA) Conference, Gresham, Oregon

Cumbie WP, Eckert A, Wegrzyn J, Whetten R, Neale D, Goldfarb B (2011) Association genetics of carbon isotope discrimination, height and foliar nitrogen in a natural population of *Pinus taeda* L. Heredity (Edinb) 107:105–114. doi:10.1038/hdy.2010.168

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, Genomes Project Analysis G (2011) The variant call format and VCFtools. Bioinformatics 27:2156–2158. doi:10.1093/bioinformatics/btr330

Eckert AJ, Bower AD, Wegrzyn JL, Pande B, Jermstad KD, Krutovsky KV, St Clair JB, Neale DB (2009a) Association genetics of coastal Douglas fir (*Pseudotsuga menziesii var. menziesii*, Pinaceae). I. Cold-hardiness related traits. Genetics 182:1289–1302. doi:10.1534/genetics.109.102350

Eckert AJ, Pande B, Ersoz ES, Wright MH, Rashbrook VK, Nicolet CM, Neale DB (2009b) High-throughput genotyping and mapping of single nucleotide polymorphisms in loblolly pine (*Pinus taeda* L.) Tree Genet Genomes 5:225–234. doi:10.1007/s11295-008-0183-8

Eckert AJ, van Heerwaarden J, Wegrzyn JL, Nelson CD, Ross-Ibarra J, Gonzalez-Martinez SC, Neale DB (2010) Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). Genetics 185:969–982. doi:10.1534/genetics.110.115543

Eckert AJ, Wegrzyn JL, Cumbie WP, Goldfarb B, Huber DA, Tolstikov V, Fiehn O, Neale DB (2012) Association genetics of the loblolly pine (*Pinus taeda*, Pinaceae) metabolome. New Phytol 193:890–902. doi:10.1111/j.1469-8137.2011.03976.x

Emhart VI (2005) Physiological genetics of contrasting loblolly and slash pine families and clones. Dissertation, University of Florida

Emhart VI, Martin TA, White TL, Huber DA (2007) Clonal variation in crown structure, absorbed photosynthetically active radiation and growth of loblolly pine and slash pine. Tree Physiol 27:421–430

Farquhar GD, Ehleringer JR, Hubick KT (1989) Carbon isotope discrimination and photosynthesis. Annu Rev Plant Phys 40:503–537. doi:10.1146/annurev.arplant.40.1.503

Flanagan LB, Johnsen KH (1995) Genetic variation in carbon isotope discrimination and its relationship to growth under field conditions in full-sib families of *Picea mariana*. Can J For Res 25:39–47

Flint J, Mackay TF (2009) Genetic architecture of quantitative traits in mice, flies, and humans. Genome Res 19:723–733. doi:10.1101/gr.086660.108

Frederick WJ Jr, Lien SJ, Courchene CE, DeMartini NA, Ragauskas AJ, Iisa K (2008) Production of ethanol from carbohydrates from loblolly pine: a technical and economic assessment. Bioresour Technol 99:5051–5057. doi:10.1016/j.biortech.2007.08.086

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res 31:3784–3788

Gonzalez-Martinez SC, Krutovsky KV, Neale DB (2006) Forest-tree population genomics and adaptive evolution. New Phytol 170:227–238. doi:10.1111/j.1469-8137.2006.01686.x

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. doi:10.1534/genetics.106.061127

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. Mol Ecol Resour 9:1322–1332. doi:10.1111/j.1755-0998.2009.02591.x

IPCC (2014) Climate change 2014: Sunthesis report. Contribution of Working Groups I, II AND III to the Fifth Assessment Report of the intergovernmental Panel on Climate Change. In: Core Writing Team, Pachauri RK, Meyer LA (eds.). IPCC, Geneva

Isik F (2014) Genomic selection in forest tree breeding: the concept and an outlook to the future. New Forest 45:379–401. doi:10.1007/s11056-014-9422-z

Jayaraman K (1999) A statistical manual for forestry research. Forestry research support programme for Asia and the Pacific. http://www.fao.org/docrep/003/x6831e/x6831e13.htm. Accessed March 1999

Kayihan GC, Huber DA, Morse AM, White TL, Davis JM (2005) Genetic dissection of fusiform rust and pitch canker disease traits in loblolly pine. Theor Appl Genet 110:948–958. doi:10.1007/s00122-004-1915-2

Lambeth C, Hubert D (1997) Inheritance of branching and crown traits and their relationship to growth rate in loblolly pine. Paper presented at the 24th biennial southern forest tree improvement conference, Orlando, FL, June 9–12

Ledig FT, Guries RP, Bonefeld BA (1983) The relation of growth to heterozygosity in pitch pine. Evolution 37:1227–1238. doi:10.2307/2408843

Leister RT, Katagiri F (2000) A resistance gene product of the nucleotide binding site—leucine rich repeats class can form a complex with bacterial avirulence proteins in vivo. Plant J 22:345–354

