# Local Selection Across a Latitudinal Gradient Shapes Nucleotide Diversity in Balsam Poplar, *Populus balsamifera* L

**Stephen R. Keller,* Nicholas Levsen,†,1 Pär K. Ingvarsson,‡ Matthew S. Olson,†,1 and Peter Tiffin*,2**
*Department of Plant Biology, University of Minnesota, Saint Paul, Minnesota 55108, †Institute for Arctic Biology, University of Alaska, Fairbanks, Alaska 99775 and ‡Department of Ecology and Environmental Science, Umeå Plant Science Centre, Umeå University, SE-90187 Umeå, Sweden

**ABSTRACT** Molecular studies of adaptive evolution often focus on detecting selective sweeps driven by positive selection on a species-wide scale; however, much adaptation is local, particularly of ecologically important traits. Here, we look for evidence of range-wide and local adaptation at candidate genes for adaptive phenology in balsam poplar, *Populus balsamifera*, a widespread forest tree whose range extends across environmental gradients of photoperiod and growing season length. We examined nucleotide diversity of 27 poplar homologs of the flowering-time network—a group of genes that control plant developmental phenology through interactions with environmental cues such as photoperiod and temperature. Only one gene, *ZTL2*, showed evidence of reduced diversity and an excess of fixed replacement sites, consistent with a species-wide selective sweep. Two other genes, *LFY* and *FRI*, harbored high levels of nucleotide diversity and exhibited elevated differentiation between northern and southern accessions, suggesting local adaptation along a latitudinal gradient. Interestingly, *FRI* has also been identified as a target of local selection between northern and southern accessions of *Arabidopsis thaliana*, indicating that this gene may be commonly involved in ecological adaptation in distantly related species. Our findings suggest an important role for local selection shaping molecular diversity and reveal limitations of inferring molecular adaptation from analyses designed only to detect species-wide selective sweeps.

MOST species experience variable environments across their geographic range, and many show phenotypic variation that reflects local adaptation of ecologically important traits. However, molecular population genetic analyses of selection often focus on detecting selective sweeps driven by positive directional selection in a species-wide sample (Nielsen 2005). Only recently have studies started to scan patterns of gene nucleotide diversity for signatures of selection below the species level, particularly among human regional subpopulations (Tang *et al.* 2007; Nielsen *et al.* 2009; Pickrell *et al.* 2009; Chen *et al.* 2010).


doi: 10.1534/genetics.111.128041
Manuscript received February 22, 2011; accepted for publication May 15, 2011
Supporting information is available online at http://www.genetics.org/content/suppl/2011/05/30/genetics.111.128041.DC1.
Sequence data from this article have been deposited with the GenBank Data Libraries; accession numbers provided in supporting information, Table S1.
[1]Present address: Department of Biological Sciences, Texas Tech University, Lubbock, TX 79409.
[2]Corresponding author: Department of Plant Biology, University of Minnesota, 250 Biological Sciences Building, 1445 Gortner Ave., Saint Paul, MN 55108. E-mail: ptiffin@umn.edu

The population genetic signatures of positive selection due to species-wide sweeps are quite different from those expected under local selection, used here to mean selection favoring different allelic variants in different parts of the range, even at broadly defined spatial scales. Sweeps due to positive selection can result in a species-wide sample that shows low levels of nucleotide variation (Maynard Smith and Haigh 1974), an excess of rare or derived polymorphisms (Tajima 1989), or reduced polymorphism relative to divergence (Hudson *et al.* 1987). In contrast, local selection may result in local sweeps within subpopulations and/or reduced effective migration of selected alleles between subpopulations (Charlesworth *et al.* 1997; Charlesworth 1998), leading to increased subpopulation differentiation, elevated polymorphism, or an excess of common variants in species-wide samples (*i.e.*, positive Tajima's *D*; Nordborg and Innan 2003; Innan and Kim 2008). Empirical evidence for local selection on candidate genes of ecological significance has been found by comparing nucleotide diversity and differentiation among subpopulations found in different selective environments to empirically derived or model-based

neutral expectations (Le Corre 2005; Ingvarsson *et al.* 2006; Hancock *et al.* 2008; Moeller and Tiffin 2008; Storz and Kelly 2008; Wachowiak *et al.* 2009; Turner *et al.* 2010; Ma *et al.* 2010).

For species with geographic ranges that span seasonally variable environments, the timing of development for traits related to growth and reproduction is often subject to local selection favoring different phenotypic optima in different parts of the species' range (Bradshaw *et al.* 2004). Among forest tree species, and plants in general, this is often manifest as variation among subpopulations in the seasonal timing of growth (*i.e.*, phenology) in response to light and temperature, with strong local adaptation often observed at the phenotypic level (Savolainen *et al.* 2007).

In eudicots, phenological variation frequently maps to homologs of the *Arabidopsis thaliana* flowering-time gene network (Mouradov *et al.* 2002; Simpson and Dean 2002; Ehrenreich *et al.* 2009). This network functions via chromophore signaling of light to the circadian clock, which interprets critical daylengths and controls expression of downstream genes that integrate signals from interacting pathways to control meristem development (Jackson 2009). Comparative research across diverse taxa suggests that this gene network is responsible for variation in multiple phenological traits responsive to photoperiod, including flowering, tuberization, and seasonal dormancy (Lagercrantz 2009). Based upon transgenic and QTL mapping experiments in *Populus*, genes in this network are also associated with flowering and vegetative bud development in trees (Frewen *et al.* 2000; Chen *et al.* 2002; Böhlenius *et al.* 2006; Ingvarsson *et al.* 2008; Jackson 2009; Ma *et al.* 2010; Rohde *et al.* 2010).

*A priori*, we expect genes controlling phenology to be excellent candidates for the molecular basis of local adaptation. Surveys of natural allelic variation among *A. thaliana* accessions sampled from diverse latitudes have identified two flowering-time genes, *FRIGIDA* (*FRI*) and *FLOWERING LOCUS C* (*FLC*), that show evidence of local adaptation with latitude (Caicedo *et al.* 2004; Stinchcombe *et al.* 2004; Le Corre 2005). In other species, local selection has targeted members of the phytochrome gene family (*PHYE)* in *Cardamine japonica* (Ikeda *et al.* 2009) and *PHYB2* in *Populus tremula* (Ingvarsson *et al.* 2006; Ingvarsson *et al.* 2008) and components of the circadian clock (Böhlenius *et al.* 2006; Slotte *et al.* 2007; Brachi *et al.* 2010; Ma *et al.* 2010). These studies suggest that homologs of the *A. thaliana* flowering-time gene network underlie ecologically important adaptive variation in phenological responses in other species, making them promising candidates for studying the molecular basis of local adaptation, as well as testing the repeatability of evolution across taxa in genes that influence quantitative traits.

Balsam poplar (*Populus balsamifera* L.) is a widespread forest tree that has recolonized most of boreal North America in a postglacial range expansion during the past 10,000–18,000 years. On the basis of SNP data from 412 loci assayed from 474 trees sampled from across the species range, this expansion appears to have involved the spread of populations from a southwestern refugial population into the northwest and northeast (Keller *et al.* 2010). This spread is reflected in the current diversity forming three broadly distributed genetic clusters, currently situated in the northern, central, and eastern parts of the range. These clusters are weakly but significantly differentiated at SNP loci ($F_{CT}$ = 0.04), but show much stronger regional structure at multiple ecophysiological traits related to growth and phenology (Keller *et al.* 2011). The covariance between SNPs, ecophysiology, and environment strongly suggests that latitude or other climate-related aspects of the environment have driven adaptation in phenology traits during and since the northward expansion.

Here, we present analyses of the molecular evolution of 27 candidate genes homologous to the *A. thaliana* flowering-time network and associated pathways (hereafter referred to collectively as "phenology genes") in balsam poplar. We test for evidence of nonneutral evolution in our species-wide sample by comparing (i) nucleotide diversity and the site frequency spectrum of phenology genes to a set of 219 background reference loci and to neutral expectations based on an approximate Bayesian computation (ABC) model of demographic history (Beaumont *et al.* 2002) and (ii) diversity and divergence of phenology genes against a reduced set of reference loci and to neutral coalescent expectations. Finally, we test for evidence of local adaptation by comparing nucleotide differentiation between high- and low-latitude accessions to neutral expectations simulated under a demographic model that reflects balsam poplar's recent history of postglacial expansion.

## Methods

### *Sampling and DNA sequencing*

We isolated DNA from a range-wide sample of 24 accessions (Figure 1) and sequenced each sample for a set of 27 phenology-associated candidate genes: 24 genes from the flowering-time network and 3 abscisic acid (ABA) genes that cosegregate with QTL for dormancy phenotypes in a *P. trichocarpa* × *P. deltoides* mapping population (Frewen *et al.* 2000). The 24 sampled accessions are part of the AgCanBaP collection growing in Indian Head, Saskatchewan, Canada (for collection details, see Soolanayakanahally *et al.* 2009). We identified *P. balsamifera* homologs of *A. thaliana* phenology genes using BLASTN searches against the *P. trichocarpa* genome (Tuskan *et al.* 2006). Primers (supporting information, Table S2) were designed to amplify and sequence the exons of each gene, as these are most likely to contain causative polymorphisms for adaptive protein evolution, but for most genes some intron sequence was also captured. The majority of sequence data were obtained by direct sequencing of PCR products. For two genes (*PiF3.1* and *TOC1*), some samples were sequenced from PCR products that had been cloned into pGemT vectors. For the

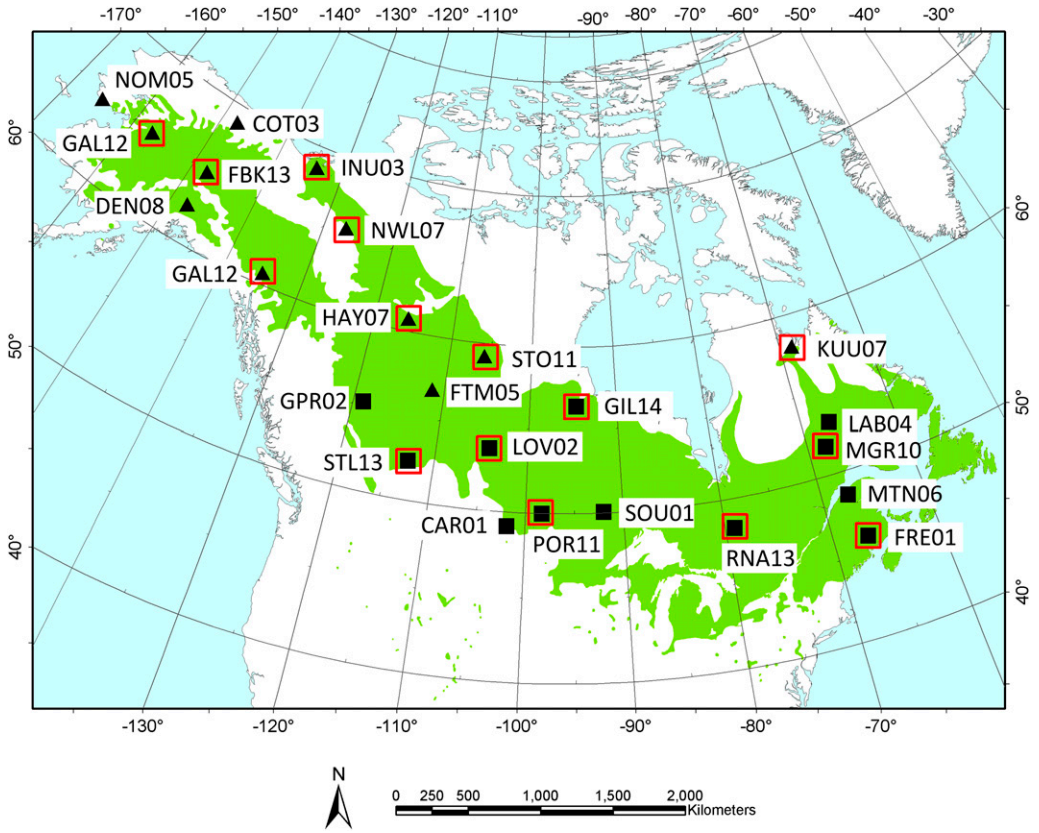


**Figure 1** Sampling locations and IDs for accessions sequenced in this study. Accessions in the northern group are symbolized by triangles, and those in the southern groups as squares. Red boxes denote those accessions included in the reference gene sequences (from Olson *et al.* 2010) used for generating the empirical distributions of diversity. The geographic range of *P. balsamifera* is shown in green.

cloned sequences, all variants present in only one sample were assumed to be due to misincorporation during amplification if they could not be verified by direct sequencing of PCR products from an independent reaction. For *TOC1*, the first four exons were exclusively from cloned sequence, while sequences from exons 5–6 were direct sequenced from PCR products. This necessitated phasing one part of the alignment and not the other (see below); therefore, we report analyses for these two regions separately.

To conduct analyses that require an outgroup, we obtained sequence data for 24 of the candidate genes as well as 16 reference loci from either previous studies (Ingvarsson *et al.* 2006; Hall *et al.* 2007; Neiman *et al.* 2009; Ma *et al.* 2010) or PCR sequences from one to four *P. tremula* individuals. For sites that were polymorphic among *P. tremula* alleles, we assumed the ancestral state was identical to the base found in the *P. balsamifera* sample.

