Signatures of natural selection and local adaptation in *Populus trichocarpa* and *Populus deltoides* along latitudinal clines

Rajesh K. Bawa

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science
In
Forestry

Jason A. Holliday

Eric M. Hallerman

Khidir W. Hilu

Boris A. Vinatzer

December 13, 2012

Blacksburg, Virginia

Signatures of natural selection and local adaptation in *Populus trichocarpa* and *Populus deltoides* along latitudinal clines

Rajesh K. Bawa

ABSTRACT

Trees, like many other organisms, decrease their rate of metabolic activities to cope up with harsh environments. This stage of 'dormancy' is marked by shedding of leaves and bud-set in deciduous trees. Recent studies have revealed the role of the circadian clock in synchronizing the timing of dormancy and physiology for conferring fitness in trees. To better understand the possible role of natural selection on circadian clock-related genes in climatic adaptation, I took a candidate gene approach, selecting circadian clock genes, some of which had been functionally validated, and others hypothesized, to identify signatures of natural selection in *Populus trichocarpa* and *P. deltoides*. Using both frequency spectrum based tests and tests of heterogeneity, I identified genetic variants deviating from selective neutrality. Results reveal that photoreceptors and dormancy regulator genes may have been the targets of natural selection. Nearly the same levels of selective constraints were found in different functional groups of genes irrespective of pleiotropy. Further, upstream regions of all genes showed high selective constraint, with some of them (FT-2, PIF-4, FRIGIDA) showing significantly higher variation than the other genes, hinting at the role of non-coding regulatory regions in local adaption. In some cases, the same genes in both species appeared as outliers, including PIF-6, FRI, FT-2, SRR1, TIC, and CO, which might reflect their common role in adaptation across species boundaries. All of these results indicate a complex nature of phenology regulation and local adaptation in *Populus* species with photoreceptors and dormancy regulator genes playing key roles.

Acknowledgments

Thanks go to all members of the Holliday laboratory, for being very helpful and supportive during the whole degree program. In particular, I would like to thank Haktan, Xiaoyan and Lecong for helping with scripts and logistics. I would also like to thank members of Brunner lab for providing their laboratory supplies in need. I also thank all of my family members for encouraging and supporting me during these three long years. Special thanks to Daljit, Umesh, Amit and other friends for their support and help in odd hours whenever needed. Thanks also to all my committee members, Dr. Eric Hallerman, Dr. Boris A. Vinatzer, and Dr. Khidir Hilu. Finally, I would like to thank my supervisor, Dr. Jason Holliday for all his guidance, support and patience in this whole process.

This thesis research was supported by graduate research assistantship, funded by the Department of Forest Resources and Environmental Conservation and Dr. Holliday's research funds.

Table of Contents

ABSTRACT

Acknowledgments iii

Table of Contents iv

List of figures vi

List of tables ix

List of supplementary information x

Introduction 1

Literature Review 2

Objective 7

Material and Methods 8

Plant material 8

Candidate genes selection 8

DNA extraction, PCR amplification, and sequencing 9

Demographic history and population structure 10

Sequence analysis and tests of neutral evolution 12

Results 14

Gene amplification and sequencing 14

Population structure 14

Intergenic regions and demographic history 14

Level and pattern of nucleotide diversity 17

P. trichocarpa level: 18

P. deltoides level: 20

Neutrality tests 21

DH, HEW and DHEW test: 26

Tests for heterogeneity 27
Discussion 29
Population structure and demography 29
Patterns of nucleotide diversity and neutrality tests 31
P. trichocarpa: 32
P. deltoides: 37
Conclusion 38
References 41
Supplementary Information 51

List of figures

Figure 1. Map showing positions of all the 45 samples used for study. The green pins represent *P. trichocarpa* and the yellow pins represent *P. deltoides.* 10

Figure 2. Criteria for classification of genes into different groups or classes. The arrows represent the stages and interaction between them. 13

Figure 3. Results of STRUCTURE program from *Populus trichocarpa* control sequences. The top figure shows results at K=3 and second figure at K=2. Both of these results were reported as likelihood of both K's was very high. 15

Figure 4. Plot of Evanno's mean likelihood of K and δ K calculated using Structure Harvester program. 15

Figure 5. Relationship between nucleotide diversity and recombination in control regions of *Populus trichocarpa* 17

Figure 6. Pairwise nucleotide diversity (π) in phenology genes of *Populus trichocarpa.* The black bars represent synonymous and grey bars represent non-synonymous substitutions. 19

Figure 7. Pairwise nucleotide diversity (π) in upstream regions of candidate genes. 19

Figure 8. Box plots of different genes of circadian clock pathway and control sequences, for various nucleotide polymorphism statistics in *P. trichocarpa*. Box edges represent the upper and lower quantile with median value shown as bold line in the middle of the box. Whiskers represent range from maximum to minimum. The table on right shows p-values for non-parametric pairwise Wilcoxon rank test. 20

Figure 9. Pairwise nucleotide diversity in phenology genes of *P. deltoides.* The black bars represent synonymous and grey bars represent non-synonymous sites. 21

Figure 10. Box plots of different genes of circadian clock pathway and control sequences, for various nucleotide polymorphism statistics in *P. deltoides*. . Box edges represent the upper and lower quantile with median value shown as bold line in the middle of the box. Whiskers

represent range from maximum to minimum. The table on right shows p-values for non-parametric pairwise Wilcoxon rank test. 22

Figure 11. Comparison between mutation rate per site and pairwise nucleotide diversity in *P. trichocarpa*..... 22

Figure 12. . Comparison between mutation rate per site and nucleotide diversity in *P. deltoides.* 23

Figure 13. Pattern of Tajima's D statistics in *P. trichocarpa*. 23

Figure 14. Pattern of Tajima's D statistics in *P. deltoides*..... 24

Figure 15. Box plots between different classes of genes in circadian clock and control sequences for neutrality tests in *P. trichocarpa*. Box edges represent the upper and lower quantile with median value shown as bold line in the middle of the box. Whiskers represent range from maximum to minimum. The table below figures shows p-values for non-parametric pairwise Wilcoxon rank test. 24

Figure 16. Box plots between different classes of genes in circadian clock and control sequences for neutrality tests in *P. deltoides*. Box edges represent the upper and lower quantile with median value shown as bold line in the middle of the box. Whiskers represent range from maximum to minimum. The table below figures shows p-values for non-parametric pairwise Wilcoxon rank test. 25

Figure 17. Scatter plot for Fu and Li's D statistics in *P. trichocarpa*. The dashed line represents cut-off and any outliers above the line are significant. Scatter-plot on left is for upstream regions, while the one on right, for gene regions. 26

Figure 18. Scatter-plot for Fu and Li's D statistics in *P. deltoides*. The dashed line represent cutoff, so any outliers above the dashed line are significant. 26

Figure 19. 3-D scatter plot for DHEW test in *P. trichocarpa*. Each axis represents p-value for the gene fragments plotted. 27

Figure 20. HKA test for selection at phenology candidate genes. The solid line represents the best fit and dashed lines represent prediction intervals. Significant outliers were the ones outside these intervals and were labeled. The figure on left is for *P. deltoides* and on right for *P. trichocarpa.*..... 28

Figure 21. Neutrality Index, calculated using the CMH method for phenology genes in *P. trichocarpa.* The solid represent NI = 1, which means neutral evolution. 29

Figure 22. Direction of selection (DOS) estimates for all gene fragments in *P. trichocarpa*. Red line at reference point represents neutrality. 30

Figure 23. Box plots showing direction of selection estimates for all classes of circadian clock genes. 30

List of tables

Table 1. Summary statistics of inferring effects of selection on transcriptionally control regions in *P. trichocarpa*. 16

Table 2. Results of Approximate Bayesian Computation (ABC) for three different demographic models. P-values reflect the likelihood of each model given the data. Hence, larger values reflect a model that is more likely. 17

Table 3. One–tailed unpaired student t-test: Differences in diversity between species..... 18

Table 4. Gene fragments showing significant p-value in different compound tests..... 27

List of supplementary information

Table S 1. Location and name of all the *P. trichocarpa* samples with their geographical gradients. 51

Table S 2. Location and name of all the *P. deltoides* samples with their geographical gradient. 52

Table S 3. Gene name, length and their function with number of fragments in which they were amplified. 54

Table S 4. The distance matrix of F_{ST} between populations. The population 1 is northern Canada, 2 as California-Sierra and 3 as BC-Washington samples 54

Table S 5. Average F statistics over all loci in the population. F_{IS} is measure of variation within population, F_{ST} is measure of variation between population and F_{IT} as total variation of population. 54

Table S 6. Nucleotide polymorphism statistics and tests of selection in *Populus trichocarpa*. The values in green represent the small proportion of values which show negative Tajima's D. 55

Table S 7. Nucleotide polymorphism statistics and tests of selection in *P. deltoides*. The values in green represent small proportion of values which show negative Tajima's D. 56

Introduction

Local adaptation is the selective process by which populations can track changes in their environment (Savolainen et al., 2007). Such adaptive challenges drive genetic changes and affect the persistence of species and populations in a given environment. Theoretical work provides the foundation to understand the genetic dynamics and architecture of adaptation in general (Orr, 1998) and local adaptation in particular (Barton 1999; Kawecki and Ebert 2004), and the availability of new statistical and genomic tools have facilitated empirical studies of these phenomena.

Local adaptation in tree species has been well studied, as field performance of particular populations or clones depends on correctly matching seed sources to the planting environment. Early investigations using common garden experiments have often shown substantial genetic differentiation among tree populations, and that this differentiation is of recent origin (Langlet, 1971; Mátyás, 1996; Neale and Kremer, 2011). Range-wide collection of tree populations typically show continuous variation for phenological traits along environmental gradients, which are known as clines (Savolainen et al., 2007). Understanding the geographical and spatial distribution of populations is an important goal for conservation purposes (Neale and Kremer, 2011), and though common garden experiments have been informative, they pose logistical and practical limitations and do not provide information about the underlying genetic architecture of the traits of interest. The most common approaches used to identify genes and variation underlying the adaptive phenotype are QTL mapping (Bradshaw, H.D., Otto, K.G., et al. 1998, Colosimo, 2004) and use candidate gene known to control the trait of interest (Nachman, M.W., Hoekstra, H.E., et al. 2003, Wittkopp, P.J., Vaccaro, K., et al. 2002). QTL mapping requires crossing experiments and further use of molecular markers to narrow down genomic regions to a single gene level. It however poses some challenges, with the major ones being long time for crossing and experimentation in study systems like trees and scoring of adaptive phenotypes, which may or may not be totally known (Li, Y.F., Costello, J.C., et al. 2008, Pooni, H.S., Coombs, D.T., et al. 1987).The other alternative is to use genomic approaches to identify molecular targets of natural selection through studies of re-sequencing datasets. Knowledge of the genetic underpinnings of adaptation can be applied to such diverse goals as conservation

genetics, genome-enabled breeding, and predicting capacity of tree species to adapt to climate change.

In this study, I employ resequencing of candidate genes involved in the circadian clock pathway and its putative targets to understand genetic underpinnings of local adaptation in *Populus* sp. I use molecular population genetic approaches and coalescent methods to identify targets of local adaptation.

Literature Review

Forest trees are not typically considered to be model systems in plant biology. Various characteristics — such as long generation time for full phenotypic expression and the extensive area required for field plantations — make them difficult organisms to study. However, some of their characteristics, such as lack of domestication bottleneck and large open-pollinated natural populations, make them ideal organisms for discovering the molecular basis of natural selection and adaptive divergence (Neale and Ingvarsson, 2008; González-Martínez et al., 2006). Many species have had little or no human disturbance, meaning that extant populations are the result of natural evolutionary forces and that questions related to adaptation, speciation, and demography are not confounded by anthropogenic factors (Savolainen and Pyhäjärvi, 2007). By contrast agricultural species are difficult to find in a natural undisturbed state, and their genetics has been severely affected by domestication bottlenecks.

Trees are a highly diverse, polyphyletic life form (Petit and Hampe, 2006). The most obvious manifestation of this is between deciduous (angiosperm) trees (e.g. *Populus*) and evergreens (gymnosperms, e.g., *Pinus*). There are large differences in genetics and life history of these two groups that affect their suitability for genetic studies. The commercially and ecologically dominant genus *Pinus*, belonging to the family Pinaceae, consists of approximately 100 species, all of which are monoecious with mixed mating systems, primarily outcrossing (White et al., 2007). Pines have a very long generation time, becoming reproductively mature at 10 years of age, and typically have very large genomes (10-40 GB). The genus *Populus* has 29 species and belongs to the family Salicaceae. Poplars are dioecious and obligate outcrossers (Tuskan et al., 2006), with a time of reproductive maturity of around four to five years. The poplar genome is fairly small at ~ 450 Mb, and fully sequenced for *P. trichocarpa* (black cottonwood). Some

poplar species and hybrids are amenable to genetic transformation, making them a useful model for both functional and comparative genomic approaches. These factors make poplars in general, and black cottonwood in particular, strong candidates for genetic dissection of adaptation.

Natural genetic variation in temperate plant species has traditionally been investigated using common garden experiments and provenance tests (Langlet, 1971; Mátyás, 1996). These studies have often shown clinal variation in adaptive traits in range-wide collections (Wright, 1976; Morgenstern, 1996), the result of natural selection and local adaptation (Campbell, 1979; García-Gil et al., 2003). These clines have been found along geographical gradients (latitude, elevation and longitude) suggesting directional selection in response to environmental gradients (Savolainen et al., 2007). For example in a classical study, Clausen et al. (1948) reciprocally transplanted *Achillea* populations to low–, mid– and high–elevation sites and measured their fitness in terms of various morphological and physiological attributes such as height, growth rate, timing of dormancy and frost sensitivity. They found significant differences in phenotypic and physiological responses in different *Achillea* ecotypes, giving us first evidence of local adaptation. Similar experiments with northern temperate conifers (such as spruce and pines) showed that populations from northern latitudes set buds and flush earlier than southern populations. Steep clines were observed in species extending over large latitudinal ranges for example *Pinus sylvestris* in Scandinavia, Giertych (1991); *Picea sitchensis* in N. America, Mimura and Aitken, (2007). Similar types of clines were found for bud burst in north temperate angiosperms. In beech (*Fagus sp*) (vonWuehlisch et al., 1995), birch (*Betula sp)* (Worrell et al., 2000) and sessile oak (*Quercus petraea*) (Liepe, 1993), southern population flush earlier than the northern populations, and oaks from higher elevations flushed earlier than those from lower elevations (Vitasse et al., 2009) . These clines in phenological traits suggest they are driven in trees by abiotic factors such as light and temperature (Pruneda-Paz and Kay, 2010).

Although these common garden experiments are very useful in characterizing broad patterns of adaptive genetic variation, they provide no information of the genetic variants that underlie them. I therefore, like other recent molecular studies (Dvornyk et al., 2002; Brown et al., 2004; Pot et al., 2005; García-Gil et al., 2003) in perennials focus on exploring the molecular basis of local adaptation in populations of *Populus trichocarpa* and *P. deltoides* originating from across their respective native ranges, using modern statistical tests on variation in candidate genes related to

bud phenology and dormancy. A suite of neutrality tests have been developed in recent years based on theoretical expectations about the frequency distribution of mutations and their properties (e.g., synonymous/non-synonymous), and these tests are increasingly used in non-model species to identify genes potentially under natural selection (Dvornyk et al., 2002; Brown et al., 2004; Pot et al., 2005; García-Gil et al., 2003; Heuertz et al., 2006 and Garcia and Ingvarsson, 2007).

Natural selection comes in many forms – directional, purifying, balancing, and divergent (Nielsen, 2005). Positive selection is defined as any selection that increases the frequency of an advantageous mutation. Similarly, negative selection tends to reduce or eliminate deleterious allele/mutation. Balancing selection tends to maintain multiple alleles, therefore increasing variability in populations. Overdominance can be referred to as special case of balancing selection, as the heterozygote has more variability than its homozygous counterparts (Nielsen, 2005). Diversifying selection, on another hand is a geographically restricted selection. It tends to maintain two extreme phenotypes simultaneously and is caused by spatial variation in selective factors e.g. climate, latitude, elevation etc.

Various statistical tests can be used to test neutrality and identify different forms of selection. Statistical tests of neutral evolution fall into four broad categories: F_{ST} outlier–based tests, those based on the frequency distribution of mutations (the site frequency spectrum), intra– and inter–specific comparisons of diversity vs. divergence, and those that test for deviations from neutral expectations for the proportion of synonymous vs. non-synonymous mutations. Wright's F_{ST} (Wright, 1950) was originally devised, along with F_{IT} and F_{IS}, as a way to partition variation into within– and between–population components. Lewontin and Krakauer (Lewontin and Krakauer, 1975) subsequently proposed that this statistic could be used to identify polymorphisms that have been targets of diversifying selection by comparing locus specific estimates of F_{ST} to the genome-wide average. Unusually high F_{ST} is indicative of diversifying selection, while low F_{ST} is suggestive of stabilizing (balancing) selection (Holsinger and Weir, 2009). Locus specific estimates of F_{ST} have been widely used in various studies (Akey et al., 2002; Weir et al., 2005; Hallerman, E.M. 1984) to detect divergent selection.

Selection affects the frequency distribution of new mutations and other variants in the segregating population. Changes in frequency of these variants therefore can be employed to detect signatures of selection in the population. The tests which employ the frequency distribution of mutations to detect selection are frequency spectrum–based tests. When a mutation occurs, it occurs in the background of other variants, and when it gets selected for over generations, linked non-adaptive mutations increase in frequency along with it, therefore giving a signature of selection around the adaptive allele. This phenomenon is termed as "selective sweep" or "genetic hitchhiking". Selective sweeps can be used to detect positive selection, whereas in negative (or purifying) selection, deleterious mutations are kept at low frequency (or eliminated), leading to an increase in the number of mutations of low frequency (Nielsen, 2005). Various classical tests use this information to detect departures from neutrality, with the most common one being Tajima's D.

Tajima's D (Tajima, 1989) is a summary statistic based on the site frequency spectrum. This test measures the standardized difference between two estimators of nucleotide diversity – the average number of nucleotide differences between two pair of sequences (known as Tajima's π) and the number of segregating sites (known as Watterson's θ) (Nielsen, 2005). Even though this test is still widely used to test neutrality, it poses some limitations in certain scenarios, for example, rejecting the standard neutral model (SNM) in samples showing population growth, and inability to distinguish between background selection and selective sweeps. The Hudson-Kreitman-Aguade (HKA) test (Hudson et al., 1987) compares within–species polymorphism and between–species divergence. This test depends on both θ ($\theta = 4N\mu$) and time of divergence ($T = t_{div}/2N$) and uses goodness–of–fit to detect outliers from the null hypothesis of no selection. Under neutral evolution, diversity at a locus should be proportional to divergence at the locus. High diversity relative to divergence suggests genetic hitchhiking associated with a selective sweep, whereas high divergence relative to diversity suggests purifying selection. This test is fairly robust, but is somewhat conservative as it assumes complete linkage within each gene complex and free recombination between them (Hartl, 2000), and may detect false positives due to effects of migration and demography. The McDonald-Kreitman test (McDonald and Kreitman, 1991) uses the ratio of non-synonymous and synonymous mutations within and between species to detect selection, and is most robust to departures from demographic equilibrium.

