American Genetic Association

Journal of Heredity, 2016, 1–10
doi:10.1093/jhered/esw077
Original Article

OXFORD

Original Article

Combining Genotype, Phenotype, and Environment to Infer Potential Candidate Genes

Benoit Talbot, Ting-Wen Chen, Shawna Zimmerman, Stéphane Joost, Andrew J. Eckert, Taylor M. Crow, Devrim Semizer-Cuming, Chitra Seshadri, and Stéphanie Manel

From the Department of Biology, University of Western Ontario, 1151 Richmond Street, London, Ontario, Canada N6A 3K7 (Talbot); J. F. Blumenbach Institute of Zoology and Anthropology, Georg-August-Universität Göttingen, Göttingen, Germany (Chen); Department of Ecosystem Science and Sustainability, Colorado State University, Fort Collins, CO (Zimmerman); Laboratory of Geographic Information Systems (LASIG), School of Architecture, Civil and Environmental Engineering (ENAC), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland (Joost); Department of Biology, Virginia Commonwealth University, Richmond, VA (Eckert); Department of Ecosystem Science and Management, University of Wyoming, Laramie, WY (Crow); Department of Forest Genetics and Forest Tree Breeding, Georg-August-Universität Göttingen, Göttingen, Germany (Semizer-Cuming); Center for Environmental Studies, Virginia Commonwealth University, Richmond, VA (Seshadri); and EPHE, PSL Research University, CNRS, UM, SupAgro, IRD, INRA, UMR 5175 CEFE, Montpellier, France (Manel).

Address correspondence to B. Talbot at the address above, or e-mail: btalbot4@uwo.ca.

Received June 10, 2016; First decision August 5, 2016; Accepted October 30, 2016.

Corresponding Editor: Mark Chapman

Abstract

Population genomic analysis can be an important tool in understanding local adaptation. Identification of potential adaptive loci in such analyses is usually based on the survey of a large genomic dataset in combination with environmental variables. Phenotypic data are less commonly incorporated into such studies, although combining a genome scan analysis with a phenotypic trait analysis can greatly improve the insights obtained from each analysis individually. Here, we aimed to identify loci potentially involved in adaptation to climate in 283 Loblolly pine (*Pinus taeda*) samples from throughout the species' range in the southeastern United States. We analyzed associations between phenotypic, molecular, and environmental variables from datasets of 3082 single nucleotide polymorphism (SNP) loci and 3 categories of phenotypic traits (gene expression, metabolites, and whole-plant traits). We found only 6 SNP loci that displayed potential signals of local adaptation. Five of the 6 identified SNPs are linked to gene expression traits for lignin development, and 1 is linked with whole-plant traits. We subsequently compared the 6 candidate genes with environmental variables and found a high correlation in only 3 of them ($R^2 > 0.2$). Our study highlights the need for a combination of genotypes, phenotypes, and environmental variables, and for an appropriate sampling scheme and study design, to improve confidence in the identification of potential candidate genes.

Subject areas: Genomics and gene mapping; Molecular adaptation and selection
Key words: environmental variation, genome scan, local adaptation, phenotypic traits, *Pinus taeda* (Loblolly pine), potential candidate gene

Within species of large distribution, distinct populations are subjected to various evolutionary processes. Through the influence of gene flow, genetic structure among populations often mirrors geography (Novembre *et al.* 2008), such that individuals that are closer geographically are usually more genetically similar than individuals that are far apart. This leads to the isolation by distance pattern. Also, depending on local environmental conditions, distinct populations can respond to different selective pressures, which leads to adaptation to the local conditions (Kawecki and Ebert 2004). Local adaptation occurs when individuals from a particular population have higher fitness in their home population compared to individuals from other populations (McKay *et al.* 2005).

The detection of local adaptation in the field, which is a major goal in evolutionary biology, remains difficult. An ideal situation is to be able to show that individuals have higher fitness in their home population, and that this higher fitness is genetically driven and heritable. Assembling all these criteria is usually nearly impossible however. Nonetheless, a number of different types of approaches exist for studying potentially adaptive variation in natural populations, each providing different and independent insights into local adaptation.

Quantitative trait loci (QTL) analyses were developed to study the association between phenotypic traits and genotypes, using pedigree information (Kearsey and Pooni 1996; Barrett and Hoekstra 2011; Neale and Kremer 2011). They allow one to deconstruct the proportion of variation in phenotypic traits due to heredity, environmental conditions, and, most importantly in a local adaptation context, to an interaction between heredity and environmental conditions. However, by comparing genotypic frequency at a suite of genetic markers and the variation at a phenotypic trait (i.e., association studies), it is also possible to find markers correlated with a phenotypic trait without any pedigree information. These markers are either inside of genes or promoters that directly affect the trait, or are at linkage disequilibrium with such a gene. Since the early 2000s, association studies have been a popular method to investigate adaptation in tree populations (Savolainen *et al.* 2007; Aitken *et al.* 2008; Rellstab *et al.* 2015; Pluess *et al.* 2016), and more specifically in wild conifer trees (reviewed in Neale and Savolainen 2004). For example, Westbrook *et al.* (2015) found 16 candidate regulators of resin canal number in Loblolly pine, including genes associated with oleoresin flow and xylem growth.

The advancement of high throughput sequencing technology has allowed the tracking of a large number of loci including neutral and potentially adaptive loci in nonmodel species, with few genomic resources (Segelbacher *et al.* 2010; Schoville *et al.* 2012). Therefore, interest in genome scans, the survey of genetic variability across whole genomes or across a large number of loci from distinct environments (Luikart *et al.* 2003; Pritchard *et al.* 2010) is growing (Haasl and Payseur 2016; Jensen *et al.* 2016). In this context, genome scan analyses eliminate the need for prior knowledge of candidate genes or phenotypes (Bonin *et al.* 2007). They offer an efficient strategy for identifying loci that are potential targets of selection (Gonzalez-Martinez *et al.* 2006). The signal of selection leading to local adaptation in a species can be inferred in a few ways from genome scans, one being a locus, or suite of loci, having significantly greater differentiation than expected by chance alone (Schoville *et al.* 2012; Bragg *et al.* 2015; Rellstab *et al.* 2015). Loci with greater relative differentiation are referred to as outlier loci. One of the most popular methods to detect such outlier loci has been to compare F_{ST} values of individual loci to an expected F_{ST} value estimated under neutral demographic models (Beaumont and Nichols 1996). If the individual locus F_{ST} values are significantly higher than values estimated under neutral demographic models, this is an indication that such loci may be located in a gene or physically linked to a gene under selection, and we consider such loci to be potential candidate genes. There are now many methods available to researchers to identify loci with non-neutral divergence from the screening of large spatial datasets of individuals or populations genotyped at a large number of markers (Manel *et al.* 2016). Some recent applications include the identification of domestication and improvement genes in cultivated sunflower (Baute *et al.* 2015) and loci responding to spatially varying selection in white poplar (Stölting *et al.* 2015). It is important to acknowledge that confounding effects (genetic drift, demography, and background selection, etc.) may produce a signal similar to selection, and thus lead to false positives (Schoville *et al.* 2012). It is therefore imperative to account for these confounding effects using analyses specifically designed to do so (Lotterhos and Whitlock 2014), and to add additional resources and validation steps to genome scan analyses (Manel *et al.* 2016). Specifically, adding phenotypic trait data to genomic and environmental data in such analyses can support and greatly improve the inference of potential candidate genes (Stinchcombe and Hoekstra 2008; Haasl and Payseur 2016). The number of studies that have connected potential candidate genes inferred through genome scan analysis in a meaningful way to phenotypic traits is limited (Jensen *et al.* 2016). The objective of this study is to provide an example analysis that fills this gap, and to discuss the advantages and limitations of such approaches combining genomic, phenotypic, and environmental datasets to detect potential candidate genes.