Sequences were trimmed of poor flanking sequence, aligned, and edited manually in Aligner v. 3.0.0 (Codon Code Corporation, Dedham, MA). Sites that were heterozygous, polymorphic, or of low sequence quality were visually examined for base call quality. Insertion–deletion polymorphisms were reconstructed from forward and reverse sequences and were removed from the alignments prior to further analysis. One gene, *HY2.2*, showed an extremely divergent haplogroup suggestive of introgression from another *Populus* species (Figure S1); thus we confined analysis of *HY2.2* to a reduced set of samples that omitted the divergent haplogroup. No other genes showed evidence of introgression. Gene annotations were based on the *P. trichocarpa* genome build v. 2.0 available at Phytozome (http://www.phytozome.net/poplar.php). All new sequences have been deposited in GenBank (accession numbers in Table S1)

### *Data analysis*

Diploid sequences were disambiguated into haplotypes using Phase v. 2.1 (Stephens *et al.* 2001; Stephens and Scheet 2005). We ran 10,000 iterations of the Bayesian MCMC chain, sampling every 10 iterations after a burn-in of 100 iterations. We then used the phased haplotypes to estimate the number of segregating sites, nucleotide diversity as Watterson's $\theta_W$, and pairwise divergence ($\pi$) at replacement, synonymous, and silent (synonymous and noncoding) sites, and the site frequency spectra based on Tajima's *D* (Tajima 1989) and Fay and Wu's *H* (Fay and Wu 2000), using *P. tremula* as the outgroup for calculating *H*. Summary statistics were calculated using DNASP v. 5 (Librado and Rozas 2009).

To characterize range-wide diversity and search for evidence of selection in the species-wide sample, we compared synonymous nucleotide diversity ($\theta_W$ and $\pi$) and site frequency spectra (using Tajima's *D*) for phenology genes to both genome-wide empirical distributions of synonymous diversity and simulated neutral distributions. The empirical distributions were derived from a set of nuclear

reference loci (Olson *et al.* 2010) sequenced in 15 accessions that represent a subset of the 24 accessions sequenced for phenology genes (Figure 1). We computed distributions of summary statistics based on synonymous sites, restricting the analysis to just the 219 reference loci with ≥100 synonymous sites (mean length = 130.3 synonymous sites). Defining expected patterns of diversity from empirical distributions of reference genes has the advantage of being free from specifying a particular model of demographic history or defining *a priori* how indirect selection may affect the variance among sites across the genome (Tenaillon and Tiffin 2008; Garrigan *et al.* 2010). However, because our reference loci were considerably shorter than our phenology genes (mean ± SD = 617 ± 63 bp compared to 2535 ± 1137 bp) direct comparison of the phenology to reference loci may be confounded by sampling effects, causing the reference loci to have greater variance of summary statistics and a large number of sequences with zero segregating sites. For this reason, we also used ABC ( Beaumont *et al.* 2002) to develop a demographic model that captured the pattern of diversity found at reference loci and could subsequently be used to generate distributions of expected neutral diversity for loci that were the same average length as our phenology genes.

The ABC modeling process involved simulating data under three demographic models: a standard neutral model, an instantaneous size change model, and an exponential growth model. We refrained from testing more complicated models that include population subdivision and migration among regional groups (*e.g.*, Keller *et al.* 2010) because our initial simple models captured the major features of diversity found in the range-wide sample of 15 individuals (see *Results*). In other words, these ABC models should not be viewed as accurate descriptors of the demographic history as much as simplified models that capture the patterns of nucleotide diversity found in the reference loci from a widespread sample of a limited number of individuals. For each model, we generated $10^6$ simulated data sets of the neutral coalescent with recombination using Hudson's ms (Hudson 2002). Each simulated data set consisting of 219 loci containing 130.3 neutrally evolving (synonymous) sites, a fragment length for recombination of 617 bp, and a sample size of 28 chromosomes (values reflect averages for the reference loci). Demographic parameters were drawn at random from prior distributions. For all three models, the population mutation rate per locus ($\theta = 4N_e\mu$) had a uniform prior of $\theta \sim [0.1–4]$ and values for the population recombination rate ($\rho = 4N_e r$) were drawn from a uniform prior of $\rho \sim [0–0.25]$. For the instantaneous size change model, looking backward in time we estimated the first size change ($T_B$) going from the current population size ($N_0$) to the previous size ($N_B$), and then the second size change at time ($T_A$) to the ancestral population size ($N_A$). Prior distributions on timing and size parameters were uniform, and on the intervals of $T_B \sim [0.01–4]$, $N_B \sim [0.01–2]$, $T_A \sim [T_B + 0.01–1]$, and $N_A \sim [N_B. + 0.5–4]$. Because the instantaneous size change model allows for either increases or decreases in population size at each time this model can recreate, but is not limited to, a traditional bottleneck model. The exponential growth model followed the form $N_t = N_0^{-\alpha t}$, where $N_0$ is the current effective population size, $t$ is time in $4N_0$ generations, and $\alpha$ is the growth parameter (positive values indicate increasing population size looking forward in time). Prior values on growth were drawn from a uniform distribution of $\alpha \sim [1–20]$, and the timing of the start of growth ($T_G$) was drawn from a uniform prior of $T_G \sim [0.01–10]$.

We compared the results from the simulated data to the empirical data from our 219 reference loci using five summary statistics based on only synonymous sites: mean and standard deviation of $\pi$, mean and standard deviation of Tajima's $D$, and the proportion of loci with zero segregating sites. From the pooled set of $3 \times 10^6$ simulations, we retained the 3000 simulations (*i.e.*, 0.1% of the pooled set) that were closest to the summary statistics from our reference loci using the Euclidian rejection algorithm, msreject, distributed with the MSBayes package (Hickerson *et al.* 2007), and calculated the percentage of simulations contributed by each model as our estimate of that model's posterior probability. We then used ABCREG (Thornton 2009) with the tangent transformation to perform local regression estimation of the posterior distributions of model parameters. From these posteriors, we obtained the Bayesian most probable estimate of each parameter from the median value of the distribution. These estimates were then used in a final round of coalescent simulations where the simulated length was adjusted to match the phenology genes by increasing the per-gene $\theta$ and fragment length for recombination to reflect the longer candidate gene sequences (average of 302 synonymous sites and a total length of 2535 sites). Under these conditions, we simulated 500,000 replicates of the coalescent and obtained the shape of the distribution of neutral diversity under the best fitting demographic model. Simulated distributions under neutrality were then compared with the observed values for the phenology genes. Note that although the simulated distributions were generated using the best-fit parameters obtained from the ABC model of the reference loci, because we simulated regions that are approximately four times longer than the length of the loci used to fit the model, we no longer expect the new simulated distributions to match the reference loci. In particular, because of the longer length of the simulated sequences, we no longer expect the inflation of loci with zero polymorphism that is observed in the reference loci data. The full ms command line input for ABC model development and the final simulated distributions are provided in File S1.

In addition to examining patterns of intraspecific diversity, we looked for signatures of selection in our range-wide samples using two approaches for comparing patterns of nucleotide diversity within *P. balsamifera* to divergence from *P. tremula*. First, we used J. Hey's multilocus HKA program to evaluate whether any of the candidate genes

deviated from neutral expectations of diversity and divergence, using 10,000 coalescent simulations for significance testing. Second, we used the hierarchical Bayesian implementation of the MacDonald–Kreitman Poisson random fields model (MKPRF) to estimate the strength of selection ($\gamma = 2N_e s$) on amino acid changes (Bustamante *et al.* 2002, 2005). We defined a hierarchical model that allowed selection to act differently on the phenology genes than the reference loci. For the MKPRF, we omitted eight phenology genes that had fewer than four replacement site mutations (either polymorphic or fixed differences) because they would provide little statistical power for detecting selection (*e.g.*, Bustamante *et al.* 2005), leaving 16 phenology genes and 16 reference loci. We ran 10 independent mcmc chains, drawing 6000 samples from each following a burn-in of 1000 steps, and sampling every 10 steps. Input parameter values were set to default values except for $\rho$, the ratio of $N_e$ in species 2 relative to species 1, which we set to 5.0 on the basis of previous work that showed $\theta$ ($= 4N_e\mu$) is roughly five times larger in *P. tremula* than *P. balsamifera* (Ingvarsson 2008; Olson *et al.* 2010). For both HKA and MKPRF analyses, we used sequences from *P. tremula*, instead of *P. trichocarpa*, as an outgroup, because *P. trichocarpa* and *P. balsamifera* are recently diverged and share abundant ancestral polymorphism (Levsen *et al.*, unpublished data).

To test if nucleotide diversity was structured by local adaptation to environmental conditions that covary with latitude (*e.g.*, photoperiod, growing season length, summer and winter temperatures; Keller *et al.* 2011), we assigned each of the 24 individuals into either northern or southern subgroups, depending on whether individuals had been sampled from high (>56° N) or low (<56° N) latitude sites. The cutoff between north and south was chosen from the median latitude among our sampled individuals, rather than an *a priori* expectation of a threshold latitude effect. We then tested if the empirical differentiation between subgroups was greater than expected in the absence of selection, using an index of absolute nucleotide differentiation, $\pi_{T\text{-}S} = \pi_T - \pi_S$, in which $\pi_T$ is the total nucleotide diversity across the pooled sample and $\pi_S$ is the average nucleotide diversity within each of the two latitudinal groups (Charlesworth 1998). Because we lacked the sampling to explicitly model a demographic scenario of latitudinal divergence among three regional clusters (Keller *et al.* 2010), we simulated the expected distribution of $\pi_{T\text{-}S}$ under a neutral model that captured the major biological features of balsam poplar's northward expansion from a southern refugial population. Our coalescent model consisted of a single ancestral population of $N_0 = 20{,}000$ that split into two equal sized subpopulations ($N_1 = N_2 = 10{,}000$) 650 generations before present. Assuming a 15-year generation time (Ingvarsson 2008), this corresponds to ~10,000 years ago, when a large proportion of the glaciated area of *P. balsamifera*'s current range started to become ice free and available for colonization. For three reasons our demographic model should produce conservative estimates of differentiation for testing local adaptation: there was no migration between subpopulations (allowing that migration narrowed the confidence intervals on $\pi_{T\text{-}S}$; results not shown), unaccounted-for population substructure will elevate $\pi_S$, thereby lowering the empirical estimates of $\pi_{T\text{-}S}$, and finally the northern subgroup we used for our empirical data contained individuals from all three regional genetic clusters (Keller *et al.* 2010), which should increase empirical estimates of $\pi_S$ and lower $\pi_{T\text{-}S}$ over expectations from a simple north–south demographic split. We simulated 50,000 replicates of the coalescent with Hudson's MS, conditioned on the total number of segregating sites for each phenology gene, and compared the observed $\pi_{T\text{-}S}$ to the 95% confidence interval under neutrality. In addition, we scanned for regions of elevated differentiation within genes by conducting sliding window analyses of $\pi_{T\text{-}S}$ using 300-bp windows with 100-bp step increment.

## Results

### *Nucleotide diversity at phenology genes vs. expectations from demographic history*

We sequenced 876–5434 bp from each of the 27 phenology candidate genes, resulting in ~71 kb of sequence from each of 24 individuals, sampled across the range of *P. balsamifera* (Figure 1). Nucleotide diversities based on segregating sites and pairwise divergence were highly variable among phenology genes ($\theta_W = 0.000 - 0.0146$; $\pi = 0.000 - 0.0186$; Figure 2 and Table S3) and most phenology genes had synonymous site $\theta_W$, $\pi$, and $D$ that were well within the range of values from the genome-wide empirical distributions of reference loci (Figure 3 and File S2). However, two phenology genes, *LFY* and *FRI*, had high pairwise diversity ($\pi_{syn} = 0.0186$ and $\pi_{syn} = 0.0114$, respectively, Figure 3A) within the upper 2% and 7% tail of the empirical distribution. These two genes also harbored more segregating sites ($\theta_{syn}$) than other candidates and were in the upper tail of the empirical distribution (*LFY* $\theta_{syn} = 0.0146$, upper 4%; *FRI* $\theta_{syn} = 0.0088$, upper 17.5% tail; Figure 3B). In contrast, four phenology genes had zero synonymous polymorphism (*ABI1D*, *HY2.2*, *TFL1.1*, and *TOC1*). Because a large number of reference loci lacked polymorphism, these were not outliers in the empirical distributions (Figure 3). Only one phenology gene, *ZTL2*, exhibited an excess of rare variants ($D_{syn} = -1.72$) relative to reference loci (lower 3% tail).