The extent of variation in natural populations due to systematic and stochastic evolutionary forces is generally confounded by demographic patterns of variation. Therefore, separation of selection signals from demography is important to identify underlying true adaptive genetic variation (Thornton et al., 2007). One way of inferring demographic signals is by using putatively neutral genetic markers (Neale and Kremer, 2011). For example, a study of European oaks used chloroplast DNA to understand the post-glacial migration of species and to partition various sources of extant variation, including adaptive differentiation (Derory et al., 2010). Similar studies have been done in *Fagus sativa* (Beech) *(Magri et al., 2006)* and *Pinus sylvestris* (Scots Pine) (Cheddadi et al., 2006). Several resequencing–based approaches in different species have shown that extant variation in candidate genes for adaptive traits show confounding effects of demographic and selective factors (Savolainen et al., 2007 ; Neale and Ingvarsson, 2008). Much of the genomic variation I see today is due to bottlenecks associated with Pleistocene glaciation (*Picea abies:* Heuertz, De Paoli et al. (2006); *Pinus sylvestris:* Pyhäjärvi et al., (2007); Aleppo pine: Grivet et al., (2009)). Demographic events such as population expansion and bottlenecks leave a genome–wide signature, whereas the selection only affects fitness-related and closely linked loci. Thus, one way of unraveling the molecular signals left by evolutionary forces is the use of genome–wide datasets, which allows us to disentangle the overall signal of demography and to detect the loci showing deviation from the background variation (Derory et al., 2010). Ultimately genome-wide scans are preferable to candidate gene–based approaches, but due to their cost, have only been carried out so far in humans (Biswas and Akey, 2006) and model species *Arabidopsis* (Toomajian et al., 2006) and *Drosophila* (Andolfatto and Przeworski, 2000; Begun et al., 2007; Hartl et al., 1994; Sawyer et al., 2007; Sella et al., 2009; Wall et al., 2002).

Candidate genes in forest trees are usually selected based on their putative function (from homology searches), expression profiles, or tight linkage with quantitative trait loci (Sorkekh et al., 2008). There are abundant expressed sequence tag (EST) resources available in *Populus* (Sterky et al., 2004) with about 360,000 ESTs from 10 different species. Moreover, with the complete genome sequence of *Populus trichocarpa* having been published (Tuskan et al., 2006), the genus *Populus* has become a model plant from an evolutionary perspective (Jansson and Douglas, 2007). The relatively close relation of poplar to the model plant *Arabidopsis thaliana* (Tuskan et al., 2006) has given a boost to various functional genomics approaches for gaining

insights into various developmental and functional aspects (Brunner et al., 2004). Given the different life histories of *Populus* and *Arabidopsis*, with the former being a long–lived woody perennial and the latter a very short–lived annual, comparative genomics can provide interesting insights into the various mechanisms being preserved and how different evolutionary lineages have used particular developmental pathways to solve different evolutionary problems (Neale and Ingvarsson, 2008). A recent comparative study of *Populus* and *Arabidopsis* showed that components of the same pathway control photoperiodic control of flowering in *Arabidopsis* (Pruneda-Paz and Kay, 2010) and seasonal growth cessation in *Populus* (Ibáñez, 2010; Rohde and Bhalerao, 2007). Some of the genes involved in this process have been also implicated in control of growth cessation of conifers (Gyllenstrand et al., 2007).

While most of the studies described above have focused on the protein encoding genes and ESTs, there is growing interest in rapid evolution of non-coding regions, particularly in cis-regulatory and un-translated regions in or near core promoters (Suzuki et al., 2011; Torgerson et al., 2009; Andolfatto, 2005 and Haygood et al., 2007). Moreover the high degree of protein sequence similarity between phenotypically diverged species (for example, only 9% of genes show no similarity between *Arabidopsis* and *Populus trichocarpa*) (Tuskan et al., 2006) suggests that regulatory evolution may be considerably more important than protein evolution. Promoters integrate internal (endocrine) and external (environmental) signals in gene expression, and a slight change in the environment may have a physiologically and developmentally important impact on gene expression (Cheung, V.G., Spielman, R.S., et al. 2005, DeCook, R., Lall, S., et al. 2006, Keurentjes, J.J.B., Fu, J., et al. 2007), thus making them important candidates for study of local adaptation. It is therefore important to consider the role that variation in regulatory regions may play in shaping evolution.

Objective

My study species, *P. trichocarpa and P. deltoides,* are distributed widely across North America. *P. trichocarpa* is distributed over a wide latitudinal range along the west coast, from Alaska to California, and *P. deltoides* can be found in the central and southeastern United States (Stettler and National Research Council, 1996). I acquired the samples for *P. trichocarpa* from Canada and the western United States and in the mid-west and south-east United States for *P. deltoides*.

The main objective of study was to assess whether candidate genes in the photoperiodic pathway have been targets of natural selection that may be related to climatic adaptation. To do this, I resequenced these genes, which were selected on the basis of homology to photoperiod–related genes in *Arabidopsis*, assessed their nucleotide diversity, and conducted a variety of neutrality tests.

Material and Methods

Plant material

Clones for this study were gathered from along the west coast for *Populus trichocarpa*, and from the midwest and southeast of the United States for *Populus deltoides*. The *P. trichocarpa* collection came in two parts, 18 samples from a collection made by the Ministry of Forestry, British Columbia, Canada and 15 samples from along the west coast of the United States (Washington to California) which was originally collected by Greenwood Resources, Inc. (http://www.greenwoodresources.com/). The 33 plants were sampled so as to encompass most of the latitudinal and longitudinal gradients (Figure 1 and Table S1), uniformly across the range. Similarly the 12 *P. deltoides* clones were sampled to include much of the species range (Figure 1 and Table S2), and also were provided by Greenwood Resources.

Candidate genes selection

Candidate genes for bud phenology and dormancy were selected on the basis of numerous recent studies (Pruneda-Paz and Kay, 2010; Ibáñez et al., 2010; Rohde and Bhalerao, 2007) that show that this pathway is a central regulator of the seasonal growth and dormancy cycle in *Populus*. I included all genes known to be involved in the photoperiodic pathway based on functional studies in *A. thaliana* (Pruneda-Paz and Kay, 2010), although not all the genes have been functionally characterized in perennials (Böhlenius et al., 2006; Rohde and Bhalerao, 2007 and Ibáñez et al., 2010). The DNA sequences of genes involved were selected from the *Arabidopsis* genome database (Swarbreck et al., 2008), and BLAST was used to retrieve the orthologus poplar sequences from Joint Genome Institute's (JGI) *Populus trichocarpa* genome database (Tuskan et al., 2006). In total, 26 genes were selected and amplified in fragments with lengths of approximately 1kb (Table S3). As promoter sequences may be important targets of selection, 1

Kb upstream regions for all the candidate genes were selected in order to include the promoter sequences within them.

DNA extraction, PCR amplification, and sequencing

Leaf tissues were sampled for DNA extraction, and the genomic DNA was extracted from all 45 samples using the Qiagen Maxi kit (Hsu et al., 2006). The extracted DNA was stored at - 20°C. Primers were designed for all the 26 candidate genes using Primer 3 (Rozen and Skaletsky, 1999) and Net Primer (http://www.premierbiosoft.com/netprimer/index.html) software. All primers were designed to amplify the genes in fragments of 1-1.5 Kb in size, with at least a 100bp overlap to facilitate the post-sequencing joining of fragments. An oligo-extension of universal primers M13 forward (GTAAAACGACGGCCAGT) and M13 reverse (GCGGATAACAATTTCACACAGG) was added to forward and reverse primers, respectively, of all the genes to facilitate direct sequencing of PCR products. The same primers were used for both *P. trichocarpa* and *P. deltoides* samples to amplify the genes. The normal PCR conditions used were — 95°C for 4 min., followed by 30 cycles of 95°C for 45 sec., 58-60°C for 45 sec., 72°C for 1min. 30sec; and 72°C for 7min. The PCR products were purified using Agentcourt magnetic bead purification kit (Agencourt AMPure XP) and then sequenced using Sanger sequencing method (Sanger et al., 1977). The sequencing procedures were performed on ABI-310 sequencer by MCLAB Inc. (http://www.mclab.com/home.php).

In addition to conventional Sanger-sequencing of gene targets described above, I employed data obtained by a recently initiated sequence capture project (Gnirke et al., 2009) to make inferences about population structure and demography. The sequence capture method is a high-throughput genome complexity reduction method used to develop genome-wide markers, and involves use of oligonucleotide 'baits' to retrieve and sequence particular regions of the genome. In this case, I targeted 320 inter-genic control regions, each of 120bp in length. The availability of sequence-capture data from the same *P. trichocarpa* population used for candidate gene sequencing made them an attractive addition to this study.

Sequences of all gene fragments were assembled using CodonCode Aligner (CodonCode Corp. Dedham, MA) software, and mutations were identified using the embedded algorithm. All single nucleotide polymorphisms (SNPs) were manually verified. Conventional diversity statistics were

computed using Arlequin (Excoffier et al., 2005) and DnaSP (Rozas and Librado, 2009) software. These statistics included nucleotide diversity (θ_π and θ_W), haplotype diversity and number of haplotypes.

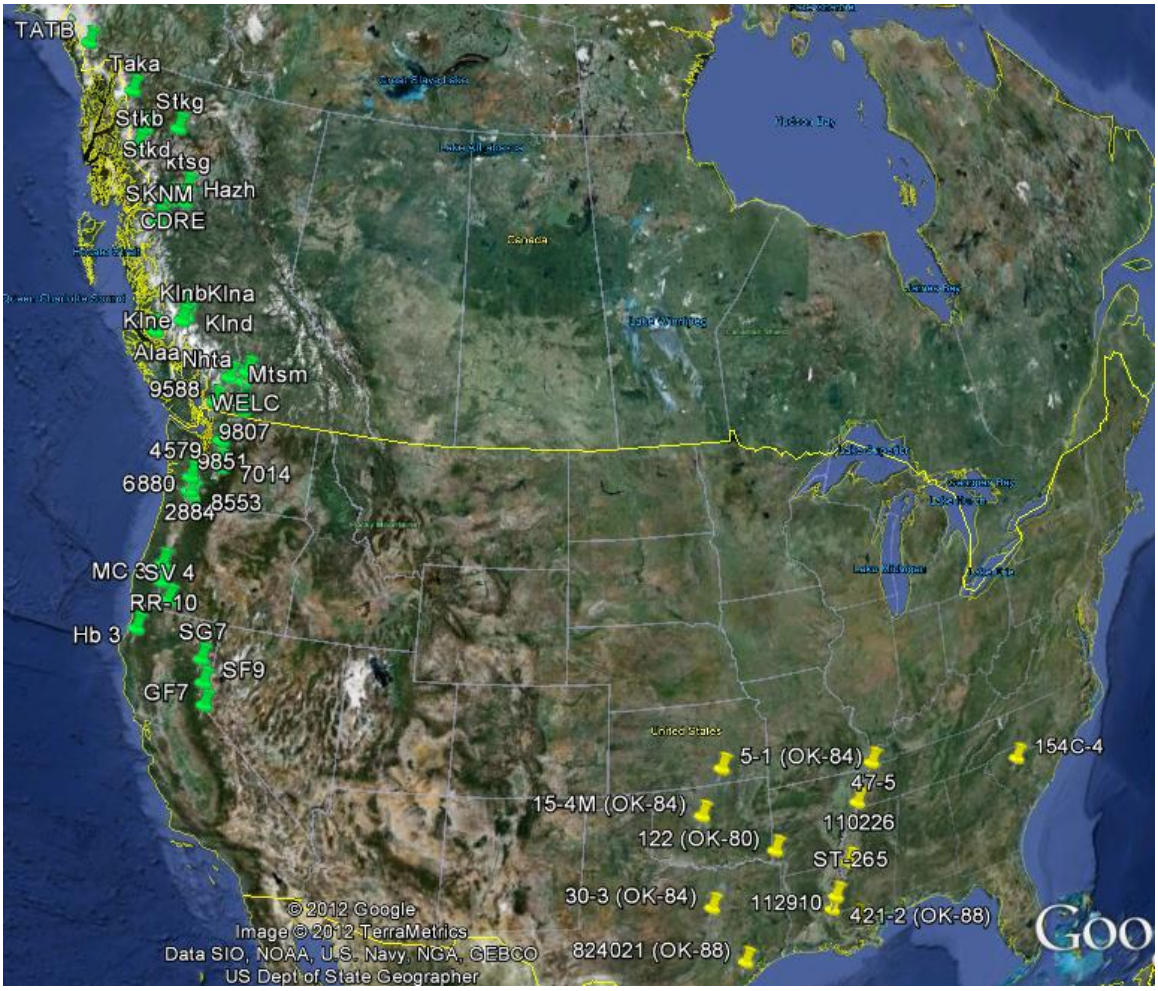


Figure 1. Map showing positions of all the 45 samples used for study. The green pins represent *P. trichocarpa* and the yellow pins represent *P. deltoides.* (Source: Google Earth /version/ 6.2.2.6613)

Demographic history and population structure

I used the program STRUCTURE /version 3.0/ (Pritchard and Donnelly, 2001) to determine whether there was evidence of population structure in the samples. Structure is a widely used Bayesian clustering algorithm and uses genotypic information to infer population structure. I used admixture as my ancestry model, and correlated allelic frequency and lambda were set to 1. I used 100,000 burn-in steps followed by 100, MCMC iterations, with assumed number of clusters between 1 and 10 tested for three replications per cluster.

Evanno's Δ K (Evanno et al., 2005) was used to infer the optimum number of genetic clusters in the population. Optimum number of clusters can be inferred by looking at the likelihood values for each run of STRUCTURE conditioned on various assumed number of populations (clusters). Simulation studies have shown that the 'true' number of populations in a sample can be approximated by examining the rate of change in likelihood values from successive runs of Structure assuming different number of clusters. The program 'Structure Harvester' (Earl and vonHoldt, 2012) was therefore used to estimate this parameter from the data.

To assess whether past demographic events have shaped the frequency spectrum of mutations in *P. trichocarpa*, I used Approximate Bayesian Computation (Beaumont et al., 2002) to assess alternative models, including population growth, population bottleneck, and neutral equilibrium. For each of these models, per-site diversity and recombination were drawn from uniform prior distributions with range U (0, 0.01). The neutral model is comprised of only these parameters, while the bottleneck model involves three additional parameters: timing of the contraction, which was drawn from a log-uniform prior distribution with range $T_b \sim U(0.0001,1)$; reduction in population size (bottleneck severity), which was drawn from log-uniform prior distribution with range $S_b \sim U(0.0001,0.5)$; and duration of the bottleneck, which was fixed at 0.015. The bottleneck timing, severity, and duration parameters are in units of 4Ne generations. The assumption underlying this model is that a population of the same size as the contemporary one sampled contracted at time T_b to a new size defined by S_b, and remained at that size for 0.015N generations, after which it expanded to the current size. For the growth model, the population was assumed to have begun growing at a time in the past defined by U(0.0001,1) from an ancestral population with size defined by U(0.0001,0.5). To compare parameters estimated for the pooled range-wide sample with those for individual populations, I also conducted these simulations conditional on statistics from the pooled sample (i.e., 48 haploid genomes). Coalescent simulations were performed using the program mlcoalsim (Ramos-Onsins and Mitchell-Olds 2007) kindly provided by Dr. Sebastian Rasmos-Onsins, and subsequent ABC analyses were conducted using R scripts provided by Dr. Mark Beaumont. A total of 500,000 realizations were performed for each model, and summary statistics used for each ABC step were Tajima's π, Watterson's θ, Tajima's D, and the standard deviations for each of these.

Sequence analysis and tests of neutral evolution

Each of the approaches to detect selection described in the introduction was applied to the sequence data. For frequency spectrum–based neutrality tests, I calculated Tajima's D, Fu and Li's D* and Fu and Li's F* statistics using DNAsp. These are the most commonly used statistics to compare observed allele frequency spectrum against the null model of neutral evolution, and tend to be negative in the case of a selective sweep and positive in case of balancing selection. However, they are fairly sensitive to demographic signals and, therefore, were used in conjunction with tests of heterogeneity. There are two types of heterogeneity tests: McDonald-Kreitman test and Hudson-Kreitman-Aguade, which use two or more loci to test neutrality and assess diversity by divergence statistics to test for selection acting on specific loci in the population.

The McDonald-Kreitman test (1991) was used to infer intra- and inter-specific levels of diversity in non–synonymous and synonymous mutations. Under the neutral hypothesis, the ratio of number of non–synonymous polymorphisms to fixed non–synonymous differences (P_N:D_N) should be equivalent to the number of synonymous polymorphism to fixed synonymous differences (P_S:D_S) (Hartl 2000; Li et al., 2008). An excess of non–synonymous fixed differences is suggestive of positive selection, while the excess of non–synonymous polymorphism represent weakly segregating deleterious mutations or in some cases a result of strong balancing selection (Weinreich and Rand, 2000; Li et al., 2008 and Hall et al., 2011).

I also applied the Hudson-Kreitman-Aguade (HKA) (Wright and Charlesworth, 2004) test. The rationale behind this test is that with no selection, equilibrium levels of polymorphism depend on θ ($\theta=4N_e\mu$), but the divergence depends on θ and T (where T is time of divergence in units of 2N generations, $T=t_{div}/2N$). Excess diversity or divergence is suggestive of positive or purifying selection in the focal species, respectively (Hartl, 2000).

Finally, I used DHEW test (Zeng et al., 2007) to find signatures of adaptive evolution in my sequences. This test is a combination of three neutral statistics, Tajima's D, Fay-Wu's H and Emmen-Watterson test statistics. Each of these tests can be and are independently used to test neutrality of sequences, but are sensitive to historical demography (e.g., population size changes), which may lead to false positives. Using these tests together results in a test that is

more robust and less sensitive to processes like drift and demography (Zeng et al., 2007), and unlike other tests such as HKA and Bayesian procedures, can easily be used to find positive selection at single loci.

In addition to coding regions, I sequenced putative promoter regions for each of the 26 genes. These sequences were aligned and then tested for significant differences in evolutionary rates between promoters of different categories (samples and genes), using the same methods as described above, except that in this case I compared the overall rate of evolution in promoter regions with the "neutral rate" estimated from control sequences obtained using sequence capture, as the dn/ds ratio is not defined for non-coding regions.

