

Environmental Genome-Wide Association Reveals Climate Adaptation Is Shaped by Subtle to Moderate Allele Frequency Shifts in Loblolly Pine

Amanda R. De La Torre ^{1*}, Benjamin Wilhite¹, and David B. Neale²

¹School of Forestry, Northern Arizona University

²Department of Plant Sciences, University of California-Davis

*Corresponding author: E-mail: Amanda.de-la-torre@nau.edu.

Accepted: October 7, 2019

Abstract

Understanding the genomic basis of local adaptation is crucial to determine the potential of long-lived woody species to withstand changes in their natural environment. In the past, efforts to dissect the genomic architecture in gymnosperms species have been limited due to the absence of reference genomes. Recently, the genomes of some commercially important conifers, such as loblolly pine, have become available, allowing whole-genome studies of these species. In this study, we test for associations between 87k SNPs, obtained from whole-genome resequencing of loblolly pine individuals, and 270 environmental variables and combinations of them. We determine the geographic location of significant loci and identify their genomic location using our newly constructed ultradense 26k SNP linkage map. We found that water availability is the main climatic variable shaping local adaptation of the species, and found 821 SNPs showing significant associations with climatic variables or combinations of them based on the consistent results of three different genotype–environment association methods. Our results suggest that adaptation to climate in the species might have occurred by many changes in the frequency of alleles with moderate to small effect sizes, and by the smaller contribution of large effect alleles in genes related to moisture deficit, temperature and precipitation. Genomic regions of low recombination and high population differentiation harbored SNPs associated with groups of environmental variables, suggesting climate adaptation might have evolved as a result of different selection pressures acting on groups of genes associated with an aspect of climate rather than on individual environmental variables.

Key words: GEA, loblolly pine, climate adaptation, ultradense linkage map.

Introduction

Local adaptation may arise by differential selection pressures across heterogeneous environments leading to increased fitness in the local environment compared to the nonlocal environment. Although of great interest by population geneticists, the genomic architecture of local adaptation remains largely unsolved in natural populations of nonmodel species (Anderson et al. 2013). A majority of studies aiming to dissect the genetic architecture of local adaptation have focused at detecting signals of selection in which a new advantageous mutation in a single gene is rapidly driven to fixation, also known as the “hard sweep” model (Smith and Haigh 1974). In contrast, recent genome-wide association studies in humans and forest trees species have suggested a largely polygenic basis of local adaptation (Hancock et al. 2010; Pritchard et al. 2010; Neale and Kremer 2011; Le Corre and

Kremer 2012). If a population that is well adapted to a geographic location moves to a new environment, natural selection will increase the frequency of certain alleles until the typical phenotype in the population matches the phenotype optimum in the new environment (Pritchard et al. 2010). This type of adaptation also called “polygenic adaptation” is characterized by subtle to moderate shifts in allele frequencies and may be frequent in traits that have standing genetic variation for selection to act on, are highly heritable and controlled by many loci of small effect (Pritchard et al. 2010; Pritchard and Di Rienzo 2010; Berg and Coop 2014).

A common approach to find genes contributing to local adaptation has been based on the idea that genes under selection should be more genetically differentiated among populations than a neutral locus, and will therefore have high F_{st} values (Cavalli-Sforza 1966; Whitlock and Lotterhos 2015).

© The Author(s) 2019. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Loci mostly affected by spatially heterogeneous selection will have high F_{st} values, whereas the ones under spatially uniform balancing selection will show lower than neutral F_{st} values (Whitlock and Lotterhos 2015). Due to a wide distribution of F_{st} values in neutral markers, only advantageous alleles with high frequencies can be detected under the F_{st} outlier approach (Pritchard et al. 2010). This problem is exacerbated by the lack of power of individual tests when correcting for multiple comparisons over thousands of loci. As a result, weakly selected loci, characteristic of polygenic adaptation, will unlikely to be detected (Le Corre and Kremer 2012; Whitlock and Lotterhos 2015; Yeaman 2015). In contrast, genotype–environment association (GEA) studies are more likely to detect signals associated with smaller allele frequency shifts (Hancock et al. 2010; Forester et al. 2018).

Lately, increasing numbers of genome-wide markers have enabled the study of the genomic architecture of local adaptation in natural populations. The geographic and genomic distribution of adaptive alleles can give us insights into the evolutionary forces that had shaped adaptation in a species. For example, alleles associated with day length in natural populations of *Arabidopsis thaliana* showed narrow geographic distribution in which one allele had rapidly driven to high frequency in the population as a result of a hard-selective sweep, whereas SNPs associated with relative humidity had widespread distributions (Hancock et al. 2011). Also, the clustering of adaptive alleles in genomic regions of low recombination due to linkage or divergence hitchhiking can give us insights into the maintenance of population differentiation and local adaptation in the face of gene flow (Via 2012; Yeaman 2013).

In the past, efforts to dissect the genomic architecture of local adaptation in gymnosperms species have been limited due to the absence of reference genomes. Recently, the genomes of some commercially important conifers have become available, allowing whole-genome studies of these species (De La Torre, Birol, et al. 2014; Neale et al. 2017). Current assemblies of reference genomes are still quite fragmented and do not allow the location of genes in the genomes unless high-density linkage maps are available. Loblolly pine (*Pinus taeda*) is a widely distributed species in the southeastern United States, characterized by its outcrossing mating system, large population sizes, weak population structure, and rapid decay of linkage disequilibrium (Eckert, van Heerwaarden, et al. 2010). Phylogeographic studies of unglaciated North America suggested loblolly pine follows the Mississippi River discontinuity, which is consistent with a dual Pleistocene refugial model, and has been used to explain differences in growth, disease resistance, drought tolerance, and genetic differentiation between eastern and western populations (Teskey et al. 1987; Schmidting 2001; Soltis et al. 2006; Eckert, Bower, et al. 2010; Eckert, van Heerwaarden, et al. 2010). In this study, we aimed to dissect the genomic architecture of climate adaptation in loblolly pine. We tested for associations between 270 environmental variables and 87k

SNPs obtained from widely distributed coding and noncoding regions across the 22-Gb genome of the species. We then determined the geographic and genomic location of significant alleles using our newly constructed, ultradense, 26k SNP linkage map. In addition to identify the main climatic variables driving the adaptation of the species, we were also interested in the following questions: (1) Are adaptive alleles globally occurring alleles with varying frequencies or localized ones? (2) Do adaptive alleles have narrow or widespread genomic distributions? (3) Does local adaptation occur by large or subtle shifts in allele frequencies?

Materials and Methods

Sample Collection and SNP Genotyping

Needle tissue from 377 outcrossing, unrelated individuals distributed across the species' natural range were collected from the ADEPT2 common garden located in Mississippi, southeast United States (fig. 1A). Ten populations were assigned based on individuals' geographical proximity within geopolitical states, following Eckert, van Heerwaarden, et al. (2010). In addition, 2 three-generation full-cross outbred pedigrees constructed and maintained by the Weyerhaeuser Company were used to collect 192 needle samples (Sewell et al. 1999). From these, 92 full-sib progeny samples came from the *qt*/pedigree, and 100 full-sibs came from the *base* pedigree. DNA was extracted using a protocol that included one day of tissue lysis and incubation at 96°C, followed by several steps of precipitation and filtering using the Qiagen DNeasy mini-prep Plant kit with an Eppendorf automated pipetting workstation. DNA concentration and quality were evaluated using picogreen on a Qubit Fluorometer. Raw reads from whole-genome resequencing data for ten individuals were used to call a large number of SNPs (455 M) that were later scored, filtered and included in an Affymetrix Axiom myDesign species-specific and customized SNP array comprising 635k SNP markers [full description of this procedure can be found in De La Torre et al. (2019)]. After removing SNPs that did not pass the genotyping quality control criteria and those that were monomorphic, we kept 84,738 high-resolution SNPs. In addition, 3,087 gene-based SNPs previously reported by Eckert, van Heerwaarden et al. (2010) were added to the data set, resulting in a total of 87,825 SNPs. From these SNPs, 20,367 matched genes, exons, transcripts, or a combination of those. Minor allele frequency distribution of all SNPs can be found in [supplementary figure S1, Supplementary Material](#) online.

