Genomics of Climatic Adaptation in *Populus trichocarapa*

Man Zhang

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

Doctor of Philosophy
In
Forestry

Jason Holliday
Amy Brunner
Ina Hoeschele
Mohammad Saghai-Maroof

July 14th 2016
Blacksburg, VA

Keywords: Climatic adaptation; Population genomics; *Populus trichocarpa*; deleterious load

Genomics of Climatic Adaptation in *Populus trichocarapa*

Man Zhang

Abstract (academic)

Temperate tree species exhibit seasonal growth cycling, and the timing of such transition varies with local climate. Under anthropogenic climate change, the local pattern of growth and dormancy in tree populations is expected to become uncoupled with shifting seasonal environmental signals, particularly temperature. Thus, an understanding of the genetic underpinnings of local adaptation is key to predicting the fate of tree populations in the future. In this thesis, we coupled sampling of range-wide natural accessions of *P. trichocarpa* with adaptive trait phenotyping and genome-wide genotyping to uncover relationships between genotype, phenotype, and environment. We detected strong correlations between adaptive phenotypes, climate, and geography, which suggested climatic selection driving adaptation of these populations to local environments. We subsequently combined genotype-phenotype association tests with sliding window analysis and identified regions strongly associated with these adaptive traits. We also compared adaptive markers identified in two independent GWAS on samples across latitude and altitude transects and found a set of associated variants shared across both transects. We further scanned the genome with three selection tests to identify regions showing evidence of recent positive and divergent selection. By comparing candidate selection regions across altitude and latitude, we detected a set of overlapping regions showing differentiation across gradients of the same climate variables. We validated the functional imortance of these selection regions by combining GWAS and showed that selection regions contain a strong signature of phenotypic associations. We also studied the distribution of deleterious allels across genome and natural populations, and found that deleterious alleles preferentially accumulate in regions of low recombination and hithihking regions. Finally, marginal populations contained more deleterious alleles compared with central populations, which is likely due to ineffective selection in small populations and recent bottlenecks associated with postglacial recolonization. These findings provide new insights into the genomic architecture underlying climatic adaptation and how selection drives adaptive evolution of tree species.

Genomics of Climatic Adaptation in *Populus trichocarapa*

Man Zhang

Abstract (public)

Temperate tree species cycle between periods of growth in summer and dormancy in winter. The timing of these transitions is tightly linked with local climate conditions, which are predicted to be altered by climate change. Thus, tree populations must either adapt to these novel environments or migrate to new areas. It is unlikely that either of these response will happen fast enough to keep up with the predicted pace of climat echange, and knowledge of the genetic mechanisms underlying climatic adaptations is therefore key to maining forest health in a warmer future. In this study, we first demonstrated that climate is a major environmental factor that drives differentiation in growth and dormancy traits among populations of black cottonwood, a broadleaf tree native to western North America. We identified a number of genes that likely control this variation in adaptation-related traits. Interestingly, we found the genetic mechanism underlying adaptation to two independent geographical gradients (latitude and altitude transects) is partially conserved. As adaptation-related genes are likely to be targeted by natural selection, we further scanned the poplar genome for genomic regions showing evidence of recent selection. We identified significant level of overlap in candidate selection regions in both transects, indicating parallel adaptation to similar climatic gradients. These results suggest novel genetic targets relevant to local adaptation, and provide a basis for assessing the 'adaptive health' of populations. On the other hand, deleterious mutations that reduce the viability of individuals and populations is another evolutionary force affecting population fitness and adaptive evolution. We studied the distribution of deleterious alleles and found that purifying selection mostly keeps deleterious alleles at low frequnecy. However, selection efficiency is limited in genomic regions of low recomibnation, and deleterious alleles also accumulate in regions that have been targeted by natural selection due hithhiking with favorable genetic variants. In addition, we found an excess of deleteirous alleles in marginal populations compared with central populations, which may impede the ability of the former to adapt to future climates. These findings together provide new insights into the genomic architecture underlying climatic adaptation and how selection drives adaptive evolution of tree species.