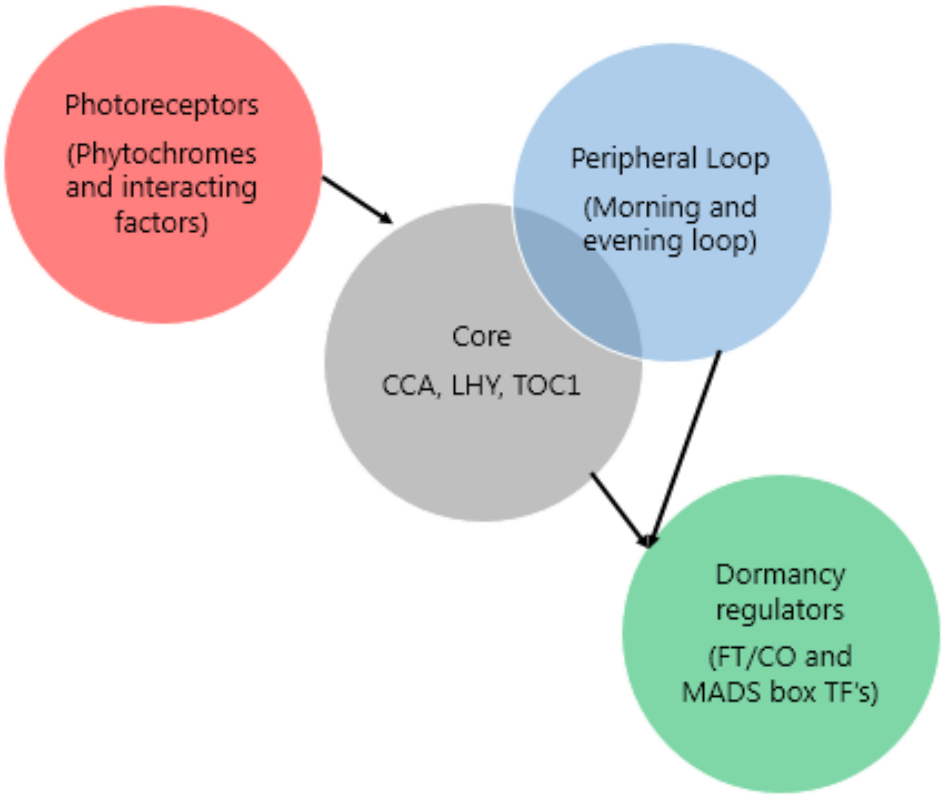


Figure 2. Criteria for classification of genes into different groups or classes. The arrows represent the stages and interaction between them.

Results

Gene amplification and sequencing

All candidate genes (Figure S3) were amplified using PCR and sequenced using the standard Sanger sequencing method (Sanger et al., 1977). All genes were amplified in 1–kb fragments with about 100 bp overlap with the next and previous fragments to facilitate assembly of total gene sequence. Even though the primers were designed to retrieve all fragments, some fragments were not amplified in some samples. This was more problematic in *P. deltoides*, as I did not have any reference sequence for the species, and therefore had to rely on *P. trichocarpa* primers. Some of the genes that were partially amplified and sequenced in *P. trichocarpa* were PIF-6, CONSTANS, CONSTANS-1, LUX, ABI3 and FT-1, and the genes that were partially amplified in *P. deltoides* were ELF-4, CONSTANS, and ABI3.

Population structure

A total of 71.7 kb of non-coding regions were sequenced, which revealed 2215 segregating sites. I used these SNP data to run the program Structure (Pritchard and Donnelly, 2001) and to determine whether there was evidence of population structure in my samples. After running STRUCTURE, Evanno's δ K (Evanno et al., 2005) was used to test the likelihood of different population clusters in the sample and I used the Structure harvester program (Earl and vonHoldt, 2012) to determine the most likely K. I found three population groups in the sample (Figure 3 and 4); the first being samples from northern Canada; the second from southern Canada, Washington and Oregon; and the third ones from California and the Sierra Nevada populations. The second population group was larger than other two groups.

Intergenic regions and demographic history

Summary statistics showed much higher polymorphism in control regions (Table 1) than in protein–encoding regions, and neutrality tests for these regions mostly adhered to neutral expectations. In order to test the neutrality of these sequences, I compared the nucleotide diversity with Kelly's ZnS (Kelly, 1997), which is a measure of intralocus recombination or linkage disequilibrium. It is believed that the Zns value tends to inflate when the locus departs from neutrality (Kelly, 1997). Most of the loci showed an extremely low ZnS values, suggesting

neutrality (Figure 5). Loci showing ZnS value higher than 0.2 were eliminated for the purposes of demographic modeling, as they would increase the average of all other statistics.

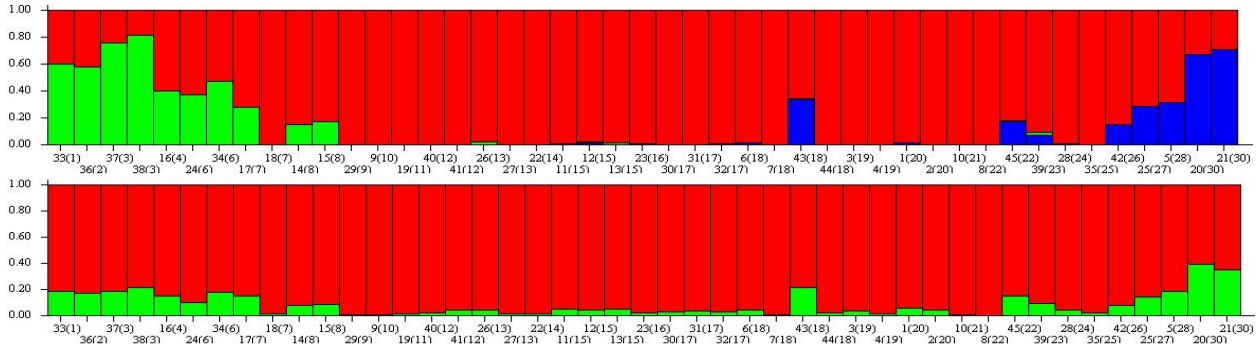


Figure 3. Results of STRUCTURE program from *Populus trichocarpa* control sequences. The top figure shows results at K=3 and second figure at K=2. Both of these results were reported as likelihood of both K's was very high.

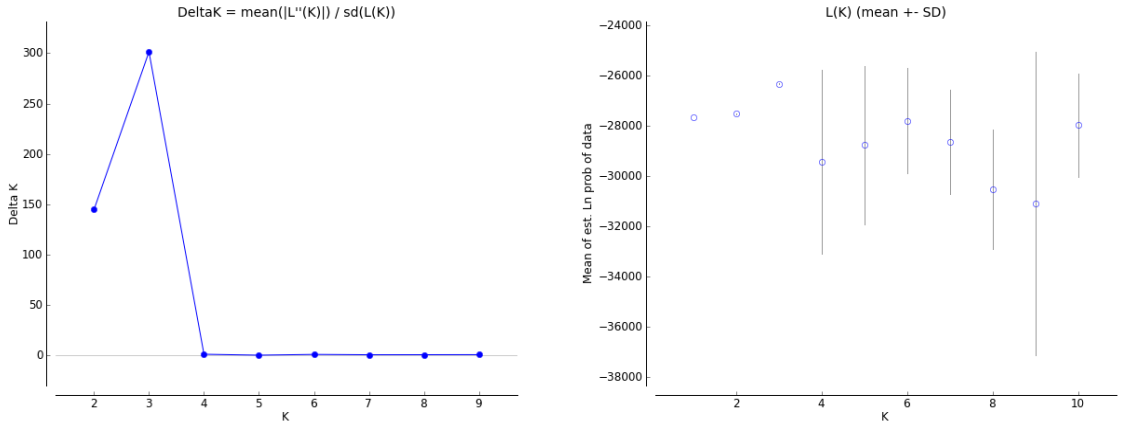


Figure 4. Plot of Evanno's mean likelihood of K and δ K calculated using Structure Harvester program.

The reference or neutral data were available only for *P. trichocarpa* data sets, and as the reference sequences were very few as compared to genic regions, I used the Approximate Bayesian Computation approach to estimate the most fitting demographic model for the data sets. This approach is more widely used these days (Keller et al., 2011; Ingvarsson, P.K. 2010) and allows one to tease apart confounding demographic forces from selection signals.

Species	Parameters	All
P. trichocarpa	Total Sites	71776
	Segregating Sites	2215
	Watterson θ per site	0.0076
	Tajima's D	-0.3570
	Nucleotide diversity (π)	0.00642
	Fu and Li's D	-0.0817

Table 1. Summary statistics of inferring effects of selection on transcriptionally control regions in *P. trichocarpa*.

I used the Approximate Bayesian Computation approach to determine whether a departure from neutrality was evident in the data. Temperate and boreal tree populations are not static in terms of their size, in particular due to range expansion and contraction associated with Pleistocene glaciation (Holliday et al., 2010). Understanding whether this is also the case in *Populus* will inform the choice of neutrality tests for candidate genes. That is, given a demographic signal in intergenic sequence data, I may wish to choose tests that are known to be less sensitive to such a signal.

The posterior probability — that is, the probability of a particular demographic model given the data - was calculated for neutral equilibrium, bottleneck and growth models. The bottleneck model showed the best fit to the data, with a posterior probability of 0.76, although the growth model could not be rejected (Table 2). The neutral model was completely rejected, with an extremely low probability of 4.34×10^{-71}. Although I did not use these ABC simulations to correct for background skew in the frequency spectrum–based tests, the results informed my choice of neutrality tests.

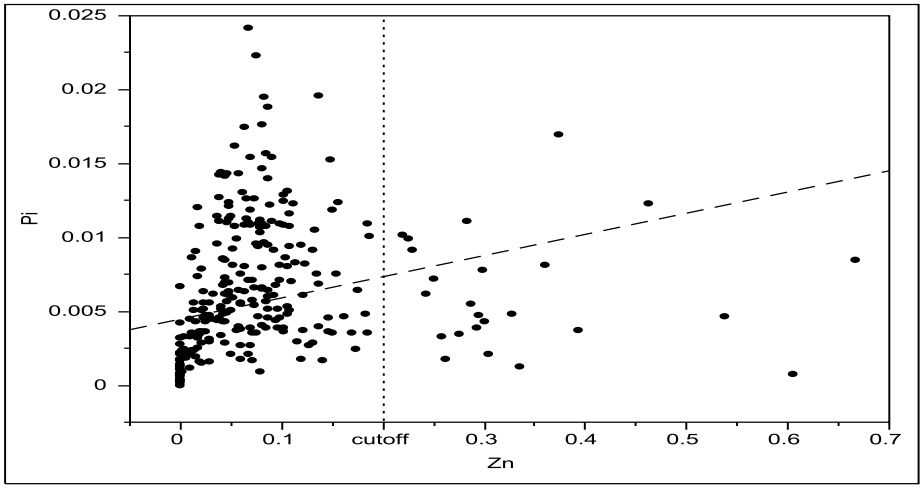


Figure 5. Relationship between nucleotide diversity and recombination in control regions of *Populus trichocarpa*.

Models	Pr (model) | data
Neutral equilibrium	4.348547 e-71
Growth	0.24
Bottleneck	0.76

Table 2. Results of Approximate Bayesian Computation (ABC) for three different demographic models. P-values reflect the likelihood of each model given the data. Hence, larger values reflect a model that is more likely.

Level and pattern of nucleotide diversity

A total of 110,159 bps of was sequenced for candidate genes for each *P. trichocarpa* individual, and 67,421 bps in *P. deltoides*, yielding 929 and 626 segregating sites, respectively. Diversity varied at both the species and gene level, and on a whole average nucleotide diversity was very similar to that estimated in past studies of trees, with more specifically for *Populus* species. (Hall et al., 2011; Keller et al., 2011). The average nucleotide diversity (π) in *P. trichocarpa* was 0.0028 and in *P. deltoides* was 0.0020. Related diversity statistics (such as haplotype diversity, and theta/site) were also lower in *P. deltoides* than in *P. trichocarpa* (Table 3).

Diversity Statistics	t-stats	p-value
Nucleotide diversity (π)	2.8	0.0039**
Haplotype diversity (H_d)	0.23	0.41
Theta per site (θ/site)	1.98	0.027*
Synonymous nucleotide diversity (π ds)	0.77	0.22
Non-synonymous nucleotide diversity (π dn)	0.23	0.41
Student t-test on different diversity statistics in two species, *P. trichocarpa* and *P. deltoides* (** represent significance of results).		

Table 3. One–tailed unpaired student t-test: Differences in diversity between species.

P. trichocarpa level:

The level of nucleotide diversity (π) was highly variable among genes as well as between non-synonymous and synonymous polymorphism types. Some of the genes, such as ELF3, ZTL, FT-1, showed very low to no synonymous nucleotide diversity, while others, such as LHY, TOC-1, PRR-5 and FT-2, had considerably higher synonymous nucleotide diversity as compared to non-synonymous nucleotide diversity (Figure 6). All genes except LHY showed some level of non-synonymous polymorphism. FRI, PIF4 and CO-1 in particular had very high nonsynonymous diversity. Whole gene nucleotide diversity was calculated separately for genic regions and promoter regions. The pattern of nucleotide diversity mostly remained the same for only gene regions, with some slight change in rank order, but the pattern was completely different with respect to promoter regions (Figure 7). The FT2 and FRI genes among other outliers had very high nucleotide diversity in the promoter regions.

I grouped the genes based on the importance and position in the circadian clocks. The core genes, LHY, TOC1 and CCA1 were grouped as core loop, whereas the genes which were involved in morning– and evening–phased loops were grouped as peripheral loops. The remaining genes were either photoreceptors or downstream targets regulating dormancy and were grouped based on their function.

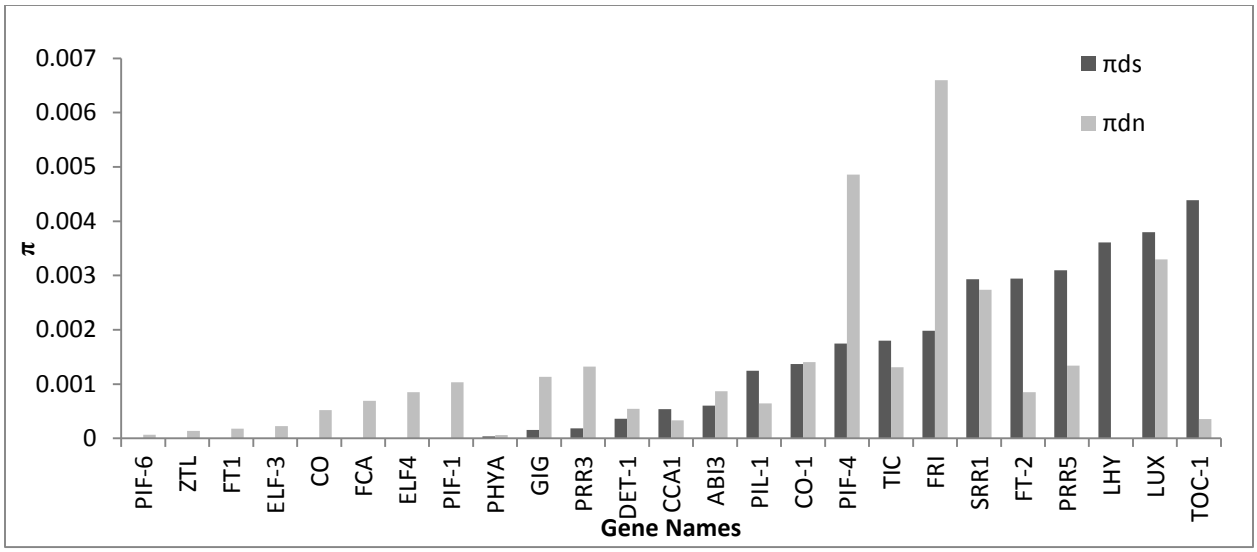


Figure 6. Pairwise nucleotide diversity (π) in phenology genes of *Populus trichocarpa.* The black bars represent synonymous and grey bars represent non-synonymous substitutions.

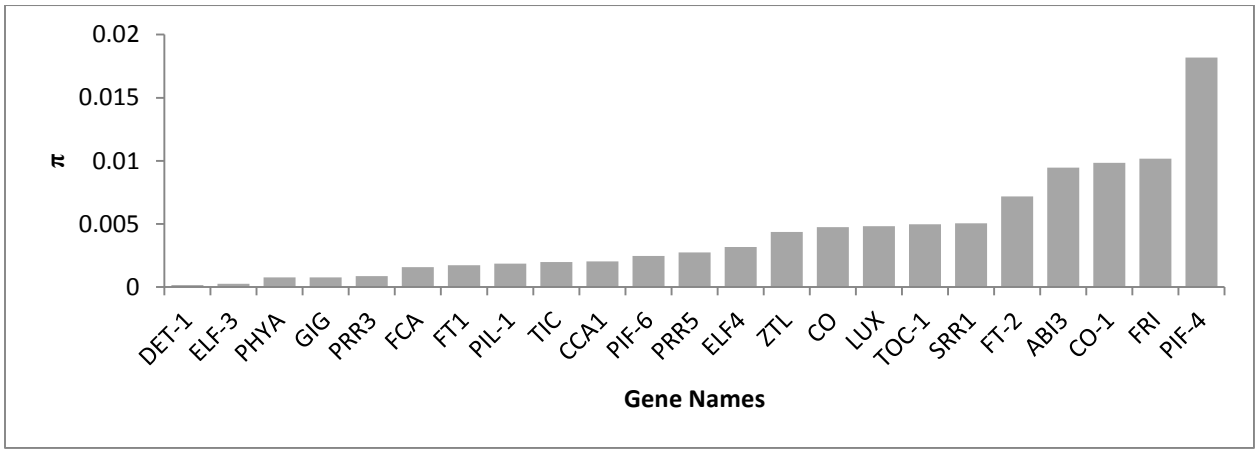
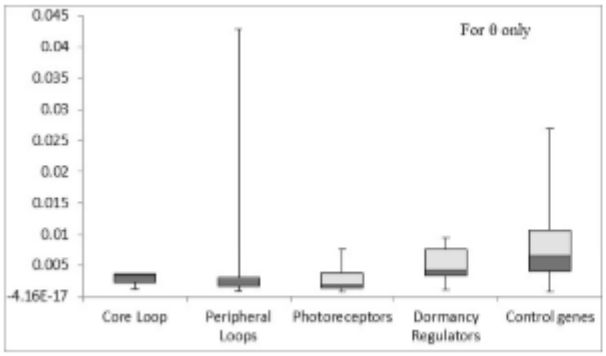


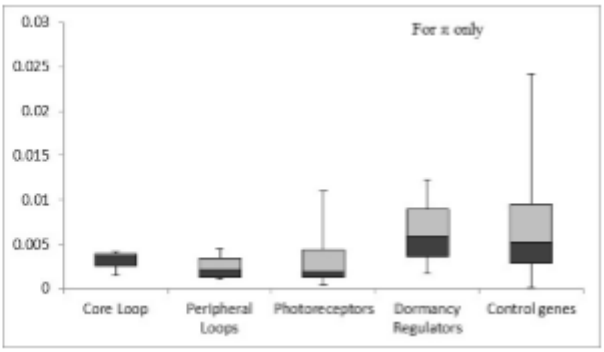
Figure 7. Pairwise nucleotide diversity (π) in upstream regions of candidate genes.

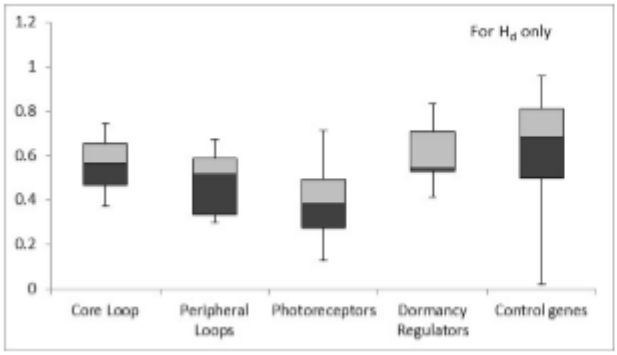
The photoreceptors and transcription factors involved in light signaling pathway were grouped as photoreceptors whereas the downstream targets of circadian clock, which played a major role in dormancy and bud phenology regulation, were grouped as dormancy regulators. This grouping was basically done to get a better picture of level and amount of variation in each class of genes, and whether those differences were significant enough to make any functional and biological significance.