Population Structure, Diversity Estimates, and F_{st} Outlier Test

Population structure in the SNP data set was evaluated using the Python2.x fastStructure algorithm based on a variational framework for posterior inference of K clusters (Raj et al. 2014). Models in fastStructure were replicated 10 times with K from 1

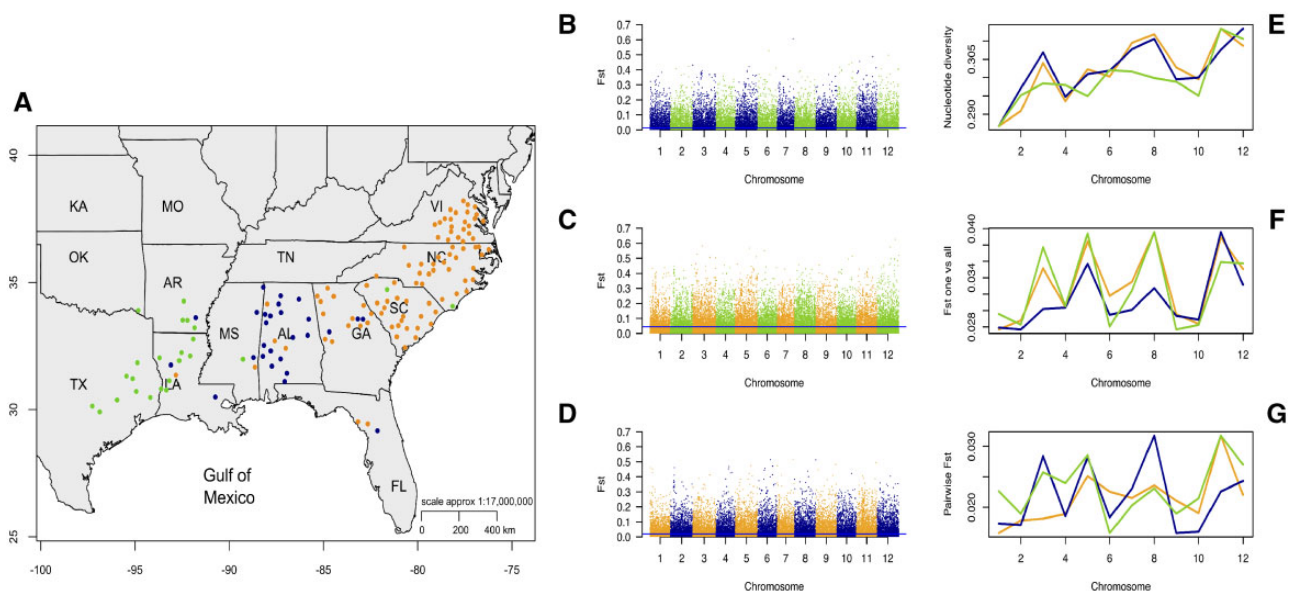


Fig. 1.—Population structure and nucleotide diversity of loblolly pine based on 87k genome-wide SNP markers. Results of PCA and Structure suggest the presence of three different genetic clusters along longitude (A). Manhattan plots show pairwise F_{st} distributions between west and center (B), west and east (C), and center and east genetic clusters (D). Horizontal blue line represents the mean pairwise F_{st} for each pair of comparisons. Average nucleotide diversity, F_{st} one versus all, and pairwise F_{st} among genetic clusters are shown in (E–G). Colors in all figures match genetic clusters in (A).

to 10 using the default prior; seeds for random number generators were modified for each run. The chooseK.py python script in fastStructure was used to estimate the model complexity that maximizes marginal likelihood and the model components were used to explain structure in the data. Main pipeline and Distruct package in CLUMPAK (Kopelman et al. 2015) were used for the summation and graphical representation of FastStructure results. In addition, we did a PCA analysis using the Adegnet v2.0.1 R package (Jombart 2008; Jombart and Ahmed 2011). Outlier SNPs showing higher F_{st} than neutral loci were identified using the OutFLANK R package. OutFLANK infers the null distribution by removing loci in the top and bottom 5% of the distribution, and it is suggested to be robust to demographic history (Whitlock and Lotterhos 2015). Genetic clusters identified by the population structure analyses were then used to estimate nucleotide diversity and pairwise F_{st} . Nucleotide diversity for each SNP in each of the 12 linkage groups was estimated with the R Package PopGenome v.2.2.4 (Pfeifer et al. 2014). Pairwise fixation index (F_{st}) was estimated by comparing all possible pairs of genetic clusters, and by comparing each of them against all other individuals (F_{st} one vs. all). F_{st} values for each SNP in each linkage group were plotted in figure 1B–D; and average F_{st} values were displayed in figure 1E–G.

Genotype–Environment Associations

We used a combination of 248 monthly, seasonal, and annual variables obtained from climate normal data from 1961–1990 in ClimateNA v5.41 (Wang et al. 2016). In addition, we used

19 GIS-derived bioclimatic variables from WorldClim 2.5-min (www.worldclim.org; last accessed September 2019), and aridity index by quarter (every 4 months starting January), as previously calculated by Eckert, van Heerwaarden, et al. (2010). Geographical variables for each individual tree (latitude and longitude), and combinations of environmental data in the form of principal components, were also added to the analysis. All 87k SNPs were tested for associations with 270 environmental variables and their 3 first principal components using three different GEA methods: linear mixed model regressions implemented in GAPIT, latent factor mixed models implemented in LFMM, and a Bayesian approach implemented in Bayenv2.

GEAs were identified for each of the 270 climatic variables with 87,859 SNPs with compressed mixed linear model (Zhang et al. 2010) implemented in the GAPIT R package (Lipka et al. 2012). To reduce the chance of identifying false-positive associations as a result of population structure, we conducted the association analysis with only those SNPs having a minor allele frequency (maf) higher than 3% and used principal components of genetic data as covariates. Manhattan plots were built using the SNP locations in our newly constructed ultradense linkage map for loblolly pine with the R package qqman (Turner 2014). SNP functional annotations were obtained from the annotated genome of loblolly pine v2.01 in TreeGenes (<https://treegenesdb.org>; last accessed September 2019). For the SNPs matching transcripts, we aligned them against the nonredundant protein sequences database using BLASTX 2.8.0 (e value $1e-10$) (Zheng et al. 2000).

In addition, we used a Bayesian bootstrap approach implemented in LFMM command line version 1.5 (Frichot et al. 2013). LFMM accounts for random effects due to population structure and spatial autocorrelation with the use of latent factors (k) (Frichot et al. 2013). Each run was repeated five times with random seeds using the following parameters: $k=3$, 10,000 iterations, and 5,000 burning length. Correction for multiple testing was done by adjusting P -values with the genomic inflation factor ($\lambda = \text{median}(z\text{-score}^2)/0.456$) after combining z -scores obtained from multiple runs with the LEA R package (Frichot and Francois 2015). Runs were repeated with $k=2$ and $k=5$, to check for sensitivity of results to the number of latent factors in the model. To address the potential correlations among environmental variables, we run a PCA analysis using the `prcomp` function in R package. The first three principal components resulting from this analysis were tested for associations with all the 87k SNPs in LFMM. Only SNPs with a minor allele frequency $>3\%$ were included. We also tested for the presence of genomic clusters [more than 10 SNPs within a 1 cM window following Renaut et al. (2013)] for all SNPs significantly associated with any climatic variable or principal component of them.

Finally, we used the software Bayenv2 (Günther and Coop 2013), which implements a Bayesian approach developed by Coop et al. (2010). Bayenv2 first simulates a null model of neutral genetic structure, represented as a covariance matrix of estimated allele frequencies. The null model is then compared to a linear model between allele frequencies and an environmental variable to see if the linear model has an improved fit over the null. The software delivers Bayes factors (BFs) for each locus. We also used the nonparametric extension of Bayenv2, which calculates Spearman's rank correlation coefficient (ρ) and Pearson's correlation coefficient (r_S). Populations were assigned a priori as detailed above in the methods section. Eight runs (100,000 iterations each) were carried out during the matrix estimation step to ensure an accurate covariance matrix was used. SNPs with BF equal or higher than 3 were considered candidates for divergent selection, following Eckert, van Heerwaarden, et al. (2010) and De La Torre, Roberts, et al. (2014).

Environmental Coassociation Network Analysis

In addition to the univariate GEA analyses implemented in LFMM, Gapit, and Bayenv, we also tested for the multivariate response of groups of genes associated with the environment, using the environmental coassociation network analysis as described in Lotterhos et al. (2018). SNPs showing univariate associations in the GEA Gapit analysis were used to construct a network analysis using a hierarchical clustering of the associations between SNP allele frequencies and environmental variables using the `reshape2` and `gplots` R packages in RStudio (RStudio Team 2016).

Construction of Individual Pedigree Linkage Maps

Qtl and *base* pedigrees' pseudotestcrosses were used as *R/qtl* (Broman et al. 2003) objects to construct ultradense linkage maps using the MSTmap algorithm implemented in the ASMap v.0.4 R package with default P -value (Taylor and Butler 2017). Pairwise recombination (r) and Logarithm of the odds (LOD) scores (obtained from linkage disequilibrium test) were estimated for each pair of SNP markers. Markers showing an $r < 0.5$ were considered located in the same linkage group. SNP markers with pairwise recombination frequency estimated as zero (colocating markers) were placed into a recombination bin with ASMap. Several rounds of stringent filtering included the removal of markers with 30% or more missing data, duplicated individuals, double crossovers, markers with distorted segregation patterns, and markers not mapping well in any of the linkage groups. The presence of switched alleles (wrong phase) was also corrected during mapping. JointMap v5.0 (Van Ooijen et al. 2017) was used for fine mapping and ordering of bins obtained from ASMap.