Acknowledgements

I firstly acknowledge my advisor, Jason Holliday, whose patience, advice, and knowledge in forestry genomics supported me throughout my graduate life. I thank all my committee members Amy Brunner, Ina Hoeschele, Mohammad Saghai-Maroof for their advice on my research and generosity in sharing lab equipment. I also thank Lecong Zhou for his effort in DNA extraction and generating sequences. I want to thank current and former Holliday lab members, Rajesh Bawa, Regis W. Oubida Kaplan Abdullah, Dash Gantulga, Xiaoyan Sheng, Mihir Mandal for assistance in phenotyping trees, and Haktan Suren for bioinformatics support. I also like to thank Cees van Oosten, Kyle Peer, Clay Sawyers, and Deborah Bird (Virginia Tech Reynold's Homestead Forestry Research Station) for assistance with installation and maintenance of the common garden. I also thank MPS and Advanced Research Computing at Virginia Tech for providing computational resources and technical support related some of the analyses described here. This work was supported by the National Science Foundation Plant Genome Research Program (IOS:1054444) and the USDA National Institute of Food and Agriculture (McIntire Stennis Project 10005394).

Attribution

This thesis was written as a series of manuscripts on the advice of my supervisory committee. Chapter 1 provides the introduction, context, and research objectives. Chapter 2, 3, 4 are manuscripts to be submitted to journals and Chapter 5 is currently in review. Chapter 6 summarizes results and discussion of this study and discussed future research directions. For all chapters of this thesis, I took the lead in developing the ideas, collecting and analyzing data, and producing draft manuscripts, figures, and tables. However, my advisor, Jason Holliday initiated concept for association mapping in studying local adaptation and deleterious load, and provided guidance throughout all chapters. I would like to recognize him as co-author on all chapters. The chapter 2 would not be possible without the help of all Holliday lab members, particularly Rajesh Bawa, Regis W. Oubida, Kaplan Abdullah, Dash Gantulga, Mihir Mandal. They are included as co-authors on Chapter 2. To recognize Lecong Zhou' effort in producing sequencing data, he is included as co-authors on Chapter 3, 4, and 5. Hakan Suren provided help in generating sequencing data with GATK and Rajesh Bawa provided calculation of effective population size for Chapter 5. So they are included as co-authors on Chapter 5.

Table of Contents

List of Figures

List of Tables

1. General Introduction and Research Objectives

1.1 Tree populations under climate change

Anthropogenic climate change is reshaping bioclimatic zones across the United States and the world. Primarily due to the use of fossil fuels, atmospheric concentration of CO_2 and CH_4 in 21st century far exceeded the natural range over last 650,000 years (IPCC 2007). The increasing concentrations of greenhouse emissions affect the absorption and scattering of atmospheric radiation, resulting in an average earth surface temperature increase of 1.5 to 2.5 °C globally (IPCC 2007). Global warming is also disrupting seasonal patterns of temperature, precipitation, and causing extreme weather events such as drought, storms, etc (Saxe, et al. 2001). As a result of these impacts on climate, global warming is imposing acute threats on human society, food production, and terrestrial ecosystems (Parry, et al. 2004; Grimm, et al. 2013).

Rapid warming and associated disturbance to ecosystems is re-shaping the distribution of terrestrial plants across the world (Grimm, et al. 2013). It is predicted that the warming-driven shift of species range may lead to change of biome composition across 5-20% of United States land area (Grimm, et al. 2013), of which forests comprise approximately 34%. Temperature increase makes it possible for tree species to migrate northwards, however precipitation and soil nutrient sufficiency can limit tree growth (Saxe, et al. 2001). Historical spatial analysis of climate change suggests that more than 1/7 of North America is vulnerable to warming-induced biome shift (Gonzalez, et al. 2010). However, rates of contemporary climate change mean that there is likely to be a lag between changes in bioclimactic niches and the ability of individual species and populations to occupy those niches (Gonzalez, et al. 2010). In addition, the interaction between invasive species and local vegetation may alter the competitive hierarchy and structure-function relationships of existing ecosystem (Sturm, et al. 2001; Grimm, et al. 2013). Globally warmer conditions also increase the extent of plant drought stress, pest and wildfire outbreaks (Littell, et al. 2009). Collectively, these disturbances are expected to cause extensive tree mortality and reduced forest productivity (Zavaleta 2005; Grimm, et al. 2013).