Materials and Methods

We first used a quantitative genomic association analysis that calculates the correlation between single nucleotide polymorphism (SNP) data and phenotypic trait data to identify SNP loci whose variation is strongly associated with variation of phenotypic traits. We then compared the identified loci with those inferred to be potentially under local adaptation using a population genomic method that uses only SNP data to identify loci whose degree of divergence over space is distinct from that of loci evolving under neutral conditions. We subsequently tested whether loci identified by both methods were correlated with environmental variation (see Figure 1 for a conceptual description of our approach). Such significant correlations can indicate a role of the environmental variation as a selective pressure on Loblolly pine throughout its range.

More specifically, using a quantitative genomic association analysis, we first identified SNP loci whose intraspecific divergence disproportionately contributes to phenotypic variation using a redundancy analysis (RDA) (these loci are referred to as RDA-outlier loci). We considered phenotypic traits that have a potential genetic basis and are susceptible to being affected by environmental conditions as well, such as traits related to development and resistance to environmental stressors. Next, we applied an F_{ST}-based population genomic method to identify outlier SNP loci potentially under selection, or linked to a locus under selection (referred as F_{ST}-outlier loci). We then compared the RDA- and F_{ST}-outlier SNP loci and kept only those loci identified in both analyses, referred to as potential candidate genes. We subsequently used linear regression models to assess the association between genotypes at these potential candidate genes (detected at the individual level) and environmental variation (estimated at the level of the county), and we compared the spatial distribution of the candidate gene allele frequencies to the spatial distribution of environmental variables. Ultimately having genotypes, phenotypes, and

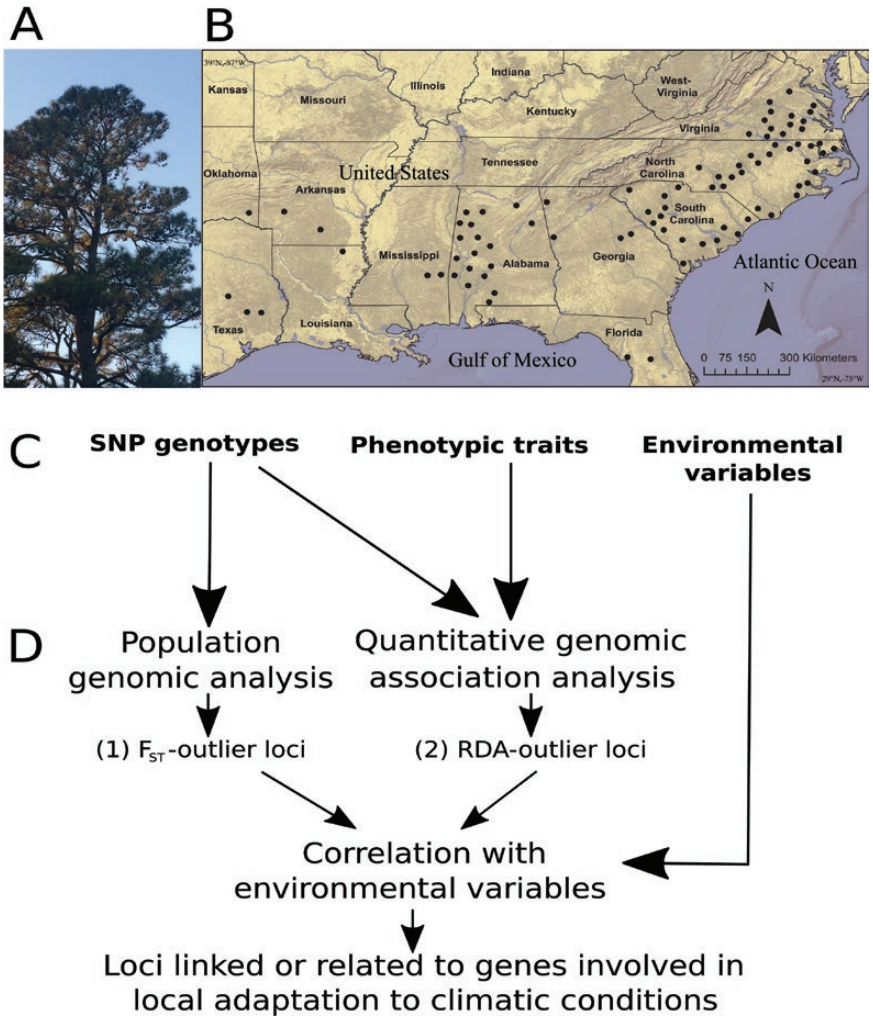

Figure 1. Summary of approaches used in this study to find loci relevant to local adaptation of *Pinus taeda* throughout its range in the southeastern United States: (**A**) Photograph of a *P. taeda* individual (taken by Andrew J. Eckert). (**B**) Map of southeastern United States, showing locations of the 86 counties where *P. taeda* samples were collected (Created with ArcGIS v10.3, ESRI, Redlands, USA). (**C**) Data available from the dataset of 283 *P. taeda* sampled individuals. (**D**) Schematic summarizing the 2 different approaches we used to discover loci linked or related to genes involved in local adaptation.

environmental variables for a single population or species allows for identification of potential candidate genes with more confidence.

Model Species and Previous Knowledge

Loblolly pine (*Pinus taeda* L.) is native to the eastern United States with a range spanning 370 000 km^2 (Schmidtling 2001), from Delaware to Texas. The large geographic range of this species exposes individuals to substantial environmental variation. Previous studies have identified geographic variation in genetic structure (Segelbacher *et al.* 2010), disease resistance (Dorman 1976), secondary metabolites (Squillace *et al.* 1980), growth form (Dorman 1976), and tolerance to aridity (Eckert et al. 2010a, 2010b) in Loblolly pine. In addition, 2 previous range-wide landscape genomic studies have been conducted on this species, using a large SNP genotype data set (Eckert et al. 2010a, 2010b). These studies correlated SNP genotype data at the individual level to environmental variables summarized at the county level. Eckert et al. (2010a, 2010b) also found population genetic structure corresponding to 3 geographic clusters: the Atlantic Coast extending from Florida to Virginia, along the Gulf Coast (Mississippi and Alabama), and west of the Mississippi (Louisiana, Texas, and Arkansas). However, these previous landscape genomic studies did not attempt to correlate variation at SNP markers with variation at phenotypic traits. Here, we used the same SNP dataset in our study, along with phenotypic data obtained from the same individual trees, as a case study to investigate association between SNP, phenotypic, and environmental data.

Sample Collection, SNP Data, and Quantitative Traits Loci

We used published genomic data from a total of 283 Loblolly pines, as described by Eckert et al. (2010a, 2010b). They collected needle tissue samples over the species' natural range in the southeastern United States (Figure 1b), in 86 counties, with 1 to 11 individual trees sampled in each county. Unfortunately, precise geographic coordinates of samples were not available, and for that reason we were unable to perform any analysis of spatial variation in our study. Eckert *et al.* (2010b) isolated total genomic DNA from each needle tissue sample at the USDA National Forest Genetics Laboratory (NFGEL, Placerville, CA) using DNeasy Plant Kits (Qiagen, Valencia, CA). They then selected one SNP from each of 3082 sequence fragments obtained through RADSeq (Davey and Blaxter 2010), to avoid problems related to linkage disequilibrium among SNP markers.