The parameters used for comparison between different gene groups were nucleotide diversity (Watterson θ, Tajima's π) and haplotype diversity (H_d). I used the non-parametric Kruskal–wallis

test and pairwise Wilcoxon test to identify significant differences between nucleotide diversity statistics of functional gene groups and control sequences. There was a significant difference between some functional gene groups compared with control sequences, but no pairwise tests between functional groups except for π between dormancy regulators and photoreceptors (Figure 8).







Pair-wise Wilcoxon test

Measure	Level	Level	p-value
Theta	Peripheral loops	Control sequences	0.0309*
Theta	Photoreceptors	Control sequences	0.0044*
Pi	Peripheral loops	Dormancy regulators	0.0298*
Pi	Photoreceptors	Control sequences	0.039*
Pi	Peripheral loops	Control sequences	0.0099*
Hd	Peripheral loops	Control sequences	0.0408*
Hd	Photoreceptors	Control sequences	0.0088*

Figure 8. Box plots of different genes of circadian clock pathway and control sequences, for various nucleotide polymorphism statistics in *P. trichocarpa*. Box edges represent the upper and lower quantile with median value shown as bold line in the middle of the box. Whiskers represent range from maximum to minimum. The table on right shows p-values for non-parametric pairwise Wilcoxon rank test.

P. deltoides level:

The average level of nucleotide diversity patterns in *P. deltoides* differed from that of *P. trichocarpa*, although they were of similar magnitude and similar to that observed in previous studies (Ismail et al., 2012; Hall et al., 2011; Keller et al., 2011), but at gene–to–gene level, the patterns of variation were very different. Fourteen genes showed no synonymous polymorphism, and six genes that showed neither synonymous nor non-synonymous polymorphism (Figure 9). On the other hand, PIF4, PIF1, LUX, FCA, and ZTL had a fairly high level of synonymous

polymorphism. Non-synonymous polymorphism was considerably less, with only PIF4, PIF-1, and DET-1 showing a very high level of non-synonymous polymorphism. Genes were grouped according to their function in circadian clock and light signaling as mentioned above for *P. trichocarpa,* and diversity statistics were compared both among these groups and with the control sequences (Figure 2). Results were similar to those seen in *P. trichocarpa* – there were no differences between functional gene groups, but for both 'π' and 'θ', diversity of all functional groups was significantly different from control regions (Figure 10). For haplotype diversity, diversity of photoreceptors and peripheral loops was also significantly different from control regions.

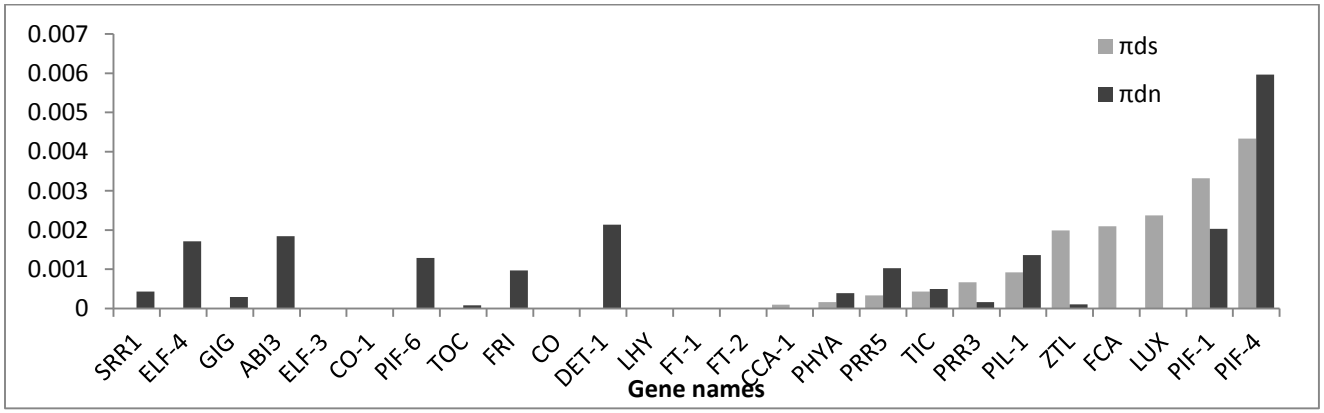
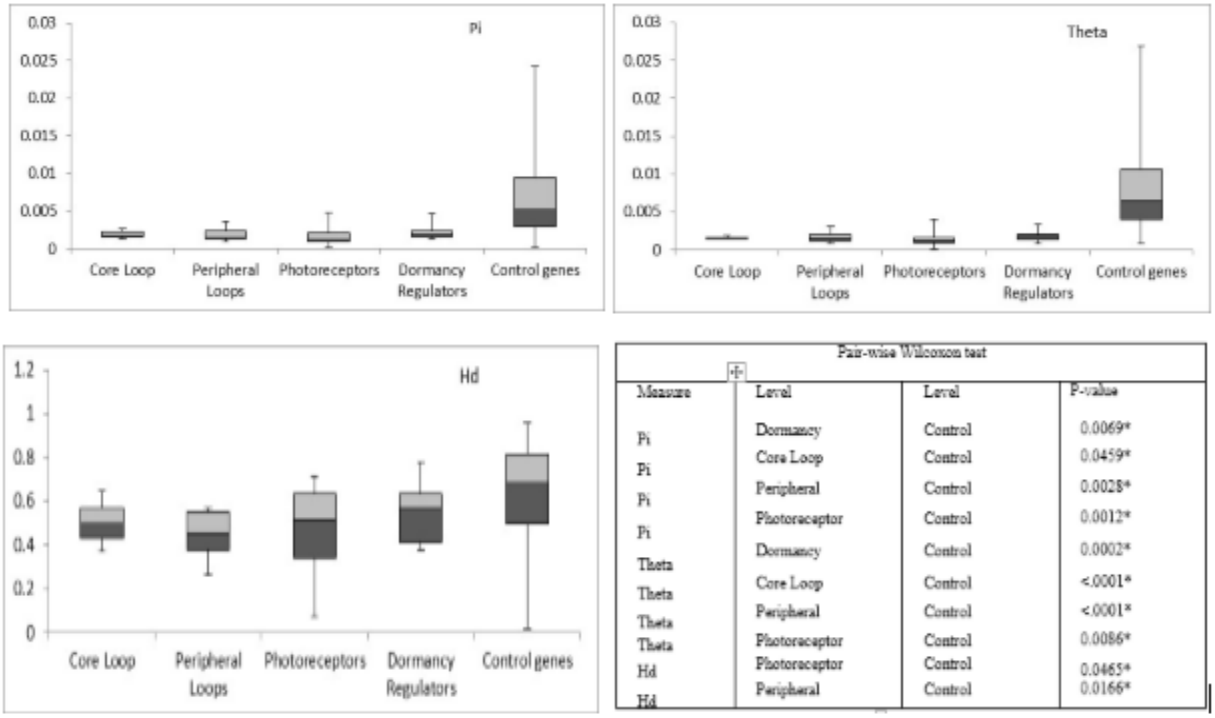


Figure 9. Pairwise nucleotide diversity in phenology genes of *P. deltoides.* The black bars represent synonymous and grey bars represent non-synonymous sites.

Neutrality tests

I compared nucleotide diversity (π) and theta (θ) for all sequences in *P. trichocarpa* and *P. deltoides*. Only one gene, DET-1, showed a significant difference among these statistics in *P. trichocarpa,* whereas no significant differences were seen in any of the genes in *P. deltoides* (Figure 11 and 12).Tajima's D, one of the statistics most commonly used to test of neutrality, and was calculated both for individual genes and for functional gene groups compared with control sequences in both species. Control sequences, on average, had a negative Tajima's D.



Pair-wise Wilcoxon test

Measure	Level	Level	P-value
Pi	Dormancy	Control	0.0069*
Pi	Core Loop	Control	0.0459*
Pi	Peripheral	Control	0.0028*
Pi	Photoreceptor	Control	0.0012*
Theta	Dormancy	Control	0.0002*
Theta	Core Loop	Control	<.0001*
Theta	Peripheral	Control	<.0001*
Theta	Photoreceptor	Control	0.0086*
Hd	Photoreceptor	Control	0.0465*
Hd	Peripheral	Control	0.0166*

Figure 10. Box plots of different genes of circadian clock pathway and control sequences, for various nucleotide polymorphism statistics in *P. deltoides*. . Box edges represent the upper and lower quantile with median value shown as bold line in the middle of the box. Whiskers represent range from maximum to minimum. The table on right shows p-values for non-parametric pairwise Wilcoxon rank test.

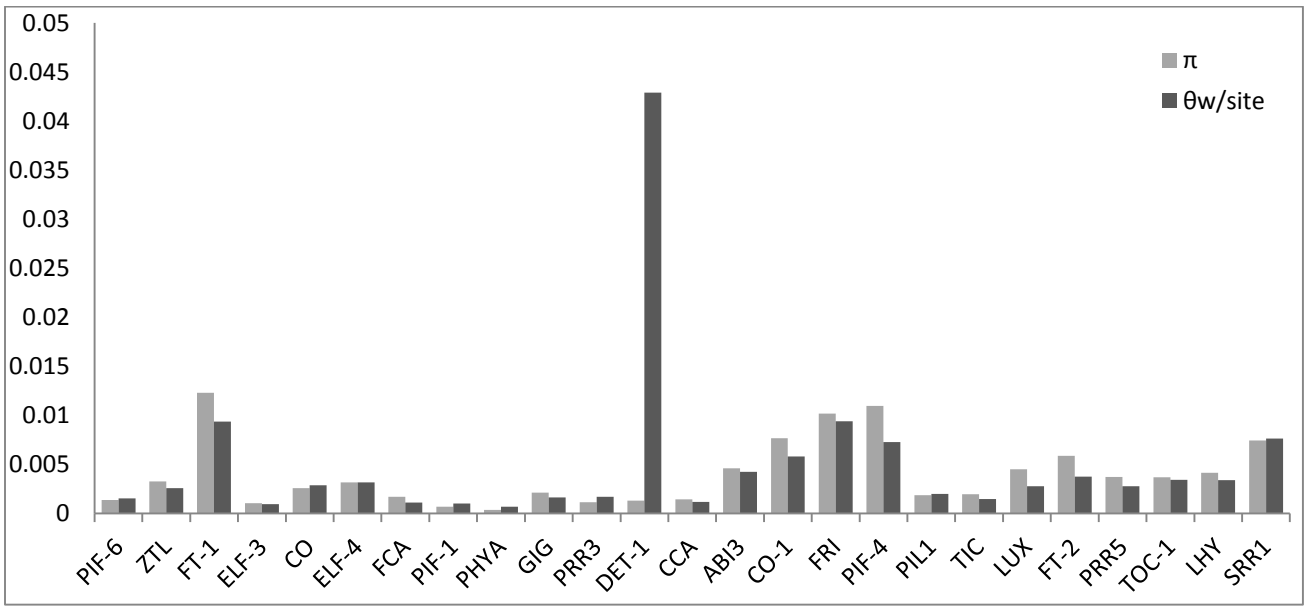


Figure 11. Comparison between mutation rate per site and pairwise nucleotide diversity in *P. trichocarpa*.

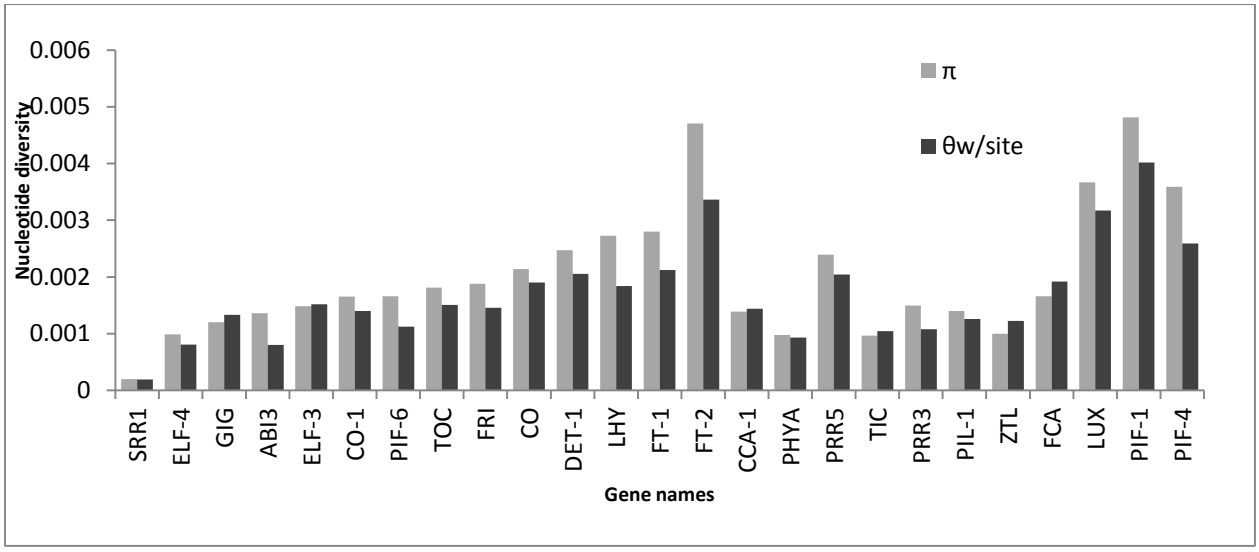


Figure 12. . Comparison between mutation rate per site and nucleotide diversity in *P. deltoides.*

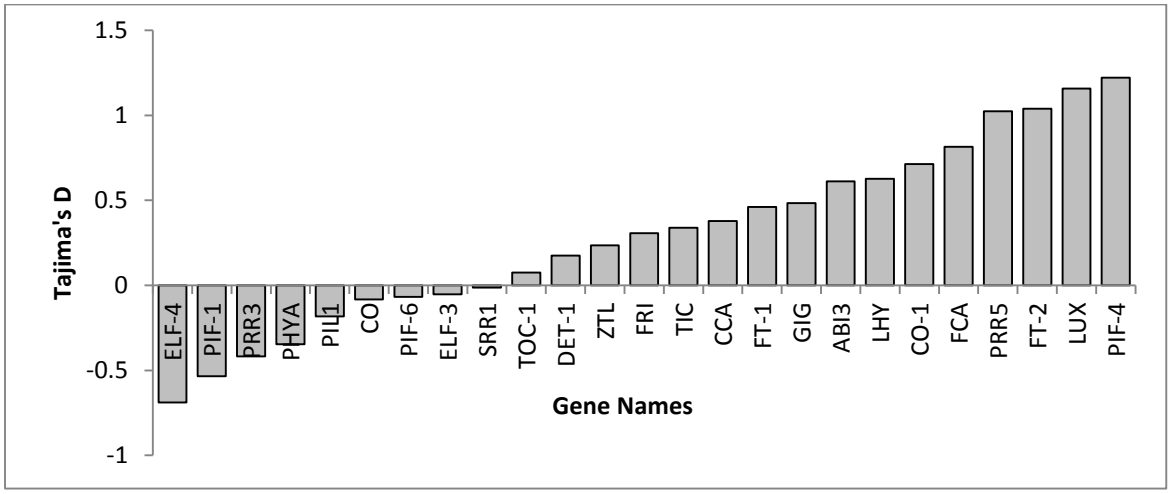


Figure 13. Pattern of Tajima's D statistics in *P. trichocarpa*.

In *P. trichocarpa* circadian clock genes on an average showed a positive Tajima's D (Figure 13). There was no change in pattern of Tajima's D, when whole gene was split into gene only and promoter regions. All functional gene groups except photoreceptors when compared to control sequences showed significant differences in the level of Tajima's D (at $\alpha = 0.05$), but no significant differences were found when functional gene groups were compared with each other (Figure 15).

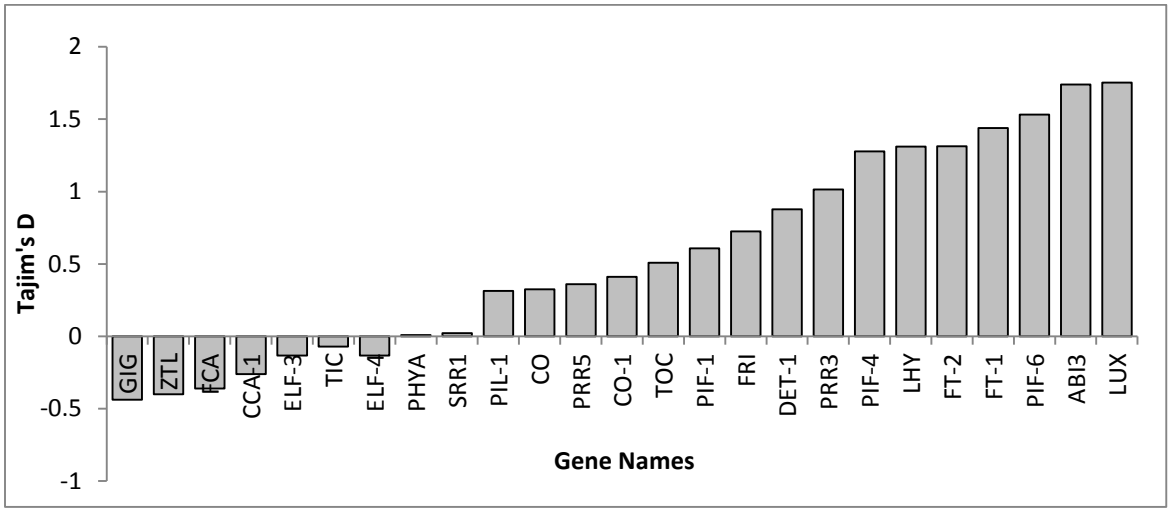
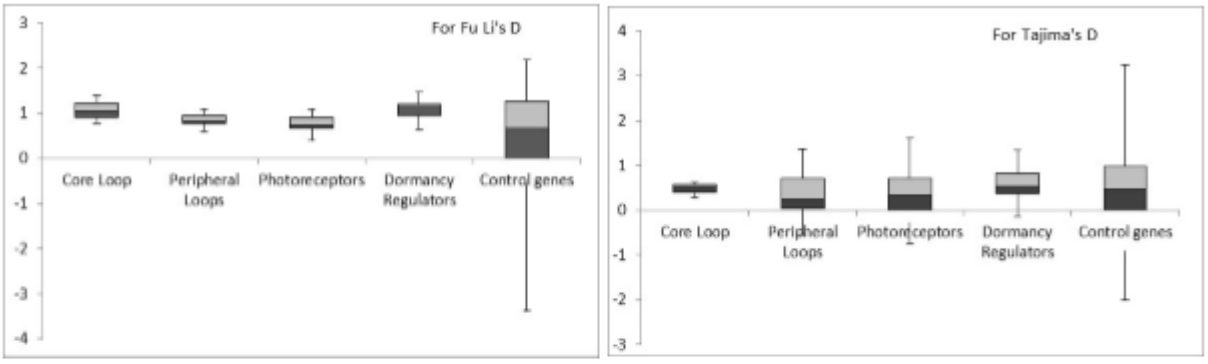
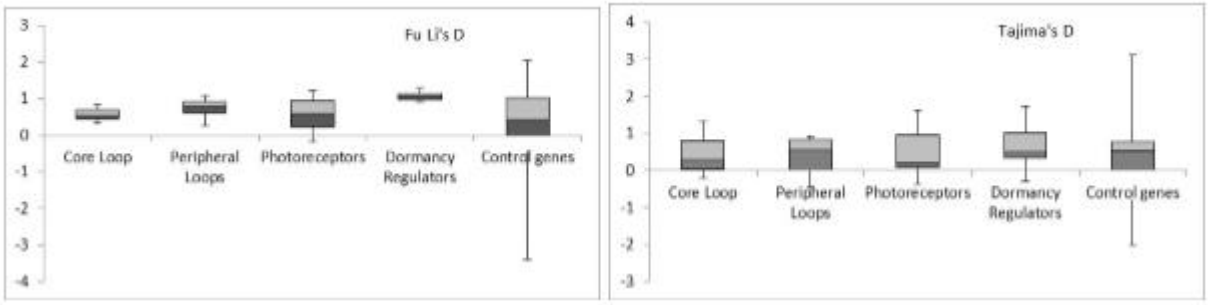


Figure 14. Pattern of Tajima's D statistics in *P. deltoides*.