Though elevated CO2 and rising temperature may increase photosynthetic rate and concomitant biomass production, as well as enhance soil organic matter decomposition and rates of nutrient mineralization and availability (Saxe, et al. 1998; Saxe, et al. 2001), temperate and boreal tree

species are expected to suffer from disrupted seasonal phenology and incomplete cold hardening (Saxe, et al. 2001). Unlike annuals and biennials, trees in temperate regions alternate between vegetative growth and endodormancy that closely match the local seasonal oscillations each year (Chen, et al. 2002). In response to environmental signals such as short photoperiod and low temperature, trees transit from active growth to the protected state of cold hardiness, enabling the fully acclimated meristems survive the harsh winter (Howe, et al. 2003). Cold hardening requires induction by long nights and subsequent chilling temperatures below ~5°C. For many species, substantial cold hardiness can be only acquired by this synchronization of cooling and long nights, with the former likely to be disrupted by climate change (Saxe, et al. 2001; Norgaard Nielsen and Rasmussen 2009). If trees delay their bud set and the level of cold hardiness is not adequate, meristems can be damaged by fall frost and poor growth and productivity can occur (Frewen, et al. 2000). Endodormant buds require chilling and subsequent accumulative thermal time in order to break the state of rest and resume growth (Cannell and Smith 1986; Hänninen 1995). Elevated temperature can cause insufficient cold hardiness, hasten dehardening, and cause early bud burst, leaving trees at the risk of frost damage (Repo, et al. 1996; Leinonen 1997). Furthermore, fluctuating temperature can reduce growth competence of dormant buds (Hanninen, et al. 1990; Partanen, et al. 1998). Thus, climate change may disturb the synchronization between endodormancy and environmental cues, causing shifted seasonal development, maladaptation, or even extinction of tree species (Saxe, et al. 2001; Hamann and Wang 2006; St Clair and Howe 2007).

1.2 Importance of local adaptation

With increases in earth surface temperatures and projected further increases (IPCC 2007), in the absence of adaptation or migration, populations of many native plant species are projected to decline or be extirpated (Aitken, et al. 2008). The changes in climate envelopes will lead to forest decline in many areas, and decreased productivity in others. Where local populations persist, their average fitness may drop, physiology and growth may be disrupted, and population size and productivity will likely decline (Saxe, et al. 2001; Anderson, et al. 2012; Aitken and Whitlock 2013). Historical evidence of glacial and post-glacial migration of plant species and recolonization suggested high capacity of plants for migration and adaptation to new environment (Davis and Shaw 2001; Aitken, et al. 2008). Strategies including migration, local

adaptation, and phenotypic plasticity may enable species to persist through anthropogenic climate change (Aitken, et al. 2008; Anderson, et al. 2012). The relative contribution each strategy in mitigating risk of changing climate varies with species life characteristics, gene flow capacity, and strength of climate selection (Anderson, et al. 2012).