Because reference loci sequences were considerably shorter than our phenology genes, the empirical distribution was skewed toward values of zero diversity (Figure 3). For this reason, we generated distributions of expected neutral diversity using an ABC-derived demographic model. After rejecting all but 0.1% of the coalescent simulations that most closely matched our empirical data, 57.2% of our ABC simulations were from a sequential size change model, 42.2% from a standard neutral model, and 0.6% from a growth model. While the growth model was clearly rejected, the standard neutral model with a current effective population size of $N_e = 32{,}000$ was almost as good a fit to

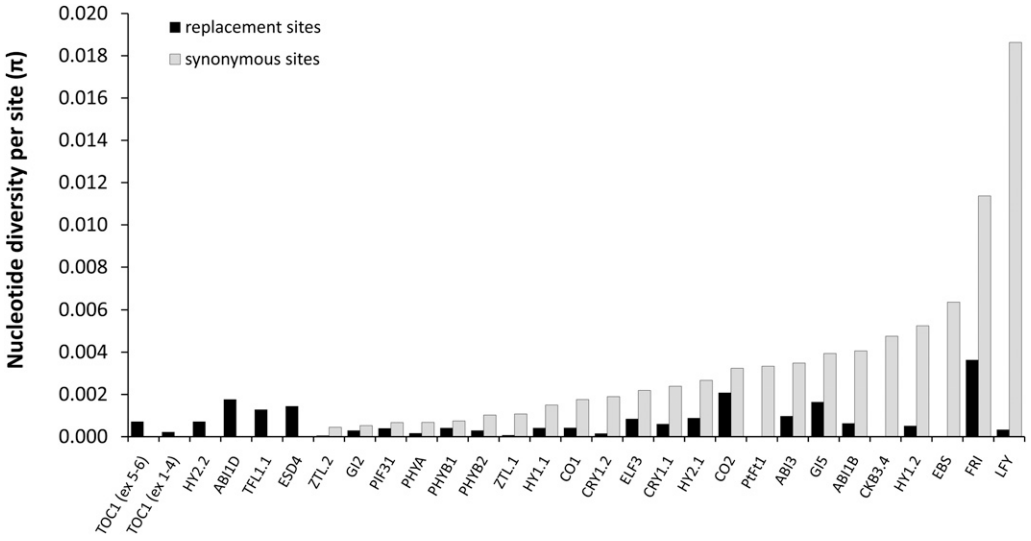


**Figure 2** Phenology gene nucleotide diversity from pairwise divergence ($\pi$) estimated for replacement and synonymous sites. Genes are ranked from lowest to highest on the basis of $\pi_{syn}$.

the reference loci as the size change model, which described a large ancestral population of $N_e$ = 148,000, bottlenecked ~1.7 MYA to an $N_e$ = 36,600, and then reduced in size again ~639 KYA to an $N_e$ = 32,200 (Table 1). Because the timing of the oldest bottleneck far exceeds the expected coalescent time within *P. balsamifera*, as well as the estimated divergence time between *P. balsamifera* and *P. trichocarpa* (Levsen *et al.*, unpublished data), this model likely is affected by the recent speciation of *P. trichocarpa* and *P. balsamifera* and possibly introgression of alleles from more distantly related *Populus* species.

Distributions of nucleotide diversity from ABC demographic models were still skewed toward low values, similar to the empirical distributions, but there was no longer a peak at zero diversity, reflecting that the simulated sequences were considerably longer than the reference loci. Moreover, *LFY* and *FRI* remained outliers in the simulated distributions; values of synonymous nucleotide diversity at *LFY* and *FRI* were in the upper 0.5% and 5% tails of $\pi_{syn}$, respectively, and in the upper 1.5% and 11.5% tails of $\theta_{syn}$ (Figure 3, A and B). The four phenology genes with zero synonymous polymorphism (*ABI1D*, *HY2.2*, *TFL1.1*, and *TOC1*) fell within the lower 1% of simulated values of $\pi_{syn}$ and $\theta_{syn}$. Similar to the empirical distribution, only *ZTL2* showed significantly negative Tajima's *D*, falling in the lower 2.5% tail of simulated values (Figure 3C).

### *Tests of selection based on comparisons of diversity and divergence*

The average silent site divergence of phenology genes ($K_S$) was 0.033, and the ratio of replacement to silent site divergence ($K_A/K_S$) was 0.29 (Table S4). Four phenology genes (*CRY1.1*, *EBS*, *ELF3*, and *ZTL2*) showed strongly negative values of *H* (all values $P < 0.05$) indicating an excess of high frequency derived sites consistent with positive selection (Table S4). HKA tests comparing diversity within *P. balsamifera* to divergence from *P. tremula* revealed significant heterogeneity among phenology genes ($\chi^2$ = 67.49, d.f. = 23, $P < 0.001$), with the largest deviances from neutrality driven by an excess of polymorphism in *LFY* and *FRI*, and a deficit of polymorphism in *PHYB2*, relative to divergence from *P. tremula* (Figure 4 and Table S4).

Estimates of the scaled selection coefficient on replacement sites $\gamma$ ($2N_e s$) revealed similar patterns of replacement site polymorphism relative to divergence at phenology and reference genes (Figure 5). Two phenology genes showed evidence of nonneutral evolution at replacement sites. *FRI* had a significantly negative selection coefficient ($\gamma = -0.626$, $P < 0.001$), indicating an abundance of replacement site polymorphism relative to neutral expectations, likely caused by either balancing/local selection maintaining polymorphism or segregating weakly deleterious mutations. The other gene, *ZTL2*, showed an excess of fixed replacement sites between *P. balsamifera* and *P. tremula* ($\gamma$ = 0.533, $P$ = 0.045) consistent with positive selection.

### *Local selection and latitudinal adaptation*

The elevated range-wide diversity at *LFY* and *FRI* suggests the possibility of geographically variable local selection. To test this, we compared differentiation between high and low latitude accessions ($\pi_{T\text{-}S}$) for the phenology genes relative to expectations from a neutral model of range expansion following Pleistocene glacial retreat. This demographic model produced null distributions that frequently overlapped empirical estimates of $\pi_{T\text{-}S}$ from phenology genes. At the whole-gene level, only *FRI* showed $\pi_{T\text{-}S}$ in excess of the demographic model, falling in the upper 1% of neutral values (Figure 6).

Sliding window analyses of $\pi_{T\text{-}S}$ identified windows of elevated $\pi_{T\text{-}S}$ within four phenology genes. These included 13 windows in *FRI* (upper 0.2–2% of simulated values) which also showed excess north–south differentiation in the whole-gene analysis, as well as *LFY* (three windows, upper 0.2–2%), *EBS* (two windows, upper 2%), and *ZTL1* (two windows, upper 1–2%) (Figure 7 and Figure S2). Due to the large number of comparisons, some of these windows

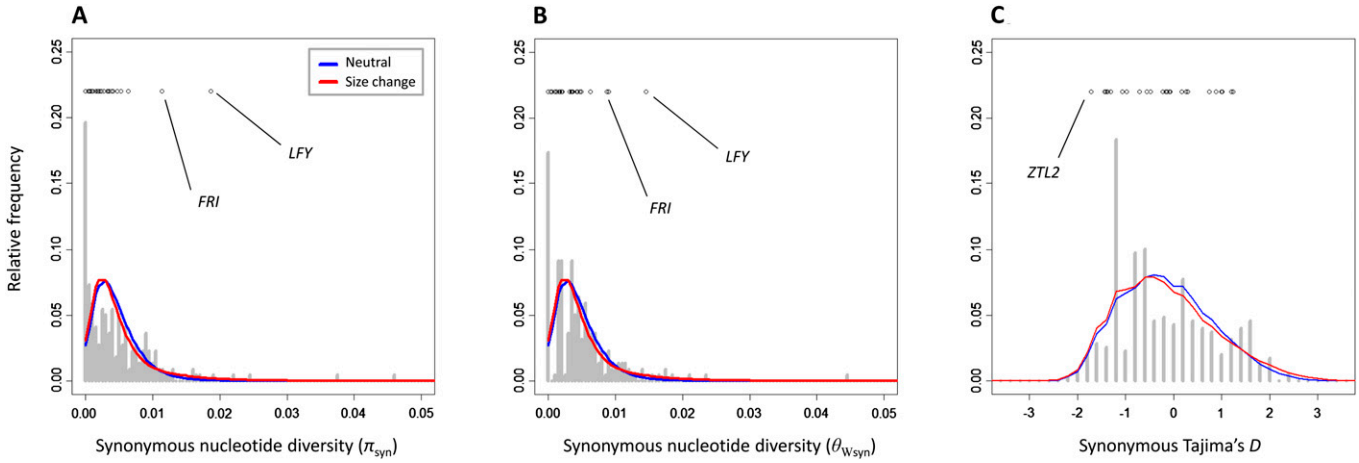


**Figure 3** Comparison of nucleotide diversity (A and B) and site frequency spectrum (C) at synonymous sites for phenology genes (points), background reference loci (shaded bars), and coalescent simulations from approximate Bayesian computation (ABC) demographic models (colored lines). The two ABC models with high posterior probability are shown: a standard neutral model (blue) and a size change model (red).

may be significant by chance alone. We reiterate, however, that our tests of differentiation should be conservative since our demographic null model simulated drift between subgroups with no migration and no population structure within subgroups. Many windows of elevated differentiation were composed predominately, or at times exclusively, of polymorphisms at synonymous or intron sites that are typically assumed to be silent with regard to selection. For example, in *FRI*, latitudinal differentiation was most pronounced across the first intron (Figure 7). Similarly, in *EBS* and *ZTL1*, latitudinal differentiation was exclusively at synonymous SNPs in the regions sequenced.

## Discussion

### *Molecular adaptation of phenology network genes*

Among the 27 candidate genes for adaptive phenology, *FRI* was repeatedly identified as a potential target of selection. In particular, *FRI* harbored high intraspecific nucleotide diversity, was among the upper tail of diversity values from the empirical genome-wide and simulated ABC distributions, and showed elevated nucleotide differentiation between northern and southern accessions. The functional role of *FRI* in temperature-induced phenological responses makes this gene a particularly strong candidate for local adaptation to regional differences in growing season length. Among natural accessions of *A. thaliana*, *FRI* is a major regulator of flowering-time variation by controlling expression of *FLC* to produce a vernalization requirement in high latitude accessions, permitting flowering only after a prolonged exposure to cold temperatures (Johanson *et al.* 2000; Mouradov *et al.* 2002; Putterill *et al.* 2004). In balsam poplar, our analyses implicate *FRI* in the response to local selection at different latitudes, probably reflecting a role of *FRI* in the temperature-sensitive timing of seasonal development. Because flowering-time homologs in *Populus* can affect a range of developmental traits, including flowering and bud dormancy (Böhlenius *et al.* 2006; Ingvarsson *et al.* 2008; Lagercrantz 2009; Ma *et al.* 2010), the link between allelic variation in *FRI* and variation in specific phenotypes in poplar awaits further study.

Interestingly, perhaps the strongest evidence for local selection in *FRI*, excess north–south differentiation, is largely due to polymorphism at synonymous and intron sites within our sequenced region, and these are separated from the

**Table 1 Approximate Bayesian computation (ABC) analysis of the two best fitting models of species-wide demographic history in *P. balsamifera***

| | Standard neutral model | Bottleneck model |
|---|---|---|
| % of retained simulations | 42.2 | 57.2 |
| Theta ($\pi_s$) $\times 10^{-3}$ | 4.79 (4.34–5.30) | 4.83 (4.10–6.12) |
| Recombination ($\rho_s$) $\times 10^{3-}$ | 0.66 (0.028–1.86) | 1.11 (0.078–1.89) |
| Current population size | 32,000 (28,900–35,300) | 32,200 (27,300–40,800) |
| Time to size change[a] | – | 639,000 (503,000–3,320,000) |
| Size change population size | – | 36,600 (7,100–50,500) |
| Time to ancestral population[1] | – | 1,740,000 (637,000–4,490,000) |
| Ancestral population size | – | 148,000 (76,700–194,000) |

A third model of exponential growth provided a poor fit to the observed data (0.6% of retained simulations), and so was not included for parameter estimation.

[a] Time estimates were converted from coalescent units (= $4N_0$ generations) to absolute time (in years), assuming a neutral mutation rate per site per year of $2.5 \times 10^{-9}$ (Tuskan *et al.* 2006).

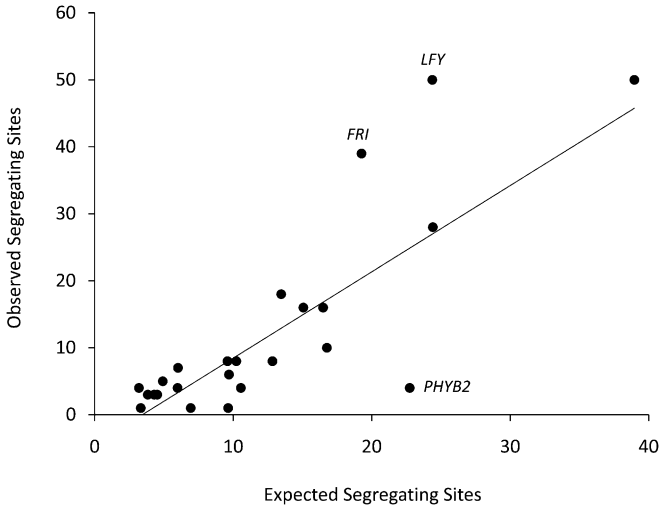


**Figure 4** HKA test of genetic hitchhiking. Values are observed numbers of segregating sites at each gene *vs.* expected values based on 10,000 neutral coalescent simulations in J. Hey's HKA software. The line is the best fit excluding genes showing evidence of selection (labeled).

nearest nonsynonymous polymorphism by windows of non-significant differentiation. It is possible that some of the silent site variation in *FRI* results in alternate splice variants, changes to gene regulatory regions, or RNA processing. In this context, it is relevant to note that much of the natural phenotypic variation in *Arabidopsis* flowering time is associated with loss of function mutants that result in premature truncation of the *FRI* protein, rather than nonsynonymous polymorphisms (Johanson *et al.* 2000; LeCorre 2002).