Measure	Level	Level	P-value
Tajima's D	Core Loop	Control genes	0.0209*
Tajima's D	Dormancy	Control genes	0.0010*
Tajim's D	Peripheral	Control genes	0.0133*
Fu Li's D	Dormancy	Control genes	0.0002*
Fu Li's D	Core Loop	Control genes	0.0162*
Fu Li's D	Peripheral	Control genes	0.0021*
Fu Li's D	Photoreceptor	Control genes	0.0025*

Figure 15. Box plots between different classes of genes in circadian clock and control sequences for neutrality tests in *P. trichocarpa*. Box edges represent the upper and lower quantile with median value shown as bold line in the middle of the box. Whiskers represent range from maximum to minimum. The table below figures shows p-values for non-parametric pairwise Wilcoxon rank test.



Measure	Level	Level	P-value
Fu Li's D	Dormancy	Control	0.0011*
Fu Li's D	Photoreceptor	Control	0.0022*
Fu Li's D	Peripheral	Control	0.0098*
Tajima's D	Core	Control	0.0011*
Tajima's D	Dormancy	Control	0.0022*
Tajima's D	Peripheral	Control	0.0098*

Figure 16. Box plots between different classes of genes in circadian clock and control sequences for neutrality tests in *P. deltoides*. Box edges represent the upper and lower quantile with median value shown as bold line in the middle of the box. Whiskers represent range from maximum to minimum. The table below figures shows p-values for non-parametric pairwise Wilcoxon rank test.

Similar patterns were seen in *P. deltoides*. Most of the genes in *P. deltoides* had a positive Tajima's D (Figure 14). For functional gene groups in comparison with control sequences, most functional groups except photoreceptors showed significant differences, and matched the pattern same as in case of *P. trichocarpa* samples. This might reflect similar dynamics of control of same class of genes in both species (Figure 16).

The second neutrality test applied was Fu and Li's D statistic. In *P. trichocarpa* all genes on average had positive values for Fu and Li's D. When D values were calculated for each gene fragment, very few gene fragments showed a negative value (Figure 17). Some upstream regions of the genes showed significant values of D (Figure 17). Functional gene groups were compared among each other and with control sequences. All functional gene groups showed significant differences with respect to Fu and Li's D statistics when compared with the control sequences, but none of the functional groups had significant differences among one other. (Figure 15).

Similar patterns were seen in *P. deltoides*. All genes and most of the gene fragments had a positive Fu and Li's D, with some fragments showing significant values (Figure 18). In *P. deltoides* however, no promoter and gene-only divisions were done, as only few genes had their

upstream regions amplified and sequenced. All functional groups showed significant differences compared with control regions (Figure 16).

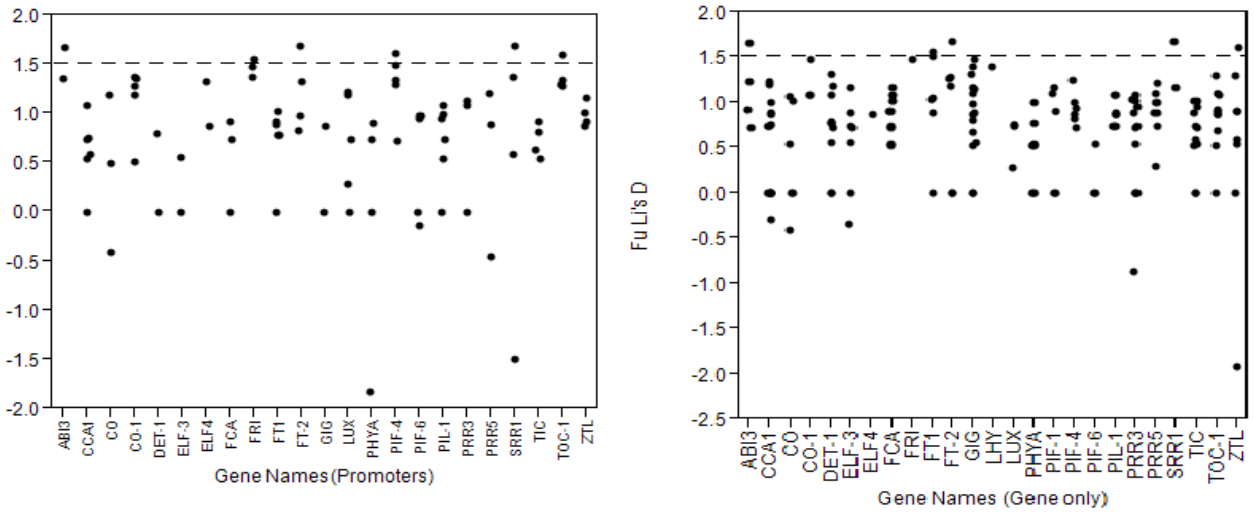


Figure 17. Scatter plot for Fu and Li's D statistics in *P. trichocarpa*. The dashed line represents cut-off and any outliers above the line are significant. Scatter-plot on left is for upstream regions, while the one on right, for gene regions.

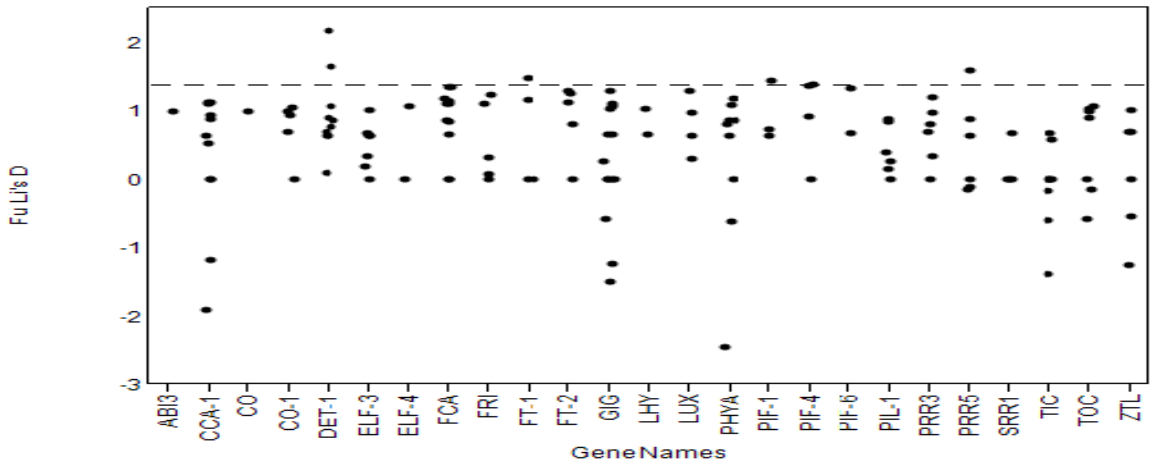


Figure 18. Scatter-plot for Fu and Li's D statistics in *P. deltoides*. The dashed line represent cutoff, so any outliers above the dashed line are significant.

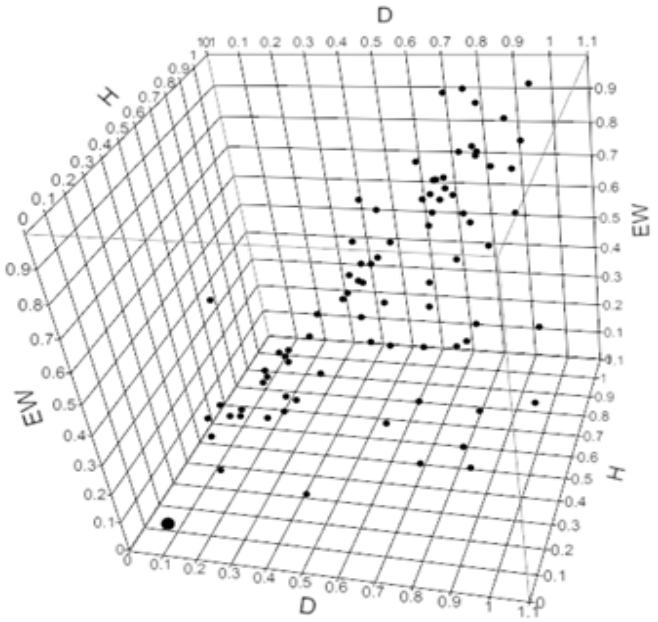
DH, HEW and DHEW test:

I performed the DH, HEW, and DHEW tests on all *P. trichocarpa* sequences. These tests were first performed on all gene fragments using mostly *P. deltoides* as the outgroup, and *P. tremula*

for those sequences where *P. deltoides* sequences were not available. This test was not completed in *P. deltoides*, as the numbers of sequences in each gene fragment were not large enough to do a robust test. PIF-1 was the only gene fragment with a significant p-value for all of these three tests, indicating of possible selective sweep at this sequence. Other than PIF-1, only the DET-1 gene fragment gave a significant HEW test result (Figure 19 and Table 4). None of the test values were significant when fragment values were averaged for each gene.

Tests for heterogeneity

I performed the HKA test on both *P. trichocarpa* and *P. deltoides* sequence, using them as reciprocal outgroups for one other. This test compares the level of intraspecific polymorphism with the level of between-species divergence, and was calculated both for gene fragments and total values calculated for the entire genes. Four genes and eight gene fragments in *P. trichocarpa* departed from neutral expectations (Figure 20). Interestingly all of these genes were either dormancy regulators or photoreceptors.



Name	DH	HEW	DHEW
DET_2T	0.225	0.045	0.177
PIF1_1T	0.014	0.0083	0.0072

Table 4. Gene fragments showing significant p-value in different compound tests.

Figure 19. 3-D scatter plot for DHEW test in *P. trichocarpa*. Each axis represents p-value for the gene fragments plotted.

P. deltoides, when compared with *P. trichocarpa* showed different outliers in terms of functional groups of genes in results from the HKA test. In *P. deltoides*, all of the outlier functional gene groups were photoreceptors, peripheral loops and core loop (Figure 20). At the gene level I saw

FCA, which is a dormancy regulator as an additional outlier. SRR1, FT2, and PIF-1 were outliers for both of these species.

I performed the MK test on *P. trichocarpa* sequences, and mostly used *P. deltoides* as the outgroup unless *P. deltoides* sequences were not available for specific gene fragments; in that case, *P. tremula* sequences were used. The results included neutrality index and direction of selection tests.

I calculated both the Neutrality index by Tarone and Greenland (NI_{TG}), which is a weighted measure of neutrality index and direction of selection (DOS) (Stoletzki and Eyre-Walker, 2011) estimates for each gene and its gene fragments. Most of the genes showed neutrality index less than or equal to one, but four genes, FRI, PRR5, TIC1, and CO showed a very high value of NI (Figure 21) indicating strong purifying selection in these genes.

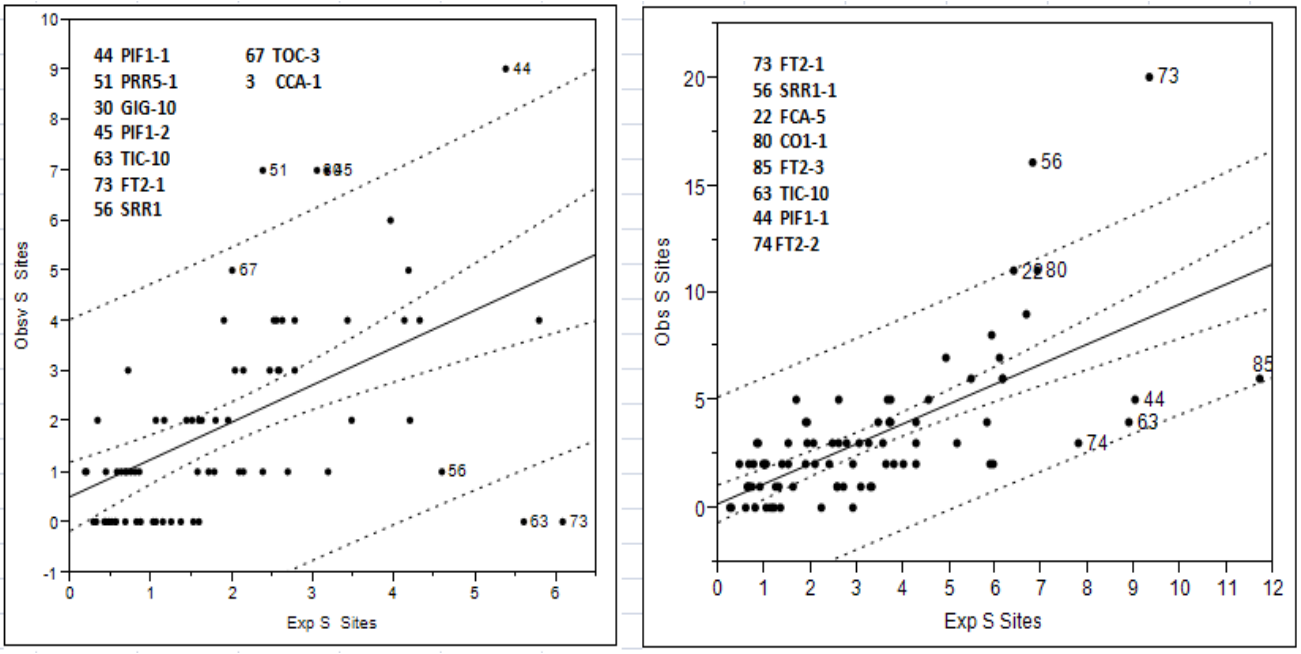


Figure 20. HKA test for selection at phenology candidate genes. The solid line represents the best fit and dashed lines represent prediction intervals. Significant outliers were the ones outside these intervals and were labeled. The figure on left is for *P. deltoides* and on right for *P. trichocarpa.*

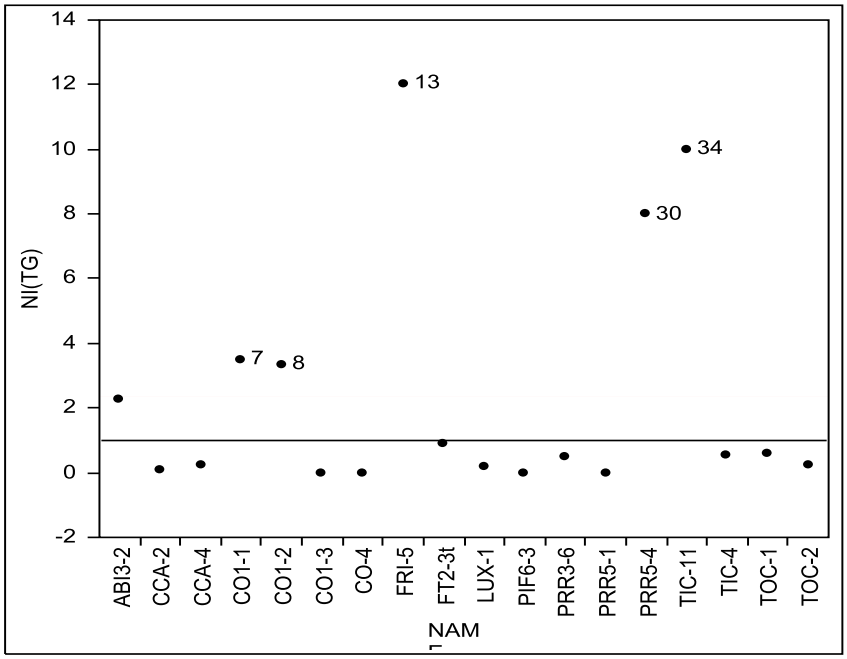


Figure 21. Neutrality Index, calculated using the CMH method for phenology genes in *P. trichocarpa.* The solid represent NI = 1, which means neutral evolution.

I also calculated direction of selection for each gene. This measure tends to be less biased than neutrality index and, being the difference between proportion of substitutions and proportion of polymorphism, is less biased than NI (Stoletzki and Eyre-Walker, 2011). In *P. trichocarpa*, many loci showed a negative DOS, indicating the presence of segregating deleterious mutations and thus purifying selection. Some loci had a positive value of DOS indicating adaptive evolution in the sequences (Figure 22). The two note-worthy outliers were PRR5 (which had a very high value of DOS>1) and CCA-1 (with DOS of -1).

Discussion

Population structure and demography

Population structure was not pronounced in *P. trichocarpa* samples, and assignment test analysis revealed three population groups or sub-populations among sampled poplars. F_{ST} estimates were very low among these sub-populations with values ranging from 0.05-0.1 (Figure S4 and S5). *Populus* species. are wind-pollinated with high levels of dispersal (Stettler and National Research Council, 1996), and the relative homogeneity of populations across broad distances indicates high levels geneflow in my populations.

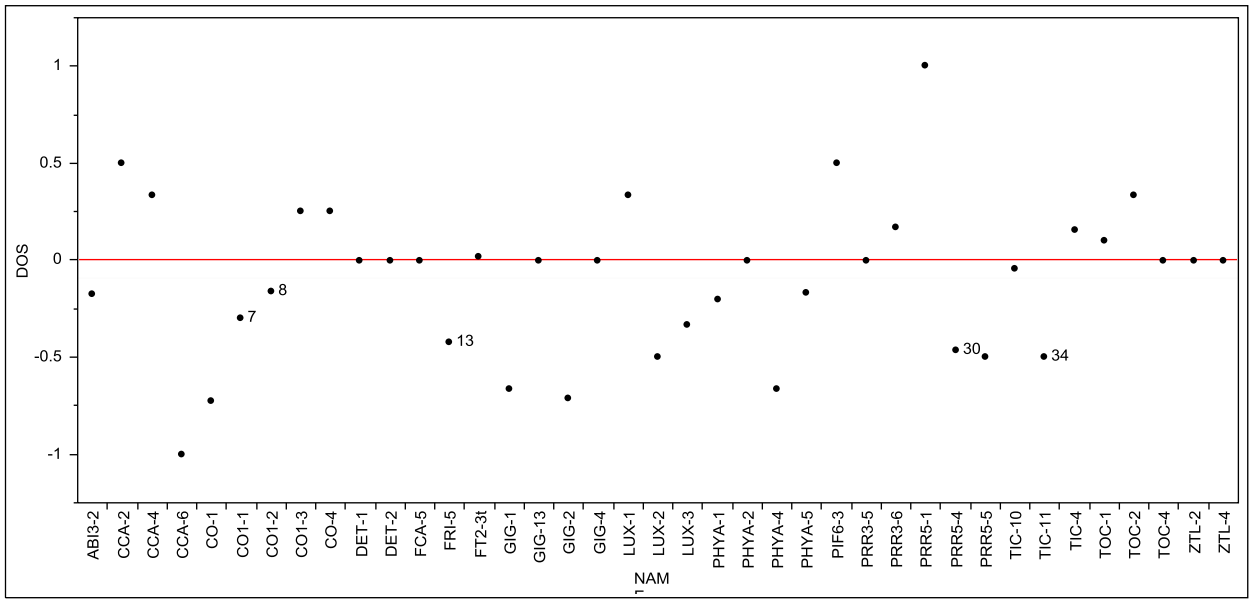


Figure 22. Direction of selection (DOS) estimates for all gene fragments in *P. trichocarpa*. Red line at reference point represents neutrality.