Phenotypic plasticity enables a single genotype to produce different phenotypes under different environmental conditions, allowing individuals to match their phenotypes to spatial or temporal environments in the short term (Sultan 2003). However, adaptive plasticity is energy-intensive, requiring accurate tracking and forecasting environmental information, maintaining developmental capacity for facultative responses, and is therefore evolutionarily constrained (Dewitt, et al. 1998; Sultan 2003; Reed, et al. 2010). Gene flow by pollen and seed dispersal could increase genetic variation and adaptive potential as well as promote migration and colonization of new habitat (Davis and Shaw 2001; Anderson, et al. 2012). The migration capacity of northern tree species, however, is predicted to be 25% less than the rate of climate change, and it is therefore unlikely to meet the requirement for rapid range shift predicted by the climate-based species distribution model (Davis and Shaw 2001; Loarie, et al. 2009; Anderson, et al. 2012). Due to the larger population size of central populations, gene flow from central populations may introduce maladapted alleles into trailing edge, resulting in greater chance of extinction of marginal populations (Aitken, et al. 2008; Holliday, et al. 2011). Assisted gene flow has been proposed to aid the introduction of pre-adapted individuals or populations to new local climates to mitigate local maladaptation (Weeks, et al. 2011; Aitken and Whitlock 2013). Such strategy could ease the ecological pressure by supplying more fit genotypes, thereby maintaining fecundity and survival of resident populations (Aitken and Whitlock 2013). However, this approach is limited to movement across few degrees of latitude because large distance transfer can cause outbreeding depression, disrupted seasonal growth cycles, and loss of original lineages (Aitken and Whitlock 2013; Aitken and Bemmels 2016). In addition, assisted migration requires information on genetic clines and population adaptation capacity to match genetic groups with new climates (Aitken and Bemmels 2016).

Local adaptation is an important alternative strategy for the survival of species with broad latitudinal ranges as climate change continues challenging their genetic tolerance (Aitken, et al.

2008; Anderson, et al. 2012). Local adaptation is defined as population's ability to appropriately tune their phenotype to environmental conditions (Savolainen, et al. 2007). Reciprocal transplant experiments have shown that many plants display relatively higher fitness at or near their home sites, indicating local adaptation (Kawecki and Ebert 2004; Ågren and Schemske 2012). Selection pressure from novel environment can lead to increased mortality, as well as decreased fitness and fecundity, forcing the local populations to evolve traits that match the current habitat (Kawecki and Ebert 2004). Adaptation capacity of species determines how well populations respond to rapid environmental change in their own habitat, as well as how far populations can colonize new habitats out of their local zone (Davis and Shaw 2001; Grassein, et al. 2014). The degree of adaptation is constrained by many factors such as spatially heterogeneous habitats, population size, gene flow, and species habitat fidelity (Buckling, et al. 2008). Large population size, sufficient genetic diversity, and increasing spatial variation in selection favor the potential of local adaption (Galloway and Fenster 2000; Buckling, et al. 2008).

1.3 Local adaptation of temperate trees

Tree species often occupy wide natural habitats and go through range shifts facilitated by extensive pollen and seed mediated gene flow (Savolainen, et al. 2007). Dispersal into new environments is required to develop new tolerances to bioclimatic variables (Davis and Shaw 2001). Given their long generation time, trees must often tolerate altered fitness optimum across their life span due to changing environmental conditions (Davis and Shaw 2001). In response to climate change during last glacial maxima and glacial-interglacial period, many plant species have survived by dispersal and different coping strategies (Davis and Shaw 2001; Clark, et al. 2012). The historical range shifts and subsequent adaptation to a range of latitudes and elevations resulted in significant adaptive divergence among modern populations of tree species (Davis and Shaw 2001). Common garden studies and provenance tests highlighted high levels of genetic diversity and strong adaptation to clinally varying environmental gradients among populations in many tree species, including *Picea glauca* (Furnier, et al. 1991), *Pinus contorta* (Yang, et al. 1996; Green 2005)*, Abies lasiocarpa* (Green 2005), *P. trichocarpa* (McKown, et al. 2014).

Tree species within temperate or boreal zones have distinct seasonal growth in face of temperature and photoperiod seasonality. Temperate trees adopt physiological mechanisms to

avoid and tolerate unfavorable temperatures by inducing cold acclimation, which comprises growth cessation, low metabolic activity, and high freezing tolerance during winter (Vitasse, et al. 2014). The survival-adaptation responses were first evolved from the stochastic occurrence of freezing temperature during interglacial period and are essential for species to tolerate temperature extremes (Caffarra and Donnelly 2010; Vitasse, et al. 2014). Cold hardiness increases with decreasing photoperiod and freezing temperature and reaches maximal levels by late autumn (Weiser 1970). Chilling in early spring induces trees to deacclimate, leaf out and resume meristem growth when the requirement for warm temperature days is met (Cooke, et al. 2012). This complex process requires precise timing of growth-dormancy phase transition to maximize annual growth as well as avoid cold injury (Weiser 1970). For example, the requirement for chilling prevents premature bud-break especially during mild winters and reduces risk of spring frost damage (Cooke, et al. 2012).

