MBE Advance Access published May 9, 2012

Local Adaptation in the Flowering-Time Gene Network of Balsam Poplar, *Populus balsamifera* L.

Stephen R. Keller,*,†,[1] Nicholas Levsen,[2] Matthew S. Olson,[2,3] and Peter Tiffin*,†,[1]

[1]Department of Plant Biology, University of Minnesota

[2]Department of Biology, Texas Tech University

[3]Institute of Arctic Biology, Department of Biology and Wildlife, University of Alaska Fairbanks

†Present address: Appalachian Laboratory, University of Maryland Center for Environmental Science

***Corresponding author:** E-mail: skeller@umces.edu; ptiffin@umn.edu.

Associate editor: John Novembre

Abstract

Identifying the signature and targets of local adaptation is an increasingly important goal in empirical population genetics. Using data from 443 balsam poplar *Populus balsamifera* trees sampled from 31 populations, we tested for evidence of geographically variable selection shaping diversity at 27 homologues of the *Arabidopsis* flowering-time network. These genes are implicated in the control of seasonal phenology, an important determinant of fitness. Using 335 candidate and 412 reference single nucleotide polymorphisms (SNPs), we tested for evidence of local adaptation by searching for elevated population differentiation using F_{ST}-based outlier analyses implemented in BayeScan or a Hierarchical Model in Arelquin and by testing for significant associations between allele frequency and environmental variables using BAYENV. A total of 46 SNPs from 14 candidate genes had signatures of local adaptation—either significantly greater population differentiation or significant covariance with one or more environmental variable relative to reference SNP distributions. Only 11 SNPs from two genes exhibited both elevated population differentiation and covariance with one or more environmental variables. Several genes including the abscisic acid gene *ABI1B* and the circadian clock genes *ELF3* and *GI5* harbored a large number of SNPs with signatures of local adaptation—with SNPs in *GI5* strongly covarying with both latitude and precipitation and SNPs in *ABI1B* strongly covarying with temperature. In contrast to several other systems, we find little evidence that photoreceptors, including phytochromes, play an important role in local adaptation. Our results additionally show that detecting local adaptation is sensitive to the analytical approaches used and that model-based significance thresholds should be viewed with caution.

Key words: ecological genomics, F_{ST}, selection, population structure, landscape genomics.

Introduction

Species with ranges spanning strong environmental gradients experience spatially heterogeneous selection pressures, often leading to local adaptation of ecologically important traits. For decades, the study of local adaptation has been a cornerstone of ecological quantitative genetics (Clausen and Hiesey 1958; Ford 1964; Antonovics 1976; Kawecki and Ebert 2004), and substantial evidence supports the conclusion that populations often attain the highest mean fitness in their native environment (Leimu and Fischer 2008; Hereford 2009). Local adaptation is thus highly likely to shape nucleotide and genomic diversity, yet it is still understudied in population genetics relative to searches for species-wide selection. Recently, however, more attention has been focused on discovering and identifying the genetic targets of local selection. For example, in humans, genes underlying important traits, such as disease resistance, lactose tolerance, and skin pigmentation, have been identified as targets of local selection (Tishkoff et al. 2006; Voight et al. 2006; Tang et al. 2007; Hancock et al. 2008; Coop et al. 2009; Pickrell et al. 2009), and several recent studies in *Arabidopsis* and forest trees have identified genes involved in local adaptation in plants (Wachowiak et al. 2009; Eckert et al. 2010; Ma et al. 2010; Turner et al. 2010; Fournier-Level et al. 2011). Identifying the genes underlying local adaptation will lead to a greater understanding of how selection shapes functional genomic variation within species, and the degree to which ecological adaptation is achieved through convergent or parallel genetic architecture (Stern and Orgogozo 2009; Elmer and Meyer 2011).

Population genomic signatures of local adaptation include elevated differentiation among populations (F_{ST}), decreased polymorphism within populations (expected H_e), or extended linkage disequilibrium relative to neutral expectations (Lewontin and Krakauer 1973; McKay and Latta 2002; Schlotterer 2002; Beaumont 2005; Tang et al. 2007). Correctly identifying targets of local adaptation necessitates differentiating the effects of selection from demographic history. Doing so can be challenging because demographic history can be complex and produce patterns of diversity at some neutral genes that resemble expectations from selection, especially for species that have experienced rapid range expansion, recent admixture, or strong changes in effective population size (Nielsen 2005; Excoffier and Ray 2008; Excoffier et al. 2009). Moreover, although identifying putative targets of local adaptation is sometimes accomplished by identifying loci in the tails of the

distribution of a test statistic, such as F_{ST} (Beaumont and Balding 2004), the problem of sifting false positives from the true targets of local selection is still challenging. Model-based approaches that separate locus-specific from population-specific effects are effective in reducing the false-positive rate, although the robustness of these approaches are dependent upon the underlying demographic model (Foll and Gaggiotti 2008; Excoffier et al. 2009). An alternative, and perhaps more robust, approach for identifying targets of adaptation is to use empirical distributions that account for the variability across the genome (Akey et al. 2002; Nordborg et al. 2005; Garrigan et al. 2010), for example, using genome-wide reference loci to define the null distribution against which the a priori identified candidate genes are compared. The advantage of this approach is it is free of any particular demographic model.

Balsam poplar, *Populus balsamifera* L. Salicaceae, is a widely distributed forest tree with a range that spans strong environmental gradients of photoperiod, growing season length, temperature, and precipitation. Across this range, balsam poplar exhibits strong signals of differentiation in phenology and ecophysiological traits that are greater than expected given the demographic history of postglacial range expansion (Keller et al. 2010; Keller, Soolanayakanahally, et al. 2011). Analyses of range-wide sequence diversity in *Populus* homologs of the *Arabidopsis* flowering-time gene network—a key group of genes involved in ecological responses to light, temperature, and growing season length (Mouradov et al. 2002; Lagercrantz 2009), uncovered signals of selection in at least one gene and possibly up to four genes consistent with local adaptation between high- and low-latitude environments (Keller, Levsen, et al. 2011). This gene network is of great interest to plant evolutionary and molecular biologists since it forms the functional basis of many important plant traits, including germination, flowering, and seasonal dormancy (Mouradov et al. 2002; Flowers et al. 2009; Jackson 2009; Lagercrantz 2009); thus, understanding how selection acts on these genes at local scales and across environmental gradients is of high priority. However, range-wide samples that are too small to allow for assessment of local adaptation within species, as used in Keller, Levsen, et al. (2011) and characteristic of many samples used to characterize diversity in plant species, are likely to miss many targets of selection among local populations.