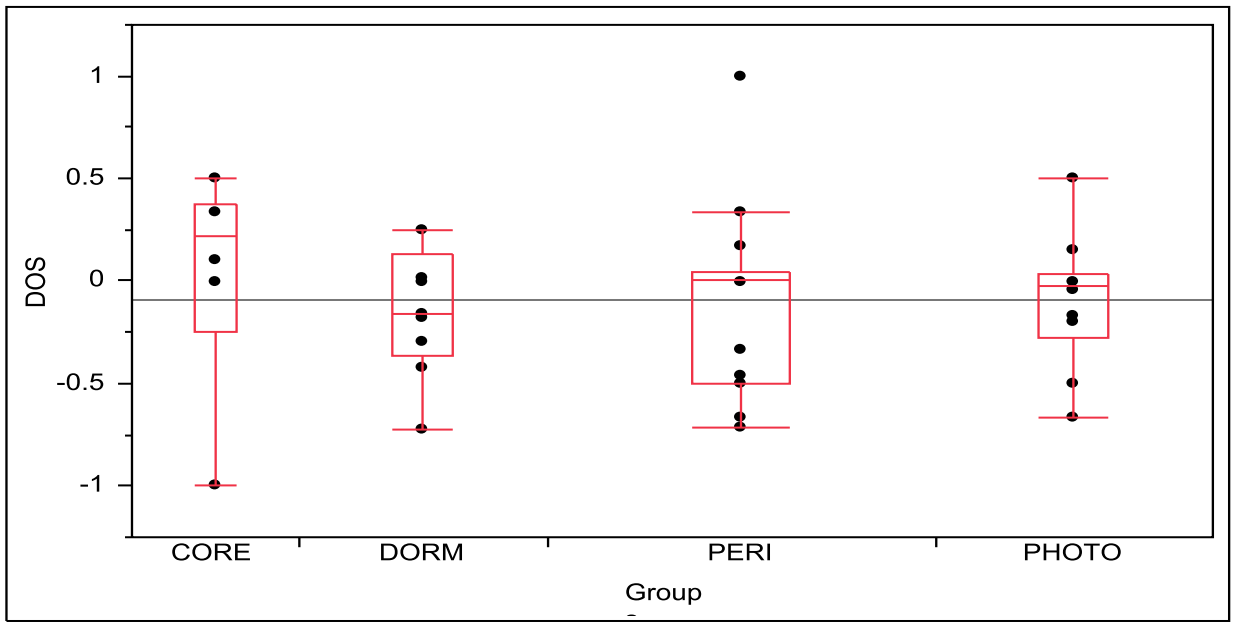


Figure 23. Box plots showing direction of selection estimates for all classes of circadian clock genes.

Demography, like population structure, is one of the major confounding forces with natural selection, and can lead to various false positives, as it leaves very similar genetic signatures (Thornton et al., 2007).

My ABC simulations suggested that a bottleneck model is most probable, with a posterior probability of 0.76. The probability of the growth model was also fairly high, suggesting that the signal in the data may not be strong enough to differentiate these models. The empirical data used are drawn from samples along a continuous geographical gradient from Alaska to the Sierra Nevada populations. Theoretically, the populations in the south and center of the range acted as refugia for northern populations, and should to some extent have recovered from the bottleneck caused by glacial maxima, therefore leading to high fit for both models. I simulated a single population, and this may not be a sufficient approximation of natural circumstances. In the future, a more complex refugial model involving multiple populations with gene-flow and multiple bottleneck/growth events might provide a better fit to the data. Nevertheless, these coalescent simulations suggest that neutral equilibrium is a very unlikely scenario, at least in *P. trichocarpa,* this outcome indicates that it was important to use tests robust to these demographic effects, and interpret with care those tests which are subject to skew by demography.

Patterns of nucleotide diversity and neutrality tests

I investigated patterns of nucleotide diversity in 25 candidate genes, in both *P. trichocarpa* and *P. deltoides*. *P. trichocarpa* showed more nucleotide diversity than *P. deltoides* populations. This could be in part because I had more samples for *P. trichocarpa,* or because the geographical range of *P. trichocarpa* is much wider than *P. deltoides,* successively leading to more genetic variation in *P. trichocarpa*. Also, the two species have experienced different demographic histories, which plays a major role in divergence of species (Ismail et al., 2012). *P. trichocarpa* is restricted to the western coast of North America, whereas *P. deltoides* occurs mostly in the mid-west and southeastern United States. These ecological differences may influence the targets of selection in the two species. Whereas *P. trichocarpa* is largely restricted to coastal temperate habitat, the continental habitat of *P. deltoides* is hotter and more arid. By contrast, *P. trichocarpa* inhabits a more diverse latitudinal range that may lead to selective pressures related to low temperatures and more extreme changes in day length.

These differences are also prominent in other *Populus sp*, e.g. *P. tremula* (European aspen) has much higher nucleotide diversity, nearly 3.5 times more than its North American counterparts, *P. trichocarpa* and *P. balsamifera* (Ismail et al., 2012). Similar results were obtained in *Arabidopsis*, in which its wild relative *Arabidopsis lyrata*, which occupies different and broader geography than *Arabidopsis thaliana*, showed higher diversity than *A. thaliana* (Foxe et al., 2008). Moreover, *P. trichocarpa* went through a massive range expansion, and is still expanding, while *P. deltoides* has not expanded to this extent. This might lead to increase in number of segregating sites and thus increase in diversity estimates in *P. trichocarpa*.

P. trichocarpa:

I calculated nucleotide polymorphisms of both genes and functional gene groups in *P. trichocarpa*. Classification of gene groups was based on their functions and positions in the circadian clock pathway. Even though functions of most of the genes have been validated in past molecular studies, they still to a large extent remain ambiguous, as function is validated in *Arabidopsis thaliana* and other annuals, with only a few studies based on perennials. The basic criterion behind classification of these genes was to test for selective constraints on each gene group. It is believed that genes which are in core loop are most important genes and should have more linkages to other pathways, or in other words are more pleiotropic than other genes, thus should show more selective constraints as compared to other genes (Hall et al., 2011; Ramsay et al., 2009).

Interestingly, there was no significant difference in nucleotide diversity among the gene groups, but all of the genes showed significant difference in diversity when compared with control sequences. Similar results were obtained in *P. tremula* (Hall et al., 2011), where all classified gene groups showed nearly same amount of selective constraints irrespective of their pleiotropy. This might indicate the nature of adaptation among gene families and its role in differences of phenological responses among different geographical phenotypes.

Looking at the synonymous and non-synonymous nucleotide diversity among 25 candidate genes revealed SRR1, PIF1, FRI and CO-1 as having high non-synonymous nucleotide diversity. On the other hand, synonymous nucleotide diversity was more evenly spread in circadian clock genes, with the highest levels in LHY, TOC1 and PRR5. Non-synonymous mutations are

typically understood to be more deleterious than synonymous mutations, as they lead to amino acid changes that may affect the functioning of proteins. Non-synonymous mutations should therefore face more selective constraint compared to synonymous mutations. SRR1 and PIF1 are Phytochrome–interacting factors; with SRR1 being involved in mediating the red–light signaling pathway induced by PHYB (Staiger et al., 2003), and PIF1 interacts with phytochromes to regulate shade avoidance responses in plants (Franklin and Whitelam, 2005; Leivar et al., 2012). So, amino acid changes in these genes might have a role in gating the circadian clock to different photoperiods and develop shade-avoidance responses in the natural environment. FRI and CO-1 on the other hand are well–defined downstream targets which are fed by circadian clock pathways and are involved in regulating growth cessation and dormancy in perennials (Rohde and Bhalerao, 2007). FRIGIDA (FRI) is a flowering time gene involved in temperature induced phenological responses (Lizal, Pavel, 2005) and CO-1 is one of the paralogs of the CONSTANS gene, central in mediating short-day signals in perennials (Böhlenius et al., 2006). Various studies in *Arabidopsis* (Fournier-Level et al., 2011; Matthew et al., 2012; Le Corre et al., 2002) and the recent ones in *Populus* have shown FRI to be a candidate gene for local adaptation (Keller et al., 2011), with different haplotypes in different temperature zones, putatively reflecting its role in local adaptation. So, these non-synonymous changes probably can be expected to have a role in growth cessation and cold acclimation.

Upstream regions of genes showed very high selective constraint, with most of the genes showing no to very low nucleotide diversity. Three genes (FT2, FRI, and GIGNATEA) showed very high nucleotide diversity in the first 100–bp region of a gene. Interestingly, FT2 showed very high nucleotide diversity in upstream region, whereas in gene region the diversity was extremely low in both synonymous and non-synonymous variants types. Upstream regions normally harbor promoter regions, responsible for transcriptional regulation of gene, and this variation could play a role in regulating transcription of FT2, thus affecting regulation of dormancy in perennials (Chen et al., 2011; Chung et al., 2006). Moreover, promoters have a special role to play in the circadian clock and its connected pathways. Circadian clock and dormancy regulation pathways have transcription factors, which bind to the promoter regions of other genes to regulate feedback loops. These genes need to be phosphorylated for their proper functioning, decreasing the availability of binding sites, and subsequently having an affect on the whole biochemical pathway (McClung, 2011). FT2 recently was shown to have a role in warmer

temperatures and long growing season's post-dormancy, therefore is believed to have some role in regulating vegetative growth (Hsu et al., 2011). FT2, moreover, was reported to interact with the TFL1 gene to regulate flowering in *Arabidopsis* (Hanano and Goto, 2011). TFL1 represses bud flush in *Populus* species; therefore, its promoter variation might signify the role in TFL1-FT2 interactions affecting the bud flush in *Populus* species as it was shown in *Arabidopsis*.

I investigated neutrality of genes using a variety of tests, including those based on allele frequency spectrum, haplotype frequency-based tests and tests of heterogeneity which use two more genes and compare polymorphism and divergence patterns. The most common neutrality test is Tajima's D (Tajima, 1989), which compares two estimators of diversity - theta and pi - which should be equal under neutral equilibrium. I compared both of these statistics and found DET-1 to be the only gene showing very high value of θ_w/site as compared to value of π, therefore showing positive selection on a variant of this gene.

Tajima's D was calculated for all sequences, and interestingly most of the genes gave a positive value of D, whereas past studies have found the genome-wide average for this statistics to be negative in trees (Ismail et al., 2012; Holliday et al., 2010; Keller et al., 2011), which was the case of control sequences. This could be attributed to two possible reasons, the first one being as an artifact of pooling the samples together. Pooling different geographical samples that experience different selection pressures together could result in a positive Tajima's D due to different alleles being present at high frequency in alternative populations (Städler et al., 2009). In this way, a positive Tajima's D in a range-wide sample can be a signature of divergent selection across the range. Another reason could be a strong demographic signature left by post-glacial migration. Most of the populations except those in the south migrated to newly ice-free habitat within the last 15,000 years. These populations had very limited founders in the beginning and restricted gene-flow, which leads to an apparent excess of medium-frequency mutations and positive Tajima's D. As these species in Canada and the Northern U.S. are comprised of fairly recent stands, they still are suspected to have that genetic imprint of bottleneck, thus leading to positive Tajima's D.

Some gene fragments had a significant Tajima's D and interestingly all of them were positive. . All of the gene outliers were similar to those for nucleotide diversity (ZTL-5, PRR5, TOC1, PIF4, and LUX). These could be a selection signature, but could equally be suspected from the

pooling effect, which might have led to an increase in medium–frequency alleles. Similar results were obtained when I compared this statistic among classified gene groups and to those of control sequences. None of the groups showed significant differences among themselves, but all of them showed significant differences from control sequences. Nothing could be interpreted from this significance, owing to the simplicity of Tajima' D and its sensitivity to demography and pooling effects.

Fu and Li's D is another allele frequency spectrum–based test, and has the same basis behind interpreting neutrality as Tajima' D, but is less sensitive to demography (Städler et al., 2009). I obtained similar results with Fu and Li's D both with respect to gene fragments and gene groups.

To assess possible selection signatures with more robust tests, I used DH, HEW and DHEW tests. These are compound tests that use a combination of neutrality tests to detect selection. So, DH is combination of Tajima's D and Fay Wu's H, HEW is Fay Wu's H and Emmen-Watterson test and DHEW is combination of DH and HEW tests together (Zeng et al., 2007). Simulation studies show these statistics to be more robust than each of these single tests, as they combine frequency spectrum– and haplotype spectrum–based tests, and are not sensitive to recombination.

I found two outliers, PIF1 and DET1, with extremely low p-values. PIF1 was significant in all three combination tests and DET1 only in the HEW test. The HEW test is sensitive to recent bottleneck events and considering my most probable model using ABC simulations being bottleneck model, DET1 variation could be a demographic signature rather than a selective sweep. PIF1, on another hand, is a photoreceptor with its role in optimizing photomorphogenesis in *Arabidopsis* (Shen et al., 2005). It is antagonistic to light and modulates absisic acid and gibberellic acid pathways. It interacts with the PHYA and PHYB pathway and regulates light signaling, therefore should have an important role in regulation of phenological responses in perennials.

Tests of heterogeneity

The HKA test showed SRR1 and FT2 to have an excess of segregating sites over expected. Both of these gene fragments were near upstream regions of the gene, thus might signify balancing selection near promoter regions. These two genes even showed significantly high non-

synonymous polymorphism, and this variation might signify both gating of light signals to regulate circadian clock, and regulating FT2-TFL1 to regulate bud–flush timings. There were six other outliers departing from neutrality, mostly in photoreceptor and dormancy regulator classes. In order to verify the credibility of these outliers, I used the McDonald-Kreitman test (1991), which uses the synonymous and non-synonymous ratio in the coding region both within and between species to detect selection (Nielsen, 2005). I calculated the neutrality index for all the gene fragments using the Cochran-Mantel-Haenszel (CMH) procedure. This procedure is preferred to normal averaging of the neutrality index (NI) for all gene fragments, as the latter will overestimate the value of NI leading to false positives (Stoletzki and Eyre-Walker, 2011). I also calculated the direction of selection (DOS) estimate for each gene. This measures both the direction and extent of selection, and can be defined as proportion of substitutions to proportion of polymorphisms that are non-synonymous.

Three genes – FRI, PRR5 and TIC – showed strong negative selection (NI >8). Similarly, in DOS estimates, the majority of the genes showed negative selection, with only five showing signatures of positive selection. DOS is an unbiased measure as it is a difference in proportions and is considered robust when data are sparse (Stoletzki and Eyre-Walker, 2011). I compared the DOS within the classified gene groups and found no significant difference among the gene groups (Figure 23). A surprising observation was that same gene's different gene fragments showed different directions of selection. For example in PRR5, regions near promoters (150 bp) showed a significantly positive selection, but its other gene fragments, PRR5-4 and PRR5-5 showed very strong negative selection. Similar patterns could be seen in CCA-1, which is a core loop gene, and its 5'UTR linked gene region showed a very strong negative selection, while the rest of the gene showed positive selection. Very similar patterns were seen in CONSTANS where first 200–bp region showed strong negative selection, while the rest of the gene region showed a weak but positive selection signature

This body of findings gives strong evidence of the complex nature of adaptation in circadian clock genes. There is difference in selection pressure within the genes, with the transcription regulatory regions or regions near them (linked) showing strong negative selection, while other gene fragments show a positive or negative selection, depending on the gene. PHYA was the only gene among photoreceptors that showed consistent strong negative selection, while other

photoreceptors interacting factors like PIF-6 and TIC showed varying selection pressure. These kinds of patterns might reflect strong selection in core photoreceptor genes like PHYA, which regulate the red light–induced light signaling pathway (Shen et al., 2009), and more selective freedom to peripheral connections. This could possibly be interpreted as peripheral photoreceptor–interacting factors having more functional importance in geographical phenological responses. PIF-6 has been reported to modulate shade-avoidance responses in plants (Franklin and Whitelam, 2005; Leivar et al., 2012), and TIC likewise is involved in modulating light signaling based on timing of day in plants (McClung, 2011). These genes (PIF-6 and TIC) role in entrainment of circadian clock to light might play a key role in determining the optimum time for dormancy in plants.

P. deltoides:

Like *P. trichocarpa*, diversity in *P. deltoides* was evenly distributed across gene classes, suggesting little selective constraint based on pleiotropy. There were no significant differences between different functional classes of genes, but they showed significant differences when compared with that of control sequences (Figure 10).

Average nucleotide diversity in *P. deltoides* was low compared to *P. trichocarpa*. Two genes, PIF1 and PIF4 showed high levels of both synonymous and non-synonymous nucleotide diversity. These two genes were outliers even in *P. trichocarpa*, which might indicate their adaptive significance in both species.

I used both allele frequency spectrum-based and heterogeneity-based tests to investigate neutrality of candidate genes in *P. deltoides*.

Tajima's D and Fu and Li's D revealed similar patterns, as were seen in *P. trichocarpa*. Most of the genes showed a positive Tajima's D on an average. This is interpreted most likely as artifact of pooling all of the samples together, as explained in the case of *P. trichocarpa* samples. Another possible reason could be a strong demographic signature, i.e., of a bottleneck in my populations which might have inflated the statistics.

At the gene level, results were different among Fu and Li's D and F statistics. These tests are less sensitive to demography and revealed more outliers, with notable ones being PRR5, DET1, PIF1

and PIF4. These targets were similar to ones seen in *P. trichocarpa*. The significantly positive value of Fu and Li's D in these genes fragments indicates balancing selection upon them.

Test of heterogeneity

The HKA test showed many outliers with most of them showing negative or purifying selection. PIF1, PRR5, and GIGANTEA showed higher number of segregating sites than expected, indicating excessive polymorphism, signifying balancing selection. On the other hand, the upstream region of the FT2 and TIC genes showed significantly lower numbers of segregating sites than expected hinting of a recent selective sweep. However, the HKA test is sensitive to demography, though somewhat less so than Tajima's D and related measures, and some the outliers could be false positives and might be more of a demographic signature than of natural selection. However, it is also worth considering that most of the HKA outliers were also outliers in Fu and Li's and Tajima's D statistics and were similar to the ones seen in *P. trichocarpa* studies, suggesting that they have some role in adaptation. The McDonald-Kreitman test could not be used because of outgroup constraints.