For the individual trees represented in the SNP dataset of Eckert et al. (2010a, 2010b), we collected information from 3 phenotypic trait categories: expression levels at genes related to lignin and cellulose (gene expression), primary and secondary metabolite concentration (metabolites), and drought tolerance and disease resistance (whole-plant traits). Authors of several studies collected the phenotypic trait data. First, Palle *et al.* (2011) mapped the expression of 111 genes involved in the production of lignin, which has important consequences for tree growth and survival (Palle *et al.* 2011) for 246 trees (87% of the trees in the SNP dataset; between 1 and 10 per county). We used these gene expression data, along with the SNP genotypic data for the same individual trees, in our study as a first dataset (referred to as EXP for ease of reference). We expected differences in gene expression in the Loblolly pine in response to environmental differences, as suggested by Palle *et al.* (2011). We also expected some of these differences to be driven by a genetic response (Gibson 2008; Whitehead *et al.* 2011). Second, Eckert *et al.* (2012) measured the concentration of 292 types of metabolites from wood tissue using established or repeatable spectra (e.g., basic sugars and sugar derivatives, lipids, etc.), resulting from a diversity of cellular processes or being putative secondary compounds, for 224 trees (79% of the trees in the SNP dataset; between 1 and 11 per county). We used metabolite data, along with SNP genotypic data for the same individuals, as a second dataset (referred to as MET for ease of reference). We expected the variation in metabolites to have an underlying genetic component and also to be related to the environment (Chan *et al.* 2010). Lastly, Cumbie *et al.* (2011) measured root biomass, length of the pitch canker (a diseased part of the tree caused by the fungus *Fusarium circinatum*; Quesada *et al.* 2010), nitrogen and carbon isotope discrimination, foliar nitrogen concentration, and total height after 2 growing seasons (see Cumbie *et al.* 2011, for complete protocol), for 242 trees (86% of the trees in the SNP dataset; between 1 and 11 per county). These traits give a broad idea of growth in each individual, and thus are considered together as whole-plant traits. We used whole-plant trait data, along with SNP genotypic data for the same individuals, as a third dataset (referred to as WPT for ease of reference). Cumbie *et al.* (2011) demonstrated that nitrogen and carbon isotope discrimination have nonzero heritabilities. We expected growth to vary depending on local temperature and precipitation throughout the range of the Loblolly pine, and to have a genetic component. We thus conducted a separate analysis using each of the 3 different datasets that combined measurement from a particular phenotypic trait category with corresponding SNP data on individual trees (EXP, MET, and WPT). All the analyses were conducted at the individual level.

Quantitative Genomic Association Analysis: RDA-Outlier Detection

We used a principal component analysis (PCA) to reduce the number of genotypic variables, separately for each genotypic dataset (EXP, MET, and WPT). We reduced the count of the dataset's least common allele in each individual (either 0, 1, or 2) for all of the 3082 SNP loci to a number of principal components (PCs) governing the majority of the variation observed in the genetic data. We applied the *prcomp* function in the *stats* package in R (version 3.1.0, R Development Core Team, Vienna, Austria) with the standardized option for genotypic data. We selected the 50 PCs with the highest eigenvalues. Including fewer PCs lessens the representativeness of each variable and including more PCs reduces the amount of degrees of freedom used in further analyses. We refer to PCs generated with genotypic data as genotypic PCs. We compared the explained variation of each genotypic PC with that predicted by the Broken Stick model (i.e., explaining more variation than expected by chance), by using the *PCAsignificance* function in the *BiodiversityR* R-package (Kindt 2016).

We tested the effect of genotypic variables (i.e., genotypic PCs) on each type of phenotypic trait (i.e., gene expression, metabolite, and whole-plant), each of which is itself composed of multiple variables (Methods section). We applied the RDA, a direct extension of multiple regression for multivariate response data (Legendre and Gallagher 2001), with the *rda* function in the *vegan* package; Dixon 2003) in R (version 3.1.0, R Development Core Team), separately to each of the 3 datasets corresponding to each type of phenotypic trait (EXP, MET, and WPT). The RDA accounts for multiple response variables in the analysis: it allows one to determine the effect of the total genotypic variation on each type of phenotypic trait as a whole. We attributed to each individual tree/genotype a value of 1, 2, or 3 corresponding to each of the 3 previously identified genetic clusters in Eckert *et al.* (2010b), and used that variable as a conditioning factor in each RDA. This corrects for the confounding effect of population structure in the 3 RDA analyses. Finally, we used a global permutation test (9999) on response values to determine if some individual genotypic PCs, rather than all of them, significantly explain the response variables ($\alpha = 0.05$) in each of the RDA analyses. The procedure also computes an R^2 between the fitted values of the model, for each genotypic PC, and the randomized response values. This test is implemented in the *vegan* R-package (Dixon 2003) under the *envfit* function. For each genotypic PC found to have a significant effect on the response variables in each of the RDAs, we returned to the initial genotypic PCA to extract SNP loci in the tails of the distribution of genotypic PC loadings. We extracted the loci loaded in the 2.5% and the 97.5% quantiles of the axis. We referred to those loci as RDA-outlier loci.

Population Genomic Analysis: F_{ST}-Outlier Detection

We applied Bayescan v2.1 (Foll and Gaggiotti 2008), which uses a Bayesian logistic regression approach to find loci with the most marginal F_{ST} while correcting for spatial genetic structure. Bayescan assumes a model in which subpopulation allele frequencies are correlated, through a common migrant gene pool from which they differ in varying degrees, thus taking differentiation due to geographic distance into account. We used default parameters for the chain and the model. It is important to acknowledge that even using the default prior odds of 10:1, which is intended to reduce the rate of false positives, can still produce a notable false positive rate (Lotterhos and Whitlock 2014). We conducted the analysis using counties as populations; to account for finer genetic structure than can be found among genetic clusters. We conducted 3 separate analyses in Bayescan, each time using the genetic data that correspond to 1 of our 3 phenotypic datasets. We used Q values, which correct for false discovery rate (Storey 2003) and that were calculated directly in Bayescan, as an alternative to P values. We extracted loci with Q values that are lower than 0.1,and we referred to such loci as F_{ST}-outlier loci.

Overlap between F_{ST}-Outlier Loci and RDA-Outlier Loci

We then compared the list of the F_{ST}-outlier loci with the list of RDA-outlier loci to identify any overlap. We can be fairly confident that any loci that are identified by both methods are in a gene, or are physically linked to a gene, that is under selection and triggering a

fitness related phenotypic response in the Loblolly pine. We named these loci potential candidate genes.

Overlaying Spatial Distribution of Potential Candidate Genes on County-Level Environmental Variation

We used the same environmental variables as in Eckert et al. (2010a, 2010b). As variation within counties has been shown to be small (Eckert *et al.* 2010b), we used averaged values for each of the 86 counties. The dataset of Eckert et al. (2010a, 2010b) includes 58 climatic variables, obtained from the PRISM Group (Oregon State University), yielding monthly minimum and maximum temperature and precipitation, and averaging data from years in the time period between 1971 and 2000. Eckert et al. (2010a, 2010b) constructed 2 additional variables using climate and elevation data. First, they calculated the accumulated growing degree-days above 5 °C on a monthly basis according to Rehfeldt (2006), and secondly, they calculated the aridity index on a monthly basis according to Eckert *et al.* (2010a). An in-depth description of all environmental variables is available in supplementary files (Supplementary Table 1).