Three other genes, *LFY*, *ZTL1*, and *EBS*, also showed evidence of local selection. *LFY*, a transcription factor that is part of the integrative pathway downstream from *FLC* and *FRI* that is responsible for floral meristem and organ development (Moyroud and Tichtinsky 2009), harbored an excess of intraspecific diversity. Regions of exon 3 in *LFY* also showed greater than expected latitudinal differentiation. Multiple functional studies have shown that *LFY* affects flowering and other developmental phenotypes, including in *Populus* (Rottman *et al.* 2000). In contrast to the very high levels of segregating variation in *Populus LFY*, *Arabidopsis LFY* harbors low diversity consistent with a recent selective sweep (Olsen *et al.* 2002). The protease-encoding *EARLY BOLTING IN SHORT DAYS* (*EBS*) that regulates the transition from juvenile to reproductive maturity in *Populus* through transcriptional control of *FT* (Böhlenius *et al.* 2006) also showed an excess of north–south differentiation.

Finally, both *ZEITLUPE* paralogs (*ZTL1* and *ZTL2*) showed evidence of selection across multiple analyses. *ZTL1* showed a combination of low diversity ($\theta_{syn}$ = 0.00057; <5% of values from ABC simulations), and evidence of a balanced polymorphism ($D_{syn}$ = 1.24; >85% of values from ABC simulations). This was due to a single synonymous SNP at the end of exon 2 segregating at high frequency, which also showed differentiation between northern and southern accessions (Figure 7). This could reflect genetic hitchhiking during local selective sweeps, or drift during expansion (Excoffier and Ray 2008). Diversity at *ZTL2* was consistent with a recent selective sweep – including an excess of rare polymorphisms and derived variants (negative Tajima's *D* and Fay and Wu's *H*) and evidence of positive selection on replacement sites (positive γ from MKPRF). *ZEITLUPE* genes are hypothesized to be blue light photoreceptors, known to interact functionally with the circadian clock genes *GIGANTEA* (*GI*) and *TIMING OF CAB 1* (*TOC1*) (Somers *et al.* 2000; Kim *et al.* 2007).

Several recent studies have investigated evidence for adaptation in plant genomes (Gossman *et al.* 2010), including

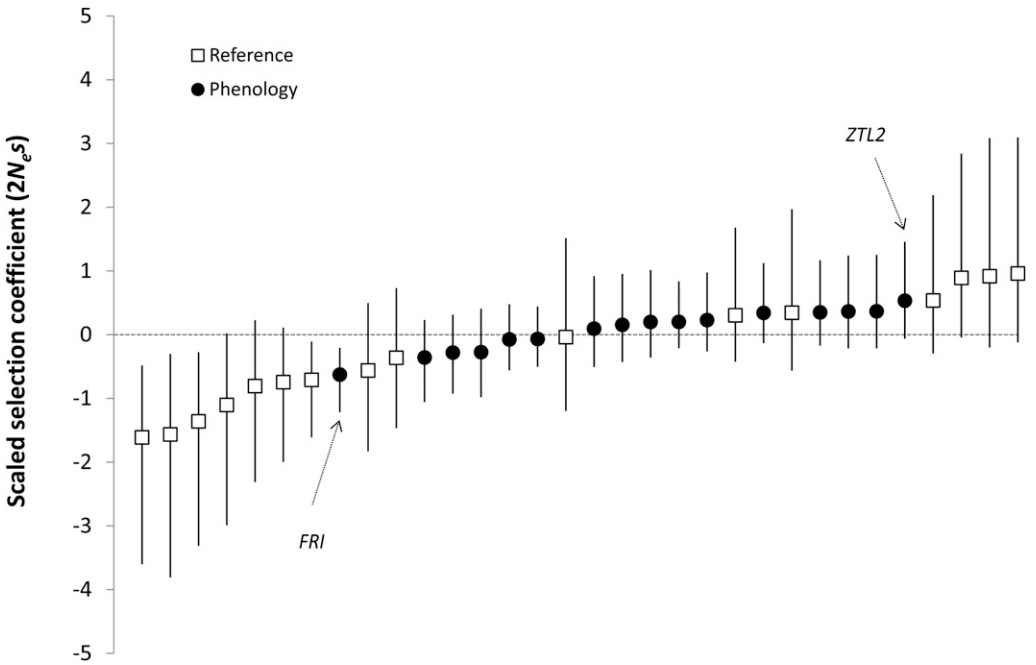


**Figure 5** Strength of selection ($\gamma = 2N_es$) on replacement site variation. Error bars are 95% credible intervals from a hierarchical Bayesian MKPRF analysis. Genes showing significant deviation from zero are labeled with arrows.

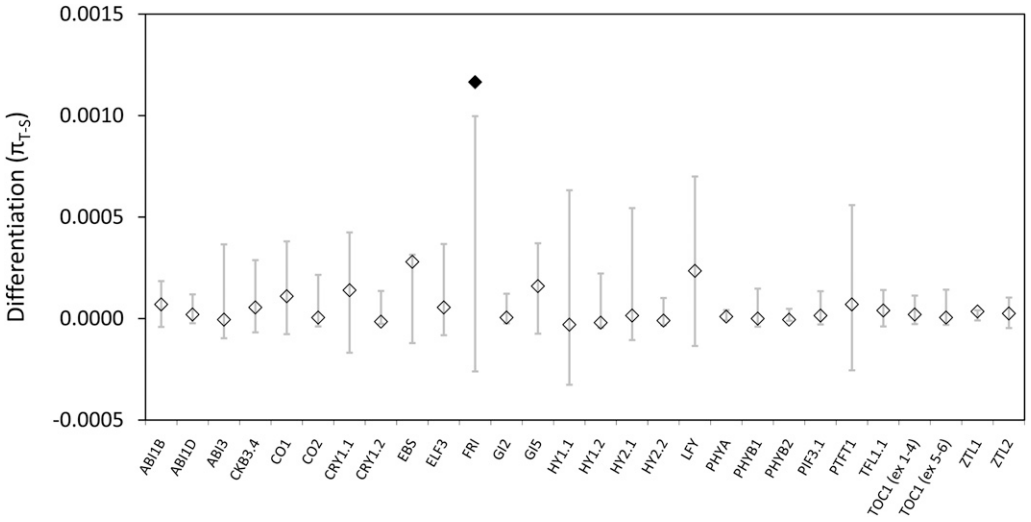


**Figure 6** Coalescent analysis of nucleotide differentiation between high and low latitude samples. Shaded bars denote 95% confidence intervals around the simulated distribution from the neutral demographic model. Diamonds are observed values of differentiation, with values falling outside of the simulated distributions depicted by solid diamonds.

specifically on the flowering-time gene network (Flowers *et al.* 2009, Hall *et al.* 2011), and have found scant evidence for adaptive evolution (but see Ingvarsson 2010 and Hall *et al.* 2011 for *Populus*, and Slotte *et al.* 2010 for *Capsella*). However, the generally employed sampling schemes and analytical methods in these and similar studies are designed to detect positive selection on a species-wide scale, and less attention has been paid to scanning sequence data for signatures of local selection within a species (Tenaillon and Tiffin 2008; Siol *et al.* 2010). In our species-wide tests, we found only weak evidence of selective sweeps in *P. balsamifera*. This is in contrast to *P. tremula*, which shows evidence of positive selection at a large fraction of substitutions (Ingvarsson 2010). Thus, had we confined our analyses in *P. balsamifera* to just the species-wide scale, there would have been little to suggest adaptation in phenology genes, despite their prominent role in controlling important life-history traits. Theoretical studies of geographically variable selection on quantitative traits indicate that local adaptation may be difficult to detect from QTL or DNA sequence data, unless adaptation involves genes of major effect (Latta 2003; Kelly 2006). Our findings of local selection in *Populus*, along with several recent studies in other plant species, contribute empirical evidence that geographically variable selection can be detected at the level of DNA sequence (LeCorre 2002; Ingvarsson *et al.* 2006; Moeller and Tiffin 2008; Wachowiak *et al.* 2009; Ma *et al.* 2010; Turner *et al.* 2010), and may constitute a major cause of adaptive evolution in plant genes that underlie ecologically important traits. Local selection on poplar phenology genes is also consistent with previous results showing elevated phenotypic differentiation along a latitudinal gradient for traits such as bud set and height growth (Keller *et al.* 2011).

### *Common targets of selection on plant phenology genes*

In addition to our study, the molecular population genetics of the phenology network have been investigated in *P. tremula* (25 genes; Hall *et al.* 2011) and *A. thaliana* (52 genes; Flowers *et al.* 2009), and many of the individual genes involved in developmental phenology have been investigated in related species in the Brassicaceae. These data provide an opportunity to investigate the repeatability of ecological adaptation at the molecular level (Stern and Orgogozo 2009), and in a gene network that is known to control important life-history phenotypes (Mouradov *et al.* 2002; Putterill *et al.* 2004; Lagercrantz 2009). In both *P. balsamifera* and *A. thaliana*, the strongest candidate for local adaptation to conditions that covary with latitiude is *FRI*, which harbors high polymorphism as well as elevated subpopulation differentiation between northern and southern accessions in both species (LeCorre 2002, 2005; Caicedo *et al.* 2004; Stinchcombe *et al.* 2004; this study). *FRI* alleles also display an excess of long-range haplotypes in *A. thaliana*, consistent with a recent history of selection (Flowers *et al.* 2009). Population genetic analysis of *FRI* has not yet been done in *P. tremula*. Two aspects of *FRI* may explain why it has been the target of independent local selection in these phylogenetically distant species. First, *FRI* shows relatively low "connectedness" with other genes in the network, especially compared to core circardian clock genes that participate in many epistatic interactions, which may lessen the amount of constraint of *FRI*'s rate of evolution (Rausher *et al.* 1999; Yukelivich *et al.* 2008; Stern and Orgogozo 2009). Studies have shown that *FRI* interacts directly with *FLC* to affect flowering in *Arabidopsis*, but *FRI* does not appear to interact with any other core components of the network. Second, at least in *A. thaliana*, loss-of-function mutations hasten development and produce earlier flowering and potentially provide adaptive phenotypic variation (Johanson *et al.* 2000; Scarcelli and Kover 2009). Given that loss-of function mutations are much more common than gain-of function, this may result in an evolutionary bias toward ecologically relevant allelic variation in *FRI* underlying adaptation to local variation in growing season. At this stage, it is not possible to identify which trait(s) mediate local selection in *FRI*, as multiple different developmental phenotypes are probably affected by the

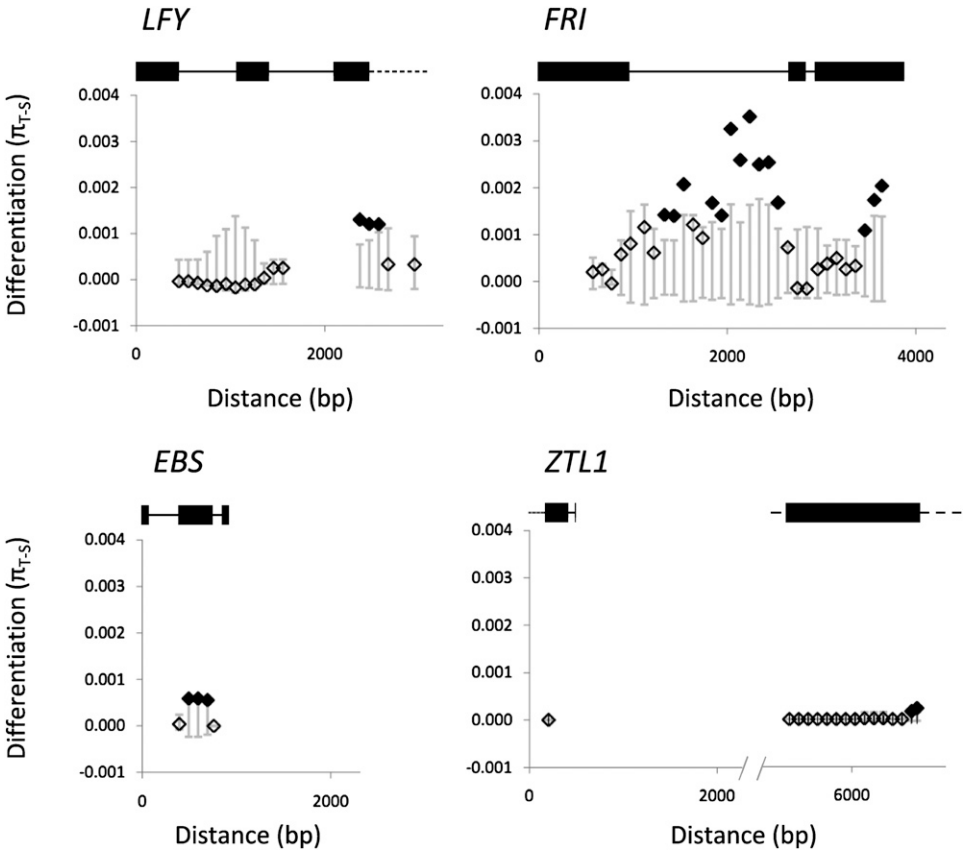


**Figure 7** Sliding window analysis of latitudinal differentiation within phenology candidate genes. At the top of each plot is shown the gene boundaries, including exons (rectangles), introns (thin solid lines), and UTR. Note that the current gene model from v. 2.0 of the *P. trichocarpa* genome divides *ZTL1* into two separate coding regions separated by intergenic space. See Figure 6 legend for additional details.

outputs from this network, including, but not limited to, flowering time and bud traits.