Construction of Averaged-Sex and Consensus Maps

To allow the construction of averaged-sex maps for each pedigree, a set of anchor markers composed by 131 fragment-based markers (RAPD, RFLPs, ESTs, and SSRs) and 2804 SNPs (Eckert, van Heerwaarden, et al. 2010; Martinez-Garcia et al. 2013) were added to our data set. SNPs with suspected linkages as defined by a recombination frequency higher than 0.6 and a LOD score higher than 1 were excluded in further analyses. Forty-eight individual maps were merged to create 24 averaged-sex maps for each pedigree. Averaged-sex maps were merged using maximum intervals (K) from 1 to 8, generating eight consensus maps for each of the 12 linkage groups, with the R package LPmerge (Endelman and Plomion 2014). Consensus maps with the lowest root mean-squared error (RMSE) standard deviation (mapping conflicts between individuals maps and consensus) were selected for each linkage group (Endelman and Plomion 2014). Graphical display was done with Circos (Krzywinski et al. 2009). All markers were anchored to the reference genome of loblolly pine v2.01 (<https://treegenesdb.org>; last accessed September 2019) with the BWA-MEM algorithm in BWA (<http://bio-bwa.sourceforge.net>; last accessed September 2019). Convergence and map accuracy was evaluated by comparing genomic physical location (scaffold ID) and linkage group; and by comparing our maps with previously published maps in the species (Martinez-Garcia et al. 2013; Westbrook et al. 2015).

Results

Population Structure and Genetic Diversity Levels

Results of the PCA analysis with 87k SNPs implemented in the Adegenet and Gapit R packages suggest the presence of

three major genetic clusters (east, center, west) that extend longitudinally across the species' natural range (fig. 1A and [supplementary fig. S2, Supplementary Material](#) online). When using the posterior inference of clusters based on variational Bayesian framework implemented in fastStructure, we found two major clusters when K varies from 2 to 4. When $K=5$, there is a third smaller cluster ([supplementary fig. S3, Supplementary Material](#) online). However, our fastStructure test of the model complexity that maximizes marginal likelihood suggested $K=2$ better explains the genetic structure of the species. When $K=2$, the center and eastern clusters are differentiated from the western clusters, suggesting the Mississippi river as the major barrier for gene flow, as previously observed in De La Torre et al. (2019). We found that eastern and western clusters present higher pairwise F_{st} estimates and were therefore more genetically distant than eastern and center, and center and western clusters (P -value < 0.001 ; fig. 1B–D and [supplementary table S1, Supplementary Material](#) online). For example, from the 622 SNPs with a pairwise $F_{st} > 0.3$, 394 SNPs differentiated the western and eastern genetic clusters. Pairwise F_{st} was found to significantly vary in all linkage groups with the exception of LG 7 and 10 (P -value < 0.001 ; fig. 1F and G, and [supplementary table S1, Supplementary Material](#) online). Nucleotide diversity across all linkage groups was significantly different among genetic clusters (P -value < 0.01), and also differed between the center and west, and the east and west in linkage groups 3, 4, 7, 9, and 12 (P -value < 0.001 ; fig. 1E). Average nucleotide diversity based on SNP data was 0.296, average F_{st} values (one vs. all others) was 0.029, and average pairwise F_{st} was 0.027 ([supplementary table S1, Supplementary Material](#) online).

Fst Outlier Analysis

When including all 87k SNPs in the analysis, we identified 205 SNPs with P -values < 0.001 ; however, only 3 of them passed the threshold after correction for multiple testing (q value < 0.01). Because low heterozygosity SNPs may have confounding effects on the F_{st} distribution, we screened out those SNPs and ended up with a data set of 68k SNPs. Our results identified 330 SNPs with P -values < 0.001 , and confirmed the same top 3 outliers as in the 87k SNP data set after correction for multiple testing ([table 1](#)). When increasing the RightTrimFraction (parameter that removes the loci mostly affected by selection before estimating the shape of the F_{st} distribution through likelihood), the fit of the null distribution model was increased in the 87k data set but not in the 68k data set (default Left and RightTrimFractions were used in this analysis). OutFLANK identified a large number of SNPs with moderate to high F_{st} values; however, the wide distribution of F_{st} values did not allow a clear distinction between putatively neutral and putatively under selection SNPs, as it is expected in an F_{st} outlier distribution ([supplementary fig. S4, Supplementary Material](#) online).

GEA and Environmental Coassociation Network Analysis

After comparing the results of the three univariate GEA analyses (Bayenv, Gapit, and LFMM), we found that 821 SNPs showed significant associations in at least two of the three analyses ([supplementary table S3, Supplementary Material](#) online). Number of shared SNPs across the different GEA methods can be found in [figure 2A](#). From the 821 SNPs, 131 SNPs came from coding regions, and 376 were located in linkage groups. In the Gapit results, CMD_wt (winter Hargreaves Climatic Moisture Deficit, which is the difference between a reference evaporation and precipitation during winter) and CMD02 (Hargreaves Climatic Moisture Deficit in February) showed the highest number of associated markers, with 72 SNPs associated to each of them. Only a small group of SNPs matching transcripts aligned to known proteins at the NCBI nonredundant protein sequences database. In contrast, environmental variables showing the highest number (> 1000) of associated SNPs in the Bayenv results were all related with temperature, evaporation and radiation during the summer (Tmax06, Tmax07, Rad07, Rad08, Rad_sm, Eref07, Eref08, Eref_sm, and EXT). Both CMD and Temperature-related associated SNPs were strongly represented in the combined results among GEA analyses which include a large number of environmental variables ([supplementary tables S2 and S3, Supplementary Material](#) online). The location of SNPs associated with Radiation during August (Rad08) can be found in [figure 2B](#). Associated SNPs were mainly found to be involved in transport, stress response, transcriptional activity, and enzymatic functions. SNPs associated with any of the climatic variables were widely distributed across all 12 linkage groups in the genome of loblolly pine.

Hierarchical clustering of significant associations between SNP allele frequencies and environmental variables suggests the presence of two main modules, one related to aridity (temperature variables, radiation, and degree-days above 18°C) and another mainly associated with humidity [climate moisture deficit during winter (CMDwt), NFFD, Annual Heat-Moisture index (AHM)] ([fig. 2B](#)). In the first module, six SNPs (AX-172791235, AX-173011888, AX-172909447, AX-173250534, AX-173348038, and AX-173361850) showed strong associations with a large number of temperature-related environmental variables in the Gapit analysis. These SNPs were also found to be associated with PC1 in both the Gapit and LFMM analysis. Interestingly, even though we were not able to map any of these SNPs in our linkage map, we know that three of them are located in the same scaffold (super 3645) based on the latest genome assembly of the species. The second module is composed by several submodules or subgroups. In one of them, all SNPs were associated with CMDwt and CMD02 cluster together. Smaller submodules were associated with NFFD during April, Eref during the summer; and annual heat moisture index. Although our results suggest pleiotropic effects of the some of the SNPs,

Table 1

Results of Outlier SNP Analysis Implemented by OutFlank for 87k SNPs Obtained from Whole-Genome Resequencing Individuals in Loblolly Pine

SNP	He	Fst	maf	q values	P values	P values Right Tail
AX-173368010	0.209176214	0.348875729	0.118670886	0.00129688	7.53E-08	3.76E-08
AX-173042514	0.340850969	0.316549401	0.217910448	0.003090746	3.59E-07	1.79E-07
AX-173175402	0.235345516	0.292451521	0.136231884	0.009047309	1.58E-06	7.88E-07

NOTE.—Heterozygosity (He), Fixation index (Fst), mean allele frequency (maf), and P-values are shown.