Conclusion

Among 25 candidate genes involved in light perception, circadian clock and dormancy regulation, *P. trichocarpa* showed many candidate targets of selection. Eight candidate genes were consistent outliers among all neutrality tests, suggesting that these are credible targets of natural selection. These were TIC, CO, FT2, FRI, PRR5, SRR1, PIF1and PIF6, and belong to all four classified gene groups. This suggests that phenological adaptation likely involves multiple genes, a result that is corroborated by other recent studies (Hall et al., 2011; Keller et al., 2011). Moreover when genes were examined at the fragment level, I found several instances of varying selection pressure along the gene, with upstream regions being under strong selective constraint. This suggests that adaptive responses are complex and localized at various positions within a single gene. Gene groups showing a fairly even distribution of diversity, irrespective of their position in the clock and their functional importance may hint of co-adaptation of these gene complexes, rather than each gene being independent targets (Hall et al., 2011) .

Similar putative targets of selection were obtained in *P. deltoides. P. deltoides* occupies a non-overlapping and quite different climatic range compared with *P. trichocarpa*, although the need

for adaptation along climatic gradients, using day-length as a proxy for the seasons, remains. It is therefore not surprising that I saw similar genetic outliers in both species. This similarity might indicate that the genes have a universal role in phenology and circadian cycle regulation. This similarity already was seen in previous studies of *Arabidopsis thaliana* and *Arabidopsis lyrata*, two species that occupy very different ranges but show similar targets of selection (Foxe et al., 2008; Hancock et al., 2011).

This study has revealed several candidate targets of natural selection in two cottonwood species, some of which already have been validated and some which were brought in attention with this study. Although I applied a suite of tests to identify targets of natural selection, validation on a larger scale is warranted owing to the complexity of circadian clock and phenological responses in trees. Recently, various population genomic studies have revealed circadian clock genes to be targets of local adaptation in annuals and perennials, but the targets differ hugely with respect to the geographical locations of the populations. For example studies of *Pinus sp.* in southern North America have revealed genes related to heat and aridity (Eckert et al., 2010), whereas coastal and northern latitude studies on perennials revealed genes with photoperiod and cold response as outliers (Holliday et al., 2010; Keller et al., 2011). Moreover recent studies have shown an intercellular, cytosol– and nuclear–based receptor sensitive regulation of dormancy in trees (Rinne et al., 2011), suggesting much more complex and sophisticated regulation of dormancy in trees. Hormones have a significant role in regulating growth in plants, and are vital for all sorts of metabolic activities. As dormancy is not complete shutdown of metabolic machinery, and in more precise words is a decrease in metabolic rate of activity, there should be some hormonal flow within a plant (Rinne et al., 2011; van der Schoot and Rinne, 2011). This generality does not hold because same hormones like gibberellins regulate growth which is ceased in dormancy. Recent studies have shown a role of plasmodesmata or inter-cellular connections to play role in affecting hormonal transfer (Rinne et al., 2011; van der Schoot and Rinne, 2011).

In addition, the circadian clock has a various stages and sub-stages of regulation, with major ones being: environment, transcriptional, post-transcriptional, and post-translational (McClung, 2011). Troein et al., (2009) showed that multiple photoperiod inputs, cues from seasons and weather, and noise in light signals all favor evolution of a complex network. Similar results were seen by Edwards et al. (2010), whose simulation study tested network complexity for efficient

tracking of dawn and dusk. If all these studies are further expanded with biotic and abiotic stresses taken into consideration, they will demand a more complex network. Moreover, these mathematical analyses have suggested that increase in complexity favors stability and flexibility, with more robust entrainment to environmental signals (McClung, 2011). Therefore, a more comprehensive study at the genome-wide level, with recombination and haplotype parameters taken into consideration, will give us better insight into new targets of natural selection, which play a significant role in phenology and local adaptation of perennials.

References

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

Andolfatto P. Adaptive evolution of non-coding DNA in *Drosophila*. Nature (2005) 437:1149-1152.

Andolfatto P, Przeworski M. A Genome-Wide departure from the standard neutral model in natural populations of *Drosophila*. Genetics (2000) 156:257-268.

Barton NH. Clines in polygenic traits. Genet Res (1999) 74:223-236.

Beaumont MA, Zhang W, Balding DJ. Approximate Bayesian Computation in Population Genetics. Genetics (2002) 162:2025-2035.

Begun DJ, et al. Population Genomics: Whole-Genome analysis of polymorphism and divergence in *Drosophila simulans*. In: PLoS Biol (2007): Public Library of Science. e310.

Biswas S, Akey JM. Genomic insights into positive selection. Trends in Genetics (2006) 22:437-446.

Böhlenius H, et al. CO/FT regulatory module controls timing of flowering and seasonal growth cessation in trees. Science (2006) 312:1040-1043.

Bradshaw, H.D., Otto, K.G., Frewen, B.E., McKay, J.K. and Schemske, D.W. 1998 Quantitative Trait Loci Affecting Differences in Floral Morphology Between Two Species of Monkeyflower (Mimulus). Genetics, 149, 367-382.

Brown GR, Gill GP, Kuntz RJ, Langley CH, Neale DB. Nucleotide diversity and linkage disequilibrium in loblolly pine. Proceedings of the National Academy of Sciences of the United States of America (2004) 101:15255-15260.

Brunner AM, Busov VB, Strauss SH. Poplar genome sequence: functional genomics in an ecologically dominant plant species. Trends in Plant Science (2004) 9:49-56.

Campbell RK. Genecology of Douglas Fir in a watershed in the Oregon Cascades. Ecology (1979) 60:1036-1050.

Cheddadi R, et al. Imprints of glacial refugia in the modern genetic diversity of *Pinus sylvestris*. Global Ecology and Biogeography (2006) 15:271-282.

Chen C-H, Lin H-Y, Pan C-L, Chen F-C. The genomic features that affect the lengths of 5' untranslated regions in multicellular eukaryotes. BMC Bioinformatics (2011) 12:S3.

Cheung, V.G., Spielman, R.S., Ewens, K.G., Weber, T.M., Morley, M. and Burdick, J.T. 2005 Mapping determinants of human gene expression by regional and genome-wide association. Nature, 437, 1365-1369.

Chung B, Simons C, Firth A, Brown C, Hellens R. Effect of 5'UTR introns on gene expression in *Arabidopsis thaliana*. BMC Genomics (2006) 7:120.

Clausen J. Experimental studies on the nature of species. III. Environmental responses of climatic races of *Achillea*. Carnegie Inst. Wash. Publ. (1948) 581:1.

Colosimo, P.F., Peichel, C.L., Nereng, K., Blackman, B.K., Shapiro, M.D., Schluter, D. et al. 2004 The Genetic Architecture of Parallel Armor Plate Reduction in Threespine Sticklebacks. PLoS Biol, 2, e109.

DeCook, R., Lall, S., Nettleton, D. and Howell, S.H. 2006 Genetic Regulation of Gene Expression During Shoot Development in Arabidopsis. Genetics, 172, 1155-1164.

Derory J, et al. Contrasting relationships between the diversity of candidate genes and variation of bud burst in natural and segregating populations of European oaks. Heredity (2010) 104:438-448.

Dvornyk V, Sirviö A, Mikkonen M, Savolainen O. Low nucleotide diversity at the pal1 locus in the widely distributed *Pinus sylvestris*. Molecular Biology and Evolution (2002) 19:179-188.

Earl D, vonHoldt B. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources (2012) 4:359-361.

Eckert AJ, Bower AD, Gonzalez-Martinez SC, Wegrzyn JL, Coop G, Neale DB. Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). Molecular Ecology (2010) 19:3789-3805.

Edwards KD, et al. Quantitative analysis of regulatory flexibility under changing environmental conditions. Mol Syst Biol (2010) 6:424.

Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software structure: a simulation study. Molecular Ecology (2005) 14:2611-2620.

Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): An integrated software package for population genetics data analysis. Evol Bioinform (2005) 1:47-50.

Franklin KA, Whitelam GC. Phytochromes and shade-avoidance responses in plants. Annals of Botany (2005) 96:169-175.

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

Foxe JP, Dar V-u-N, Zheng H, Nordborg M, Gaut BS, Wright SI. Selection on amino acid substitutions in *Arabidopsis*. Molecular Biology and Evolution (2008) 25:1375-1383.

García-Gil MR, Mikkonen M, Savolainen O. Nucleotide diversity at two phytochrome loci along a latitudinal cline in *Pinus sylvestris*. Molecular Ecology (2003) 12:1195-1206.

Garcia MV, Ingvarsson PK. An excess of nonsynonymous polymorphism and extensive haplotype structure at the PtABI1B locus in European aspen (*Populus tremula*): a case of balancing selection in an obligately outcrossing plant? Heredity (2007) 99:381-388.

Giertrych M, (1991) Provenance variation in growth and phenology. In "genetics of Scots Pine", Giertych M., Matyas Cs eds, Elsevier, pp 87-98

González-Martínez SC, Krutovsky KV, Neale DB. Forest-tree population genomics and adaptive evolution. New Phytologist (2006) 170:227-238.

Google Inc. (2012). Google Earth (Version 6.2.2.6613) software. Available from http://www.google.com/earth/download/ge/agree.html

Gnirke A, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol (2009) 27:182-189.

Grivet D, Sebastiani F, González-Martínez SC, Vendramin GG. Patterns of polymorphism resulting from long-range colonization in the Mediterranean conifer Aleppo pine. New Phytologist (2009) 184:1016-1028.

Gyllenstrand N, Clapham D, Källman T, Lagercrantz U. A Norway Spruce FLOWERING LOCUS T homolog is implicated in control of growth rhythm in conifers. Plant Physiology (2007) 144:248-257.

Hall D, Ma X-F, Ingvarsson PK. Adaptive evolution of the *Populus tremula* photoperiod pathway. Molecular Ecology (2011) 20:1463-1474.

Hallerman, E.M. 1984 Isozyme Allele Frequencies in Five Strains of Channel Catfish, Ictalurus Punctatus, Subjected to Selection for Rapid Growt. PhD, Auburn University.

Hanano S, Goto K. *Arabidopsis* TERMINAL FLOWER1 is involved in the regulation of flowering time and inflorescence development through transcriptional repression. The Plant Cell Online (2011) 23:3172-3184.

Hancock AM, et al. Adaptation to climate across the *Arabidopsis thaliana* Genome. Science (2011) 334:83-86.

Hartl DL. A primer of population genetics (2000) Sunderland, Mass.: Sinauer Associates.

Hartl DL, Nurminsky DI, Jones RW, Lozovskaya ER. Genome structure and evolution in *Drosophila*: applications of the framework P1 map. Proceedings of the National Academy of Sciences, USA (1994) 91:6824-6829.

Haygood R, Fedrigo O, Hanson B, Yokoyama K-D, Wray GA. Promoter regions of many neural- and nutrition-related genes have experienced positive selection during human evolution. Nat Genet (2007) 39:1140-1144.

Heuertz M, et al. Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway Spruce [*Picea abies* (L.) Karst]. Genetics (2006) 174:2095-2105.

Holliday JA, Yuen M, Ritland K, Aitken SN. Postglacial history of a widespread conifer produces inverse clines in selective neutrality tests. Molecular Ecology (2010) 19:3857-3864.

Holsinger KE, Weir BS. Genetics in geographically structured populations: defining, estimating and interpreting F_{ST}. Nat Rev Genet (2009) 10:639-650.

Hsu C-Y, Liu Y, Luthe DS, Yuceer C. Poplar FT2 shortens the juvenile phase and promotes seasonal flowering. Plant Cell Online (2006) 18:1846-1861.

Hsu C-Y, et al. FLOWERING LOCUS T duplication coordinates reproductive and vegetative growth in perennial poplar. Proceedings of the National Academy of Sciences, USA (2011) 108:10756-10761.

Hudson RR, Kreitman M, Aguadé M. A test of neutral molecular evolution based on nucleotide data. Genetics (1987) 116:153-159.

Ibáñez C, Kozarewa I, Johansson M, Ögren E, Rohde A, Eriksson ME. Circadian clock components regulate entry and affect exit of seasonal dormancy as well as winter hardiness in populus trees. Plant Physiology (2010) 153:1823-1833.

Ingvarsson, P.K. 2010 Natural selection on synonymous and nonsynonymous mutations shapes patterns of polymorphism in *Populus tremula*. Molecular Biology and Evolution, 27, 650-660.

Ismail M, et al. Comparative nucleotide diversity across north american and european *Populus* species. Journal of Molecular Evolution (2012) 74:257-272.

Jansson S, Douglas CJ. *Populus*: A model system for plant biology. Annual Review of Plant Biology (2007) 58:435-458.

Kawecki TJ, Ebert D. Conceptual issues in local adaptation. Ecol. Lett. (2004) 7:1225-1241.

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

Kelly JK. A Test of neutrality based on interlocus associations. Genetics (1997) 146:1197-1206.

Keurentjes, J.J.B., Fu, J., Terpstra, I.R., Garcia, J.M., van den Ackerveken, G., Snoek, L.B. et al. 2007 Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. Proceedings of the National Academy of Sciences, 104, 1708-1713.

Langlet O. Two hundred years genecology. Taxon (1971) 20:653.

Lewontin RC, Krakauer J. Testing the heterogeneity of F values. Genetics (1975) 80:397-398.

Le Corre V, Roux F, Reboud X. DNA Polymorphism at the FRIGIDA Gene in *Arabidopsis thaliana*: Extensive Nonsynonymous Variation Is Consistent with Local Selection for Flowering Time. Molecular Biology and Evolution (2002) 19:1261-1271.

Leivar P, et al. Dynamic Antagonism between Phytochromes and PIF Family Basic Helix-Loop-Helix Factors Induces Selective Reciprocal Responses to Light and Shade in a Rapidly Responsive Transcriptional Network in *Arabidopsis*. The Plant Cell Online (2012) 24:1398-1419.

Li YF, Costello JC, Holloway AK, Hahn MW. "Reverse Ecology" and the Power of Population Genomics. Evolution (2008) 62:2984-2994.

Liepe K. Growth-chamber trial on frost hardiness and field trial on flushing of sessile oak (*Quercus petraea* Liebl). Ann. For. Sci. (1993) 50:208s-214s.

LÍZAL, Pavel - DADEJOVÁ, Martina - CETKOVSKÁ, Kateřina - ŘEPKOVÁ, Jana - RELICHOVÁ, Jiřina. Molecular analysis of FRI (FRIGIDA) gene in early and late-

flowering ecotypes and late-flowering mutants in Arabidopsis thaliana. In From laboratory to business. České Budějovice : Institute of Plant Molecular Biology ASCR, 2005. ISBN 80-86778-16-9, p. 51-51. 12.9.2005, České Budějovice.

Magri D, et al. A new scenario for the Quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. New Phytologist (2006) 171:199-221.

Matthew WH, et al. Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. Nature Genetics (2012) 44:212-216.

Mátyás C. Climatic adaptation of trees: Rediscovering provenance tests. Euphytica (1996) 92:45.

McClung CR. The genetics of plant clocks. Adv Genet (2011) 74:105-139.

McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in *Drosophila*. Nature (1991) 351:652-654.

Mimura M, Aitken SN. Adaptive gradients and isolation-by-distance with postglacial migration in *Picea sitchensis*. Heredity (2007) 99:224-232.

Morgenstern EK. Geographic variation in forest trees : genetic basis and application of knowledge in silviculture. (1996) Vancouver: UBC Press.

Nachman, M.W., Hoekstra, H.E. and D'Agostino, S.L. 2003 The genetic basis of adaptive melanism in pocket mice. Proceedings of the National Academy of Sciences, 100, 5268-5273.

Neale DB, Ingvarsson PK. Population, quantitative and comparative genomics of adaptation in forest trees. Current Opinion in Plant Biology (2008) 11:149-155.

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

Nielsen R. Molecular signatures of natural selection. Annu Rev Genet (2005) 39:197-218.

Orr HA. The population genetics of adaptation: The distribution of factors fixed during adaptive evolution. Evolution (1998) 52:935-949.

Petit RJ, Hampe A. Some evolutionary consequences of being a tree. Annual Review of Ecology, Evolution, and Systematics (2006) 37:187-214.

Pooni, H.S., Coombs, D.T., Virk, P.S. and Jinks, J.L. 1987 Detection of epistasis and linkage of interacting genes in the presence of reciprocal differences. Heredity, 58, 257-266.

Pot D, et al. Nucleotide variation in genes involved in wood formation in two pine species. New Phytologist (2005) 167:101-112.

Pritchard JK, Donnelly P. Case-control studies of association in structured or admixed populations. Theoretical Population Biology (2001) 60:227-237.

Pruneda-Paz JL, Kay SA. An expanding universe of circadian networks in higher plants. Trends Plant Sci (2010) 15:259-265.

Pyhäjärvi T, García-Gil MR, Knürr T, Mikkonen M, Wachowiak W, Savolainen O. Demographic history has Influenced nucleotide diversity in european *Pinus sylvestris* populations. Genetics (2007) 177:1713-1724.

Ramos-Onsins SE, Mitchell-Olds T. Mlcoalsim: multilocus coalescent simulations. Evol Bioinform Online (2007) 3:41-44.

Ramsay H, Rieseberg LH, Ritland K. The correlation of evolutionary rate with pathway position in plant terpenoid biosynthesis. Molecular Biology and Evolution (2009) 26:1045-1053.

Rinne PLH, et al. Chilling of dormant buds hyperinduces FLOWERING LOCUS T and recruits GA-inducible 1,3-β-glucanases to reopen signal conduits and release dormancy in *Populus*. The Plant Cell Online (2011).

Rohde A, Bhalerao RP. Plant dormancy in the perennial context. Trends in Plant Science (2007) 12:217-223.

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

Rozen S, Skaletsky H. Primer3 on the www for general users and for biologist programmers (1999). 365-386.

Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences (1977), USA 74:5463-5467.

Savolainen O, Pyhäjärvi T. Genomic diversity in forest trees. Current Opinion in Plant Biology (2007) 10:162-167.

Savolainen O, Pyhäjärvi T, Knürr T. Gene flow and local adaptation in trees. Annual Review of Ecology, Evolution, and Systematics (2007) 38:595-619.

Sawyer SA, Parsch J, Zhang Z, Hartl DL. Prevalence of positive selection among nearly neutral amino acid replacements in *Drosophila*. Proceedings of the National Academy of Sciences (2007), USA 104:6504-6510.

Sella G, Petrov DA, Przeworski M, Andolfatto P. Pervasive natural selection in the *Drosophila* genome PLoS Genet (2009) 5:e1000495.

Shen H, Moon J, Huq E. PIF1 is regulated by light-mediated degradation through the ubiquitin-26S proteasome pathway to optimize photomorphogenesis of seedlings in *Arabidopsis*. The Plant Journal (2005) 44:1023-1035.

Shen Y, et al. Phytochrome A mediates rapid red light–induced phosphorylation of *Arabidopsis* FAR-RED ELONGATED HYPOCOTYL1 in a low fluence response. Plant Cell Online (2009) 21:494-506.

Sorkheh K, Malysheva-Otto LV, Wirthensohn MG, Tarkesh-Esfahani S, Martinez-Gomez P. Linkage disequilibrium, genetic association mapping and gene localization in crop plants. Genet Mol Biol (2008) 31:805-814.

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

Staiger D, et al. The *Arabidopsis* SRR1 gene mediates phyB signaling and is required for normal circadian clock function. Genes & Development (2003) 17:256-268.

Sterky F, et al. A *Populus* EST resource for plant functional genomics. Proceedings of the National Academy of Sciences of the United States of America (2004) 101:13951-13956.