We reduced the number of environmental variables with a PCA to a few main principal components governing the majority of the environmental variation observed in the data. We selected all PCs with an eigenvalue larger than 1 in the analysis, so that all variables account for as much variance as a typical variable (see Kachigan 1991, for additional details). We performed all analyses with the *prcomp* function in the *stats* package in R (version 3.1.0, R Development Core Team). We refer to PCs generated with environmental data as environmental PCs.

Finally, we applied a linear univariate regression analysis, using the *lm* function in R (version 3.1.0, R Development Core Team), between the frequency of the least frequent allele among individuals within each county (referred to as minor allele frequency, or MAF), for each locus, and environmental values for each county, for each environmental PC. We derived adjusted R^2 from each linear regression analysis (i.e., for each locus and environmental PC combination). We then compared the adjusted R^2 values obtained for potential candidate genes (detected with previous analyses) that show an adjusted R^2 higher than 0.1, to the distribution of the adjusted R^2 of the other (noncandidate) loci by calculating their percentile rank, to check that the higher adjusted R^2 values for the candidate genes were not obtained by chance. Finally, we reported the distribution of the MAF of each candidate gene, and of the environmental PC that explains it the most (highest adjusted R^2 value), on a geographic map of the study's counties (Figure 3). This procedure allowed for a visual comparison of allelic frequencies of candidate genes and environmental variation.

Results

Quantitative Genomic Association Analysis

All 50 selected genotypic PCs for each genotypic dataset (EXP, MET, and WPT) explained more variation than predicted by chance. Results from the RDA analysis showed a significant effect ($\alpha = 0.05$) of genotypic variation on phenotypic data relative to the gene expression variables (EXP), while no significant effect was observed on the 2 other datasets (MET and WPT) (Table 1; Supplementary Figure 1). Using the same RDA analysis scores, we wanted to know if some genotypic PCs, rather than all of them, had a significant relationship ($\alpha = 0.05$) with phenotypic trait variation, when corrected for genetic structure. The RDA analyses detected a total of 11 genotypic PCs significantly associated with a single trait dataset, or 2 of them, but none with all 3 of the phenotypic trait datasets (Table 2). We found 4 genotypic PCs significantly related to gene expression (EXP dataset), 3 to metabolite (MET dataset), and 6 to whole-plant traits, respectively (WPT dataset) (Table 2). Loci loaded in those genotypic PCs can be associated with candidate genes. Most of the R^2 values were very low (85% were <0.1), demonstrating that the majority of loci loaded in the axis of each genotypic PC only show a weak correlation, indicating most loci loaded in each genotypic PC probably have minimal or no effect on genes related to phenotypic traits. We detected 624 SNP loci in the EXP dataset (or about 20.2% of all analyzed loci), 468 SNP loci in the MET dataset (or about 15.2% of all analyzed loci), and 936 SNP loci in the WPT dataset (or about 30.4% of all analyzed loci) falling within the 2.5% and the 97.5% quantiles of the distribution of genotypic PC axis loadings. We refer to those loci as RDA-outliers.

Population Genomic Analysis

We found only approximately 0.6% of all loci had F_{ST} values that allowed them to be identified significantly as outliers (19 SNP loci total): 13 of these were in the gene expression dataset, 3 were in the metabolites dataset, and 3 were in the whole-plant traits dataset. We refer to those outlier loci as F_{ST}-outliers.

Overlap between F_{ST}-Outlier Loci and RDA-Outlier Loci

We found 5 SNP loci that were identified as both RDA- and F_{ST}-outliers with the EXP dataset (Figure 2a), and 1 locus that was identified as both an RDA- and F_{ST}-outlier with the WPT dataset (Figure 2c). However, we found no SNP locus that was identified as both RDA- and F_{ST} outlier with the MET dataset (Figure 2b). Overall, we found a total of 6 SNP loci that were detected as both types of outliers, referred to as potential candidate genes. They are listed, with their annotations, in Table 3. For reference, we assigned a name to each 1 of those 6 potential candidate genes. Since no annotation has been given to 4 of those candidate genes, we referred to them using the dataset in which they were identified (i.e., *wpt* for the only SNP locus identified from the WPT dataset, and *exp1-5* for the 5 SNP loci identified from the EXP dataset).

Relationship between Potential Candidate Genes and Environmental Variation at the County Level

From the PCA analysis on the environmental variables, we selected PCs with an eigenvalue larger than 1, so that all variables explain

Table 1. Results of the RDA used in the study's quantitative association analysis—Fisher's *F*, degree of freedom (df), *P* value (bold for significant values at $\alpha = 0.05$) and R^2—on the effect of genotypic variation and genetic structure on 3 types of phenotypic traits: gene expression, metabolites, and whole-plant traits, in Loblolly pine (*Pinus taeda*) throughout its range in southeastern United States

	Gene expression	Metabolites	Whole-plant traits
F	1.200	0.973	0.980
df	50	50	50
P	**0.004**	0.781	0.571
R^2	0.040	–0.006	–0.005

Table 2. Regression parameters—*P* value and R^2—for each significant (at $\alpha = 0.05$, ns = non significant) genotypic principal component (PC) variable in the 3 redundancy analyses of the study's quantitative association analysis, on the effect of genotypic variation and genetic structure on 3 types of phenotypic traits: gene expression, metabolites, and whole-plant traits, in Loblolly pine (*Pinus taeda*) throughout its range in southeastern United States

Genotypic PC variable	Gene expression		Metabolites		Whole-plant traits	
	R^2	*P*	R^2	*P*	R^2	*P*
PC5	—	ns	0.039	0.013	0.032	0.043
PC12	—	ns	—	ns	0.071	0.001
PC16	—	ns	—	ns	0.052	0.005
PC17	—	ns	—	ns	0.058	0.003
PC18	—	ns	—	ns	0.033	0.037
PC19	0.036	0.015	—	ns	—	ns
PC24	0.027	0.046	—	ns	—	ns
PC28	0.103	<0.001	0.062	0.001	—	ns
PC30	0.036	0.015	—	ns	—	ns
PC41	—	ns	—	ns	0.057	0.003
PC45	—	ns	0.125	<0.001	—	ns

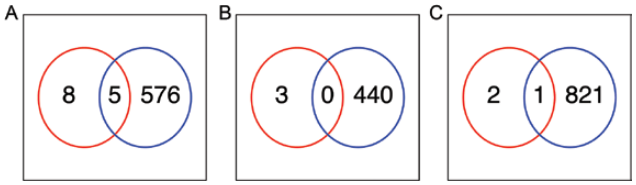

Figure 2. Venn diagrams showing overlap between outlier SNP counts from the genome scan Bayescan approach (left) and the RDA on each of the 3 measured phenotypic traits (right): (**A**) gene expression data, (**B**) metabolite data, and (**C**) whole-plant trait data, in Loblolly pine (*Pinus taeda*) throughout its range in southeastern United States.

more variance than a typical variable (Kachigan 1991). We thus selected 5 environmental PCs that explain 96.2% of the total variance. We named the first PC "Temperature," because it represents most of the variation in mean monthly maximal and minimum temperature, and growing degree-days above 5 °C (GDD5). We named the second PC "Humidity" because it mainly represents mean monthly precipitation and aridity index. The third PC represents some variation in summer (June–September) mean monthly maximal and minimum temperature, aridity index and precipitation, and we thus named it "Summer." The fourth PC mainly represents mean monthly precipitation and aridity index, and we thus named it "Humidity2." The fifth PC represents summer mean monthly minimum temperature, and summer and winter mean GDD5, and we thus named it "Temperature2."