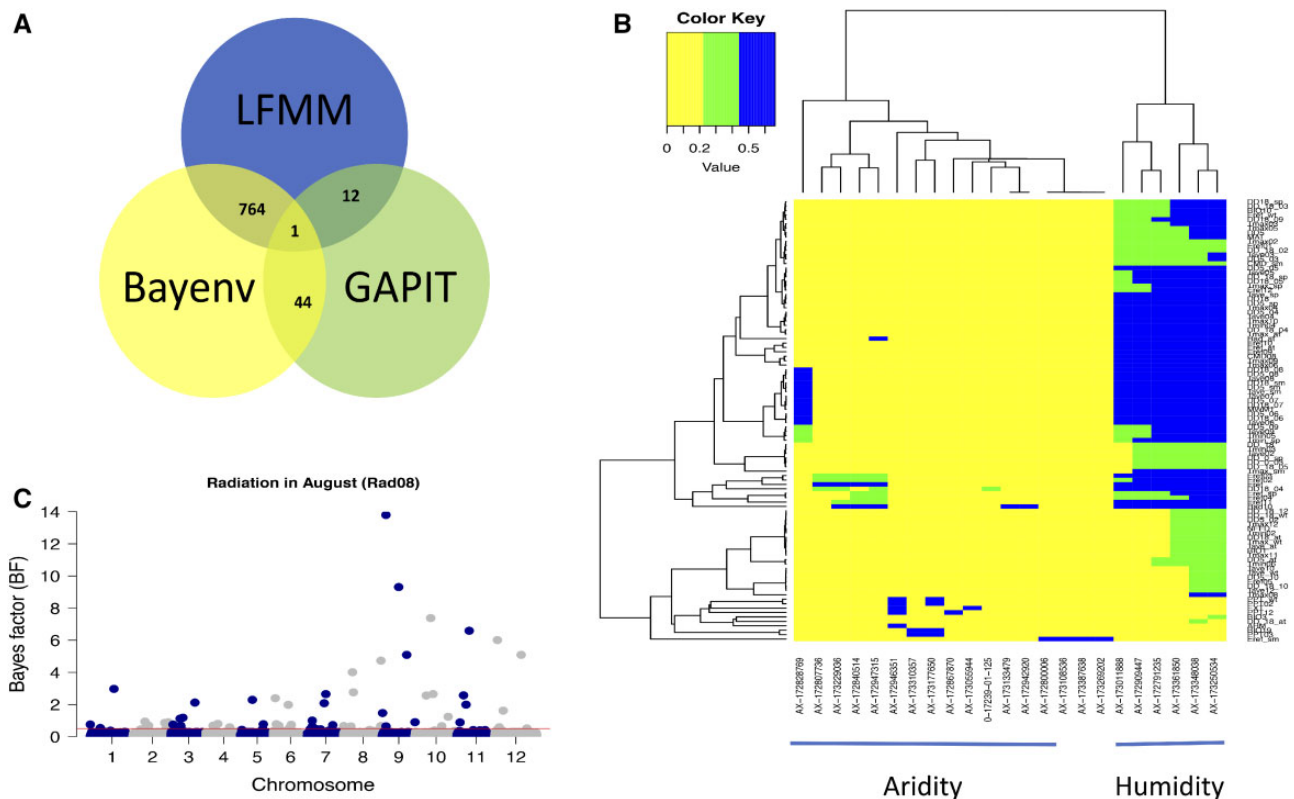


FIG. 2.—Results of the univariate GEA and environmental co-association analyses with 87k SNPs and 270 environmental variables. (A) The number of SNPs that were significant in each of the GEA analyses using three different methods: LFMM, Bayenv, and GAIT. (B) The results of environmental coassociation network analysis. Hierarchical clustering of top-significant ($R^2 > 0.4$) associations between SNP allele frequencies and environmental variables shows two main clusters, right cluster: related to aridity (temperature variables, radiation, and degree-days above 18°C) and a left cluster: mainly associated with humidity. (C) The genomic location of SNPs associated with Radiation in August (Rad08), one of the environmental variables with more associated SNPs based on the results of two or more GEA methods. Y-axis indicates Bayes factor (BF) divided by 10.

our conclusions are limited by the confounding effects of environmental correlations (supplementary fig. S5, Supplementary Material online).

PCA analysis of environmental variables showed PC1 explains 66.56% of the variation in the data set, PC2 explains 14.49%, and PC3 6.51%. PC1 negatively correlates with Temperature-derived variables, Relative Humidity (RH), and Radiation (Rad); and positively with Degree-days below 0°C (DD_0), Degree-days below 18°C (DD_18), Precipitation as snow (PAS), Frost-Free period (FFP) and Temperature seasonality variables such as Isothermality (BIO3); Temperature seasonality (BIO4); Annual Temperature Range (BIO7), and

Continentalty (TD). PC2 positively correlates with RH, and negatively with Temperature seasonality variables (TS) (supplementary figs. S6 and S7, Supplementary Material online). LFMM results showed 716 SNPs associated with PC1, 444 SNPs with PC2, and 741 SNPs with PC3 (Bonferroni-Holmes-Adjusted P -value < 0.01 , $maf > 0.1$). Changing the number of latent factors (k) that account for population structure mainly identified the same group of associated SNPs (data not shown). Bayenv identified 950 SNPs when BF was equal or larger than 3; 498 SNPs when $BF \geq 3$ and $\rho > 0.1$; and 52 SNPs when $BF \geq 3$ and $\rho > 0.2$. Finally, GAIT identified 131 SNPs after Bonferroni-Holmes correction for multiple

testing (supplementary table S4, Supplementary Material online). Convergent results among GEA analyses suggested the presence of 25 common significant SNPs among LFMM, Gapit and Bayenv. SNPs associated with any of the PCs were widely distributed across all 12 linkage groups in the genome (supplementary table S5, Supplementary Material online).

From the individual SNPs showing associations with any climatic variable in two or more GEA methods, we selected a random subset and tested for changes in allele frequency. SNPs showed clinal patterns with increased or decreased subtle to moderate allele frequency shifts (mean = 0.14 ± 0.07) along the longitudinal species' natural range. Allele frequency of the minor allele increased from east to west in SNPs associated with Climatic Moisture Deficit during winter (CMD_wt), and February (CMD02), and with AHM (fig. 3). SNPs associated with Hargreaves reference evaporation during August (Eref08), Degree-days above 18°C, and Number of frost free days during autumn and spring showed an increase in the frequency of the minor allele from west to east. In SNPs associated with Radiation during August (Rad08), SNPs showed either increased or decreased allele frequency of the minor allele along longitude (supplementary table S6, Supplementary Material online).

Linkage Maps

As a result of mapping with ASMap and JoinMap, 17,924 SNPs from the *base* pedigree and 10,995 SNPs from the *qtl* pedigree were mapped in 12 linkage groups in loblolly pine. Pairwise recombination and LOD scores for the *Qtl* and *base* pedigrees' pseudotestcrosses can be found in supplementary figure S8, Supplementary Material online. Anchors allowed the construction of average-sex maps using JoinMap. Total lengths were 2158.662 cM for the *base* linkage map and 2141.44 cM for the *qtl* linkage map. The consensus map had a length of 2270.41 cM across 12 linkage groups, and was built with 26,360 SNPs (table 2, fig. 4, and supplementary table S8, Supplementary Material online). These results represent the most complete and dense map ever built for the species [previous map had 3,856 markers in Westbrook et al. (2015)], and one of the few ultradense linkage maps available to date in gymnosperms (others include Norway spruce, Bernhardsson et al. 2019; and *Ginkgo biloba*, Liu et al. 2017). The mapped SNPs were distributed in 18,163 scaffolds in the genome of the species (loblolly pine v2.01 in TreeGenes, treegenesdb.org). Consensus maps had variable measurements of the lowest RMSE standard deviation (mapping conflicts between individuals maps and consensus), ranging from 0.09 to 20.78. Scaffolds' information of collocated SNPs and assuming all SNPs in the same scaffold were also in the same linkage group, we identified the location of 18,362 more SNPs, resulting in 44,722 SNPs with known positions in linkage groups (17,486 scaffolds) (supplementary

table S8, Supplementary Material online). Average number of SNPs among linkage groups was 3,726, with the number of SNPs in each linkage group ranged from 3,179 to 4,148 markers (supplementary table S9, Supplementary Material online). Convergence and map accuracy was evaluated by comparing genome physical location (scaffold ID) and linkage group and by comparing our maps with previously published maps in the species (Martinez-Garcia et al. 2013; Westbrook et al. 2015). Our results indicate that 4% of SNPs had conflicting positions between the physical and linkage positions, suggesting a very small number of SNPs within the same scaffolds were assigned to different linkage groups. These SNPs locations were not considered in further analysis. Comparison between 715 common SNPs in our map and previous maps revealed a convergence of 95.3% of linkage group assignment with Martinez-Garcia et al. (2013) map and 94.2% with Westbrook et al. (2015). Linkage group numbers were chosen to match those at the Martinez-Garcia et al. (2013) linkage map.

Genomic Clusters

We looked for the presence of genomic clusters at four different levels: (1) SNPs associated with individual climatic variables; (2) SNPs associated with principal components of climate variables; (3) SNPs showing high pairwise F_{st} among populations; and (4) SNPs located in the same environmental coassociation modules. When evaluating SNPs that were significantly associated with groups of environmental variables in two or more GEA analyses, we found two small clusters, one in linkage group 3 (9 SNPs at 48.76 cM) and another in linkage group 9 (12 SNPs at 19.28 cM). Both of these clusters contained SNPs associated with temperature, evaporation, and radiation during the summer (Tmax06, Tmax07, Rad07, Rad08, Rad_sm, Eref07, Eref08, Eref_sm, and EXT) (supplementary table S3, Supplementary Material online). The evaluation of genomic clusters was not possible in SNPs showing significant associations with principal components in two or more GEA analyses because of the small number of associated SNPs (supplementary table S5, Supplementary Material online). In contrast, when evaluating SNPs associated with principal components from the results of only one GEA method, we found larger numbers of collocated SNPs. This was observed in linkage groups 2 (10 SNPs at 190.9 cM), 5 (15 SNPs at 48.76 cM), and 10 (30 SNPs at 19.28 cM). SNPs in LG10 genomic cluster were also included among the top 20% most significant SNPs in the LFMM results data set (supplementary table S4, Supplementary Material online and fig. 4).