Members of the phytochrome family of light-signaling genes also appear to be repeat targets of local selection in some taxa, although they do not appear to be under local selection in *P. balsamifera*. In *P. tremula*, *PHYB2* shows excess polymorphism as well as strong latitudinal clines (Ingvarsson *et al.* 2006; Hall *et al.* 2007; Ma *et al.* 2010; Hall *et al.* 2011) and in *Cardamine nipponica*, a close relative of *Arabidopsis*, *PHYE* shows elevated replacement site divergence between northern and southern populations, and a significant excess of intermediate frequency polymorphism in species-wide samples (*i.e.*, positive Tajima's *D*) (Ikeda *et al.* 2009). In *P. tremula*, a role in local adaptation is bolstered by an association between *PHYB2* alleles and adaptive variation in bud set—a key phenotypic trait that controls local adaptation to photoperiod (Ingvarsson *et al.* 2008; Ma *et al.* 2010). In contrast, the phytochromes in *P. balsamifera* (this study) show low nucleotide diversity and little evidence of positive selection (Table S4 and Figure 2), nor do phytochrome genes cosegregate with bud set QTL in *Populus trichocarpa* × *P. deltoides* mapping populations (Chen *et al.* 2002; Rohde *et al.* 2010). Although congenerics, *P. balsamifera* and *P. tremula*, are geographically isolated and estimated to have last shared a common ancestor ~40 million years ago (Hamzeh and Dayanandan 2004); thus they have presumably migrated and adapted to northern latitude environments independently. The difference in selection on phenology genes suggests that despite showing convergent phenotypic clines in bud set with latitude, these two poplar species differ in the genetic basis of local adaptation.

In summary, we found evidence for local selection driving patterns of polymorphism and latitudinal differentiation in at least one and up to four members of the phenology gene network of *Populus balsamifera*, while evidence for species-wide selection was confined to a single gene. The finding of local selection on the phenology gene network of both balsam poplar and European aspen, along with the apparently different genic targets of selection in each species, suggests that *Populus* may be a promising system for understanding the genetic basis of local adaptation, and how different forms of selection, along with the historical contingencies of population demographic history, combine to affect the evolution of genes underlying ecologically important phenotypes.

## Acknowledgments

We thank Bill Schroeder for kindly sharing tissue samples for DNA extraction, Jennifer Reese and Riya Jayachandran for assistance with sequencing, Amanda Robertson and Naoki Takebayashi for discussions and programming advice, and computational support from the Life Science Informatics cluster at the University of Alaska, Fairbanks, and the Minnesota Supercomputing Institute at the University of

Minnesota. We also appreciate the comments of Outi Savolainen and two anonymous reviewers. This work was funded by National Science Foundation Plant Genome award DBI-0701911 to M.S.O. and P.T., and the Swedish Research Council to P.K.I.

## Literature Cited

Beaumont, M. A., W. Y. Zhang, and D. J. Balding, 2002 Approximate Bayesian computation in population genetics. Genetics 162: 2025–2035.

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

Brachi, B., N. Faure, M. Horton, E. Flahauw, A. Vazquez *et al.*, 2010 Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. PLoS Genet. 6: e1000940.

Bradshaw, W. E., P. A. Zani, and C. M. Holzapfel, 2004 Adaptation to temperate climates. Evolution 58: 1748–1762.

Bustamante, C. D., R. Nielsen, S. A. Sawyer, K. M. Olsen, M. D. Purugganan *et al.*, 2002 The cost of inbreeding in *Arabidopsis*. Nature 416: 531534.

Bustamante, C. D., A. Fledel-Alon, S. Williamson, R. Nielsen, M. T. Hubisz *et al.*, 2005 Natural selection on protein-coding genes in the human genome. Nature 437: 1153–1157.

Caicedo, A. L., J. R. Stinchcombe, K. M. Olsen, J. Schmitt, and M. D. Purugganan, 2004 Epistatic interaction between Arabidopsis *FRI* and *FLC* flowering time genes generates a latitudinal cline in a life history trait. Proc. Natl. Acad. Sci. USA 101: 15670–15675.

Charlesworth, B., 1998 Measures of divergence between populations and the effect of forces that reduce variability. Mol. Biol. Evol. 15: 538–543.

Charlesworth, B., M. Nordborg, and D. Charlesworth, 1997 The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genet. Res. 70: 155–174.

Chen, T. H. H., G. T. Howe, and H. D. Bradshaw, 2002 Molecular genetic analysis of dormancy-related traits in poplars. Weed Sci. 50: 232–240.

Chen, H., N. Patterson, and D. Reich, 2010 Population differentiation as a test for selective sweeps. Genome Res. 20: 393–402.

Ehrenreich, I. M., Y. Hanzawa, L. Chou, J. L. Roe, P. X. Kover *et al.*, 2009 Candidate gene association mapping of Arabidopsis flowering time. Genetics 183: 325–335.

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

Fay, J. C., and C.-I. Wu, 2000 Hitchhiking under positive Darwinian selection. Genetics 155: 1405–1413.

Flowers, J. M., Y. Hanzawa, M. C. Hall, R. C. Moore, and M. D. Purugganan, 2009 Population genomics of the *Arabidopsis thaliana* flowering time gene network. Mol. Biol. Evol. 26: 2475–2486.

Frewen, B. E., T. H. H. Chen, G. T. Howe, J. Davis, A. Rohde *et al.*, 2000 Quantitative trait loci and candidate gene mapping of bud set and bud flush in Populus. Genetics 154: 837–845.

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

Gossman, T. I., B.-H. Song, A. J. Windsor, T. Mitchell-Olds, C. I. Dixon *et al.*, 2010 Genome wide analyses reveal little evidence for adaptive evolution in many plant species. Mol. Biol. Evol. 27: 1822–1832.

Hall, D., V. Luquez, M. V. Garcia, K. R. St Onge, S. Jansson *et al.*, 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., X. F. Ma, and P. K. Ingvarsson, 2011 Adaptive evolution of the Populus tremula photoperiod pathway. Mol. Ecol. 20: 1463–1474.

Hamzeh, M., and S. Dayanandan, 2004 Phylogeny of Populus (Salicaceae) based on nucleotide sequences of chloroplast *trnT-trnF* region and nuclear rDNA. Am. J. Bot. 91: 1398–1408.

Hancock, A. M., D. B. Witonsky, A. S. Gordon, G. Eshel, J. K. Pritchard *et al.*, 2008 Adaptations to climate in candidate genes for common metabolic disorders. PLoS Genet. 4: e32 10.1371/journal.pgen.0040032.

Hickerson, M. J., E. Stahl, and N. Takebayashi, 2007 msBayes: a flexible pipeline for comparative phylogeographic inference using approximate Bayesian computation (ABC). BMC Bioinformatics 8: 268.

Hudson, R. R., 2002 Generating samples under a Wright–Fisher neutral model. Bioinformatics 18: 337–338.

Hudson, R. R., M. Kreitman, and M. Aguade, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics 116: 153–159.

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

Ingvarsson, P. K., 2008 Multilocus patterns of nucleotide polymorphism and the demographic history of *Populus tremula*. Genetics 180: 329–340.

Ingvarsson, P. K., 2010 Natural selection on synonymous and nonsynonymous mutations shapes patterns of polymorphism in *Populus tremula*. Mol. Biol. Evol. 27: 650–660.

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

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

Innan, H., and Y. Kim, 2008 Detecting local adaptation using the joint sampling of polymorphism data in the parental and derived populations. Genetics 179: 1713–1720.

Jackson, S. D., 2009 Plant responses to photoperiod. New Phytol. 181: 517–531.

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

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

Keller, S. R., R. Y. Soolanayakanahally, R. D. Guy, S. N. Silim, M. S. Olson *et al.*, 2011 Climate-driven local adaptation of ecophysiology and phenology in balsam poplar, *Populus balsamifera* L. (Salicaceae). Am. J. Bot. (in press).

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

Kim, W.-Y., S. Fujiwara, S.-S. Suh, J. Kim, Y. Kim *et al.*, 2007 *ZEITLUPE* is a circadian photoreceptor stabilized by *GIGANTEA* in blue light. Nature 449: 356–360.

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, R., 2003 Gene flow, adaptive population divergence and comparative population structure across loci. New Phytol. 161: 51–58.

Le Corre, V., 2002 DNA polymorphism at the *FRIGIDA* gene in *Arabidopsis thaliana*: extensive nonsynonymous variation is consistent with local selection for flowering time. Mol. Biol. Evol. 19: 1261–1271.

Le Corre, V., 2005 Variation at two flowering time genes within and among populations of *Arabidopsis thaliana*: comparison with markers and traits. Mol. Ecol. 14: 4181–4192.

Librado, P., and J. Rozas, 2009 DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451–1452.

Ma, X. M., D. Hall, K. R. St. Onge, S. Jansson, and P. Ingvarsson, 2010 Genetic differentiation, clinal variation and phenotypic associations with growth cessation across the *Populus tremula* photoperiodic pathway. Genetics 10.1534/genetics.110.120873.

Maynard Smith, J., and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. Camb. 23: 23–35.

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

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

Moyroud, E., and G. Tichtinsky, 2009 The *LEAFY* floral regulators in Angiosperms: conserved proteins with diverse roles. J. Plant Biol. 52: 177–185.

Neiman, M., M. S. Olson, and P. Tiffin, 2009 Selective histories of poplar protease inhibitors: elevated polymorphism, purifying selection, and positive selection driving divergence of recent duplicates. New Phytol. 183: 740–750.

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

Nielsen, R., M. J. Hubisz, I. Hellmann, D. Torgerson, A. Am. Andrés *et al.*, 2009 Darwinian forces affecting human protein coding genes. Genome Res. 19: 838–849.

Nordborg, M., and H. Innan, 2003 The genealogy of sequences containing multiple sites subject to strong selection in a subdivided population. Genetics 163: 1201–1213.

Olsen, K. M., A. Womack, A. R. Garrett, J. I. Suddith, and M. D. Purugganan, 2002 Contrasting evolutionary forces in the *Arabidopsis thaliana* floral developmental pathway. Genetics 160: 1641–1650.

Olson, M. S., A. L. Robsertson, N. Takebayashi, S. Silim, W. R. Schroeder *et al.*, 2010 Nucleotide diversity and linkage disequilibrium in balsam poplar *Populus balsamifera*. New Phytol. 186: 526–536.

Pickrell, J. K., G. Coop, J. Novembre, S. Kudaravalli, J. Z. Li *et al.*, 2009 Signals of recent positive selection in a worldwide sample of human populations. Genome Res. 19: 826–837.

Putterill, J., R. Laurie, and R. Macknight, 2004 It's time to flower: the genetic control of flowering time. BioEssays 26: 363–373.

Rausher, M. D., R. E. Miller, and P. Tiffin, 1999 Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. Mol. Biol. Evol. 16: 266–274.

Rohde, A., V. Storme, V. Jorge, M. Gaudet, N. Vitacolonna *et al.*, 2011 Bud set in poplar: genetic dissection of a complex trait in natural and hybrid populations. New Phytol. 189: 106–121.

Rottman, W. H., R. Meilan, L. A. Sheppard, A. M. Brunner, J. S. Skinner *et al.*, 2000 Diverse effects of overexpression of *LEAFY* and *PTLF*, a poplar (*Populus*) homolog of *LEAFY/FLORICAULA*, in transgenic poplar and Arabidopsis. Plant J. 22: 235–245.

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

Scarcelli, N., and P. X. Kover, 2009 Standing genetic variation in *FRIGIDA* mediates experimental evolution in flowering time in Arabidopsis. Mol. Ecol. 18: 2039–2049.

Simpson, G. G., and C. Dean, 2002 Arabidopsis, the Rosetta Stone of flowering time? Science 296: 285–289.

Siol, M., S. I. Wright, and S. C. H. Barrett, 2010 The population genomics of plant adaptation. New Phytol. 188: 313–332.

Slotte, T., K. Holm, L. M. McIntyre, U. Lagercrantz, and M. Lascoux, 2007 Differential expression of genes important for adaptation in *Capsella bursa-pastoris* (Brassicaceae). Plant Physiol. 145: 160–173.

Slotte, T., J. P. Foxe, K. M. Hazzouri, and S. I. Wright, 2010 Genome-wide evidence for efficient positive and purifying selection in *Capsella grandiflora*, a plant species with a large effective population size. Mol. Biol. Evol. 27: 1813–1821.

Somers, D. E., T. F. Schultz, M. Milnamow, and S. A. Kay, 2000 *ZEITLUPE* encodes a novel clock-associated PAS protein from Arabidopsis. Cell 101: 319–329.

Soolanayakanahally, R. Y., R. D. Guy, S. N. Silim, E. C. Drewes, and W. R. Schroeder, 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.

Stephens, M., and P. Scheet, 2005 Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. Am. J. Hum. Genet. 76: 449–462.

Stephens, M., N. J. Smith, and P. Donnelly, 2001 A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68: 978–989.