MAF of potential candidate genes *wpt*, *exp1*, *exp2*, *exp3*, and *exp5* had the strongest correlation with the "Humidity" environmental PC (*wpt* $R^2 = 0.061$, $P = 0.014$; *exp1* $R^2 = 0.303$, $P < 0.001$; *exp2* $R^2 = 0.208$, $P < 0.001$; *exp3* $R^2 = 0.234$, $P < 0.001$; and *exp5* $R^2 = 0.033$, $P = 0.054$; Table 3), and MAF of candidate locus *exp4* had the strongest association with the "Humidity2" environmental PC ($R^2 = 0.045$, $P = 0.031$). An overlay between genotypic variation of each candidate locus and environmental variation can be viewed in Figure 3. Three loci (*exp1*, *exp2*, and *exp3*) had an R^2 over 0.1 ($P < 0.001$) with the "Humidity" environmental PC. Also, their adjusted R^2 values were all in the 99th percentile rank, when compared to the adjusted R^2 values of all noncandidate loci when regressed against the "Humidity" environmental PC. Finally, geographic areas with extreme MAF values mostly also have extreme environmental values, for these 3 loci (Figure 3).

Discussion

Global Effect of Genotypic Variation on Phenotypic Traits

Using our global RDA analysis, we found a significant relationship between genotypic variation and gene expression involved in lignin development. However, using the same analysis, we did not find an association between genotypic variation and primary and secondary metabolite concentration (metabolite data), nor between genotypic variation and root biomass, nitrogen and carbon isotope discrimination, foliar nitrogen concentration, total height after 2 growing seasons, and resistance to *F. circinatum* (whole-plant trait data).

Those results support previous studies showing that gene expression involved in lignin development is strongly determined by genotype (Gibson 2008; Whitehead *et al.* 2011). Another study also detected an effect of the environment on gene expression traits (Palle *et al.* 2011) suggesting that they can potentially be involved in local adaptation.

The lack of association with metabolite concentration is surprising, as many significant associations between metabolite concentration and genotypic variation have been observed (Eckert *et al.* 2012), and thousands of QTLs affecting one or more metabolites were found in *Arabidopsis thaliana* (Chan *et al.* 2010). Likewise, many studies have found loci associated to whole-plant trait phenotypes (Quesada *et al.* 2010; Cumbie *et al.* 2011; Eckert *et al.* 2013). For example, 15 QTLs were associated with resistance to *Pucinia hordei* in Australian barley (Ziems *et al.* 2014), and 69 QTLs were associated with oil-filling rate in soybean seed (Jiang *et al.* 2010). It seems that either the method we employed is not sensitive to the true underlying architecture of whole-plant and metabolite traits, or that the relationship between whole-plant and metabolite traits and genotype is not a linear one.

Table 3. Annotation and linear regression parameters—R^2 and P—for each environmental principal component (PC) as a separate explanatory variable, for each potential candidate gene, in Loblolly pine (*Pinus taeda*) throughout its range in southeastern United States

Potential candidate gene	0-4834-01 (*wpt*)		CL2327 Contig1-02 (*exp1*)		0-12076-01 (*exp2*)		CL528 Contig1-04 (*exp3*)		UMN5299-01 (*exp4*)		0-16138-01 (*exp5*)	
Annotation	Unknown		Transcription factor regulating root and shoot growth		Hypothetical protein		Unknown		Importin alpha re-exporter		Unknown	
Regression parameters	R^2	P	R^2	P	R^2	P	R^2	P	R^2	P	R^2	P
Environmental PC1 ("Temperature")	0.021	0.100	0.049	**0.025**	0.111	**0.001**	0.033	0.055	0.002	0.280	<0.001	0.458
Environmental PC2 ("Humidity")	0.061	**0.014**	0.303	**<0.001**	0.208	**<0.001**	0.234	**<0.001**	0.040	**0.039**	0.033	0.054
Environmental PC3 ("Summer")	<0.001	0.482	<0.001	0.335	<0.001	0.964	<0.001	0.559	<0.001	0.489	<0.001	0.858
Environmental PC4 ("Humidity2")	0.013	0.150	0.057	**0.017**	0.018	0.116	0.076	**0.007**	0.045	**0.031**	0.006	0.228
Environmental PC5 ("Temperature2")	0.024	0.086	0.053	**0.020**	<0.001	0.870	<0.001	0.305	<0.001	0.345	0.021	0.099

Bold values represent "$P \leq 0.05$."

Two potential reasons for a lack of relationship between genotypic variation and metabolite and whole-plant traits in our study are 1) unaccounted genes, and 2) coarse resolution in geographical sampling. The 3082 sequence fragments (on average of 500 bp each) used to generate this study's SNP loci represent roughly 0.00007% of its entire genome (22 Gb; Zimin *et al.* 2014). Maximizing genetic resolution in a dataset is important to capture most or all recombination events that happened in the evolutionary history of a species (Myles *et al.* 2009). In addition, an inappropriate sampling scheme consisting of few individuals across a very large spatial scale (here we had sampled 283 individuals from 86 sites across a roughly 1 000 000 km^2 study area), even through the use of a large number of SNP markers, can reduce the power of association analyses. Studies on large spatial scales are more likely to detect variation and the relevant underlying genes (Brachi *et al.* 2011). But, because of a trade-off between spatial scale versus local sample size and spatial resolution of sampling, the statistical power to detect if any of those genes are under selection is often reduced.

RDA- and F_{ST}-Outlier Loci

SNP loci related to the same genotypic PCA axis reflect loci associated with the same linear combination of allelic frequency variables. We found only 4, 3, and 6 genotypic PCA axes that were significantly associated with gene expression, metabolites, and whole-plant trait variation, respectively. Loci strongly associated to the axes of genotypic principal components significantly explaining phenotypic variables (RDA-outlier loci) are likely linked or related to genes causing variation in a phenotypic trait observed in our dataset. RDA association analysis was done over genotypic PCs, rather than individual loci. A specific genotypic PC can have a significant relationship with a phenotypic trait, but there is the potential that just a small fraction of the loci in the PC are actually influencing the relationship. For that reason, we used an independent approach to confirm identified RDA-outlier loci. We subsequently ran a Bayesian genome scan approach and found a total of 13, 3 and 3 F_{ST}-outlier loci, respective to datasets on gene expression, metabolites, and whole-plant traits. Bayescan is known to have moderate power and a moderate false discovery rate, in most simulation scenarios, as compared to other genome scan approaches (de Villemereuil *et al.* 2014). However, it is important to note that not all causative loci have an unusually large F_{ST} (Le Corre and Kremer 2003), and genome scan methods based on F_{ST} variation can miss a large proportion of loci that are experiencing local adaptation, but whose F_{ST} variation is undistinguishable from neutral loci. Additionally, we executed the RDA and Bayescan analyses considering different spatial scales. Whereas we conditioned RDA analyses with previously identified genetic clusters, we executed Bayescan analyses with reference to the counties. This procedure allows us to investigate 2 different levels of differentiation that can occur within the study area, and can further prevent identification of outlier loci due to processes other than selection. However, our ability to detect potential candidate genes may have been limited to those markers that are very variable and are of very large effect, due to the scale of the study. The 6 potential candidate genes identified by both types of analyses may be linked or related to genes that contribute to local adaption across the range of Loblolly pine. However, it is important to understand that the only way to obtain a solid cause-and-effect relationship would be to execute a controlled experiment (such as in McDermott *et al.* 2009).