Here, we test for evidence of local adaptation among 31 geographically defined subpopulations using single nucleotide polymorphism (SNP) data from 443 individuals sampled from across the species range. We genotyped each of these individuals at 747 SNPs: 335 candidate SNPs in 27 genes from the flowering-time network, and 412 reference SNPs (1 per gene) sampled randomly from across the genome. To search for signatures of local adaptation, we employed two main analytical approaches, one that infers past selection by comparing the strength of population structure among loci and a second that estimates the strength of association between an environmental variable and population SNP frequencies. In each approach, we use reference-SNP distributions to identify those candidate SNPs that show strong evidence for a departure from the genome-wide background, providing robust evidence that candidates have been shaped by past local selection.

Materials and Methods

Balsam polar, *Populus balsamifera* L., is a fast-growing forest tree that is widely distributed in northern North America. It is an ecological keystone species in boreal forests, where it grows in riparian habitats and moist uplands. In this study, we sampled 443 trees from 31 geographically defined populations (10–18 individuals per population, mean = 14 from Alaska and Canada; Soolanayakanahally et al. 2009). One population IVI in Keller et al. (2010) was represented by only three individuals and therefore was pooled with the geographically and genetically close COT population.

Candidate SNP Genotyping

We genotyped each of the 443 individuals at 335 SNPs and short indel polymorphisms (for simplicity we hereafter referred to these as SNPs, even though 13 are short indels found in 27 flowering-time genes [supplementary table S1, Supplementary Material online]), using Sequenom iPLEX genotyping or direct sequencing. These candidate polymorphisms were chosen from 824 SNPs previously identified in these genes by direct sequencing a range-wide discovery panel of 24 *P. balsamifera* genotypes (Keller, Levsen, et al. 2011). From these 824 SNPs, we designed Sequenom multiplexing assays for 176 coding region SNPs both synonymous and replacement and 198 SNPs in introns. We eliminated putative SNPs that did not amplify correctly, 36 SNPs, were monomorphic in our sample, 24 SNPs, or preliminary analyses identified as being heterozygous in nearly every individual assayed and were thus assumed to reflect amplification of paralogs, four SNPs. After filtering, we were left with 310 Sequenom SNPs, with a mean per-allele error rate of 2.1% which we estimated using 13 duplicate samples. We used direct sequencing of the *FRIGIDA FRI* gene exon 2, intron 2, and exon 3 to assay an additional 25 candidate SNPs; Sequenom assays for this gene were not effective due to the high density of polymorphisms. Reads for *FRI* were trimmed based on PHRED quality scores, aligned in CodonCode Aligner, and genotypes at 25 SNP positions were called manually from electropherograms. After combining with the Sequenom loci, we had genotype calls for a total of 335 SNPs from 443 *P. balsamifera* individuals for further analysis. We used data for 412 coding or intron SNPs representing 412 gene models that had been previously assayed on the same 443 individuals (Keller et al. 2010), as a genomic control (hereafter referred to as reference SNPs) in order to generate empirical false-positive thresholds and to evaluate the sensitivity of the local adaptation analyses to assumptions about demographic history. These 412 SNPs were originally identified from a range-wide discovery panel of 15 individuals (Keller et al. 2010). Although we only surveyed a single SNP from each reference gene region, we consider the reference SNPs to be reasonably well matched to the

candidate gene SNPs because they both derive from a similar sized discovery panel sampled from across the species' distribution, and both SNP sets assayed genic regions that included SNPs from both coding and noncoding intronic or untranslated region sequence. We thus assume that the reference SNPs effectively capture the major patterns in population structure present in the genome and serve as a neutral demographic control for identifying selection acting on the candidate gene SNPs. All reference and candidate SNP genotypes are deposited and available for download at www.popgen.uaf.edu.

Analysis

We used three approaches to scan for signatures of local adaptation. First, we looked for candidate polymorphisms with F_{ST} values greater than expected in the absence of local adaptation using a Bayesian approach described by Foll and Gaggiotti (2008) and implemented in the program BayeScan. BayeScan scans for local adaptation in multilocus allele frequency data by separately modeling a population-specific effect β assuming an island model of demography, and a locus-specific effect α, that is sensitive to the strength of selection acting on a particular locus. For each SNP, BayeScan estimates the posterior distribution under neutrality $\alpha = 0$ and separately allowing for selection $\alpha \neq 0$ and computes the posterior odds ratio (PO) as a measure of support for the model of local adaptation relative to neutral demography. We ran BayeScan analyses separately on the reference SNPs and candidate SNPs under identical run conditions, consisting of a prior odds ratio of 10 corresponding to a prior belief that a selection model is 1/10 as likely as a neutral model for any given SNP, 20 pilot runs, a burn-in of 50,000 iterations, followed by 50,000 output iterations with a thinning interval of 10 resulting in 5,000 iterations for posterior estimation. To control against false positives, we calculated the expected value of log10PO that yielded a 1% false discovery rate (FDR) based on the reference SNPs—we refer to this as the model-based significance threshold. We also define a reference SNP–based significance threshold as those candidate SNPs with greater population differentiation than all but 1% of the reference SNPs.

Our second approach for identifying targets of local selection was to identify F_{ST} outliers using the hierarchical analysis of Excoffier et al. (2009), hereafter referred to as the Hierarchical Model, which specifically accounts for coancestry among related subpopulations. We implemented this approach because BayeScan, although shown in simulations to be effective at finding true positives and minimizing the number of false positives (Narum and Hess 2011), may be biased if there is hierarchical demographic structure in the sample (Excoffier et al. 2009). This may be a concern in *P. balsamifera* because our study populations group into three regional clusters, with the overall variance among populations (6.6%) mostly partitioned among these regions (4.4%) (Keller et al. 2010). We approached the analysis in two steps. First, we generated the neutral expectation of F_{ST} conditioned on expected heterozygosity using the 412 reference SNPs. To model the hierarchical genetic structure present in *P. balsamifera*, we specified a three-group model, corresponding to the three regions identified in Keller et al. (2010). Each simulated group consisted of 100 subpopulations. Simulations were conducted in Arlequin v3.5.1.2 (Excoffier and Lischer 2010) and used 50,000 replicates of the coalescent to identify the expected distribution of F_{ST} values for the reference SNPs. We then used the same approach to calculate F_{ST} on the 335 candidate SNPs. Similar to the approach we used with BayeScan, we defined a model-based significance threshold as F_{ST} values greater than 99% of the F_{ST} values from the simulated data as well as a reference-based significance threshold (hereafter referred to as a reference threshold) as F_{ST} values greater than 99% of the reference SNPs.