Finally, we found that SNPs in the aridity coassociation module were located in the same scaffold of the species, whereas SNPs within the humidity module were in different linkage groups (fig. 2B). SNPs showing high population differentiation were also located in low recombination genomic

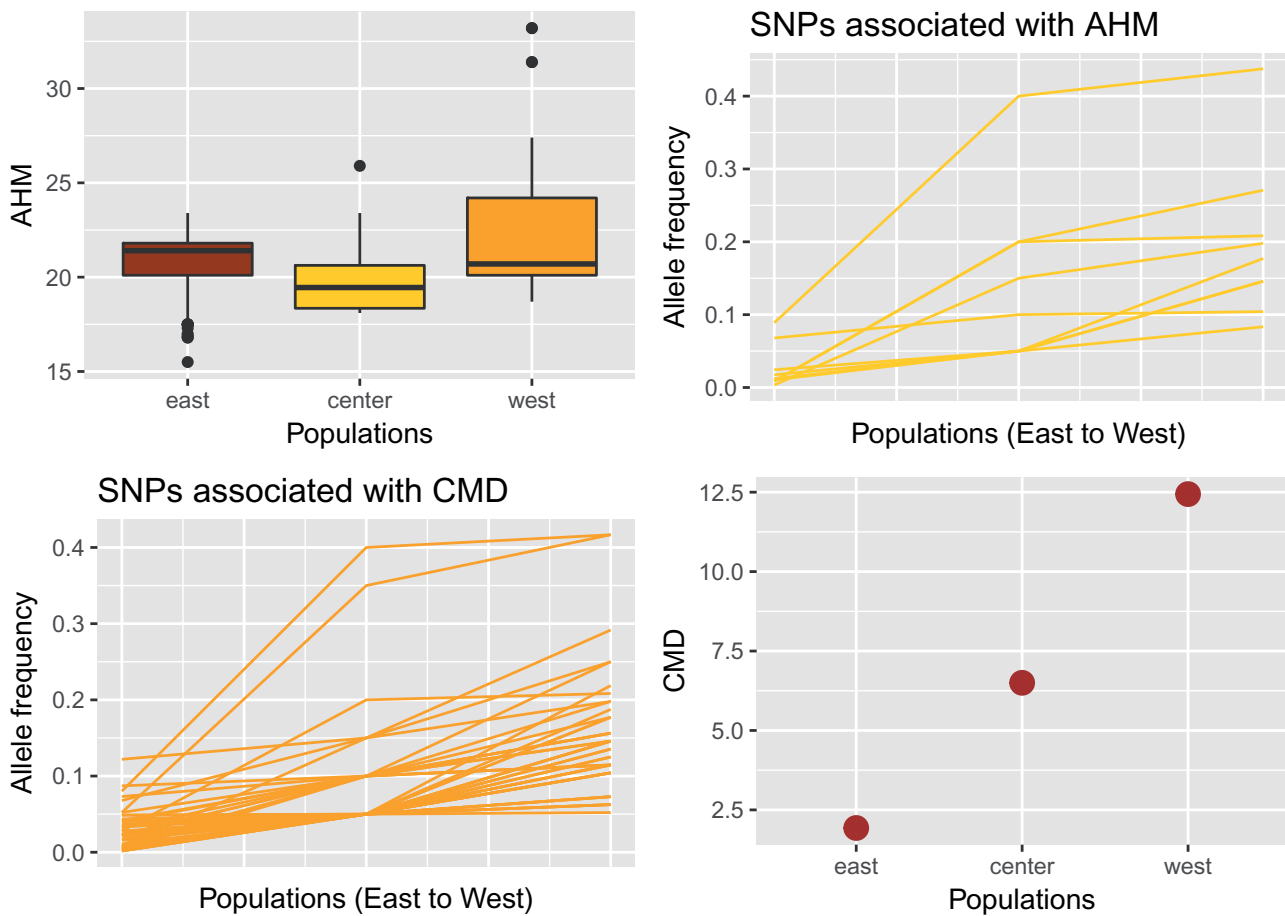


FIG. 3.—Genetic clines along longitude in which allele frequency of the minor allele increases from east to west in SNPs associated with AHM and CMD.

regions (colocated or within close proximity) (supplementary table S7, Supplementary Material online). Interestingly, even though many of these SNPs were not found to be associated with any principal component or individual climatic variables, they were in close proximity (same genomic clusters) with associated SNPs (supplementary table S5, Supplementary Material online).

Discussion

Previous studies suggest that population structure of loblolly pine has been mainly shaped by a dual-Pleistocene refugia that separated populations located east and west of the Mississippi river (Wells and Wakeley 1966; Schmidting 2003). In addition to that initial isolation, and in spite of continuous gene flow, east and west populations continued to differentiate as they became adapted to their new distinct environments. Our results suggest that adaptation to climate in the species might have occurred by many changes in the allele frequency of alleles with moderate to small effect sizes, and by the smaller contribution of large effect alleles in genes related to moisture deficit, temperature and precipitation.

Population Structure

Our results suggest the presence of two major genetic clusters (east, west) and a smaller third cluster (center) that extend longitudinally across the species' natural range (fig. 1A and supplementary Fig. S2, Supplementary Material online). When $K=2$, the center and eastern genetic clusters are differentiated from the western ones, suggesting the Mississippi river as the major barrier for gene flow. It is being suggested that the Mississippi discontinuity is consistent with a dual-Pleistocene refugial model in which populations in southern Florida and southern Texas later migrated north and expanded their distribution to the current natural distribution of the species (Wells and Wakeley 1966; Schmidting 2003). We found that eastern and western genetic clusters present higher pairwise F_{st} estimates and were therefore more genetically distant than eastern and center, and center and western groups (fig. 1B–D). Despite this differentiation, it is clear that populations were and currently are exchanging gene flow, as general genetic differentiation levels are low, suggesting a low population structure. Our results are broadly consistent with previous, smaller scale studies regarding patterns of population structure in loblolly pine (Schmidting 2003;

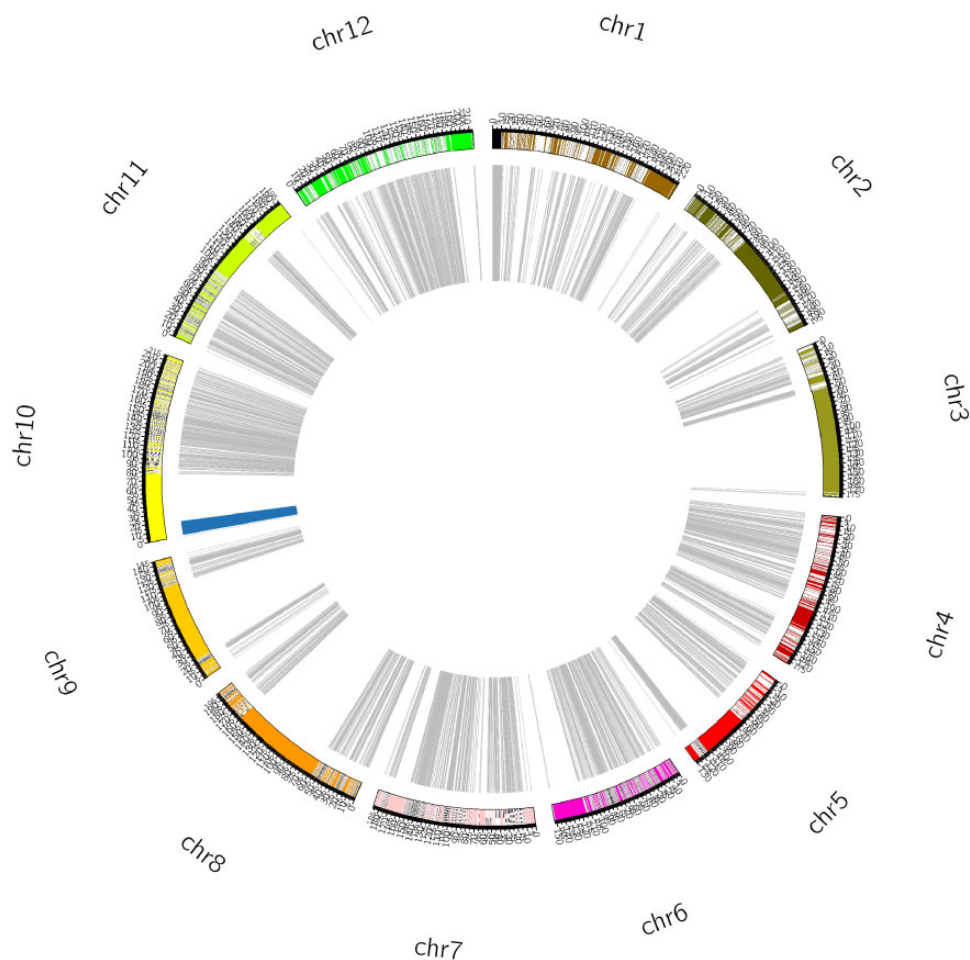


FIG. 4.—Consensus linkage map containing 26,360 SNPs for loblolly pine. Blue region in the concentric inner circle represent the location of the largest cluster of SNPs associated with principal components of climatic variables based on the LFMM results. Size of the region was enlarged for easier visualization.