Effect of the Environment on Outlier Loci

We attempted to correlate SNPs (data obtained at the individual level) with environmental variables (data obtained at the county level) after detection of outlier loci, rather than prior to detection of such loci, to address problems of lack of concordance in resolution between the genotypic and environmental datasets (Joost *et al.* 2008), which affects the precision of the modeling, possibly leading to false discovery. Such correlations, indicating a variation of allele frequency with different environmental conditions suggest potential local adaptation. We found a strong relationship between 3 candidate genes and an environmental variable, "Humidity." Geographic variation of genotypes at these 3 loci qualitatively also corresponds well with geographic variation of the "Humidity" environmental variable. The *exp1* locus is linked or related to a gene with a known function, which in corn (*Zea mays*) codes for a transcription factor regulating root and shoot growth. A study on growth in Loblolly pine seedlings in 4 locations differing in growing season temperature (from 17.6

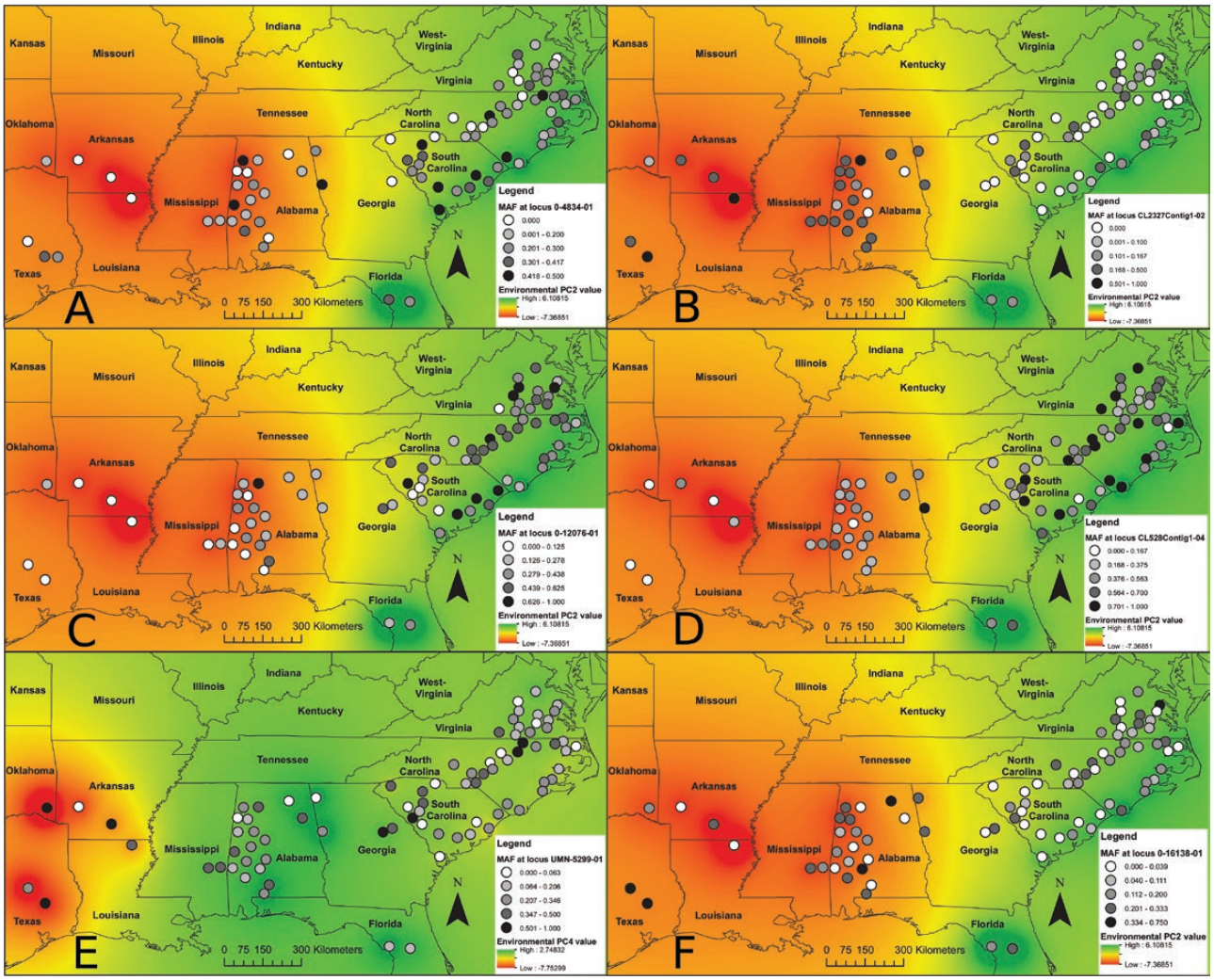

Figure 3. Geographic map of the study's counties, represented with several shades of grey corresponding to 5 minor allele frequency (MAF) value classes, for each of the study's 6 potential candidate genes: (**A**) 0-4834-01, (**B**) CL2327Contig1-02, (**C**) 0-12076-01, (**D**) CL528Contig1-04, (**E**) UMN5299-01, and (**F**) 0-16138-01, in Loblolly pine (*Pinus taeda*) throughout its range in southeastern United States. The values of the environmental PC having the strongest effect on the MAF of each locus are represented as a background gradation calculated with inverse distance weighting interpolation, using QGIS v2.12.0 (QGIS Development Team, Lyon, France). The map was created with ArcGIS v10.3 (ESRI, Redlands, USA).

to 26.3 °C), daily vapor pressure deficits (from 0.39 to 0.47 kPa), and photosynthetic photon flux density (from 38.12 to 44.96 mole/day) found significant differences in final mean component biomass of shoots, roots and leaves (Nedlo *et al.* 2009). Another study on Loblolly pine by Torreano and Morris (1998) found that availability of soil water was a strong determinant of root growth. In different treatments characterized by different soil water content, the treatment with the least amount of soil water resulted in a 41% root growth reduction, compared to the treatment with water content at near field capacity. These results suggest that allocation to root growth is likely impacted by aridity. Also, about 2445 genes were shown to up-regulate in response to severe drought stress in roots (Lorenz *et al.* 2011). Therefore, we argue that Loblolly pine from areas subjected to different climatic conditions harbour alleles coding for transcription factors that are fine-tuned for local climatic conditions. We also found a weak relationship between *exp4* and an environmental PC, "Humidity2." The locus *exp4* is related or linked to a gene coding for importin alpha re-exporter, which is a protein putatively involved in transport of proteins from the nucleus to the cytoplasm (Kutay *et al.* 1997). However, a link between the gene's function and local adaptation in Loblolly pine is unclear at the moment. Finally, a common limitation in studies on local adaptation is the scale of measurement in climatic variables (Schoville *et al.* 2012). Reducing the amount of time and space between measurements would give us a better approximation of how climate is affecting local processes. High resolution remotely sensed data is currently uncommon, though in the future is likely to be more widely available. Using an appropriate study design consisting of phenotypic, genotypic, and environmental data measured at the individual level promises much stronger inference of targets of selection in wild tree populations.