Natural selection and geography have shaped the distributions of many temperate tree species and have left signatures of local adaptation on both genetic and phenotypic diversity (Howe, et al. 2003; Savolainen, et al. 2007; Keir, et al. 2011). Many common garden studies have shown clinally varying selection and local adaptation resulted in moderate to strong clines in many phenology and physiological characters among temperate tree populations (Savolainen, et al. 2007; Aitken and Whitlock 2013). In addition to phenology traits, life history traits (body size, leaf number and area, longevity) and ecophysiology traits (leaf senescence, clonality, cold tolerance, resistance to local parasites and disease) are important manifestation of fitness and adaptability (Kawecki and Ebert 2004; Parmesan and Hanley 2015). Across latitude, growing season decreases with distance from the equator. Locally adapted populations at high latitude are more tolerant to extreme winter temperatures and require longer day length to set bud and lower chilling temperature for bud break. Thus, the clinal varying growing days and temperature leads to a steep cline in the timing of bud set and cold tolerance in tree species such as *Pinus sylvestris* (Garcia-Gil, et al. 2003; Notivol, et al. 2007). The rangewide differentiation in traits related to climatic adaptation as well as sensitivity to environmental change makes temperate tree species useful models to study local adaptation (Holliday, et al. 2016).

1.4 Genomic architecture of local adaptation

A great deal of progress has been made in dissecting the genetic basis of adaptive traits with functional genomic tools and population genetic approaches in tree species. Advanced molecular technologies have been employed in many functional studies to identify important players involved in adaptation mechanisms, including timing of reproductive development (Bohlenius 2006; Hsu, et al. 2011; Azeez, et al. 2014), endodormancy (Ruttink, et al. 2007; Bertoni 2011; Rinne, et al. 2011; Rohde, et al. 2011), and cold acclimation (Thomashow 1999) by examining gene expression and function. The photoreceptor gene PHYTOCHROME A (PHYA) was identified through overexpression of an oat PHYA in aspen, which resulted in new genotypes with altered critical day-length and hindered dormancy in short days (Olsen, et al. 1997). Circadian clock genes were also shown to synchronize bud dormancy with environmental cues through diurnal and seasonal rhythmic gene expression (Ramos, et al. 2005; Ibanez, et al. 2010). For example, winter temperature can disrupt the expression level and pattern of LATE ELONGATED HYPOCOTYL (LHY) and TIMING OF CAB 1 (TOC1), which induces growth cessation and cold hardiness (Ramos, et al. 2005). The CONSTANS (CO)/FLOWERING LOCUS T (FT) regulatory module controls photoperiodic flowering in aspen trees as well as the switch of dormancy-growth transition (Bohlenius 2006). Internal hormone and sugar signaling pathways have also been shown to be part of dormancy entrance and maintenance pathways (Horvath, et al. 2003). Chilling and photoperiod induces expression of growth-promoter genes FT, CENL1 (CENTRORADIALIS-LIKE1) and removes callose deposit at pores and plasmodesmata to reopen the signaling conduits connecting dormant bud to shoot by up-regulating gibberellins and associated cell wall 1,3-beta-glucanase genes (Bertoni 2011; Rinne, et al. 2011).