Gonzalez-Martinez et al. 2006; Eckert, Bower, et al. 2010, Eckert, van Heerwaarden, et al. 2010); and with studies in outcrossing, widespread forest tree species with large population sizes (Savolainen et al. 2007; De La Torre, Roberts, et al. 2014).

Water Availability Is the Highest Determinant for Adaptation of the Species

Water stress and temperature variation impose limitations in the survival, growth, and productivity of many forest tree species. Loblolly pine is not the exception. Adapted to long summers and mild winters, loblolly pine thrives in humid, warm-temperate environments (Baker and Langdon 1990). Soil moisture is a critical factor in seed germination and seedling establishment; whereas temperature has a dominant influence on the initiation of growth in the spring, and posterior ability to compete for light and resources (Baker and Langdon 1990).

Previous physiological and genetic studies have suggested differential responses to temperature and moisture across geographically distant populations of the species

(Teskey et al. 1987; Eckert, Bower, et al. 2010; Eckert, van Heerwaarden, et al. 2010; Lu et al. 2017). Populations east from the Mississippi river grow faster and taller and are less drought-tolerant than population in the west side (Teskey et al. 1987; Schmidting 2001).

In our study, we found SNPs associated with two main aspects of climate. One group, which we called the “aridity” module in the environmental coassociation analysis, is composed by temperature-related variables during the summer (Rad08, Eref08, Tmax07, Tmax08, AIQ3, BIO1, BIO10); and a second group, called the “humidity” module, is composed by moisture deficit and the relationship between precipitation and temperature (CMD, AHM). Results of the combined GEA analyses identified large numbers of temperature-related associated SNPs, a consistent finding with the results of the PCA analysis of environmental variables. Principal component 1, which explained 66.56% of the climatic variation in the data set, was negatively correlated with Temperature-derived variables, Relative Humidity (RH) and Radiation (Rad); and positively with Degree-days below 0°C (DD_0),

Table 2

Results of the Linkage Mapping in Loblolly Pine Showing the Length and Number of SNPs in Each Linkage Group for Individual Maps (*base* and *qtl* pedigrees) and Consensus Map

LG	Base Pedigree		qtl Pedigree		Consensus Map		
	Length	SNPs	Length	SNPs	Length	SNPs	RMSE SD
1	216.386	1520	180.793	844	216.39	2122	20.78
2	205.441	1561	196.181	972	198.23	2261	2.05
3	140.363	1538	180.977	1082	187.57	2365	1.46
4	211.516	1465	186.414	790	222.36	2022	0.09
5	200.858	1652	177.8	955	179.19	2342	1.47
6	196.016	1391	190.184	844	196.02	2046	3.17
7	178.767	1278	174	772	178.77	1880	1.21
8	211.527	1628	188.5	844	214.78	2252	2.4
9	189.513	1582	156.491	844	195.6	2176	0.35
10	136.854	1284	194.574	1065	197.53	2100	15.67
11	119.326	1463	160.416	994	131.67	2200	3.53
12	152.095	1562	155.107	989	152.3	2255	0.73
Total	2158.662	17924	2141.437	10995	2270.41	26021	NA

NOTE.—The ultradense consensus map has 26,021 SNP markers and a length of 2270.41 cM. The RMSE standard deviation gives information about potential mapping conflicts between individuals maps and the consensus map.

Degree-days below 18°C (DD₁₈), PAS, FFP, and Temperature seasonality variables such as Isothermality (BIO3), Temperature seasonality (BIO4), Annual Temperature Range (BIO7), and Continentality (TD). Functional annotation of these SNPs included sugar transport, fatty acid metabolism, auxin response, stress sensing, and signal transduction (supplementary table S3, Supplementary Material online). The GEA results also identified large numbers of SNPs associated with CMD, and AHM. Many of these SNPs showed clinal variation of allele frequencies along the longitudinal range, from the more drought tolerant western populations to the more drought susceptible eastern ones. These SNPs were involved in transport, enzymatic, and transcriptional activity (supplementary Table S2, Supplementary Material online).

These results are consistent with our previous study in the species that estimate the relationship between expression of xylem development genes and environmental variables. In that study, higher expression levels of MADS box protein, a transcription factor putatively acting as a heat shock protein binding were found with increased levels of climatic moisture deficit in May, and increased radiation in Spring. Also, higher expression levels of Xyloglucan endotransglycosylase 2 (XET-2), an enzyme involved in xylem development (and associated with Laccase 3, Laccase 7, and Phenylalanine ammonia lyase-1), were found with increased radiation during the Spring and decreased precipitation and moisture deficit during the summer (De la Torre et al. 2019).

Genomic Distribution of Adaptive Alleles

Theory predicts that due to the decrease in fitness with increasing recombination rates, clusters of alleles contributing

to adaptive trait variation are frequently located in genomic regions with low recombination (Yeaman 2013). In our study, however, we did not find evidence for genomic clustering or “genomic Islands” in SNPs associated with individual climatic variables, which is consistent with the long-standing view that in conifers linkage disequilibrium decreases rapidly due to their high outcrossing rates leading to high recombination (Neale and Kremer 2011). Most SNPs showing significant associations with climatic variables or principal components of them had a wide genomic distribution within and among linkage groups, and were present in all the 12 linkage groups of the species. Widespread genomic distributions of adaptive alleles were also found in *Picea mariana* and *Medicago truncatula* (Prunier et al. 2011; Yoder et al. 2014).

Interestingly, we found larger numbers of colocated SNPs when evaluating SNPs associated with groups of environmental variables, based on the results of the GEA, principal components, and environmental coassociation analyses. This suggests that adaptation to climate in loblolly pine may have occur as a complex process in which different selection pressures are more likely to act on groups of genes associated with an aspect of climate rather than on individual climatic variables. These “co-adapted” complexes of genes may buffer against gene flow coming from maladaptive alleles from geographically proximal but climatically different locations, maintaining polymorphisms across the species’ natural range (Holliday et al. 2016). Increased clustering of outlier loci was found across altitudinal gradients with high gene flow between populations of *Populus trichocarpa*, suggesting adaptation with gene flow might have occurred by divergence hitchhiking of physically proximate alleles in the species (Holliday et al. 2016). Evidence for recurrent hitchhiking was also found in *Capsella grandiflora*, an outcrossing species with large effective population size and low levels of linkage disequilibrium (Williamson et al. 2014).

Our results suggest that SNPs in the aridity coassociation module were located in the same scaffold of the species, whereas SNPs within the humidity module were in different linkage groups. Similarly, a different set of SNPs associated with the same aspect of climate (aridity module composed by temperature-related variables Rad08, Eref08, Tmax08) was also found in close proximity in genomic regions of low recombination, based on the combined GEA results. Physical linkage among loci adapting to different aspects of climate was also found in *Pinus contorta*, while studying modules of co-associated SNPs (Lotterhos et al. 2018). Both Lotterhos et al. (2018) and this study suggest a complex genomic architecture of local adaptation in conifer species, in which the extent of physical linkage among loci is just one of the factors contributing to the species’ evolutionary response to changes in climate.

Finally, we found genomic regions of low recombination in three of the twelve linkage groups when analyzing SNPs associated with PCs (LFMM results) and those with high

population differentiation. Increased linkage disequilibrium between these climate-associated alleles of small effect may have prevented their swamping by gene flow and may have promoted their contribution to adaptive trait divergence (Yeaman and Whitlock 2011; Yeaman 2015). In the case of the SNPs located in the genomic cluster in linkage group 5 (48.04–48.84 cM), we found a correlation between increased population differentiation and decreased diversity ($R^2 = 0.3$, $P < 0.05$). A negative relationship between recombination rate and genetic differentiation is a common signature of linked selection and has been observed in several plant species including *Populus tremula* and *P. tremuloides* (Slotte 2014; Wang et al. 2016). In addition, two of these genomic clusters (groups 2 at 189.62–191.79 cM and 5 at 48.04–48.84 cM) were previously considered as “metabolic hotspots” because they harbor SNPs associated with several metabolites such as pelargonic acid, threonine, and other metabolites of unknown origin (De La Torre et al. 2019).

It is important to mention that although our newly constructed linkage map contains the largest number of SNPs (26k) ever mapped in the genome of the species, many of the SNPs showing significant associations with environmental variables or PCs could not be located within linkage groups. An even higher density linkage map or a chromosome-scale reference genome would be required to confidently locate most or all of the SNPs associated with environmental variables.

Adaptive Alleles—Globally Occurring or Localized?