Li MC, Zhu JJ, Zhang M (2013) Foliar carbon isotope discrimination and related traits along light gradients in two different functional-type tree species. Eur J Forest Res 132:815–824. doi:10.1007/s10342-013-0723-0

Lu M, Krutovsky KV, Nelson CD, Koralewski TE, Byram TD, Loopstra CA (2016) Exome genotyping, linkage disequilibrium and population structure in loblolly pine (*Pinus taeda* L.) BMC Genomics. doi:10.1186/s12864-016-3081-8

Manna S (2015) An overview of pentatricopeptide repeat proteins and their applications. Biochimie 113:93–99. doi:10.1016/j.biochi.2015.04.004

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM (2009) Finding the missing heritability of complex diseases. Nature 461:747–753. doi:10.1038/nature08494

McAdam SA, Sussmilch FC, Brodribb TJ (2016) Stomatal responses to vapour pressure deficit are regulated by high speed gene expression in angiosperms. Plant Cell Environ 39:485–491. doi:10.1111/pce.12633

McKeand SE, Jokela EJ, Huber DA, Byram TD, Allen HL, Li BL, Mullin TJ (2006) Performance of improved genotypes of loblolly pine across different soils, climates, and silvicultural inputs. Forest Ecol Manag 227:178–184. doi:10.1016/j.foreco.2006.02.016

Mclean RA, Sanders WL, Stroup WW (1991) A unified approach to mixed linear-models. Am Stat 45:54–64. doi:10.2307/2685241

Millar CI, Stephenson NL, Stephens SL (2007) Climate change and forests of the future: managing in the face of uncertainty. Ecol Appl 17: 2145–2151

Neale DB, Wegrzyn JL, Stevens KA, Zimin AV, Puiu D, Crepeau MW, Cardeno C, Koriabine M, Holtz-Morris AE, Liechty JD, Martinez-Garcia PJ, Vasquez-Gross HA, Lin BY, Zieve JJ, Dougherty WM, Fuentes-Soriano S, Wu LS, Gilbert D, Marcais G, Roberts M, Holt C, Yandell M, Davis JM, Smith KE, Dean JF, Lorenz WW, Whetten RW, Sederoff R, Wheeler N, McGuire PE, Main D, Loopstra CA, Mockaitis K, deJong PJ, Yorke JA, Salzberg SL, Langley CH (2014) Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. Genome Biol 15:R59. doi:10.1186/gb-2014-15-3-r59

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

Nelson CD, Peter GF, McKeand SE, Jokela EJ, Rummer RB, Groom LH, Johnsen KH (2013) Pines. In: Singh BP (ed) Biofuel crops: production, physiology and genetics. Chapter 20. CABI, Wallingford, pp 427–459

Neves LG, Davis JM, Barbazuk WB, Kirst M (2013) Whole-exome targeted sequencing of the uncharacterized pine genome. Plant J 75:146–156. doi:10.1111/tpj.12193

Orians GH, Solbrig OT (1977) A cost-income model of leaves and roots with special reference to arid and semiarid areas. Am Nat 111:677–690

Palle SR, Seeve CM, Eckert AJ, Wegrzyn JL, Neale DB, Loopstra CA (2013) Association of loblolly pine xylem development gene expression with single-nucleotide polymorphisms. Tree Physiol 33: 763–774. doi:10.1093/treephys/tpt054

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Puig S (2014) Function and regulation of the plant COPT family of high-affinity copper transport proteins. Advances in Botany 2014, article ID 476917, 9 pages. doi: 10.1155/2014/476917

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575. doi:10.1086/519795

Quesada T, Gopal V, Cumbie WP, Eckert AJ, Wegrzyn JL, Neale DB, Goldfarb B, Huber DA, Casella G, Davis JM (2010) Association mapping of quantitative disease resistance in a natural population of loblolly pine (*Pinus taeda* L.) Genetics 186:677–686. doi:10.1534/genetics.110.117549

Renouard S, Tribalatc M-A, Lamblin F, Mongelard G, Fliniaux O, Corbin C, Marosevic D, Pilard S, Demailly H, Gutierrez L (2014) RNAi-mediated pinoresinol lariciresinol reductase gene silencing in flax (*Linum usitatissimum* L.) seed coat: consequences on lignans and neolignans accumulation. J Plant Physiol 171:1372–1377. doi:10.1016/j.jplph.2014.06.005

Resende MF Jr, Munoz P, Resende MD, Garrick DJ, Fernando RL, Davis JM, Jokela EJ, Martin TA, Peter GF, Kirst M (2012) Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.) Genetics 190:1503–1510. doi:10.1534/genetics.111.137026

Rodríguez-Quilón I, Santos-del-Blanco L, Grivet D, Jaramillo-Correa JP, Majada J, Vendramin GG, Alía R, González-Martínez SC (2015) Local effects drive heterozygosity—fitness correlations in an outcrossing long-lived tree. Proc Biol Sci 282:20152230. doi:10.1098/rspb.2015.2230