Though functional studies characterized many adaptive genes from a functional perspective, the genetic regulatory network underlying local adaptation remains unresolved (Howe, et al. 2003; Ehrenreich and Purugganan 2006; Holliday, et al. 2010). Adaptation mechanisms to seasonal climate are physiological processes and require complex genetic control to obtain precise responsiveness to environmental cues and endogenous signals (Howe, et al. 2003). Given the polygenic nature of adaptive traits, long generation time of trees and transformation recalcitrance of many tree species, forward genetic approaches such as QTL mapping, linkage-based mapping

complement reverse genetic screens as efficient search methods for adaptive genes on a large scale (Neale and Savolainen 2004). Adaptation-related traits are quantitative, moderately heritable, and controlled by multiple loci of small individual effect (Jermstad, et al. 2003; Neale and Savolainen 2004; Savolainen, et al. 2007). A series of QTL mapping studies were successfully used to estimate the number and location of traits for bud phenology and cold hardiness in different tree species since 1990 (Lander and Botstein 1989; Howe et al 2003; Jermstad et al., 2001; Holliday, J. A. 2009). For example, Rohde et al. identified budset-timing associated QTLs containing genes known to be differentially expressed during bud dormancy (Rohde, et al. 2011; Cooke, et al. 2012), and the PHYB2 (PHYTOCHROME B2) and ABI1B (ABI-interactor 1B) genes were shown to be associated with QTLs affecting bud set and bud break (Frewen, et al. 2000).

QTL mapping links phenotypic variation to specific genomic regions among pedigreed lines arising from parents that differ in traits of interest (Ehrenreich and Purugganan 2006). However, QTL mapping studies are limited to existing pedigrees due to the long generation time of tree species, and can identify only QTLs segregating in the parental lines (Ehrenreich and Purugganan 2006). High linkage disequilibrium among related individuals causes low mapping resolution, and most QTLs therefore span hundreds of genes (>100kb) (Chen, et al. 2002; Ungerer, et al. 2002). Association mapping is similar to QTL mapping in some respects, with main difference being the use of unstructured populations that provides much finer mapping resolution (Savolainen, et al. 2007). Association mapping employs large natural populations or populations of known pedigree to detect the co-segregation of marker alleles and phenotypic variation based on the rationale that functional loci or nearby loci (due to LD) are in non-random association with phenotypic differences caused by that locus (Pritchard and Przeworski 2001; Howe, et al. 2003). Association genetic studies can be based on either candidate genes of relevant function, or can be applied through whole genome, requiring no prior knowledge of genes. The decreasing cost of next-generation sequencing makes it possible to complete genome-wide association studies in both model and non-model organisms (Tuskan, et al. 2006). While association mapping is prone to spurious associations caused by population structure, gene flow is high in most temperate tree species leading to generally low levels of background structure

(Hamrick and Godt 1996). Nevertheless, population stratification must be controlled in association models to mitigate false positives due to population structure (Yu, et al. 2005).

1.5 *Populus trichocarpa* as model organism

Populus trichocarpa (black cottonwood) is a dioecious woody perennial tree native to North America (Slavov, et al. 2010; Levsen, et al. 2012). Typically, *P. trichocarpa* remain in vegetative growth for 7-10 years before reaching to reproductive maturity, flower in spring, and enter endodormancy in late fall (DS 1990; Slavov, et al. 2010). *P. trichocarpa* are fast-growing trees that are capable of clonal reproduction and long distance dispersal of pollen and seed (Slavov, et al. 2010). Black cottonwoods are of significant economic importance, producing light-weighted, short, fine fibers for pulp, lumber, and fiber products (DS 1990). More recently, poplars have been developed as short-rotation woody feedstocks for bioenergy. *P. trichocarpa* inhibits primarily in moist sites west of the Rocky mountains and along major streams west of the Cascade range (DS 1990). The range of *P. trichocarpa* extends from Kodiak Island in Alaska to southern California and northern Baja California, encompassing 30 degrees of latitude (DS 1990; Slavov, et al. 2012). The considerable latitudinal range of *P. trichocarpa* populations resulted from its migration history. *P. trichocarpa* initially occurred along northwest coastal areas of North America, and was restricted to coastal regions in southern refugia during glacial periods (Levsen, et al. 2012). The recolonization from refugial populations thus extended thousands of kilometers from sourthern to northern limits since the last glacial maxima, forming the spatially varying species range (Holliday, et al. 2016; Shafer, et al. 2010).