An important question in population genetics is whether alleles conferring adaptation are globally occurring alleles or localized ones. If natural selection favors specific alleles in specific locations, it is expected that these would be common in these geographic locations but rare in others. On the contrary, if natural selection removes alleles that are deleterious in one location but neutral in others, we would expect to find high-frequency alleles across the species’ range (Fournier-Level et al. 2011). Localized or private alleles with narrow geographic distribution in which one allele has rapidly driven to high frequency in the population may be a result of a hard-selective sweep, as it is being observed in natural populations of *A. thaliana* (Hancock et al. 2011). Alternate alleles might be favored in different environments leading to antagonistic pleiotropy that can result in local adaptation and the maintenance of genetic polymorphisms by natural selection. In a different scenario (conditional neutrality), alleles may be under positive selection in one environment but neutral in others (Anderson et al. 2013).

Our study found that putatively adaptive alleles in loblolly pine were widely distributed across the species’ natural range rather than localized ones. In fact, the presence of private alleles (only present in one population) was not observed in any of the SNPs showing associations with climate variables.

Globally occurring alleles had varying frequencies in which the frequency of the minor allele increased from east to west in SNPs associated with Climatic Moisture Deficit during winter (CMD_wt), and February (CMD02), and with AHM (fig. 3); whereas SNPs associated with Hargreaves reference evaporation during August (Eref08), Degree-days above 18°C, and Number of frost free days during autumn and spring showed an increase in the frequency of the minor allele from west to east. In *A. thaliana* populations, SNPs associated with relative humidity also had widespread distributions (Hancock et al. 2011). With the absence of fitness measurements, we cannot tell if loblolly pine individuals carrying these alleles are fitter in one environment or the other. However, the fact that the direction of the increase of the minor allele is coincident with increasing levels of the associated climatic variable suggests that these individuals may be locally adapted in that environment.

Allele Frequency Shifts at Many Adaptive Loci

Local adaptation in natural populations may arise by differential selection pressures across heterogeneous environments in which the targets of selection may change from one environment to another. As a consequence of this, different combinations of alleles might be favored in different environments and maintained as stable polymorphisms, or experience “partial” or “soft” sweeps due to selection acting on standing variation (Hermisson and Pennings 2005; Yoder et al. 2014). Recent studies have found a largely polygenic basis of adaptation in natural populations, in which trait variation is controlled by many loci of small effect and adaptation is characterized by subtle to moderate shifts in allele frequencies (Pritchard et al. 2010; Pritchard and Di Rienzo 2010; Berg and Coop 2014). With strong diversifying selection and high gene flow, considerable trait divergence may evolve with small allele frequency changes at individual loci (low F_{st}) but high between-population covariance in allele effect sizes (Latta 1998; Le Corre and Kremer 2003, 2012). Adaptation to climate variation via selection on polygenic traits and/or small allele frequency shifts has been observed in *M. truncatula*, *Picea glauca*, *Fagus sylvatica*, and *Maccullochella peellii* (Csillery et al. 2014; Yoder et al. 2014; Hornoy et al. 2015; Harrison et al. 2017). In *P. glauca*, small to moderate shifts in the allele frequency of putatively climate-adaptive genes was found in response to recent selection and high gene flow among populations (Hornoy et al. 2015).

In our study, we found a large number of SNPs with small to moderate effect sizes associated with climatic variables or combinations of them. In most of these SNPs, we found subtle to moderate shifts in allele frequencies across different environments, in which in many cases the increase in the frequency of the minor allele mirrored an increase of the climate variable along longitudinal gradients. These SNPs were largely captured by our GEA studies, which, in contrast

to Fst outlier tests has been suggested to detect signals associated with smaller allele frequency shifts (Hancock et al. 2010). Coincidentally, our Fst outlier test failed to detect the signal of selection in these weakly selected loci involved in local adaptation of loblolly pine, and found only three SNPs with moderate to high Fst. In polygenic traits, most loci involved in local adaptation will experience weak selection, therefore they will not be substantially more differentiated than expected of neutral loci (Le Corre and Kremer 2012; Whitlock and Lotterhos 2015). Loci highly associated with expression and disease resistance and previously suggested to be under balancing selection (De La Torre et al. 2019) were also not identified by the Fst outlier analysis, probably because the OutFLANK procedure is not accurate in the left tail of the Fst distribution (Whitlock and Lotterhos 2015). The little difference in Fst across associated versus randomly chosen SNPs may also suggest that natural selection is not driving large-scale adaptive differences among lineages of loblolly pine. Instead, genotypes are being favored by natural selection across different environments regardless of their ancestry (Eckert, van Heerwaarden, et al. 2010).

Responses to selection that arise from standing genetic variation rather than new mutations or that are relatively recent for fixation to have occurred leave a fainter molecular signature (Hermisson and Pennings 2005; Hohenlohe et al. 2010). These partial, soft sweeps often leave a signature of reduced haplotype or nucleotide diversity and extended linkage, as we found in our LG 5. Considering the low mutation rates in conifers (De La Torre et al. 2017), it is likely that many of these changes in allele frequencies might have been facilitated by the great levels of standing genetic variation rather than by de novo mutations in loblolly pine. Simulations studies have suggested that good levels of standing genetic variation are required when local adaptation occurs by alleles of small effect (Yeaman 2015). In addition, the long generation times in conifers and relatively recent migration from refugia of the species, might have contributed to the genetic architecture we see today, in which most beneficial alleles are still segregating in the populations and have not reached fixation. We therefore conclude that local adaptation to climate in loblolly pine might have occurred by many changes in the allele frequency of alleles with moderate to small effect sizes, and by the smaller contribution of large effect alleles in genes related to moisture deficit, temperature, and precipitation.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

This work was supported by the U.S. Department of Agriculture/National Institute of Food and Agriculture

[McIntire Stennis project 1020440] awarded to A.D.L.T at Northern Arizona University, and by [grant number 2011-67009-30030] awarded to D.B.N at the University of California, Davis. The authors would like to thank Chuck Burdine, Patrick Cumbie, and Dana Nelson for sample collection; and members of the Neale's Lab Annarita Marrano, Sara Montanari, and Pedro Martinez for suggestions on linkage mapping.

Author Contributions

D.N. and A.D.L.T. designed the research study, B.W. did the Bayenv analysis, A.D.L.T. generated the genomic data, did all data analyses and wrote the manuscript, all authors reviewed and commented the final version of the manuscript.

Literature Cited

- Anderson JT, Lee C-R, Rushworth CA, Colautti RI, Mitchell-Olds T. 2013. Genetic tradeoff and conditional neutrality contribute to local adaptation. *Mol Ecol.* 22(3):699–708.
- Baker JB, Langdon OG. 1990. *Pinus taeda* L. Loblolly pine. In: *Silvics of North America*, pp. 497–512.
- Broman KW, Wu H, Sen S, Churchill GA. 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19(7):889–890.
- Berg JJ, Coop G. 2014. A population genetic signal of polygenic adaptation. *PLoS Genet.* 10(8):e1004412.
- Bernhardsson C, et al. 2019. An ultra-dense haploid genetic map for evaluating the highly fragmented genome assembly of Norway spruce (*Picea abies*). G3 (Bethesda). <https://doi.org/10.1534/g3.118.200840>.
- Cavalli-Sforza LL. 1966. Population structure and human evolution. *Proc R Soc B Biol Sci.* 164:362–379.
- Coop G, Witonsky D, Rienzo AD, Pritchard JK. 2010. Using environmental correlations to identify loci underlying local adaptation. *Genetics* 185(4):1411–1423.
- Csillery K, et al. 2014. Detecting short spatial scale local adaptation and epistatic selection in climate-related candidate genes in European beech (*Fagus sylvatica*) populations. *Mol Ecol.* 23:4696–4708.
- De La Torre AR, Birol I, et al. 2014. Insights into conifer giga-genomes. *Plant Physiol.* 166(4):1724–1729.
- De La Torre AR, Roberts DR, Aitken SN. 2014. Genome-wide admixture and ecological niche modelling reveal the maintenance of species boundaries despite long history of interspecific gene flow. *Mol Ecol.* 23(8):2046–2059.
- De La Torre AR, Li Z, Van de Peer Y, Ingvarsson PK. 2017. Contrasting rates of molecular evolution and patterns of selection among gymnosperms and flowering plants. *Mol Biol Evol.* 34(6):1363–1377.
- De La Torre AR, et al. 2019. Genomic architecture of complex traits in loblolly pine. *New Phytol.* 221(4):1789–1801.
- Eckert AJ, Bower AD, et al. 2010. Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). *Mol Ecol.* 19(17):3789–3805.
- Eckert AJ, van Heerwaarden J, et al. 2010. Patterns of population structure and environmental association to aridity across the range of loblolly pine (*Pinus taeda* L, Pinaceae). *Genetics* 185(3):969–982.
- Endelman JB, Plomion C. 2014. LPmerge: an R package for merging genetic maps by linear programming. *Bioinformatics* 30(11):1623–1624.
- Fournier-Level A, et al. 2011. A map of local adaptation in *Arabidopsis thaliana*. *Science* 334(6052):86–89.
- Forester BR, Lasky JR, Wagner HH, Urban DL. 2018. Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Mol Ecol.* 27(9):2215–2233.