Ruiz-Lopez MJ, Ganan N, Godoy JA, Del Olmo A, Garde J, Espeso G, Vargas A, Martinez F, Roldan ER, Gomendio M (2012) Heterozygosity-fitness correlations and inbreeding depression in two critically endangered mammals. Conserv Biol 26:1121–1129. doi:10.1111/j.1523-1739.2012.01916.x

Schmidtling R (2001) Southern pine seed sources. In: Gen. Tech. Rep. SRS-44. U.S. Department of Agriculture, Forest Service, Southern Research Station, Asheville, NC, p 25

Seeve CM (2010) Gene expression and association analyses of stress responses in loblolly pine (*Pinus taeda* L.). Dissertation, Texas A&M University

Shmulsky R, Jones PD (2011) Forest products and wood science. Wiley, Hoboken

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal omega. Mol Syst Biol 7:539. doi:10.1038/msb.2011.75

Smith WB, Miles PD, Perry CH, Pugh SA (2009) Forest resources of the United States, 2007. In: Gen. Tech. Rep. WO-78. U.S. Department of Agriculture, Forest Service, Wahsington Office, Washington, DC, p 336

Sun X, Sun C, Li Z, Hu Q, Han L, Luo H (2016) AsHSP17, a creeping bentgrass small heat shock protein modulates plant photosynthesis and ABA-dependent and independent signalling to attenuate plant response to abiotic stress. Plant Cell Environ. doi:10.1111/pce.12683

Suren H, Hodgins KA, Yeaman S, Nurkowski KA, Smets P, Rieseberg LH, Aitken SN, Holliday JA (2016) Exome capture from the spruce and pine giga-genomes. Mol Ecol Resour. doi:10.1111/1755-0998.12570

Suzuki S, Umezawa T (2007) Biosynthesis of lignans and norlignans. J Wood Sci 53:273–284. doi:10.1007/s10086-007-0892-x

Thavamanikumar S, Southerton SG, Bossinger G, Thumma BR (2013) Dissection of complex traits in forest trees—opportunities for marker-assisted selection. Tree Genet Genomes 9:627–639. doi:10.1007/s11295-013-0594-z

Wegrzyn JL, Liechty JD, Stevens KA, Wu LS, Loopstra CA, Vasquez-Gross HA, Dougherty WM, Lin BY, Zieve JJ, Martinez-Garcia PJ, Holt C, Yandell M, Zimin AV, Yorke JA, Crepeau MW, Puiu D, Salzberg SL, Dejong PJ, Mockaitis K, Main D, Langley CH, Neale DB (2014) Unique features of the loblolly pine (*Pinus taeda* L.) megagenome revealed through sequence annotation. Genetics 196:891–909. doi:10.1534/genetics.113.159996

Wells OO, Switzer GL, Schmidtling RC (1991) Geographic variation in Mississippi loblolly pine and sweetgum. Silvae Genet 40:105–119

Westbrook JW, Chhatre VE, Wu L-S, Chamala S, Neves LG, Muñoz P, Martínez-García PJ, Neale DB, Kirst M, Mockaitis K, Nelson CD, Peter GF, Davis JM, Echt CS (2015) A consensus genetic map for *Pinus taeda* and *Pinus elliottii* and extent of linkage disequilibrium in two genotype-phenotype discovery populations of *Pinus taeda*. G3 (Bethesda) 5:1685–1694. doi:10.1534/g3.115.019588

White TL, Adams WT, Neale DB (2007) Forest genetics. Cabi, Wallingford. doi:10.1079/9781845932855.0000

Wu TD, Watanabe CK (2005) GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21:1859–1875. doi:10.1093/bioinformatics/bti310

Yamaguchi-Shinozaki K, Shinozaki K (1993) The plant hormone abscisic acid mediates the drought-induced expression but not the seed-specific expression of rd22, a gene responsive to dehydration stress in *Arabidopsis thaliana*. Mol Gen Genet 238:17–25

Zgurski JM, Sharma R, Bolokoski DA, Schultz EA (2005) Asymmetric auxin response precedes asymmetric growth and differentiation of *asymmetric leaf1* and *asymmetric leaf2* Arabidopsis leaves. Plant Cell 17:77–91. doi:10.1105/tpc.104.026898

Zhang J, Singh A, Mueller DS, Singh AK (2015) Genome-wide association and epistasis studies unravel the genetic architecture of sudden death syndrome resistance in soybean. Plant J 84:1124–1136. doi:10.1111/tpj.13069

Zheng J, Chen F, Wang Z, Cao H, Li X, Deng X, Soppe WJ, Li Y, Liu Y (2012) A novel role for histone methyltransferase KYP/SUVH4 in the control of Arabidopsis primary seed dormancy. New Phytol 193:605–616. doi:10.1111/j.1469-8137.2011.03969.x

Zhu BL, Coleman GD (2001) The poplar bark storage protein gene (Bspa) promoter is responsive to photoperiod and nitrogen in transgenic poplar and active in floral tissues, immature seeds and germinating seeds of transgenic tobacco. Plant Mol Biol 46:383–394. doi:10.1023/A:1010600504740