The massive range of *P. trichocarpa* encompasses highly heterogeneous climates that have resulted in steep genetic and phenotypic clines (Davis and Shaw 2001). A number of common garden studies have described differentiated ecophysiological and growth-related traits among *P. trichocarpa* populations (Gornall and Guy 2007; Porth, et al. 2013; McKown, et al. 2014). The life history characteristics of poplar – highly outcrossing with large effective population size – helps to maintain the abundant level of genetic variation within and among populations (Ingvarson 2010). A genome re-sequencing study of 120 individuals from 10 subpopulations of *P. trichocarpa* demonstrated the ample standing genetic variation and range-wide differentiation of allele frequencies (Slavov, et al. 2012). Regardless of high levels of genome diversity, *P.*

trichocarpa populations exhibit weak inter-population differentiation at neutral sites, suggesting low population structure (Ingvarson 2010; Slavov, et al. 2010). Natural *P. trichocarpa* populations also have relatively low level of linkage disequilibrium (LD), but with a slower LD decay (3-6 kb) than for most previously studied tree species (<2 kb) (Slavov, et al. 2012). The low LD, weak population structure along with range-wide genetic, and phenotypic differentiation make genome-wide association mapping feasible in this species (Slavov, et al. 2012; Wullschleger, et al. 2012).

P. trichocarpa was selected as a "model tree" for genome sequencing because of its modest genome size (40-50 times smaller than the pine genome), genetic transformation capability, ease of vegetative propagation and short generation time compared with other tree species (Tuskan, et al. 2006; Ellis, et al. 2010). Compared with the model annual *Arabidopsis*, *P. trichocarpa* has unique structural and physiological process such as vegetative growth to reproductive transition, dormancy traits, and wood formation (Wullschleger, et al. 2012). When long-lived woody perennials become established, they must survive environment selection throughout life, thus require adaptive strategies to cope with changing environments (Sork 2016).

1.6 Natural selection and local adaptation

Natural selection enables natural populations to adapt to local conditions by favoring genotypes of better fitness (Vitti, et al. 2013). While positive selection increases the prevalence of advantageous alleles within populations, negative selection decreases frequency of alleles and genotypes of diminished fitness (Vitti, et al. 2013). Within wide species ranges, natural populations encounter spatially heterogeneous selection resulting from a variety of environmental drivers (temperature, precipitation, radiation, atmospheric moisture), which generate locally adapted populations (Sork 2016). Selection does not act in isolation; other evolutionary forces can confound effects of natural selection. Genetic drift affects pattern of genetic variation within and among populations through random changes in allele frequencies, but is probably less effective in tree species with large population size (Ellstrand and Elam 1993; Sork 2016). Gene flow through effective dispersal mechanism in trees homogenizes and maintains connectivity among populations unless encounter geographic barriers (Sork 2016). But

selection is often strong enough not to be swamped by other evolutionary factors (Vitti, et al. 2013; Sork 2016).

Positive selection favors adaptive alleles and leaves signatures in targeted loci and adjacent chromosome regions (Bamshad and Wooding 2003). When selection acts on a beneficial mutation, it affects not only the target site but also linked neutral sites. As a result, variation at linked sites is reduced, especially in regions of low recombination, a phenomenon called selective sweep (Bamshad and Wooding 2003; Nielsen, et al. 2007). Another important result of selection is high frequency alleles with long range LD (Sabeti, et al. 2002; Bamshad and Wooding 2003). Many approaches have been developed to detect selection targets by screening for these signatures of selection across the genome and among-populations. For example, one LD-based method looks for regions of extended haplotype homozygosity (EHH) resulting from the selective sweep before recombination has time to break down the haplotype (Vitti, et al. 2013). The integrated haplotype score (iHS) is one measure of EHH that calculates the amount of EHH at a given site for ancestral alleles relative to derived allele (Hurst, et al. 2006; Vitti, et al. 2013). This method is important in detecting incomplete or ongoing selection (Hurst, et al. 2006). Selection can also result in allele frequency differentiation across populations when selection preferentially targets loci adaptive to particular environments (Barreiro, et al. 2008). F_{ST}, which characterizes the proportion of variation in allele frequencies attributable to within and among population components, is one way to identify loci that may have been targeted by divergent selection (Beaumont and Balding 2004).