- Frichot E, Schoville SD, Bouchard G, Francois O. 2013. Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol Biol Evol.* 30(7):1687–1699.
- Frichot E, Francois O. 2015. LEA: an R package for landscape and ecological association studies. *Methods Ecol Evol.* 6(8):925–929.
- Gonzalez-Martinez SC, Ersoz E, Brown GR, Wheeler NC, Neale DB. 2006. DNA sequence variation and selection of tag SNP at candidate genes for drought-stress response in *Pinus taeda* L. *Genetics*. 172:1915–1926.
- Günther T, Coop G. 2013. Robust identification of local adaptation from allele frequencies. *Genetics* 195(1):205–220.
- Hancock AM, Alkorta-Aranburu G, Witonsky DB, Di Rienzo A. 2010. Adaptations to new environments in humans: the role of subtle allele frequency shifts. *Phil Trans R Soc B.* 365(1552):2459–2468.
- Hancock AM, et al. 2011. Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* 334(6052):83–86.
- Harrison KA, et al. 2017. Signatures of polygenic adaptation associated with climate across the range of a threatened fish species with high connectivity. *Mol Ecol.* 26(22):6253–6269.
- Hermisson J, Pennings PS. 2005. Soft sweeps – molecular population genetics of adaptation from standing genetic variation. *Genetics* 169(4):2335–2352.
- Hohenlohe PA, Phillips PC, Cresko W. 2010. Using population genomics to detect selection in natural populations: key concepts and methodological considerations. *Int J Plant Sci.* 171:1059–1071.
- Holliday JA, Zhou L, Bawa R, Zhang M, Oubida RW. 2016. Evidence for extensive parallelism but divergent genomic architecture of adaptation along altitudinal and latitudinal gradients in *Populus trichocarpa*. *New Phytol.* 209(3):1240–1251.
- Hornoy B, Pavy N, Gérardi S, Beaulieu J, Bousquet J. 2015. Genetic adaptation to climate in white spruce involves small to moderate allele frequency shifts in functionally diverse genes. *Genome Biol Evol.* 7(12):3269–3285.
- Jombart T. 2008. ADEGENET: an R package for the multivariate analysis of genetic markers. *Bioinformatics* 24(11):1403–1405.
- Jombart T, Ahmed I. 2011. ADEGENET 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27(21):3070.
- Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. 2015. CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Resour.* 15(5):1179–1191.
- Krzywinski M, et al. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res.* 19(9):1639.
- Latta RG. 1998. Differentiation of allelic frequencies at quantitative trait loci affecting locally adaptive traits. *Am Nat.* 151(3):283–292.
- Le Corre V, Kremer A. 2003. Genetic variability at neutral markers, quantitative trait loci, and trait in a subdivided population under selection. *Genetics* 164:1205–1219.
- Le Corre V, Kremer A. 2012. The genetic differentiation at quantitative trait loci under local adaptation. *Mol Ecol.* 21:1548–1566.
- Lipka AE, et al. 2012. GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28(18):2397–2399.
- Liu H, Cao F, Yin T, Chen Y. 2017. A highly dense genetic map for *Ginkgo biloba* constructed using sequence-based markers. *Front Plant Sci.* 8:1041.
- Lotterhos KE, Yeaman S, Degner J, Aitken S, Hodgins KA. 2018. Modularity of genes involved in local adaptation to climate despite physical linkage. *Genome Biol.* 19(1):157.
- Lu M, et al. 2017. Association genetics of growth and adaptive traits in loblolly pine (*Pinus taeda* L.) using whole-exome-discovered polymorphisms. *Tree Genet Genom.* 13:57.
- Martinez-Garcia PJ, et al. 2013. Combination of multipoint maximum likelihood (MML) and regression mapping algorithms to construct a high-density genetic linkage map for loblolly pine (*Pinus taeda* L.). *Tree Genet Genom.* 9:1529–1535.
- Neale DB, Martínez-García PJ, De La Torre AR, Montanari S, Wei X-X. 2017. Novel insights into tree biology and genome evolution as revealed through genomics. *Annu Rev Plant Biol.* 68(1):457. 13.1–13.27.
- Neale DB, Kremer A. 2011. Forest tree genomics: growing resources and applications. *Nat Rev Genet.* 12(2):111–121.
- Pfeifer B, Wittelsburger U, Ramos-Onsins SE, Lercher MJ. 2014. PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Mol Biol Evol.* 31(7):1929–1936.
- Pritchard JK, Pickrell JK, Coop G. 2010. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr Biol.* 20(4):R208–R215.
- Pritchard JK, Di Rienzo A. 2010. Adaptation-not by sweeps alone. *Nat Rev Genet.* 11(10):665–667.
- Prunier J, Laroche J, Beaulieu J, Bousquet J. 2011. Scanning the genome for gene SNPs related to climate adaptation and estimating selection at the molecular level in boreal black spruce. *Mol Ecol.* 20(8):1702–1716.
- Renaut S, et al. 2013. Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nat Commun.* 4:1827.
- RStudio Team 2016. RStudio: integrated development for R. Boston (MA): RStudio, Inc. <http://www.rstudio.com/>
- Raj A, Stephens M, Pritchard JK. 2014. fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* 197(2):573–589.
- Savolainen O, Pyhajarvi T, Knurr T. 2007. Gene flow and local adaptation in trees. *Annu Rev Ecol Evol Syst.* 38(1):595–619.
- Sewell MM, Sherman BK, Neale DB. 1999. A consensus map for loblolly pine (*Pinus taeda* L.). I. Construction and integration of individual linkage maps from two outbred three-generation pedigrees. *Genetics* 151(1):321–330.
- Schmidting R. 2001. Southern pine seed sources. In: Gen. Tech. Rep. SRS-44. Asheville (NC): U.S. Department of Agriculture, Forest Service, Southern Research Station, p. 25.
- Schmidting RC. 2003. The southern pines during the Pleistocene. *Acta Hort.* 615:203–209.
- Slotte T. 2014. The impact of linked selection on plant genomic variation. *Brief Funct Genom.* 13(4):268–275.
- Smith JM, Haigh J. 1974. The hitch-hiking of a favourable gene. *Genet Res.* 23(1):23–35.
- Soltis DE, Morris AB, Mclachlan JS, Manos PS, Soltis PS. 2006. Comparative phylogeography of unglaciated eastern North America. *Mol Ecol.* 15(14):4261–4293.
- Taylor J, Butler D. 2017. R package ASMap: efficient genetic linkage map construction and diagnosis. arXiv: 1705.06916. doi: 10.18637/jss.v079.i06.
- Teskey RO, Bongarten BC, Cregg BM, Dougherty PM, Hennessey TC. 1987. Physiology and genetics of tree growth response to moisture and temperature stress: an examination of the characteristics of loblolly pine (*Pinus taeda* L.). *Tree Physiol.* 3(1):41–61.
- Turner SD. 2014. Qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. bioRxiv. doi: 10.1101/005165.
- Van Ooijen JW. 2017. JoinMap5, software for the calculation of genetic linkage maps in experimental populations of diploid species. Wageningen (Netherlands): Kyazma B.V.
- Via S. 2012. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Phil Trans R Soc. B* 367(1587):451–460.
- Wells OO, Wakeley PC. 1966. Geographic variation in survival, growth and fusiform infection of planted loblolly pine. *For Sci Monogr.* 11:1–40.

- Wang T, Hamann A, Spittlehouse D, Carroll C. 2016. Locally downscaled and spatially customizable climate data for historical and future periods for North America. *PLoS One* 11(6):e0156720.
- Wang J, Street NR, Scofield DG, Ingvarsson PK. 2016. Variation in linked selection and recombination drive genomic divergence during allopatric speciation of European and American aspens. *Mol Biol Evol*. 33(7):1754–1767.
- Westbrook JW, et al. 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(8):1685–1694.
- Whitlock MC, Lotterhos KE. 2015. Reliable detection of loci responsible for local adaptation: inference of a null model through trimming the distribution of *F_{st}*. *Am Nat*. 186(S1):S24–S36.
- Williamson RJ, et al. 2014. Evidence for widespread positive and negative selection in coding and conserved noncoding regions of *Capsella*. *PLoS Genet*. 10(9):e1004622.
- Yeaman S, Whitlock MC. 2011. The genetic architecture of adaptation under migration-selection balance. *Evolution* 65(7):1897–1911.
- Yeaman S. 2013. Genomic rearrangements and the evolution of clusters of locally adapted loci. *Proc Natl Acad Sci U S A*. 110(19):e1743–e1751.
- Yeaman S. 2015. Local adaptation by alleles of small effect. *Am Nat*. 186(S1):S74–S89.
- Yoder JB, et al. 2014. Genomic signature of adaptation in *Medicago truncatula*. *Genetics* 196(4):1263–1275.
- Zhang Z, et al. 2010. Mixed linear model approach adapted for genome wide association studies. *Nat Genet*. 42(4):355–360.
- Zheng Z, Schwartz S, Wagner L, Miller W. 2000. A greedy algorithm for aligning DNA sequences. *J Comput Biol*. 7(1-2):203–214.

Associate Editor: Tanja Slotte