Our third approach to identify targets of local selection was the method of Coop et al. (2010) that tests for covariance between candidate SNP frequencies and environmental variables that exceed the expected covariances estimated using reference SNPs. Specifically, we characterized the neutral covariance in population allele frequencies at the 412 reference SNPs among the 31 populations using the mcmc algorithm in the program BAYENV by Coop et al. (2010). The resulting matrix of population differences Ω is closely analogous to F_{ST} and captures the pattern of allele frequency variance among populations as expected under genetic drift. The second phase of the analysis then tests for covariance between environmental variables and the population-specific allele counts at a given SNP (β_{ENV}) while using the reference SNP–derived Ω as a covariate to control for population history. BAYENV calculates the ratio of the posterior probabilities from each model as a measure of support for the hypothesis of adaptation $\beta_{ENVP} > 0$ versus random drift $\beta_{ENV} = 0$. We analyzed each SNP individually and determined the distributions of PO separately for reference and candidate SNPs. Candidate SNPs with log10-PO greater than 99% of the reference SNP values were interpreted as showing strong support for clinal adaptation along the environmental gradient. We relied exclusively on this reference SNP–based approach for identifying outlier candidate SNPs associated with environmental variables.

We used seven environmental variables, chosen to summarize gradients in climate and photoperiod that are hypothesized to be important to fitness across *P. balsamifera*'s widespread distribution, to characterize the abiotic environment from which trees were sampled. These included three geographic variables, degrees North latitude (LT), degrees West longitude (LG), and elevation (EL), and four bioclimatic variables mean annual temperature (AT), maximum temperature of the warmest month (WT), minimum temperature of the coldest month (CT), and mean annual precipitation (AP). Correlations among these variables are in supplementary table S3 (Supplementary Material online). Bioclimatic variables derived from the Worldclim database (Hijmans et al. 2005) were obtained from the International Centre for Tropical Agriculture (http://www.ccafs-climate.org/data/) as downscaled global circulation models based on the Intergovernmental Panel on Climate Change. To reduce the

Table 1. Number of SNPs, LD Groups, and Genes Tested and the Number that Exceeded the Model-Based and Reference-Based Significance Thresholds for the BayeScan Analysis, Hierarchical Model, as well as Reference-Based Significance Thresholds for Environmental Covariance BAYENV.

	Number Tested	BayeScan		Hierarchical Model		BAYENV
Significance Threshold		Model	Reference	Model	Reference	Reference
SNPs	335	14	12	38	2	45 (14, 15)
LD groups	210	6	4	20	2	24 (8, 7)
Genes	27	4	3	11	2	13 (6, 5)

NOTE.—Values in parentheses under BAYENV report the average number of outliers per environmental variable tested (first value), and the number of outliers on the first PC from a PCA of the seven environmental variables (second value). Reference-based significance thresholds were defined as the upper 1% of empirical values for reference SNPs.

dimensionality of the climate data, we also assessed SNP-environment associations using the first principal component (PC) from a principal components analysis (PCA) that included all seven environmental variables. This new synthetic climate variable explained 59% of the variance, with latitude loading positively and strongly (0.48) on PC1, whereas all other variables loaded negatively (LG = −0.39, EL = −0.08, AT = −0.46, WT = −0.35, CT = −0.36, AP = −0.38). Thus, PC1 described an environmental gradient ranging from cold dry climates at northwestern sites (large PC scores) to warm moist climates at southeastern sites (small PC scores).

In order to more fully investigate how local adaptation may have shaped diversity within each candidate gene and to increase our chances of identifying loci that have been targets of local selection, we assayed multiple SNPs from each candidate gene. However, analyzing multiple SNPs per gene introduces the potential that SNPs identified as outliers are in linkage disequilibrium (LD) with each other and thus not independent. To account for potential nonindependence among SNPs from the same gene that show signatures of local adaptation, we calculated intragenic LD among candidate SNPs using Richard Hudson's "dipdat" program for unphased genotype data (http://home.uchicago.edu/rhudson1/source/misc/dipld/), which is based on the method of Hill (1974). This method works directly on the SNP frequencies of the genotypes and does not attempt to phase alleles when calculating LD. We considered all pairs of SNPs with $r^2 > 0.3$ to display evidence of nonindependence even if separated by intervening SNPs within a gene with $r^2 < 0.3$. Although this cutoff is somewhat arbitrary, it should provide a fairly conservative measure for how often SNPs at two positions are cosegregating. This produced 210 such LD groups, which we used to calculate the percentage of independent outliers in local adaptation scans. Intragenic LD does not affect the reference loci because only a single reference SNP was sampled from each sequenced region. Because our intent was to investigate LD between specific outlier candidate SNPs, we did not apply any minimum allele frequency filter to the SNP data when calculating LD.

Results

Divergence among Populations

BayeScan estimates of population divergence among *Populus* flowering-time homologs varied widely, with average population F_{ST} of candidate SNPs being slightly higher than reference SNPs (0.077 vs. 0.062, respectively; Mann–Whitney *U* test: $P < 0.0001$). Eight of 412 reference SNPs (1.9%) exceeding the model-based 1% FDR. In contrast, 14 of 335 candidate SNPs (4.2%), representing 6 of 210 independent LD groups, exceeded the 1% model-based false-discovery threshold for BayeScan, the majority of which ($N = 12$ SNPs) also exceeded the reference-based 1% threshold (table 1 and fig. 1). Support for selection on candidate SNPs was very strong, with 11 SNPs having log10PO ratios >3 (fig. 1). The locus-specific effect α estimated by BayeScan was also more than twice as large for candidate compared with reference SNPs (mean absolute value of $\alpha = 0.221$ vs. 0.102; Mann–Whitney *U* test: $P = 0.0003$).

Population divergence estimated from the Hierarchical Model (Excoffier et al. 2009) identified 19 reference SNPs (4.6%) and 38 candidate SNPs (11.3%) representing 20 independent LD groups (9.4%) with significantly greater F_{ST} values than all but 1% of values simulated under a neutral model (table 1 and fig. 2). Only two candidate SNPs, $GI5_{5271}$ ($F_{ST} = 0.94$) and $PIF3.1_{2601}$ ($F_{ST} = 0.64$), had F_{ST} values that greatly exceeded the highest of the reference SNP values ($F_{ST} = 0.51$). A history of regional-scale selection based on the three groups defined by Bayesian clustering in Keller et al. (2010) was also evident for 20 SNPs from eight candidate genes, *ABI1D*, *CO1*, *CRY1.1*, *CRY1.2*, *EBS*, *GI5*, *LFY*, and *PIF3.1*, which had F_{CT} values greater than all but 1% of values from the neutral simulations, with two of these equal to or greater than all the empirical reference SNP F_{CT} values ($CRY1.1_{3018}$ [$F_{CT} = 0.38$], $PIF3.1_{2601}$ [$F_{CT} = 0.65$], and highest reference [$F_{CT} = 0.38$]). The large number of reference SNPs with F_{ST} or F_{CT} values greater than expected under the simulated neutral model suggests either that the Hierarchical Model failed to properly capture the demographic history of our sample leading to a high false-positive rate or that a large portion of our reference SNPs are actually affected by local selection.