Stern, D. L., and V. Orgogozo, 2009 Is genetic evolution predictable? Science 323: 746–751.

Stinchcombe, J. R., C. Weinig, M. Ungerer, K. M. Olsen, C. Mays *et al.*, 2004 A latitudinal cline in flowering time in Arabidopsis thaliana modulated by the flowering time gene *FRIGIDA*. Proc. Natl. Acad. Sci. USA 101: 4712–4717.

Storz, J. F., and J. K. Kelly, 2008 Effects of spatially varying selection on nucleotide diversity and linkage disequilibrium: insights from deer mouse globin genes. Genetics 180: 367–379.

Tajima, F., 1989 Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585–595.

Tang, K., K. R. Thornton, and M. Stoneking, 2007 A new approach for using genome scans to detect recent positive selection in the human genome. PLoS Biol. 5: 1587–1602.

Tenaillon, M. I., and P. Tiffin, 2008 The quest for adaptive evolution: a theoretical challenge in a maze of data. Curr. Opin. Plant Biol. 11: 110–115.

Thornton, K., 2009 Automating approximate Bayesian computation by local linear regression. BMC Genet. 10: 35.

Turner, T. L., E. C. Bourne, E. J. Von Wettberg, T. T. Hu, and S. V. Nuzhdin, 2010 Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. Nat. Genet. 42: 260–263.

Tuskan, G. A., S. DiFazio, S. Jansson, J. Bohlmann, I. Grigoriev *et al.*, 2006 The genome of the black cottonwood, *Populus trichocarpa* Torr. and Gray. Science 313: 1596–1604.

Wachowiak, W., P. Balk, and O. Savolainen, 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 51: 117–132.

Yukelivich, R., J. Lachance, F. Aoki, and J. R. True, 2008 Long-term adaptation of epistatic genetic networks. Evolution 62: 2215–2235.

*Communicating editor: O. Savolainen*

GENETICS

**Supporting Information**
http://www.genetics.org/content/suppl/2011/05/30/genetics.111.128041.DC1

# Local Selection Across a Latitudinal Gradient Shapes Nucleotide Diversity in Balsam Poplar, *Populus balsamifera* L

**Stephen R. Keller, Nicholas Levsen, Pär K. Ingvarsson, Matthew S. Olson, and Peter Tiffin**


DOI: 10.1534/genetics.111.128041

**File S1**

Below is ms command line code for generating simulated datasets of neutral polymorphism when testing for selection on balsam poplar phenology candidate genes.

(ms, number of chromosomes, number of loci, options)

For ABC analysis:

(1) Standard Neutral Model:

ms 28 219 -t {0.01 4} -r {0 0.25} 617

(2) Exponential Growth Model:

ms 28 219 -t {0.01 4} -r {0 0.25} 617 -G {1 20} -eG {0.01 10} 0.0

(3) Bottleneck (Serial Size Change) Model:

ms 28 219 -t {0.01 4} -r {0 0.25} 617 -eN {0.01 4} {0.01 2} -eN {TEB + (0.01-1)} {Nb + (0.5 4)}

For simulating datasets under the best fitting demographic model from ABC to test for selection on candidate genes:

(1) Standard Neutral Model:

ms 38 500000 -t 1.4484 -r 0.4089 2534

(2) Bottleneck Model:

ms 38 500000 -t 1.4596 -r 0.5892 2534 -eN 0.3308 0.7155 -eN 0.9004 2.9013

For simulating a neutral model of population differentiation to test for local selection between northern and southern accessions:

ms NumTotal 1 –s NumSegSites –I 2 NumHigh NumLow –ej 0.01625 1 2

where, NumTotal, NumHigh, and NumLow are the chromosome counts in the pooled sample, and the high and low latitude subgroups, respectively, and NumSegSites is the number of segregating sites in the total sample.

**File S2**

File S2 is available for download as an Excel file at http://www.genetics.org/content/suppl/2011/05/30/genetics.111.128041.DC1.

These data are population genetic summary statistics for 219 of the 590 reference loci (where L >= 100 synonymous sites), sequenced in a range-wide sample of 15 individuals.

Original sequence data for these loci are available from the Poplar Population Genomics website (http://www.popgen.uaf.edu/data/data_search_seqs.php)

Functional annotation for loci in the upper 5% tail of the distribution of synonymous pairwise nucleotide diversity is based on Phytozome's genome build 2.0 for Populus trichocarpa.

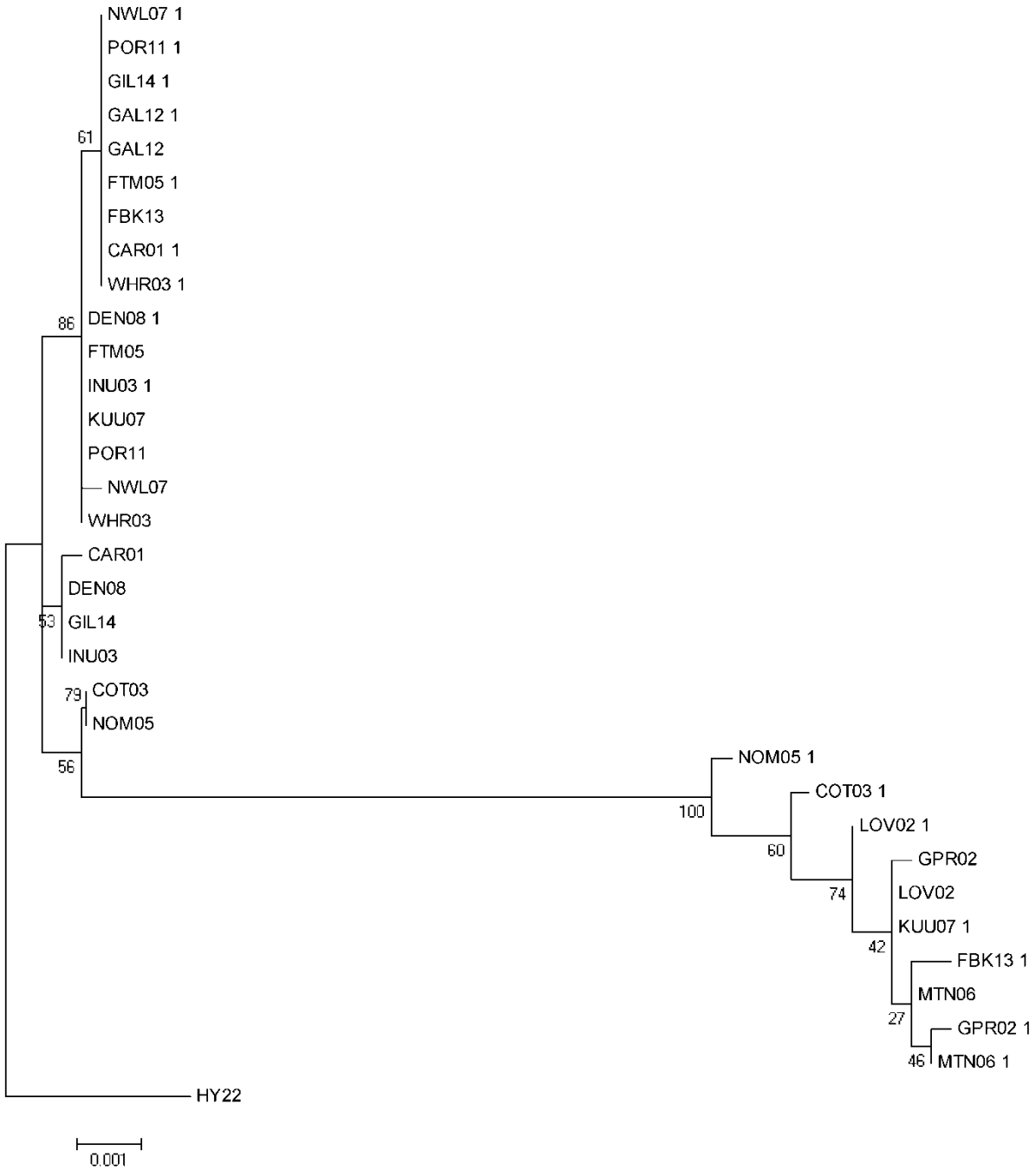


**Figure S1** Maximum-likelihood gene genealogy of *HY2.2*, rooted with *P. trichocarpa*. The model of evolution was GTR + I + Γ, and topology was optimized using a heuristic search with NNI branch swapping. Numbers at nodes are bootstrap percentages from 100 replicates.

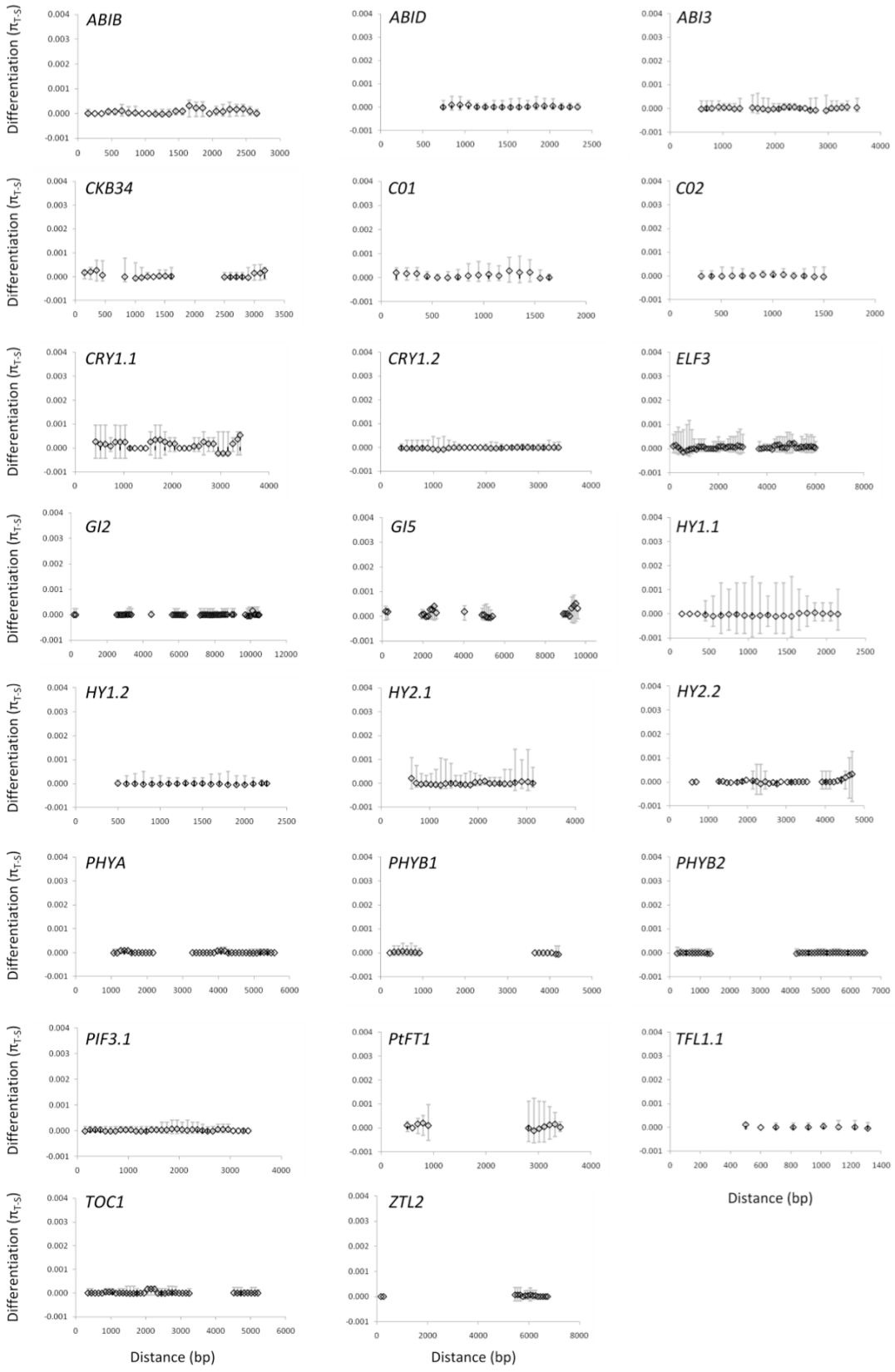


**Figure S2** Sliding window analysis of nucleotide differentiation between high and low latitude samples for additional genes in the flowering time network and abscisic acid pathways. Symbols are as described in Figure 6.