The Use of Phenotype in Studies on Local Adaptation

The first step of an investigation of genetic markers that may signal local adaptation is usually to determine if some loci in a dataset are correlated with local habitat or climatic conditions (Eckert *et al.* 2010a; Paris *et al.* 2010; Poncet *et al.* 2010; Nunes *et al.* 2011). While population genomics can give us valuable clues for which

genes are of ecological relevance, a more specific approach, using phenotype, can indicate which genes may be associated with fitness-related phenotypic traits, such as in using QTLs (Shaw *et al.* 2007; Lacaze *et al.* 2009; Du et al. 2016; McCouch *et al.* 2016). Few studies to date have considered both approaches together using a single set of samples (but see Hancock *et al.* 2011), probably because phenotypic data are not always easy to quantify. Our study is one of the few that considers a population genomics association approach along with a genotype–phenotype association analysis and environmental covariates in a tree species. The next step in this conceptual progression will inevitably consist in measuring fitness associated with different values of phenotypic traits from several local conditions (Salazar-Ciudad and Marín-Riera 2013), and comparing fitness of all possible genotypes (Hietpas *et al.* 2011).

Supplementary Material

Supplementary material can be found at http://www.jhered.oxfordjournals.org/.

Funding

This work was supported by the Natural Sciences and Engineering Research Council of Canada (BESC D3—442735—2013 to B.T.), the University of Western Ontario, and the DGS 2014 Landscape Genetics online graduate course because it was executed during BT's doctoral project.

Acknowledgments

The authors would like to thank Dr Lisette Waits and Dr Helene Wagner for setting up the DGS 2014 Landscape Genetics online graduate course that permitted the authors to meet and work together. The authors would also like to thank Dr Lisette Waits for organizing a Synthesis Meeting in May 2014 in Coeur d'Alene in Idaho, USA, which permitted the authors to put the project together for eventual publication. Finally, the authors would like to thank W. Elizabeth Jones for her help in the analyses, and Nusha Keyghobadi for her help in editing the final version of the manuscript.

References

Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S. 2008. Adaptation, migration or extirpation: climate change outcomes for tree populations: climate change outcomes for tree populations. *Evol Appl*. 1:95–111.

Barrett RDH, Hoekstra HE. 2011. Molecular spandrels: tests of adaptation at the genetic level. *Nat Rev Genet*. 12:767–780.

Baute GJ, Kane NC, Grassa CJ, Lai Z, Rieseberg LH. 2015. Genome scans reveal candidate domestication and improvement genes in cultivated sunflower, as well as post-domestication introgression with wild relatives. *New Phytol*. 206:830–838.

Beaumont MA, Nichols RA. 1996. Evaluating loci for use in the genetic analysis of population structure. *Proc R Soc B Biol Sci*. 263:1619–1626.

Bonin A, Ehrich D, Manel S. 2007. Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. *Mol Ecol*. 16:3737–3758.

Brachi B, Morris GP, Borevitz JO. 2011. Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biol*. 12:232.

Bragg JG, Supple MA, Andrew RL, Borevitz JO. 2015. Genomic variation across landscapes: insights and applications. *New Phytol*. 207:953–967.

Chan EKF, Rowe HC, Hansen BG, Kliebenstein DJ. 2010. The complex genetic architecture of the metabolome. *PLoS Genet*. 6:e1001198.

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.

Davey JW, Blaxter ML. 2010. RADSeq: next-generation population genetics. *Brief Funct Genomics*. 9:416–423.

de Villemereuil P, Frichot É, Bazin É, François O, Gaggiotti OE. 2014. Genome scan methods against more complex models: when and how much should we trust them? *Mol Ecol*. 23:2006–2019.

Dixon P. 2003. VEGAN, a package of R functions for community ecology. *J Veg Sci*. 14:927–930.

Dorman K. 1976. *The genetics and breeding of southern pines*. Washington (DC): U. S. Dept. of Agriculture, Forest Service.

Du Q, Gong C, Wang Q, Zhou D, Yang H, Pan W, Li B, Zhang D. 2016. Genetic architecture of growth traits in Populus revealed by integrated quantitative trait locus (QTL) analysis and association studies. *New Phytol*. 209:1067–1082.

Eckert AJ, Bower AD, González-Martínez SC, Wegrzyn JL, Coop G, Neale DB. 2010a. Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). *Mol Ecol*. 19:3789–3805.

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

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.

Eckert AJ, Wegrzyn JL, Liechty JD, Lee JM, Cumbie WP, Davis JM, Goldfarb B, Loopstra CA, Palle SR, Quesada T, *et al*. 2013. The evolutionary genetics of the genes underlying phenotypic associations for loblolly pine (*Pinus taeda*, Pinaceae). *Genetics*. 195:1353–1372.

Foll M, Gaggiotti O. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*. 180:977–993.

Gibson G. 2008. The environmental contribution to gene expression profiles. *Nat Rev Genet*. 9:575–581.

Gonzalez-Martinez SC, Krutovsky KV, Neale DB. 2006. Forest-tree population genomics and adaptive evolution. *New Phytol*. 170:227–238.

Haasl RJ, Payseur BA. 2016. Fifteen years of genomewide scans for selection: trends, lessons and unaddressed genetic sources of complication. *Mol Ecol*. 25:5–23.

Hancock AM, Witonsky DB, Alkorta-Aranburu G, Beall CM, Gebremedhin A, Sukernik R, Utermann G, Pritchard JK, Coop G, Di Rienzo A. 2011. Adaptations to climate-mediated selective pressures in humans. *PLoS Genet*. 7:e1001375.

Hietpas RT, Jensen JD, Bolon DNA. 2011. Experimental illumination of a fitness landscape. *Proc Natl Acad Sci U S A*. 108:7896–7901.

Jensen JD, Foll M, Bernatchez L. 2016. The past, present and future of genomic scans for selection. *Mol Ecol*. 25:1–4.

Jiang Z, Han Y, Teng W, Zhang Z, Sun D, Yang D, Li W. 2010. Identification of QTL underlying the oil-filling rate at different developmental stages of soybean seed. *Euphytica*. 176:391–402.

Joost S, Bonin A, Taberlet P, Caloz R. 2008. Un rôle pour la science de l'information géographique en écologie moléculaire: la détection de régions du génome soumises à la sélection naturelle. *Revue internationale de géomatique* 18:215–237.

Kachigan SK. 1991. *Multivariate statistical analysis: a conceptual introduction*. New York: Radius Press.

Kawecki TJ, Ebert D. 2004. Conceptual issues in local adaptation. *Ecol Lett*. 7:1225–1241.

Kearsey MJ, Pooni HS. 1996. *The genetical analysis of quantitative traits*. London: Chapman & Hall.

Kindt R. 2016. *BiodiversityR: package for community ecology and suitability analysis*. Nairobi (Kenya): World Agroforestry Center.

Kutay U, Bischoff FR, Kostka S, Kraft R, Görlich D. 1997. Export of importin alpha from the nucleus is mediated by a specific nuclear transport factor. *Cell*. 90:1061–1071.