SNP-Environment Associations

Bayesian tests of SNP-environment associations identified 45 candidate SNPs representing 24 of 210 independent LD groups. LD groups defined by SNPs with $r^2 > 0.3$ with log10PO greater than all but 1% of reference SNPs for at least one environmental variable (table 1). On average, each environmental variable was significantly associated with 14 candidate SNPs and 15 SNPs were associated with the first PC from the PCA on the seven environmental variables.

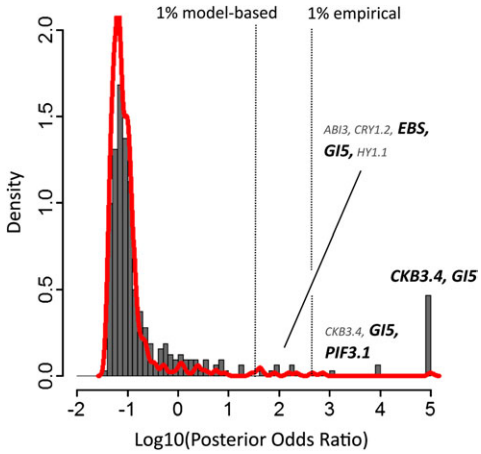

FIG. 1. The distributions of the log10PO of a selection model versus a neutral model of population divergence from the BayeScan analyses of local adaptation. The distribution based on reference SNPs is plotted in red, and the candidate phenology genes are shown in gray bars. The dotted lines show the 1% model-based significance threshold based on the reference SNPs and the 1% upper tail of the empirical reference SNP distribution. Genes with SNPs in the upper tail of the PO are labeled and those significant for local adaptation, that is, positive α are in boldface and larger font.

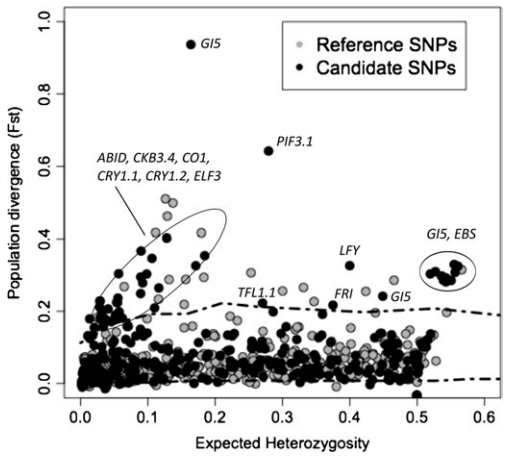

FIG. 2. Results from the Hierarchical Model analysis of differentiation among populations. Reference SNPs are in gray; candidate SNPs are in black. Dotted line indicates the 99% confidence interval from 50,000 coalescent simulations based on the reference SNPs. Circles group clusters of candidate SNPs with similar F_{ST} and H_{exp} values together for ease of labeling.

Environmental associations were especially strong for the central circadian clock gene *GI5*, with 13 of 22 *GI5* SNPs having log10PO that were outliers from the reference distribution (table 2). Environmental associations with *GI5* SNPs were especially prominent for latitude (12 SNPs) (table 2 and fig. 3) and mean annual temperature (12 SNPs) as well as longitude (11 SNPs) and precipitation (11 SNPs). Although the associations with multiple environmental variables may be due to pleiotropy or LD with other causative SNPs, it may also reflect correlations among our climate variables, as many of these SNPs were significantly associated with the synthetic PC environmental variable (table 2, supplementary table S3, Supplementary Material online). Among other candidate SNPs, several showed strong associations with latitude and temperature, especially maximum temperature of the warmest month (table 2 and supplementary fig. S1, Supplementary Material online). The strongest of these associations were between SNPs in the abscisic acid gene *ABI1B* and maximum temperature of the warmest month and also between SNPs in the *EARLY FLOWERING 3 ELF3* gene and minimum temperature of the coldest month. No reference SNPs showed strong evidence of environmental association (all log10PO < 2), and in general, reference SNPs showed less extreme values relative to candidate SNPs (supplementary fig. S2, Supplementary Material online).

Comparison of Methods

The two F_{ST} outlier approaches we used to test for local adaptation, BayeScan and the Hierarchical Model, together identified 39 locally adapted SNPs using model-based thresholds—13 of which were common to both approaches, 1 was identified only by BayeScan, and 25 were identified only by the Hierarchical Model. Several SNPs with low posterior probabilities of selection from BayeScan analyses were identified as potential targets of selection in the Hierarchical Model (supplementary fig. S3, Supplementary Material online), although none had F_{ST} values that exceeded the upper 5% of reference SNP F_{ST} values. Of the 12 candidate SNPs that showed elevated F_{ST} relative to the reference-based threshold, all but one also showed strong SNP-environment associations (BAYENV). By contrast, 34 candidate SNPs exceeded the reference-based significance threshold solely for the BAYENV test of environmental association, without being significant outliers in F_{ST}.

Although our ability to characterize the extent of local adaptation is partially complicated by LD among SNPs in our candidate genes, comparison of the apparent extent of local adaptation uncovered by each of the three methods is similar whether we look at SNPs, LD groups, or genes. For all three classes, the greatest number of genomic features identified as targets of local selection come from BAYENV and the Hierarchical Model and the fewest from BayeScan. Moreover, regardless of whether one looks at SNPs, LD groups, or genes, the number of SNPs identified as putative targets of selection by the model-based versus reference-based significance thresholds was far greater for the Hierarchical Model than for the BayeScan approach (table 1).

Discussion

Determining the genetic basis of ecological local adaptation within species is an emerging theme in population genetics. The effects of selection and demography on population genetic variation can be quite different at local scales compared with the species-wide scale (Moeller et al. 2007; Städler et al. 2009). Species-wide samples will not show standard signatures of selective sweeps when selection acts locally, especially for genes that contribute to variation in polygenic traits (Kelly 2006; Pritchard and Di Rienzo 2010; Pritchard et al. 2010). As such, range-wide samples that lack in-depth sampling of subpopulations may provide little

Table 2. Candidate Gene SNPs that Exceed Empirical Significance Threshold in Either the BayeScan or Hierarchical Model, or Significantly Covary with One or More Environmental Variables.