**Table S1 Genomic location and annotation of 27 phenology candidate genes and 16 reference genes in *Populus balsamifera* and *P. tremula***

| Gene | Locus name | Linkage group | Genbank accession # (*P. balsamifera*) | Genbank accession # (*P. tremula*) |
|---|---|---|---|---|
| *ABI1B* | POPTR_0006s24100 | VI | JN047546-JN047563 | AM690430 |
| *ABI1D* | POPTR_0018s04570 | XVIII | JN047564-JN47585 | N/A |
| *ABI3* | POPTR_0002s25330 | II | JN047586-JN047601 | JN235114 |
| *CKB34* | POPTR_0002s14080 | II | JN047602-JN047618 | JN23098-JN230400 |
| *CO1* | POPTR_0017s14410 | XVII | JN047619-JN047640 | HQ833365-HQ833366 |
| *CO2* | POPTR_0004s10800 | IV | JN047641-JN047664 | AM600895 |
| *CRY1.1* | POPTR_0005s17100 | V | JN047665-JN047672 | JN235115 |
| *CRY1.2* | POPTR_0002s09730 | II | JN047673-JN047695 | JN235116 |
| *EBS* | POPTR_0002s21570 | II | JN047696-JN047713 | JN230405-JN230408 |
| *ELF3* | POPTR_0003s04040; POPTR_0003s04030 | III | JN047714-JN047749 | JN235117 |
| *FRI* | POPTR_0015s12160 | XV | JN047750-JN047764 | JN235118 |
| *GI2* | POPTR_0002s06490 | II | JN047765-JN047786 | JN235119 |
| *GI5* | POPTR_0005s21870 | V | JN047787-JN047808 | JN235120 |
| *HY1.1* | POPTR_0006s06900 | VI | JN047809-JN047824 | JN230409-JN230412 |
| *HY1.2* | POPTR_0018s13160 | XVIII | JN047825-JN047845 | JN23094-JN23097 |
| *HY2.1* | POPTR_0009s05270 | IX | JN047846-JN047866 | JN23092-JN23093 |
| *HY2.2* | POPTR_0001s26050 | I | JN047867-JN047882 | N/A |
| *LFY* | POPTR_0015s11820 | XV | JN047883-JN047902 | JN230401-JN230404 |
| *PHYA* | POPTR_0013s00220 | XIII | JN047903-JN047925 | JN235121 |
| *PHYB1* | POPTR_0008s10470 | VIII | JN047926-JN047943 | JN235122 |
| *PHYB2* | POPTR_0010s15600 | X | JN047944-JN047961 | AM072312 |
| *PIF31* | POPTR_0013s00310 | XIII | JN047962-JN047983 | JN235123 |
| *PtFT* | POPTR_0008s07730 | VIII | JN047984-JN048000 | JN230413-JN230416 |
| *TFL1.1* | POPTR_0009s16670 | IX | JN048001-JN048018 | N/A |
| *TOC1* | POPTR_0015s07300; POPTR_0015s07310 | XV | JN048019-JN048039 | N/A |
| *ZTL1* | POPTR_0018s09800; POPTR_0018s09810 | XVIII | JN048040-JN048085 | HQ833402 |
| *ZTL2* | POPTR_0006s18000 | VI | JN048086-JN048104 | HQ833403 |
| poptrdraft_195487 | POPTR_0004s16360 | IV | GU272964-GU272993 | JN235124 |
| poptrdraft_201675 | POPTR_0009s10570 | IX | GU273220-GU273245 | JN235125 |
| poptrdraft_205970 | POPTR_0005s24980 | V | GU273336-GU273365 | JN235126 |
| poptrdraft_278850 | POPTR_0010s02000 | Scaffold 2116 | GU275328-GU275357 | JN235127 |
| poptrdraft_424878 | POPTR_0003s01410 | III | GU276088-GU276117 | JN235128 |

| | | | | |
|---|---|---|---|---|
| poptrdraft_428500 | POPTR_0006s15750 | VI | GU276326-GU276355 | JN235129 |
| poptrdraft_428609 | POPTR_0006s10810 | VI | GU276356-GU276385 | JN235130 |
| poptrdraft_557502 | POPTR_0009s09260 | IX | GU277714-GU277743 | JN235131 |
| poptrdraft_558011 | POPTR_0009s04300 | IX | GU277774-GU277803 | JN235132 |
| poptrdraft_566362 | POPTR_0010s10940 | X | GU278640-GU278669 | JN235133 |
| poptrdraft_717415 | POPTR_0006s03560 | VI | GU282644-GU282673 | JN235134 |
| poptrdraft_757865 | POPTR_0003s16660 | III | GU283954-GU283983 | JN235135 |
| poptrdraft_759621 | POPTR_0004s15720 | IV | GU284072-GU284101 | JN235136 |
| poptrdraft_762490 | POPTR_0006s10000 | VI | GU284276-GU284305 | JN235137 |
| poptrdraft_830425 | POPTR_0002s18590 | II | GU287866-GU287895 | JN235138 |
| poptrdraft_831588 | POPTR_0005s23170 | V | GU288014-GU288043 | JN235139 |

**Table S2 Primers for PCR and sequencing of 27 phenology genes in *Populus balsamifera***

| Gene | forward primer name | forward primer sequence (5'>3') | reverse primer name | reverse primer sequence (5'>3') |
|---|---|---|---|---|
| *ABI1B* | ABi1B_F1_683 | AAACCTCTTGCTCACATTCCA | ABi1B_R2_686 | TCTCGGTTTGGCTAAGATAAGA |
| | ABi1B_F3_687 | GAAAGAACCAATGGCATTATCAG | ABi1B_R4_690 | GAACTAGCCAGTAGATGGCAAC |
| *ABI3* | Abi3_F1_S1_542 | ACAGCGGCCTTGTCTTCCAC | Abi3_R1_S1_546 | TCCGAAATCCTCCATCACGTC |
| | Abi3_NEW_F1_555 | TGGACTCCAAGAAGGGGATTTC | Abi3_NEW_R1_559 | GATCCACAGCAATGCGCAAG |
| | Abi3_NEW_F2_556 | AACCACTTGGTGGGGGTGTC | Abi3_NEW_R2_560 | GAACCCCTAAGCCGGGTCTG |
| | Abi3_F1_S2_550 | GAGGGCAGGAAAATCGCAAAG | Abi3_R1_S2_551 | GCAGCAGCTGGACAATTTGC |
| *ABID* | ABI1D_F_437 | AGGTGATTAGCTTACATTCCGCTGGC | ABI1D_675254_532_R3 | GAAGGCAGATGTATGCGACA |
| | ABI1D_675254_535_F24 | TGGGCATCCAATGACTATTCG | Abi1D-675254-507-R2 | TTGAACTTCCTCTGCGCCTTCAGATC |
| *CKB34* | CKB_F1 | TTGAAACTGATAGAGAGAAG | CKB_R5 | AAGCCGTAGTCCTATTTACAACAG |
| | CKB_F3 | CATTTGCACATTCTCAGGTG | CKB_R1 | TCTTACTTCCAGCGAACACAC |
| *CO1* | CO1_F | CAATTGGGCACGCGTATGTGACACG | CO1_R | GTCTGGTCTCTGCATAGGCCTTCC |
| *CO2* | CO2_F | GCTCCGGCTGCCTTCTTRTGCAAGGC | CO2_R | ACAATGCCATATCCWGTTTCTGCCA |
| *CRY1.1* | CRY11_X1F | AAAGATGTCAGGAGGTGGG | CRY11_X2R | TTCATAAAGCAAGTCCGCG |
| | CRY11_2F | ACCCCTTGTCACTGGTTAGGGATCAC | CRY11_4R | AGTTTGAGATAGCCGCCTCCAATTGA |
| | CRY12_x1F | AGAGATGTCAGGTGGTGGC | CRY12_4R | AGTTTGAGATAGCCGCCTCCAATTGA |
| *EBS* | POP_EBS_F2 | GTCGCGGGATATTACAGCAG | POP_EBS_R2 | CTATTCAAGGGAAGCCACCA |
| *ELF3* | ELF3_F1_581 | AAATGAGAGGGGCAAAGGAT | ELF_R2_584 | AACGACCCAGATTGCTTGTC |
| | ELF_F3_585 | ACAATGCCCGACAAGACTGT | ELF_R4_588 | GCATCTGACAAGATAGGTAGCAA |
| | ELF_F5_589 | GAAAAAGGAACCTACAAGTTTCAA | ELF_R5_590 | TCCCTCAATTTTCCCTCCTAA |
| | ELF_F6_591 | CAAATTGTGCCTTCTGTTTGC | ELF_R6_592 | TTCATACAAAATGACACATGCAAC |
| | ELF_F7_593 | TCTGCTTCCAACATATCCAGT | ELF_R9_598 | CAGGCATTCCCAACAAACTA |
| *FRI* | FRI_F1 | TATCAGTGATTTACCAGCAC | FRI_R4 | AAACTTGTCCTCGATTCC |
| | FRI_F4 | TACTGCAGATATCATAGAGGG | FRI_R1 | CATGTAAGATGAATGGTGG |
| *GI2* | GI2_F1 | TTTGGTATTTTGAAGGGCAG | GI2_R1 | TTTATATCTCCATTCTACCG |
| | GI2_F2 | ACAAATTACTGCTTATGTTGAG | GI2_R2 | AACGAAAGTAGTCACTTCTG |
| | GI2_F4 | TATCGTGTTAGCTTCTTCTCGC | GI2_R4 | TTACCCTTATGCCAGATGCATG |
| | GI2_F5 | ATGTGTTCATTTGTGCAGCTTCC | GI2_R5 | AACATTGCATAGTGCGTCC |
| | GI2_F6 | ATTGCAACCTTGGATTGCC | GI2_R6 | TTGCCATAAGCTAAGACGTC |
| *GI5* | GI5_F2 | ACAAATTACTGCCTATGTCGAC | GI5_R2 | AGGTGCAGCAAGCAATATATCG |
| | GI5_F4 | AAACATCCTCAGCTCATCC | GI5_R4 | ATGCATAATCTTCAGCAGCTCG |
| | GI5_F6 | ATTGCAGCCTTGGATCGCG | GI5_R6 | TGAAATGGCTGGCAAGAGCTC |
| *HY1.1* | HY1.1_F1_703 | ATTGGCAAACATGGCTTCTC | HY1.1_R2_706 | ACCTTTATGCAAACCAGACCA |

| | | | | |
|---|---|---|---|---|
| | HY1.1_F3_707 | CGTTTACATGGATGGAGCTGT | HY1.1_R_710 | TCCTGCGTTTCAACACTACG |
| *HY1.2* | HY1.2_F1_711 | ATGGCTTCTCTCACACCAATC | HY1.2_R2_714 | TACTTGCGTGCAAACATCAA |
| | HY1.2_F3_715 | TCAAATGATTGGGAGGAAGG | HY1.2_R4_718 | TACAGATTGGCCAAATGATTGT |
| *HY2.1* | HY2.1_F1_675 | ACTGCGATGGACGGAACTAC | HY2.1_R2_678 | TCCTGACTTGGAGGAAACTTG |
| | HY2.1_F3_679 | TCAATGCCTCTTGTGGAACA | HY2.1_R4_682 | ACCACCGATGAATTCTCCTCTTG |
| *HY2.2* | HY2.2_563_F1 | GAAAACGGTGAGGTAATGTTGA | HY2.2_568_R2 | TGTGCCTGTAACTGTGTTTGC |
| | HY2.2_569_F3 | TGCGCTTGTCCTATTCATGT | HY2.2_572_R4 | CGTTTTGCAGTTGAGAGCAT |
| | HY2.2_573_F5 | GCCAATCAACTGGGCAGTAA | HY2.2_576_R6 | TGGAGTACAAAATGTGATATAGGG |
| | HY2.2_F_751 | TTTCATCACAAGGCTCATGGC | HY2.2_R_752 | GGATCAGGAAACAAACAAGTCA |
| *LFY* | LFY_F1_599 | TCGTCCTGTTAAGGGCAGTT | LFY_R2_602 | CTCTCTCCTTGGCAATGCTT |
| | LFY_F3_603 | CGGTCTTGATTACCTCTTCCA | LFY_R4_606 | ACATGGCATTTTGCATGTTGG |
| | LFY_F4_605 | TTCTGTTTTTGTTTTGTGTTTGG | LFY_R5_608 | CTGTCACGCTCAATTTCCAA |
| *PHYA* | PHYA_F1 | TGTCTTCTTCAAGGCCAAGC | PHYA_R2 | CTCGGAGAGCCAAAAAGCTA |
| | PHYA_2F1 | GCAAAGAAGAGCAAAACATACAA | PHYA_4R1 | ATGCCCTTTTTTGCGATTTA |
| *PHYB1* | PHYB1_1F1 | TGGCATCACAATCACAAAGA | PHYB1_1R2 | AACCATGAGGAGCTCGAAGA |
| | PHYB1_ex2F1 | TGAGGAGGATAAGAACGTGGA | PHYB1_ex2R1 | CCATTTTCGATGCTCTCAAG |
| | PHYB1_3F2 | TAGCAGACATATCCTAACGAGAGG | PHYB1_4R3 | TACAGTTATACCTCTCACAC |
| *PHYB2* | PHYB2_F1 | ATGGCATCACAATCACAAAGG | PHYB2_R3 | CTTTTACTGCACGATGAATAAGC |
| | PHYB2_F4 | TGTTTAACTAATGCAGGCGAAG | PHYB2_R6 | TCATGTGTGAGTTATTGATCTGGAC |
| *PIF31* | PIF31_F | ACTGAGCAGAAGCGTTGCTAACC | PIF31_Ra | AGTTGGCTTGCTCCTAGACAGTGC |
| *PtFT* | PTFT1_F1_613 | GCAACGTGTATTCTGTTGATGG | PTFT1_R1_614 | GAGAGAGAGAGAGAGCCACGTAA |
| | PTFT1_F5_621 | AAAGAGTGGACGGGGATTGT | PTFT1_R5_622 | TGATACCTTGACGTGATTAGGTG |
| | PtFT1_FOR1_741 | TGAATTTCTTGTCACGTGAACTC | PtFT1_REV2_742 | GACTGAGTACGTAGGAATGTAGACAA |
| *TFL1.1* | TFL1.1_F1_609 | AGAGGCTCCCACCCTCAAT | TFL1.1_R1_610 | CATACCGAATGTGGCATCTG |
| | TFL1.1_F2_611 | AAGGCTGATTCTATCGTTTCCA | TFL1.1_R2_612 | TTAACCCTCTCCTGCTGCAT |
| *TOC1* | TOC1_F1 | CCTTCTCCTGGTGATCCACAG | TOC1_R4 | TGTTCATGTAAGCCAAAATCTCG |
| | TOC1_F5 | AAATATCGACCTGATGTGCC | TOC1_R6 | TTGATGCTTGCTCATCACCGTCC |
| *ZTL.1* | ZTL1_F1 | AAGGAAAAGAAATGGAGTGGG | ZTL1_R1 | AGAAAACATACCAATTACGG |
| | ZTL1_F3 | ACTATGCTTGTTGTTACAGCCG | ZTL1_R3 | ATTAGGTGGCTGAACTTGC |
| *ZTL.2* | ZTL2_F1 | TTCATGGGATAAGAATTTTATTGTC | ZTL2_R1 | AACAAACAGAGTAGATTGAAGAG |
| | ZTL2_F3 | TGATACTATGCTTGTTACAGC | ZTL2_R3 | TAACTGAACTAGCCAAGGACAG |