Lacaze X, Hayes PM, Korol A. 2009. Genetics of phenotypic plasticity: QTL analysis in barley, *Hordeum vulgare*. *Heredity*. 102:163–173.

Le Corre V, Kremer A. 2003. Genetic variability at neutral markers, quantitative trait land trait in a subdivided population under selection. *Genetics*. 164:1205–1219.

Legendre P, Gallagher E. 2001. Ecologically meaningful transformations for ordination of species data. *Oecologia*. 129:271–280.

Lorenz WW, Alba R, Yu YS, Bordeaux JM, Simões M, Dean JF. 2011. Microarray analysis and scale-free gene networks identify candidate regulators in drought-stressed roots of loblolly pine (*P. taeda* L.). *BMC Genomics*. 12:264.

Lotterhos KE, Whitlock MC. 2014. Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. *Mol Ecol*. 23:2178–2192.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P. 2003. The power and promise of population genomics: from genotyping to genome typing. *Nat Rev Genet*. 4:981–994.

Manel S, Perrier C, Pratlong M, Abi-Rached L, Paganini J, Pontarotti P, Aurelle D. 2016. Genomic resources and their influence on the detection of the signal of positive selection in genome scans. *Mol Ecol*. 25:170–184.

McCouch SR, Wright MH, Tung C-W, Maron LG, McNally KL, Fitzgerald M, Singh N, DeClerck G, Agosto-Perez F, Korniliev P, *et al*. 2016. Open access resources for genome-wide association mapping in rice. *Nat Commun*. 7:10532.

McDermott R, Tingley D, Cowden J, Frazzetto G, Johnson DDP. 2009. Monoamine oxidase A gene (MAOA) predicts behavioral aggression following provocation. *Proc Natl Acad Sci U S A*. 106:2118–2123.

McKay JK, Christian CE, Harrison S, Rice KJ. 2005. 'How local is local?'—A review of practical and conceptual issues in the genetics of restoration. *Restoration Ecol*. 13:432–440.

Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES. 2009. Association mapping: critical considerations shift from genotyping to experimental design. *Plant Cell*. 21:2194–2202.

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

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

Nedlo JE, Martin TA, Vose JM, Teskey RO. 2009. Growing season temperatures limit growth of loblolly pine (*Pinus taeda* L.) seedlings across a wide geographic transect. *Trees*. 23:751–759.

Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, Indap A, King KS, Bergmann S, Nelson MR, *et al*. 2008. Genes mirror geography within Europe. *Nature*. 456:98–101.

Nunes VL, Beaumont MA, Butlin RK, Paulo OS. 2011. Multiple approaches to detect outliers in a genome scan for selection in ocellated lizards (*Lacerta lepida*) along an environmental gradient: selection in occelated lizards. *Mol Ecol*. 20:193–205.

Palle SR, Seeve CM, Eckert AJ, Cumbie WP, Goldfarb B, Loopstra CA. 2011. Natural variation in expression of genes involved in xylem development in loblolly pine (*Pinus taeda* L.). *Tree Genet Genomes*. 7:193–206.

Paris M, Boyer S, Bonin A, Collado A, David J-P, Despres L. 2010. Genome scan in the mosquito *Aedes rusticus*: population structure and detection of positive selection after insecticide treatment: population structure at a regional scale. *Mol Ecol*. 19:325–337.

Pluess AR, Frank A, Heiri C, Lalagüe H, Vendramin GG, Oddou-Muratorio S. 2016. Genome-environment association study suggests local adaptation to climate at the regional scale in *Fagus sylvatica*. *New Phytol*. 210:589–601.

Poncet BN, Herrmann D, Gugerli F, Taberlet P, Holderegger R, Gielly L, Rioux D, Thuiller W, Aubert S, Manel S. 2010. Tracking genes of ecological relevance using a genome scan in two independent regional population samples of *Arabis alpina:* tracking genes of ecological relevance. *Mol Ecol*. 19:2896–2907.

Pritchard JK, Pickrell JK, Coop G. 2010. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr Biol*. 20:R208–R215.

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.

Rehfeldt GE. 2006. *A spline climate model for the western United States*. Fort Collins (CO): U. S. Department of Agriculture, Forest Service.

Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R. 2015. A practical guide to environmental association analysis in landscape genomics. *Mol Ecol*. 24:4348–4370.

Salazar-Ciudad I, Marín-Riera M. 2013. Adaptive dynamics under development-based genotype-phenotype maps. *Nature*. 497:361–364.

Savolainen O, Pyhäjärvi T, Knürr T. 2007. Gene flow and local adaptation in trees. *Annu Rev Ecol Evol Syst*. 38:595–619.

Schmidtling R. 2001. *Southern pine seed sources*. Asheville (NC): U. S. Department of Agriculture, Forest Service.

Schoville SD, Bonin A, François O, Lobreaux S, Melodelima C, Manel S. 2012. Adaptive genetic variation on the landscape: methods and cases. *Annu Rev Ecol Evol Syst*. 43:23–43.

Segelbacher G, Cushman SA, Epperson BK, Fortin M-J, Francois O, Hardy OJ, Holderegger R, Taberlet P, Waits LP, Manel S. 2010. Applications of landscape genetics in conservation biology: concepts and challenges. *Conserv Genet*. 11:375–385.

Shaw KL, Parsons YM, Lesnick SC. 2007. QTL analysis of a rapidly evolving speciation phenotype in the Hawaiian cricket *Laupala*. *Mol Ecol*. 16:2879–2892.

Squillace A, Wells O, Rockwood D. 1980. Inheritance of monoterpene composition in cortical oleoresin of loblolly pine. *Silvae Genet*. 29:141–151.

Stinchcombe JR, Hoekstra HE. 2008. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity (Edinb)*. 100:158–170.

Stölting KN, Paris M, Meier C, Heinze B, Castiglione S, Bartha D, Lexer C. 2015. Genome-wide patterns of differentiation and spatially varying selection between postglacial recolonization lineages of *Populus alba* (Salicaceae), a widespread forest tree. *New Phytol*. 207:723–734.

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

Torreano SJ, Morris LA. 1998. Loblolly pine root growth and distribution under water stress. *Soil Sci Soc Am J*. 62:818.

Westbrook JW, Walker AR, Neves LG, Munoz P, Resende MF Jr, Neale DB, Wegrzyn JL, Huber DA, Kirst M, Davis JM, *et al*. 2015. Discovering candidate genes that regulate resin canal number in *Pinus taeda* stems by integrating genetic analysis across environments, ages, and populations. *New Phytol*. 205:627–641.

Whitehead A, Roach JL, Zhang S, Galvez F. 2011. Genomic mechanisms of evolved physiological plasticity in killifish distributed along an environmental salinity gradient. *Proc Natl Acad Sci U S A*. 108:6193–6198.

Ziems LA, Hickey LT, Hunt CH, Mace ES, Platz GJ, Franckowiak JD, Jordan DR. 2014. Association mapping of resistance to *Puccinia hordei* in Australian barley breeding germplasm. *Theor Appl Genet*. 127:1199–1212.

Zimin A, Stevens KA, Crepeau MW, Holtz-Morris A, Koriabine M, Marçais G, Puiu D, Roberts M, Wegrzyn JL, de Jong PJ, *et al*. 2014. Sequencing and assembly of the 22-gb loblolly pine genome. *Genetics*. 196:875–890.