Gene	Position[a]	F_{ST}[b]	Method[c]	Environmental Correlation[d]	Functional Class[e]
ABI1B	1468[1]	0.13		LG, WT	Abscisic acid sensitivity
	1658[1]	0.11		WT	
	1684[1]	0.11		WT	
	2055[1]	0.14		WT	
	2376[1]	0.14		LG, WT	
ABI1D	1207	0.19		CT	Abscisic acid sensitivity
	1626	0.12		LG	
ABI3	3497	0.09		LT, WT, AT	Abscisic acid sensitivity
CKB3.4	1714	0.33	B, BE		Peripheral circadian clock
ELF3	1955[1]	0.18		CT	Peripheral circadian clock
	2559[1]	0.18		CT	
	3363	0.07		LT, LG, WT, PC	
	3684	0.03		WT	
	4541[2]	0.15		EL	
	5779[1]	0.18		CT	
	5785[1]	0.20		CT	
	5834[1,2]	0.18		EL	
	5889[1]	0.23		CT	
FRI	927	0.06		WT	Vernalization response
	3020	0.09		AP, CT, AT	
	3048	0.13		AP	
GI2	10306[1]	0.02		LT, CT, WT, AT, PC	Central circadian clock
	10307[1]	0.06		LT, AT	
GI5	33[1]	0.29	B, BE	LT, LG, AP, AT, PC	Central circadian clock
	198[1]	0.32	B, BE	LT, LG, AP, CT, AT, PC	
	268	0.07		LT, WT, AT	
	1950[1]	0.30	B, BE	LT, LG, AP, AT, PC	
	2405[1]	0.33	B, BE	LT, LG, AP, CT, AT, PC	
	2612[1]	0.30	B, BE	LT, LG, AP, AT, PC	
	3966	0.24	B	LT, EL, AP, CT, AT, PC	
	5271	0.94	HE, B, BE	LT, LG, EL, AP, CT, WT, AT, PC	
	8997[1]	0.28	B, BE	LT, LG, AP, CT, AT, PC	
	9447	0.25		LG	
	9551[1]	0.31	B, BE	LT, LG, AP, CT, AT, PC	
	9585[1]	0.29	B, BE	LT, LG, AP, AT, PC	
	9659[1]	0.29	B, BE	LT, LG, AP, AT, PC	
HY2.1	138	0.00		EL	Photoreceptor
	173	0.02		EL	
LFY	2220[1]	0.09		LT, AT	Downstream target
	2905[1]	0.09		LT, WT, AT, PC	
PHYA	1344	0.11		LG	Photoreceptor
PIF3.1	2601	0.64	HE, B, BE	EL	Photoreceptor
PtFT1	461	0.12		LT, CT, PC	Downstream target
	3044	0.15		EL, CT, AT	
ZTL2	5936[1]	0.07		CT, AT	Photoreceptor
	6168[1]	0.09		CT	

NOTE.—If SNPs were also identified using the model-based significance thresholds in either the BayeScan or Hierarchical Model that is also indicated; SNPs identified only by exceeding the model-based thresholds from BayeScan or the Hierarchical Model are in supplementary table S2 (Supplementary Material online). SNPs in high LD with one another within a single gene are indicated by superscripts.

[a] Position of SNP from start of gene.

[b] F_{ST} estimate from the Hierarchical Model of Excoffier et al. (2009).

[c] Significance at the 1% level according to HE = Hierarchal Model empirical significance, shown only when an SNP was also identified by one of the other methods the 25 candidate SNPs with F_{ST} values that exceeded the model-based significance threshold are shown in supplementary table S2 (Supplementary Material online), B = simulations within BayeScan, and BE = empirical distribution from BayeScan.

[d] Abbreviations of environmental covariates: LT = Latitude, LG = Longitude, EL = Elevation, AP = Mean annual precipitation, CT = Temperature during the coldest month, WT = Temperature during the warmest month, AT = Mean annual temperature, PC = first principal component from a PCA of all seven environmental variables.

[e] Functional class groupings similar to those in Hall et al. (2011) and are based on the genes' role in the flowering-time gene network (Mouradov et al. 2002; Lagercrantz 2009).

power for detecting evidence of local adaptation, which is extremely common in plants (Leimu and Fischer 2008) and particularly forest trees (Petit and Hampe 2006; Savolainen et al. 2007; Neale and Ingvarsson 2008). However, such in-depth sampling at the population scale opens up a host of other challenges related to population structure and the covariance between demographic history, environmental gradients, and genetic divergence among populations. Addressing the question of local adaptation at the genetic level requires not only dense population-level sampling that captures how functional genetic variation shifts across ecological gradients but also a careful approach to

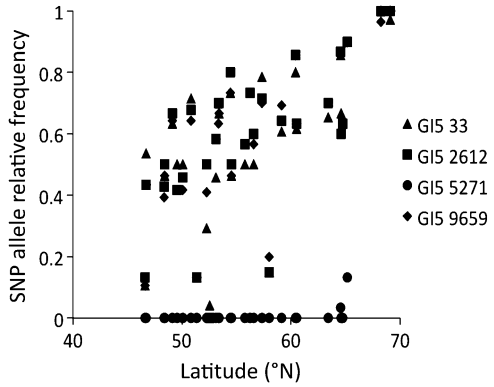

Fig. 3. Clinal variation in the circadian clock gene *GIGANTEA 5 (GI5)*. Associations with latitude at subset of four representative SNPs located across the length of the *GI5* gene; SNP frequency was scaled by the SNP that was most frequent in the northernmost population.

identification of genes under selection from the background of neutral demographic history.

Our study addresses the challenge of detecting ecologically important genetic variation under local selection by combining broad population-level sampling across the widespread range of *P. balsamifera* with paired genetic data sets that separately estimate the effects of selection on a well-studied gene network controlling phenology while independently controlling for demographic history using genome-wide reference gene SNPs. Our analyses suggest that over half of the 27 phenology candidate genes we examined harbor SNPs with frequencies that have been shaped by geographically variable selection. Linkage disequilibrium between SNPs within candidate genes generates some nonindependence among outlier SNPs, but this does not appear to affect our overall conclusions since gene-wide LD measured as Kelly's ZnS was not different among genes putatively under local selection (ZnS = 0.123) compared with those not under local selection (ZnS = 0.127 Mann–Whitney *U* test; $P = 0.751$). Regardless, our results suggest that a large portion of the flowering-time network has been subject to strong positive selection.