**Table S3 Nucleotide diversity at 27 phenology candidate genes for adaptation to latitude in *P. balsamifera*.**

| Gene | N[a] | $L_{total}$[b] | $L_{rpl}$ | $L_{syn}$ | S[c] | $\theta_{total}$ | $\theta_{rep}$ | $\theta_{syn}$ | $\pi_{total}$ | $\pi_{rep}$ | $\pi_{syn}$ | Tajima's*D* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *ABI1B* | 36 | 2553 | 1262.1 | 382.0 | 15 | 0.0014 | 0.0006 | 0.0032 | 0.0017 | 0.0006 | 0.0041 | 0.6077 |
| *ABI1D* | 44 | 1884 | 304.5 | 112.2 | 13 | 0.0016 | 0.0031 | 0.0000 | 0.0007 | 0.0018 | 0.0000 | -1.6302 |
| *ABI3* | 32 | 2858 | 1438.7 | 415.3 | 20 | 0.0017 | 0.0012 | 0.0036 | 0.0015 | 0.0010 | 0.0035 | -0.5299 |
| *CKB34* | 34 | 2150 | 654.0 | 198.0 | 20 | 0.0023 | 0.0000 | 0.0049 | 0.0019 | 0.000 | 0.0048 | -0.5285 |
| *CO1* | 44 | 1775 | 748.3 | 229.7 | 26 | 0.0034 | 0.0015 | 0.0020 | 0.0024 | 0.0004 | 0.0018 | -0.9764 |
| *CO2* | 48 | 1480 | 698.4 | 210.6 | 13 | 0.0020 | 0.0019 | 0.0043 | 0.0019 | 0.0021 | 0.0032 | -0.1327 |
| *CRY1.1* | 16 | 3205 | 1444.4 | 442.6 | 17 | 0.0016 | 0.0006 | 0.0034 | 0.0012 | 0.0006 | 0.0024 | -0.9791 |
| *CRY1.2* | 46 | 3271 | 1455.4 | 443.6 | 18 | 0.0013 | 0.0005 | 0.0021 | 0.0011 | 0.0002 | 0.0019 | -0.2836 |
| *EBS* | 36 | 634 | 378.8 | 107.2 | 4 | 0.0015 | 0.0000 | 0.0090 | 0.0011 | 0.0000 | 0.0064 | -0.7163 |
| *ELF3* | 36 | 5434 | 1537.4 | 445.6 | 66 | 0.0029 | 0.0022 | 0.0043 | 0.0019 | 0.0008 | 0.0022 | -1.3416 |
| *FRI* | 28 | 3262 | 1217.5 | 381.5 | 85 | 0.0067 | 0.0032 | 0.0088 | 0.0088 | 0.0036 | 0.0114 | 1.2116 |
| *GI2* | 44 | 4834 | 2173.7 | 715.3 | 24 | 0.0011 | 0.0010 | 0.0013 | 0.0005 | 0.0003 | 0.0005 | -1.7920 |
| *GI5* | 44 | 2679 | 997.3 | 325.7 | 39 | 0.0034 | 0.0012 | 0.0035 | 0.0039 | 0.0016 | 0.0039 | 0.4519 |
| *HY1.1* | 32 | 2287 | 554.3 | 150.7 | 19 | 0.0021 | 0.0005 | 0.0017 | 0.0019 | 0.0004 | 0.0015 | -0.3043 |
| *HY1.2* | 42 | 2019 | 559.1 | 148.9 | 17 | 0.0020 | 0.0004 | 0.0047 | 0.0019 | 0.0005 | 0.0053 | -0.1584 |
| *HY2.1* | 42 | 2698 | 462.2 | 128.8 | 61 | 0.0053 | 0.0010 | 0.0036 | 0.0042 | 0.0009 | 0.0027 | -0.7190 |
| *HY2.2* | 22 | 3338 | 728.7 | 204.3 | 7 | 0.0006 | 0.0004 | 0.0000 | 0.0006 | 0.0007 | 0.0000 | 0.2988 |
| *LFY* | 40 | 1880 | 799.4 | 241.6 | 50 | 0.0064 | 0.0003 | 0.0146 | 0.0069 | 0.0003 | 0.0186 | 0.3084 |
| *PHYA* | 46 | 3786 | 1952.0 | 580.0 | 6 | 0.0004 | 0.0002 | 0.0004 | 0.0004 | 0.0002 | 0.0007 | 0.1448 |
| *PHYB1* | 36 | 1620 | 1255.4 | 358.6 | 7 | 0.0010 | 0.0008 | 0.0020 | 0.0005 | 0.0004 | 0.0008 | -1.5197 |
| *PHYB2* | 40 | 3732 | 1725.3 | 497.7 | 6 | 0.0004 | 0.0003 | 0.0009 | 0.0003 | 0.0003 | 0.0010 | -0.5633 |
| *PIF31* | 44 | 3487 | 1371.6 | 395.4 | 18 | 0.0012 | 0.0010 | 0.0017 | 0.0006 | 0.0004 | 0.0007 | -1.5617 |
| *PtFT1* | 34 | 1397 | 360.0 | 11.0 | 20 | 0.0035 | 0.0000 | 0.0063 | 0.0033 | 0.000 | 0.0033 | -0.2035 |
| *TFL1.1* | 36 | 1071 | 396.3 | 122.7 | 4 | 0.0009 | 0.0006 | 0.0000 | 0.0013 | 0.0013 | 0.0000 | 1.0814 |
| *TOC1*(*ex1-4*) | 33 | 3182 | 277.0 | 80.0 | 11 | 0.0009 | 0.0009 | 0.0000 | 0.0005 | 0.0002 | 0.0000 | -1.3290 |
| *TOC1*(*ex5-6*) | *44* | 980 | 938.3 | 267.7 | 5 | 0.0012 | 0.0014 | 0.0000 | 0.0006 | 0.0007 | 0.0000 | -1.2813 |
| *ZTL1* | 46 | 1718 | 1187.2 | 399.8 | 2 | 0.0003 | 0.0002 | 0.0006 | 0.0003 | 0.0001 | 0.0011 | 0.2514 |
| *ZTL2* | 38 | 1755 | 1300.3 | 349.0 | 4 | 0.0005 | 0.0002 | 0.0010 | 0.0001 | 0.0001 | 0.0005 | -1.8826 |

[a]N: Number of haploid samples

[b]L: Length of sequence in bp for the entire region (total), and at just replacement (repl) or synonymous (syn) sites

[c]S: Number of segregating polymorphisms

**Table S4 Tests of selection based on *P. balsamifera* diversity and divergence from *P. tremula***

| Gene | N | $L_{total}$[a] | $K_{total}$[b] | $K_a$ | $K_s$ | $K_a/K_s$ | *H*[c] | HKA[d] | $S_s$[e] | $S_a$ | $Fixed_s$[f] | $Fixed_a$ | γ[g] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *ABI1B* | 36 | 2375 | 0.016 | 0.011 | 0.020 | 0.57 | 0.05 | 0.14 | 8 | 3 | 6 | 14 | 0.35 |
| *ABI3* | 32 | 2362 | 0.023 | 0.015 | 0.033 | 0.47 | 0.50 | 0.80 | 8 | 6 | 32 | 20 | 0.20 |
| *CKB34* | 34 | 872 | 0.015 | 0.000 | 0.022 | 0.00 | 0.70 | 0.10 | 7 | 0 | 12 | 0 | -- |
| *CO1* | 44 | 1474 | 0.027 | 0.007 | 0.045 | 0.16 | -0.82 | 0.01 | 16 | 5 | 31 | 4 | -0.36 |
| *CO2* | 48 | 1357 | 0.026 | 0.017 | 0.035 | 0.50 | 0.85 | 0.78 | 6 | 6 | 23 | 10 | -0.08 |
| *CRY1.1* | 16 | 1894 | 0.012 | 0.007 | 0.029 | 0.23 | -1.90 | 0.00 | 5 | 3 | 10 | 9 | 0.09 |
| *CRY1.2* | 46 | 1902 | 0.011 | 0.006 | 0.031 | 0.18 | 0.56 | 0.44 | 4 | 3 | 13 | 8 | 0.15 |
| *EBS* | 36 | 634 | 0.014 | 0.008 | 0.022 | 0.35 | -2.37 | 0.16 | 4 | 0 | 4 | 3 | -- |
| *ELF3* | 36 | 1213 | 0.034 | 0.028 | 0.040 | 0.71 | -4.09 | 0.25 | 8 | 7 | 22 | 13 | -0.07 |
| *FRI* | 28 | 2077 | 0.017 | 0.013 | 0.020 | 0.62 | 0.60 | 6.64 | 39 | 14 | 12 | 8 | -0.63 |
| *GI2* | 44 | 1650 | 0.027 | 0.016 | 0.046 | 0.35 | 0.40 | 2.14 | 4 | 4 | 28 | 17 | 0.34 |
| *GI5* | 44 | 2290 | 0.023 | 0.007 | 0.034 | 0.21 | 0.30 | 0.17 | 28 | 5 | 40 | 5 | -0.28 |
| *HY1.1* | 32 | 2282 | 0.013 | 0.003 | 0.017 | 0.21 | -0.18 | 0.66 | 18 | 1 | 25 | 1 | -- |
| *HY1.2* | 42 | 2003 | 0.016 | 0.002 | 0.021 | 0.10 | -0.73 | 0.03 | 16 | 1 | 26 | 1 | -- |
| *HY2.1* | 42 | 2448 | 0.029 | 0.004 | 0.034 | 0.12 | -0.46 | 0.71 | 50 | 2 | 60 | 1 | -- |
| *LFY* | 40 | 1880 | 0.021 | 0.001 | 0.036 | 0.04 | -0.32 | 8.48 | 50 | 1 | 25 | 1 | -- |
| *PHYA* | 46 | 2535 | 0.016 | 0.006 | 0.049 | 0.12 | -1.23 | 4.28 | 1 | 2 | 28 | 11 | 0.37 |
| *PHYB1* | 36 | 1620 | 0.008 | 0.003 | 0.025 | 0.13 | 0.52 | 0.13 | 3 | 4 | 9 | 4 | -0.28 |
| *PHYB2* | 40 | 3660 | 0.021 | 0.006 | 0.034 | 0.17 | 0.66 | 5.09 | 4 | 2 | 66 | 10 | 0.36 |
| *PIF31* | 44 | 981 | 0.020 | 0.016 | 0.025 | 0.63 | 0.43 | 0.28 | 3 | 3 | 10 | 9 | 0.20 |
| *PtFT1* | 34 | 1091 | 0.044 | 0.008 | 0.062 | 0.14 | -0.66 | 1.04 | 10 | 0 | 43 | 3 | -- |
| *TOC1* | 44 | 979 | 0.034 | 0.020 | 0.064 | 0.32 | 0.45 | 3.18 | 1 | 4 | 20 | 13 | 0.23 |
| *ZTL1* | 46 | 1518 | 0.007 | 0.001 | 0.020 | 0.05 | 0.57 | 1.27 | 1 | 1 | 9 | 1 | -- |
| *ZTL2* | 38 | 1497 | 0.016 | 0.012 | 0.024 | 0.52 | -4.05 | 0.36 | 3 | 1 | 10 | 12 | 0.53 |

[a]L: Length of sequence in bp

[b]K: Per-site nucleotide divergence between *P. balsamifera* and *P. tremula* at all sites ($K_{total}$), just replacement sites, ($K_a$) or just silent sites ($K_s$)

[c]H: Fay and Wu's *H* statistic (normalized).

[d]HKA: Deviance score from J. Hey's HKA test based on the difference between observed polymorphism and expectations under neutral coalescent simulations.

[e]S: Segregating sites within *P. balsamifera*

[f]Fixed: Nucleotide differences between *P. balsamifera* and *P. tremula*

[g]γ: Scaled selection coefficient ($2N_es$) on replacement sites estimated using the mkprf program. Dashes indicate omitted loci with fewer than 4 replacement site mutations (see Methods)