Stettler RF, National Research Council C. Biology of *Populus* and its implications for management and conservation. (1996) Ottawa: NRC Research Press.

Stoletzki N, Eyre-Walker A. Estimation of the Neutrality Index. Molecular Biology and Evolution (2011) 28:63-70.

Suzuki H, et al. Comparative genomic analysis of the *Streptococcus dysgalactiae* species group: gene content, molecular adaptation, and promoter evolution. Genome Biology and Evolution (2011).

Swarbreck D, et al. The *Arabidopsis* information resource (TAIR): gene structure and function annotation. Nucleic Acids Research (2008) 36:D1009-D1014.

Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. genetics (1989) 123:585-595.

Thornton KR, Jensen JD, Becquet C, Andolfatto P. Progress and prospects in mapping recent selection in the genome. Heredity (2007) 98:340-348.

Toomajian C, et al. A nonparametric test reveals selection for rapid flowering in the *Arabidopsis* genome. PLoS Biol (2006) 4:e137.

Torgerson DG, et al. Evolutionary processes acting on candidate *cis*-regulatory regions in humans inferred from patterns of polymorphism and divergence. PLoS Genet (2009) 5:e1000592.

Troein C, Locke JCW, Turner MS, Millar AJ. Weather and seasons together demand complex biological clocks. Current Biology (2009) 19:1961-1964.

Tuskan GA, et al. The Genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray). Science (2006) 313:1596-1604.

Vitasse Y, Porte AJ, Kremer A, Michalet R, Delzon S. Responses of canopy duration to temperature changes in four temperate tree species: relative contributions of spring and autumn leaf phenology. Oecologia (2009) 161:187-198.

van der Schoot C, Rinne PLH. Dormancy cycling at the shoot apical meristem: Transitioning between self-organization and self-arrest. Plant Science (2011) 180:120-131.

vonWuehlisch G, Krusche D, Muhs HJ. Variation in temperature sum requirement for flushing of beech provenances. Silvae genetica (1995) 44:343-346.

Wall JD, Andolfatto P, Przeworski M. Testing models of selection and demography in *Drosophila simulans*. Genetics (2002) 162:203-216.

Weinreich DM, Rand DM. Contrasting patterns of non–neutral evolution in proteins encoded in nuclear and mitochondrial genomes. Genetics (2000) 156:385-399.

Weir BS, Cardon LR, Anderson AD, Nielsen DM, Hill WG. Measures of human population structure show heterogeneity among genomic regions. Genome Research (2005) 15:1468-1476.

White TL, Adams WT, Neale DB. Forest genetics. (2007) Wallingford, Oxfordshire, UK; Cambridge, MA: CABI Pub.

Wittkopp, P.J., Vaccaro, K. and Carroll, S.B. 2002 Evolution of yellow Gene Regulation and Pigmentation in Drosophila. Current biology : CB, 12, 1547-1556.

Worrell R, Cundall EP, Malcolm DC, Ennos RA. Variation among seed sources of silver birch in Scotland. Forestry (Oxford) (2000) 73:419-435.

Wright JW. Introduction to forest genetics. (1976) New York: Academic Press.

Wright S. Genetical structure of populations. Nature (1950) 166:247-249.

Wright SI, Charlesworth B. The HKA test revisited. Genetics (2004) 168:1071-1076.

Zeng K, Shi S, Wu C-I. Compound tests for the detection of hitchhiking under positive selection. Molecular Biology and Evolution (2007) 24:1898-1908.

Supplementary Information

S.NO	Sample name	Location	Longitude	Latitude (°N)
1.	TATB 1.7	Alsek River, BC	137.50	59.26
2.	TAKA 3.1	Taku River, BC	133.34	58.36
3.	STKG 4.4	U. Stikine River, BC	129.40	57.57
4.	STKB 5.3	L. Stikine River, BC	131.47	56.56
5.	STKD 5.1	L. Stikine River, BC	131.47	57.20
6.	KTSG 10.5	Kitseguecla River, BC	127.55	55.06
7.	HAZH 10.5	Hazelton River, BC	127.40	55.13
8.	CDRE 10.1	Cedarvale River, BC	128.20	55.01
9.	SKNM 10.5	Skeena River, BC	128.59	54.24
10.	KLNE 20.9	Klinaklini River, BC	125.34	51.44
11.	KLND	Klinaklini River, BC	125.30	51.34
12.	KLNB 20.3	W. Klinaklini River, BC	125.46	51.18
13.	KLNA 20.3	W. Klinaklini River, BC	125.35	51.07
14.	ALAA 20.5	Kingcome River, BC	126.07	50.59
15.	NHTA	Nahatlatch River, BC	121.34	50
16.	WELC 27.4	Wellington Bar River, BC	121.25	49.40
17.	MTSM 27.2	Matsqui Is River, BC	122.20	49.07
18.	HOPG 27.2	Hope South River, BC	121.33	49.24
19.	9807	Skagit River, WA	NA	NA
20.	9588	Nooksack River, WA	NA	NA
21.	9851	Snoqualmie River, WA	NA	NA
22.	4579	Nisqually River, WA	NA	NA
23	7014	Cowlitz River, WA	NA	NA
24.	6880	L. Columbia River, W.	NA	NA
25.	8553	Willamette River, Salem, OR	NA	NA
26.	2884	L. Columbia River, E.	NA	NA
27.	MC-3	Myrtle creek, OR	123.29	43.02
28.	RR-10	Rogue River, OR	123.13	42.41
29.	SV-4	Shasta valley, OR	122.66	41.71
30	HB-3	Humboldt, CA	123.56	40.25
31	SG-7	Sierra, CA	120.46.449	39.53.797
32	SF-9	Sierra, CA	120.19.65	38.45.837
33	GF-7	Sierra, CA	120.46.889	39.17.597

Table S 1. Location and name of all the *P. trichocarpa* samples with their geographical gradients.

S.NO	Sample Name	Identity	Location	Longitude (°W)	Latitude (°N)
1.	13237	LSU-40	Mississippi, LA	90.80	30.00
2.	11480	30-3 (OK-84)	Brazos, TX	97.37	31.75
3.	11445	15-4M (OK-84)	Canadian, OK	97.92	35.40
4.	11503	5-1 (OK-84)	Arkansas, KN	97.00	37.08
5.	11563	ST-265	Mississippi, MS	90.95	32.45
6.	11408	824021 (OK-88)	Brazos, TX	95.45	29.15
7.	11499	421-2 (OK-88)	Atchafalaya, LA	91.75	30.67
8.	11444	154C-4	Saluda, SC	81.69	34.23
9.	5677	110226	Mississippi, MS	90.45	34.75
10.	11501	47-5	Mississippi, MO	89.63	36.14
11.	11362	122 (OK-80)	Red, TX	94.50	33.60
12.	13199	112910	Mississippi, MS	91.58	31.17

Table S 2. Location and name of all the *P. deltoides* samples with their geographical gradient.

S.NO	Gene Name	Length of gene (kb)	No. of fragments	Function(s)
1.	PSEUDO RESPONSE REGULATOR-3 (PRR-3)	8.96	10	Evening–phased gene, regulates post-translational TOC-1 expression and temporary entrainment of C'clock.
2.	PSEUDO RESPONSE REGULATOR-5 (PRR-5)	6.77	8	Morning–phased genes in negative feedback loop with CCA-1 and LHY and plays role in frost tolerance.
3.	CIRCADIAN CLOCK ASSOCIATED-1 (CCA-1)	8.94	10	Dawn–phased MYB-like transcription factors regulating expression of TOC-1 (-vely) and PRR's (+vely).
4.	EARLY FLOWERING-3 (ELF-3)	7.42	9	Negatively regulates light input to the clock and helps maintain robust rhythms in constant conditions.
5.	EARLY FLOWERING-4 (ELF-4)	2.36	3	Negatively regulates light input to the clock and help in maintaining robust rhythms in constant conditions.
6.	GIGANTEA (GI)	12.42	14	Gene interacting with ZTL protein in blue light to repress TOC-1
7.	PHYTOCHROME INTERACTING FACTOR LIKE-1(PIL1)	5.63	7	Transcription factor, red and far-red light signaling pathway.
8.	SENSITIVITY TO RED LIGHT REDUCED-1 (SRR-1)	2.83	3	A positive regulator to signaling in response to red and white light and plays role in setting clock pace.
9.	TIMING OF CAB EXPRESSION-1 (TOC-1)	7.42	9	Pseudo-response regulator (~~ PRR-1), evening-phased gene, regulating expression of CCA-1 and LHY.
10.	ZEITLUPE (ZTL)	8.85	13	F-box protein known to repress TOC-1 post-translationally.
11.	TIME FOR COFFEE (TIC)	8.31	10	Attenuates light signaling during the day or night for reciprocal regulation.
12.	PHYTOCHROME INTERACTING FACTOR-6 (PIF-6)	4.35	5	Transcription factor, red and far-red light signaling pathway.
13.	PHYTOCHROME INTERACTING FACTOR-4 (PIF-4)	4.75	6	Red and far-red light signaling pathway.
14.	PHYTOCHROME INTERACTING FACTOR-1 (PIF-1)	2.76	3	Shade-avoidance and red and far-red light signaling pathway.
15.	PHYTOCHROME A (PHYA)	7.98	9	Light receptor protein, phototransduction and red and far-red light signaling pathway.
16.	LUX ARRYTHMO (LUX)	2.71	3	MYB-like transcription factor, part of Circadian oscillator.
17.	GIGANTEA-1 (GIG)	12.94	14	Gene interacting with ZTL protein in blue light to degrade or repress TOC-1.

18.	FRIGIDA (FRI)	2.60	3	Up-regulates level of FCA, to maintain dormancy and repress flowering.
19.	FCA	8.52	10	Regulates dormancy, and forms complex with ABA, which prevents FCA-FY complex formation, required for flowering.
20.	DE-ETIOLATED-1 (DET-1)	7.65	9	Negative repressor of light-driven photomorphogenic development in *Arabidopsis*.
21.	CONSTANS (CO)	3.85	5	Key protein in photoperiod, which FT gene regulates flowering and dormancy.
22.	ABA INSENSITIVE-3 (ABI-3)	4.42	6	Bud repressor, acts as mediator between hormonal and transcriptional regulation of bud set and dormancy.
23.	CONSTANS-1	3.78	6	Key protein in photoperiod, which FT gene regulates flowering and dormancy.
24.	FLOWERING LOCUS T-1 (FT-1)	5.85	6	Flowering initiation and mediators of signals of growth cessation (bud burst and dormancy).
25.	FLOWERING LOCUS T-2 (FT-2)	5.18	6	Flowering initiation and mediators of signals of growth cessation (bud burst and dormancy).
26.	LATE ELONGATED HYPOCOTYL-1 (LHY1)	3.59	5	Dawn-phased MYB-like transcription factors regulating expression of TOC-1 (-vely) and PRR's (+vely).

Table S 3. Gene name, length and their function with number of fragments in which they were amplified.

	1	2	3
1	*		
2	0.01458	*	
3	0.02656	0.13550	*

Table S 4. The distance matrix of F_{ST} between populations. The population 1 is northern Canada, 2 as California-Sierra and 3 as BC-Washington samples

Average F-Statistics over all loci	
F_{IS}	0.51102
F_{ST}	0.05328
F_{IT}	0.53707

Table S 5. Average F statistics over all loci in the population. F_{IS} is measure of variation within population, F_{ST} is measure of variation between population and F_{IT} as total variation of population.

Gene Names	H_d	π	θw/site	TAJIMA'S D	FU AND LI'S D	FU AND LI'S F	Tajima'sD (WM) L	Tajima'sD (WM) S	π_{ds}	π_{dn}	π_{silent}
PHYA	0.126929	0.000372	0.000693	-0.617885	0.391015	0.104795	-0.564430644	-0.345764488	2.93E-05	4.64E-05	0.000457
ELF-3	0.2964	0.001068	0.00096	-0.03909429	0.6066	0.475391429	-0.113626055	-0.05224	0	0.000203	0.001371
PIF-1	0.276	0.000694	0.001018	-0.73514667	1.055966667	0.580593333	-0.658341716	-0.53466463	0	0.001034	0.000582
FCA	0.53475	0.001723	0.001114	0.89595455	0.837651818	0.996628182	0.727896811	0.814589121	0	0.000519	0.001806
CCA	0.372143	0.001465	0.00118	0.52377429	0.769830714	0.811079286	0.48800804	0.37878082	0.000383	0.000238	0.001587
TIC	0.3796	0.001965	0.001486	0.4769375	0.773593333	0.793925	0.217865562	0.338374796	0.001441	0.001049	0.002281
PIF-6	0.277556	0.001399	0.001547	0.01819	0.653202	0.539748	0.046390322	-0.067947194	0	3.67E-05	0.001371
GIG	0.516563	0.002138	0.001647	0.58205	0.991394615	1.010842308	0.519067157	0.484110976	0.000147	0.001063	0.002229
PRR3	0.296467	0.00117	0.001707	-0.54732727	0.753795455	0.418822727	-0.411513204	-0.417444206	0.000149	0.001057	0.000985
PIL1	0.381615	0.001872	0.002013	-0.16974667	0.8509925	0.630248333	-0.177241542	-0.182448673	0.001055	0.000547	0.001606
ZTL	0.450385	0.003278	0.002576	0.41843333	0.703008333	0.725036667	0.56714345	0.235347194	0	9.62E-05	0.003205
LUX	0.582857	0.00453	0.002779	1.366925	0.816625	1.150236667	1.327647825	1.158007647	0.001629	0.001413	0.004504
PRR5	0.6723	0.003738	0.002802	0.84332	0.784073	0.941722	0.839219824	1.023734739	0.002475	0.001073	0.004093
CO	0.413875	0.002603	0.002893	-0.14050167	0.645293333	0.480666667	-0.020565122	-0.082379302	0	0.000391	0.002994
ELF-4	0.592	0.003175	0.00318	0.2498	1.098315	0.851345	0.611197231	-0.68952125	0	0.000425	0.003245
LHY	0.746	0.00416	0.00341	0.62597	1.39913	1.35045	0.62597	0.62597	0.00361	0	0.0056
TOC-1	0.566692	0.003692	0.003427	0.27753167	1.035151667	0.917048333	0.258478637	0.075624764	0.003373	0.000275	0.005045
FT-2	0.5454	0.00587	0.003755	1.34560571	1.212578571	1.458135714	0.986388673	1.039053521	0.002061	0.000595	0.006432
ABI3	0.613	0.004616	0.004262	0.4733	1.176054	1.051094	0.484341736	0.610805725	0.000484	0.000696	0.005782
CO-1	0.8315	0.00767	0.005816	0.7661125	1.16605	1.2183875	0.606853663	0.713305596	0.000513	0.001216	0.007733
PIF-4	0.711909	0.010972	0.007271	1.32613273	1.092667273	1.339141818	1.389233957	1.221727207	0.000951	0.002651	0.010718
SRR1	0.6296	0.007462	0.007642	0.083838	0.659938	0.544884	-0.052786349	-0.013433494	0.004102	0.004806	0.007706
FT-1	0.528909	0.012303	0.009384	0.53875	1.055252222	1.049303333	0.512425012	0.4610244	0	9.73E-05	0.012301
FRI	0.804	0.010173	0.00939	0.29919333	1.463036667	1.29086	0.368471546	0.307761224	0.00066	0.0022	0.008933
DET-1	0.370636	0.001318	0.042891	0.106365	0.89535875	0.7721325	0.036857232	0.17448859	0.000327	0.000496	0.001328

Table S 6. Nucleotide polymorphism statistics and tests of selection in *Populus trichocarpa*. The values in green represent the small proportion of values which show negative Tajima's D.

Gene Name	H_d	π	θw/site	TAJIMA'S D	FU AND LI'S D	FU AND LI'S F	Tajima's D WM	π_{ds}	π_{dn}	π_{silent}
ZTL	0.3535	0.00100167	0.001225	-0.40151	0.115176	-0.02257	-0.376981612	0.001992	0.000102	0.00148
TOC	0.496142857	0.00181429	0.001507143	0.509171667	0.536428333	0.578426667	0.286810848	0	7.71E-05	0.002524
TIC	0.2958	0.000965	0.001047	-0.072112	-0.190254	-0.134678	0.118645354	0.000435	0.000499	0.001018
SRR1	0.0735	0.0001975	0.000195	0.02193	0.66689	0.56884	0.008706815	0	0.00043	0
PRR5	0.457666667	0.00239333	0.002041667	0.36086	0.559124	0.552062	0.362831943	0.000332	0.001028	0.002812
PRR3	0.561833333	0.00149667	0.00108	1.01368	0.796676	0.967456	0.775501784	0.00067	0.000162	0.002188
PIL-1	0.570166667	0.00139833	0.00126	0.31322	0.497264	0.512956	0.14794079	0.000918	0.001363	0.000692
PIF-6	0.682	0.00166	0.001125	1.532665	0.9897	1.25105	1.622791639	0	0.00129	0.001655
PIF-4	0.6155	0.00359	0.0025925	1.278613333	1.21575	1.42166	1.083164535	0.004333	0.005965	0.002953
PIF-1	0.71	0.00481667	0.00402	0.608606667	0.928776667	0.954466667	0.927884558	0.003323	0.002033	0.005343
PHYA	0.456333333	0.00097556	0.000934444	0.20546	0.2857625	0.29972875	0.247466313	0.000163	0.000389	0.001076
LUX	0.569	0.0036675	0.0031725	0.6624375	0.79403	0.82077	0.910979093	0.002375	0	0.004538
FRI	0.57	0.00188	0.001456667	1.036365	1.16462	1.26287	0.93423139	0	0.000973	0.001763
FCA	0.377909091	0.00165727	0.001917273	-0.35238	1.05985	0.767615556	-0.279174477	0.002097	0	0.00179
ELF-4	0.2665	0.00099	0.00081	0.85057	1.05247	1.02905	0.569860794	0	0.00171	0
CCA-1	0.372363636	0.00139182	0.001442727	-0.224891111	0.353997778	0.226155556	-0.203395165	9.45E-05	0	0.00159
CO-1	0.4326	0.001654	0.001398	0.4119375	0.915015	0.890525	0.360680651	0	0	0.001654
DET-1	0.53725	0.00247375	0.00205625	0.8767225	1.0804	1.020905	0.883398889	0	0.002138	0.002486
ELF-3	0.45025	0.0014875	0.0015175	-0.13257	0.65999	0.5024225	-0.454804807	0	0	0.001555
GIG	0.303928571	0.00120231	0.001332857	-0.438522	0.267523	0.077434	-0.35222589	0	0.000294	0.001039
LHY	0.646	0.002725	0.00184	1.311615	0.83656	1.102625	1.328470659	0	0	0.003415
FT-1	0.38875	0.0028025	0.0021225	1.43804	1.31067	1.448825	0.504197243	0	0	0.002803
FT-2	0.6522	0.004704	0.003366	1.3139025	1.108875	1.35074	1.110584488	0	0	0.004738
CO	0.619	0.00214	0.0019	0.32512	0.99199	0.93002	0.32512	0	0	0.00214
ABI3	0.779	0.00136	0.0008	1.73869	0.97946	1.37596	1.73869	0	0.00184	0

Table S 7. Nucleotide polymorphism statistics and tests of selection in *P. deltoides*. The values in green represent small proportion of values which show negative Tajima's D.