A major agent of the selection we detected has probably been local differences in climate and growing season length across balsam poplar's geographic range (Keller, Soolanayakanahally, et al. 2011); the majority of the SNPs we identified as targets of selection are strongly associated with latitude, temperature, or precipitation (table 2). Other environmental sources of selection (biotic or abiotic) probably also contribute to some of the population divergence observed, for example, two genes (*CKB3.4* and *EBS*) show high F_{ST} but no SNP-environment associations. Our results, along with other recent studies, suggest that comparing population-level diversity of candidate genes to variation in the genomic background can start to unravel the genetic basis of local adaptation (Hall et al. 2007; Moeller and Tiffin 2008; Baxter et al. 2010; Turner et al. 2010; Xia et al. 2010; Fournier-Level et al. 2011).

Most of the putative targets of local selection we identified were not identified as targets of selection in a previous analysis of a species-wide sample of 24 individuals (Keller, Levsen, et al. 2011), in which only three genes, *EBS*, *FRI*, and *LFY*, showed elevated nucleotide diversity or latitudinal differentiation (Keller, Levsen, et al. 2011). Furthermore, *GI5* and *PIF3.1*, the two genes with the most consistent signal of local adaptation in this study, showed no evidence in the species-wide sample of deviating from genomic background levels of nucleotide diversity, Tajima's *D*, or latitudinal differentiation. Not surprisingly, the contrasting results from the population genomic analysis of SNP genotypes used here and that of the species-wide sample of sequence diversity reported in Keller, Levsen, et al. (2011) suggest that our ability to characterize adaptive evolution in the genome is directly related to the scale of sampling. Moreover, the importance of sampling suggests that selection may be a strong force shaping plant genomic variation even if species-wide studies reveal little evidence for selection driving fixation of mutations between species (Flowers et al. 2009; Gossmann et al. 2010; Keller, Levsen, et al. 2011).

Targets of Local Selection

The gene that showed the strongest and most consistent evidence for ecological adaptation across all of our analyses was the central circadian clock gene, *GIGANTEA 5 (GI5)*. Population structure was highly elevated among *GI5* SNPs, and there were also strong patterns of covariance with multiple environmental variables, especially latitude and precipitation. Of the 13 *GI5* SNPs that were significant in one or more of the local adaptation scans, 3 were nonsynonymous changes, all of which are in strong LD with each other, $r^2 > 0.8$. In fact, the *GI5* gene contains two divergent groups of haplotypes, with LD extending up to 10 kb (supplementary fig. S4, Supplementary Material online), suggesting that one or more episodes of selection affected polymorphism along the entire length of *GI5*. *GI* is directly involved with light signaling to the circadian clock via interactions with phytochromes and is also an upstream regulator of the major floral genes *CONSTANS (CO)* and poplar *FLOWERING LOCUS T (PtFT1)* (Mizoguchi et al. 2005; Böhlenius et al. 2006; Oliverio et al. 2007; Hsu et al. 2011). These roles are consistent with the statistical association between *GI5* SNPs and latitude, which is closely correlated with seasonal photoperiod and suggests that *GI5* contributes to circadian clock-mediated adaptation through sensing of the day length (Fowler et al. 1999; Böhlenius et al. 2006; Jackson 2009; Lagercrantz 2009). Additional evidence for a role of *GI5* in local adaptation comes from cosegregation of *GI5* and bud set quantitative trait locus (QTL) in hybrid poplar mapping populations (Rohde et al. 2011), coupled with strong latitudinal clines for bud set in balsam poplar (Keller, Soolanayakanahally, et al. 2011).

The remaining genes that showed significant signals of local adaptation constituted a wider variety of functional roles, including abscisic acid production (*ABI3*, *ABI1B*, and *ABI1D*), light signaling to the circadian clock (*ZTL2*, *ELF3*, and *PIF3.1*), and temperature-dependent responses such as vernalization (*FRI*) or floral meristem development and the timing of flowering (*ELF3*, *LFY*, and *PtFT1*) (Johanson et al. 2000; Mouradov et al. 2002; Lagercrantz 2009; Hsu

et al. 2011). Interestingly, the two paralogs of *FT* in poplar *PtFT1* and *PtFT2* have recently been shown to control different phenological traits, with *PtFT1* primarily responsible for reproductive timing and sensitivity to winter temperatures (Hsu et al. 2011), consistent with the environmental association we uncovered between *PtFT1* and temperature of the coldest month.

Interestingly, of the three phytochrome genes we studied, *PHYA*, *PHYB1*, and *PHYB2*, there was little evidence of local adaptation. By contrast, phytochromes have been implicated as targets of local adaptation to different seasonal environments in several other species including the closely related *P. tremula* (Ingvarsson et al. 2006; Ma et al. 2010; Hall et al. 2011) as well as *Arabidopsis thaliana* (Samis et al. 2008), *Cardamine nipponica* (Ikeda et al. 2009), and *Arcteria nana* (Ikeda and Setoguchi 2010). The lack of evidence for local adaptation in phytochromes of *P. balsamifera* therefore suggests that convergent phenotypic adaptations in phenology have been achieved through different genetic targets, even between the closely related *P. tremula* and *P. balsamifera*.

Comparing Methods to Identify Locally Adapted Genes

The different signatures left by local selection, as well as varying assumptions about demography, have led to the development of diverse methods for detecting local adaptation from genotype data. Traditionally, tests for extreme allele frequency divergence (F_{ST}) have been used as evidence of local adaptation (Lewontin and Krakauer 1973; Beaumont 2005), and we employed two recent versions of these tests, one that explicitly models a hierarchical population structure (Excoffier et al. 2009) and another that uses a Bayesian approach to separate locus-specific effects of selection from population-specific effects of demography (Foll and Gaggiotti 2008). As an alternative to these F_{ST}-outlier tests, we also searched for associations between candidate SNPs and variation in the abiotic environment, as this approach may be more sensitive to subtle shifts in allele frequencies and has the potential to reveal ecological agents of selection (Coop et al. 2010; Hancock et al. 2010, 2011). We found large variation among these three methods (supplementary table S2, Supplementary Material online), and only 11 of the 46 SNPs, representing three LD groups, exceeded the reference threshold criterion for local selection and were identified as targets of local selection by more than one method. The Hierarchical Model identified a large number of candidate SNPs as outliers in model-based tests that were not identified by the other methods. Because the Hierarchical Model also identified a large number of reference SNPs as outliers, it is likely that this method did not adequately model the demographic structure of *P. balsamifera*, resulting in a high rate of false positives.

A large number of outliers were also identified by the SNP-environment association analysis, BAYENV. These included SNPs in the meristem identity gene *LEAFY (LFY)*, which were significantly associated with latitude and multiple temperature environmental variables and showed elevated nucleotide diversity in the species-wide sample (Keller, Levsen, et al. 2011), yet were not significant in the F_{ST} outlier analyses. The large number of outliers was probably influenced in part by correlations among climate traits, although there was still a high frequency of SNP-environment associations after accounting for nonindependence among environmental variables with PCA (table 1). Furthermore, the overall weak values of log10PO for the reference SNPs would seem to argue against a high false-positive rate. Other studies have argued that selection on individual SNPs or even QTL may be difficult to detect using traditional scans for elevated F_{ST} (Latta 2004; Kelly 2006), and recent work in humans and in *P. tremula* have found several genes that showed evidence of SNP-environment associations but were not outliers in analyses of population structure (Hancock et al. 2010; Ma et al. 2010). This may suggest that detecting subtle allele frequency clines along environmental gradients, when properly controlling for effects of neutral demography, may represent a powerful complement to F_{ST}-based outlier analyses when testing for local adaptation (Coop et al. 2010).

Supplementary Material

Supplementary tables S1–S3 and figures S1–S4 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Bill Schroeder for access to the AgCanBaP sample collection and Kenny Beckman and Dinesha Walek at the University of Minnesota BioMedical Genomics Center (BMGC) for contributions to iPLEX genotyping. The Minnesota Supercomputing Institute provided computational support. We thank the four reviewers who provided helpful comments on the manuscript. This work was funded by National Science Foundation Plant Genome award (DBI-0701911 to M.S.O. and P.T.).

References

Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. 2002. Interrogating a high-density SNP map for signatures of natural selection. *Genome Res.* 12:1805–1814.

Antonovics J. 1976. The input from population genetics: "The New Ecological Genetics." *Syst Bot.* 1:233–245.

Baxter I, Brazelton JN, Yu D, et al. (12 co-authors). 2010. A coastal cline in sodium accumulation in Arabidopsis thaliana is driven by natural variation of the sodium transporter AtHKT1;1. *PLoS Genet.* 6:e1001193.

Beaumont MA. 2005. Adaptation and speciation: what can Fst tell us? *Trends Ecol Evol.* 20:435–440.

Beaumont MA, Balding DJ. 2004. Identifying adaptive genetic divergence among populations from genome scans. *Mol Ecol.* 13:969–980.

Böhlenius H, Huang T, Charbonnel-Campaa L, Brunner AM, Jansson S, Strauss SH, Nilsson O. 2006. CO/FT regulatory module controls timing of flowering and seasonal growth cessation in trees. *Science* 312:1040–1043.

Clausen J, Hiesey WM. 1958. Experimental studies on the nature of species. IV. Genetic structure of ecological races. Washington (DC): Carnegie Institution of Washington Publication, No. 615.

Coop G, Pickrell JK, Novembre J, Kudaravalli S, Li J, Absher D, Myers RM, Cavalli-Sforza LL, Feldman MW, Pritchard JK. 2009. The role of geography in human adaptation. *PLoS Genet.* 5:e1000500.

Coop G, Witonsky D, Di Rienzo A, Pritchard JK. 2010. Using environmental correlations to identify loci underlying local adaptation. *Genetics* 185:1411.

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

Elmer KR, Meyer A. 2011. Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends Ecol Evol.* 26:298–306.

Excoffier L, Hofer T, Foll M. 2009. Detecting loci under selection in a hierarchically structured population. *Heredity* 103:285–298.

Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour.* 10:564–567.

Excoffier L, Ray N. 2008. Surfing during population expansions promotes genetic revolutions and structuration. *Trends Ecol Evol.* 23:347–351.

Flowers JM, Hanzawa Y, Hall MC, Moore RC, Purugganan MD. 2009. Population genomics of the Arabidopsis thaliana flowering time gene network. *Mol Biol Evol.* 26:2475–2486.

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.

Ford EB. 1964. Ecological genetics. London: Methuen.

Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM. 2011. A map of local adaptation in Arabidopsis thaliana. *Science* 334:86–89.

Fowler S, Lee K, Onouchi H, Samach A, Richardson K, Morris B, Coupland G, Putterill J. 1999. GIGANTEA: a circadian clock-controlled gene that regulates photoperiodic flowering in *Arabidopsis* and encodes a protein with several possible membrane-spanning domains. *EMBO J.* 18:4679–4688.

Garrigan D, Lewontin R, Wakeley J. 2010. Measuring the sensitivity of single-locus "neutrality tests" using a direct perturbation approach. *Mol Biol Evol.* 27:73–89.

Gossmann TI, Song B-H, Windsor AJ, Mitchell-Olds T, Dixon CJ, Kapralov MV, Filatov DA, Eyre-Walker A. 2010. Genome wide analyses reveal little evidence for adaptive evolution in many plant species. *Mol Biol Evol.* 27:1822–1832.

Hall D, Luquez V, Garcia VM, St Onge KR, Jansson S, Ingvarsson PK. 2007. Adaptive population differentiation in phenology across a latitudinal gradient in European aspen *Populus tremula*, L.: a comparison of neutral markers, candidate genes and phenotypic traits. *Evolution* 61:2849–2860.

Hall D, Ma X-F, Ingvarsson PK. 2011. Adaptive evolution of the *Populus tremula* photoperiod pathway. *Mol Ecol.* 20:1463–1474.

Hancock AM, Witonsky DB, Alkorta-Aranburu G, Beall CM, Gebremedhin A, Sukernik R, et al. 2011. Adaptations to climate-mediated selective pressures in humans. *PLoS Genet.* 7:e1001375.

Hancock AM, Witonsky DB, Ehler E, et al. (11 co-authors). 2010. Colloquium paper: human adaptations to diet, subsistence, and ecoregion are due to subtle shifts in allele frequency. *Proc Natl Acad Sci U S A.* 107:8924–8930.

Hancock AM, Witonsky DB, Gordon AS, Eshel G, Pritchard JK, Coop G, Di Rienzo A. 2008. Adaptations to climate in candidate genes for common metabolic disorders. *PLoS Genet.* 4:e32.

Hereford J. 2009. A quantitative survey of local adaptation and fitness trade-offs. *Am Nat.* 173:579–588.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol.* 25:1965–1978.

Hill WG. 1974. Estimation of linkage disequilibrium in randomly mating populations. *Heredity* 33:229–239.

Hsu C-Y, Adams JP, Kim H, et al. (20 co-authors). 2011. FLOWERING LOCUS T duplication coordinates reproductive and vegetative growth in perennial poplar. *Proc Natl Acad Sci U S A.* 108:10756–10761.

Ikeda H, Fujii N, Setoguchi H. 2009. Molecular evolution of phytochromes in *Cardamine nipponica* (Brassicaceae) suggests the involvement of PHYE in local adaptation. *Genetics* 182:603–614.

Ikeda H, Setoguchi H. 2010. Natural selection on PHYE by latitude in the Japanese archipelago: insight from locus specific phylogeographic structure in *Arcterica nana* (Ericaceae). *Mol Ecol.* 19:2779–2791.

Ingvarsson PK, García M, Hall D, Luquez V, Jansson S. 2006. Clinal variation in phyB2, a candidate gene for day-length-induced growth cessation and bud set, across a latitudinal gradient in European aspen *Populus tremula*. *Genetics* 172:1845–1853.

Jackson SD. 2009. Plant responses to photoperiod. *New Phytol.* 181:517–531.

Johanson U, West J, Lister C, Michaels S, Amasino R, Dean C. 2000. Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. *Science* 290:344–347.

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

Keller SR, Levsen N, Ingvarsson PK, Olson MS, Tiffin P. 2011. Local selection across a latitudinal gradient shapes nucleotide diversity in balsam poplar, *Populus balsamifera* L. *Genetics* 188:941–952.

Keller SR, Olson MS, Silim S, Schroeder WR, Tiffin P. 2010. Genomic diversity, population structure, and migration following rapid range expansion in the Balsam Poplar, *Populus balsamifera*. *Mol Ecol.* 19:1212–1226.

Keller SR, Soolanayakanahally RY, Guy RD, Silim SN, Olson MS, Tiffin P. 2011. Climate-driven local adaptation of ecophysiology and phenology in balsam poplar, *Populus balsamifera* L. Salicaceae. *Am J Bot.* 98:99–108.

Kelly JK. 2006. Geographical variation in selection, from phenotypes to molecules. *Am Nat.* 167:481–495.

Lagercrantz U. 2009. At the end of the day: a common molecular mechanism for photoperiod responses in plants? *J Exp Bot.* 60:2501–2515.

Latta RG. 2004. Gene flow, adaptive population divergence and comparative population structure across loci. *New Phytol.* 161:51–58.

Leimu R, Fischer M. 2008. A meta-analysis of local adaptation in plants. *PLoS One* 3:e4010.

Lewontin RC, Krakauer J. 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* 74:175–195.

Ma XF, Hall D, Onge KR, Jansson S, Ingvarsson PK. 2010. Genetic differentiation, clinal variation and phenotypic associations with growth cessation across the *Populus tremula* photoperiodic pathway. *Genetics* 186:1033–1044.

McKay JK, Latta RG. 2002. Adaptive population divergence: markers, QTL and traits. *Trends Ecol Evol.* 17:285–291.

Mizoguchi T, Wright L, Fujiwara S, et al. (11 co-authors). 2005. Distinct roles of GIGANTEA in promoting flowering and regulating circadian rhythms in Arabidopsis. *Plant Cell* 17:2255–2270.

Moeller DA, Tenaillon MI, Tiffin P. 2007. Population structure and its effects on patterns of nucleotide polymorphism in teosinte Zea mays ssp. parviglumis. *Genetics* 176:1799–1809.

Moeller DA, Tiffin P. 2008. Geographic variation in adaptation at the molecular level: a case study in plant immunity genes. *Evolution* 62:3069–3081.

Mouradov A, Cremer F, Coupland G. 2002. Control of flowering time: interacting pathways as a basis for diversity. *Plant Cell* 14(Suppl):S111–S130.

Narum SR, Hess JE. 2011. Comparison of F_{ST} outlier tests for SNP loci under selection. *Mol Ecol Res.* 11:184–194.

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

Nielsen R. 2005. Molecular signatures of natural selection. *Annu Rev Genet.* 39:197–218.

Nordborg M, Hu TT, Ishino Y, et al. (24 co-authors). 2005. The pattern of polymorphism in Arabidopsis thaliana. *PLoS Biol.* 3:e196.

Oliverio KA, Crepy M, Martin-Tryon EL, Milich R, Harmer SL, Putterill J, Yanovsky MJ, Casal JJ. 2007. GIGANTEA regulates phytochrome A-mediated photomorphogenesis independently of its role in the circadian clock. *Plant Physiol.* 144:495–502.

Petit RJ, Hampe A. 2006. Some evolutionary consequences of being a tree. *Annu Rev Ecol Evol Syst.* 37:187–214.

Pickrell JK, Coop G, Novembre J, et al. (11 co-authors). 2009. Signals of recent positive selection in a worldwide sample of human populations. *Genome Res.* 19:826–837.

Pritchard JK, Di Rienzo A. 2010. Adaptation—not by sweeps alone. *Nat Rev Genet.* 11:665–667.

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

Rohde A, Storme V, Jorge V, et al. (15 co-authors). 2011. Bud set in poplar—genetic dissection of a complex trait in natural and hybrid populations. *New Phytol.* 189:106–121.

Samis KE, Heath KD, Stinchcombe JR. 2008. Discordant longitudinal clines in flowering time and phytochrome C in *Arabidopsis thaliana*. *Evolution* 62:2971–2983.

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

Schlotterer C. 2002. Towards a molecular characterization of adaptation in local populations. *Curr Opin Genet Dev.* 12:683–687.

Soolanayakanahally RY, Guy RD, Silim SN, Drewes EC, Schroeder WR. 2009. Enhanced assimilation rate and water use efficiency with latitude through increased photosynthetic capacity and internal conductance in balsam poplar *Populus balsamifera* L. *Plant Cell Environ.* 32:1821–1832.

Städler T, Haubold B, Merino C, Stephan W, Pfaffelhuber P. 2009. The impact of sampling schemes on the site frequency spectrum in nonequilibrium subdivided populations. *Genetics* 182:205–216.

Stern DL, Orgogozo V. 2009. Is genetic evolution predictable? *Science* 323:746–751.

Tang K, Thornton KR, Stoneking M. 2007. A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biol.* 5:e171.

Tishkoff SA, Reed FA, Ranciaro A, et al. (19 co-authors). 2006. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Genet.* 39:31–40.

Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin SV. 2010. Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nat Genet.* 42:260–263.

Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. *PLoS Biol.* 4:e72.

Wachowiak W, Balk PA, Savolainen O. 2009. Search for nucleotide diversity patterns of local adaptation in dehydrins and other cold-related candidate genes in Scots pine *Pinus sylvestris* L. *Tree Genet Genomes.* 5:117–132.

Xia H, Camus-Kulandaivelu L, Stephan W, Tellier A, Zhang Z. 2010. Nucleotide diversity patterns of local adaptation at drought-related candidate genes in wild tomatoes. *Mol Ecol.* 19:4144–4154.