1.7 Parallel adaptation

In response to selection from biotic (e.g. interspecific competition) and abiotic (e.g. temperature, precipitation) conditions within similar environmental contexts, populations within a given species or species lineages can evolve similar morphology, behavior or physiological traits independently (Arendt and Reznick 2008). This convergence can occur by random evolution (Stayton 2008), but many cases arise due to similar natural selection and developmental mechanisms involving the same genetic pathways or variants (Arendt and Reznick 2008; Elmer and Meyer 2011; Ord and Summers 2015). Members of the same or closely related species share genome features and habitat preferences, and thus are more likely to share genetic response and

evolve similar evolutionary solutions when exposed to similar selection pressure (Stern and Orgogozo 2008; Ord and Summers 2015). For example, in *Arabidopsis thaliana*, independent mutations within gene *FRIGIDA*, the key integrator of vernalization and flowering pathway, have spread across local populations and explain 70% of flowering time variation (Shindo, et al. 2005). The likelihood that the same genes underly parallel evolution is expected to decrease as the evolutionary time separating taxa increases (Conte, et al. 2012; Ord and Summers 2015). However, there have been cases where closely related organisms evolve the same phenotype through different genetic mechanisms, and where distantly-related organism evolve similar phenotypes via same genes (Arendt and Reznick 2008). For example, Hoekstra *et al.* reported a mutation in the *Mc1r* gene causes light-coloration in Gulf Coast populations of beach mice, but was not present in Atlantic costal populations that showed similar phenotype (Hoekstra, et al. 2006). More interestingly, the same genetic polymorphism in *Mc1r* gene occurs and controls pale color in lizards (Rosenblum, et al. 2004), birds (Theron, et al. 2001) and wooly mammoths (Theron, et al. 2001).

Convergence at the gene level among independent lineages can occur under 3 possible scenarios. First, the same genetic mutations may arise independently in different populations or species. Second, the shared causative allele among different populations or species may be due to their common ancestry. Third, the common causative gene may result from introgression (Stern 2013). Populations can adapt to new environments through mutations of similar phenotypic effects (Stern 2013). The "precision" with which the same newly arisen mutations occurs and sweeps to fixation at population scale is dependent on locus-specific mutation rate, the probability and magnitude of beneficial effect on fitness, genome size and complexity, gene interactions in regulatory networks, and population parameters (Ralph and Coop 2010; Stern 2013). One example of the first case is the independent origination of an identical variant in the GABA receptor among six insect species, which is the target of insecticide (Ffrench-Constant 1994). This provides evidence that particular variants can be favored over other variants and confer similar phenotypic changes. By contrast, the other two cases are more common and faster modes of adaptation, where the shared ancestral genetic variation have been targeted by selection in the past and have fitness advantage in multiple environments (Barrett and Schluter 2008; Stern 2013). The fixation of a newly risen beneficial allele will produce selective sweep along with

low polymorphism in its vicinity in the genome before recombination obscures it (Barrett and Schluter 2008). However, standing variation with a history of recombination and selection before becoming advantageous, leaves smaller and shallower signatures (Barrett and Schluter 2008). Genetic parallelism during independent phenotypic evolution reflects the key limiting-factors for genetic and phenotypic evolutionary possibilities, including selection pressures, standing genetic variation, genetic constraints, and population features (Koonin and Wolf 2010).

In plant species, a few studies in crop domestication indicated parallel response to selection at the level of gene expression and genetic mutations (Olsen and Wendel 2013). One study compared three independently domesticated cotton species and found the similar up-regulation of the profilin gene family, which is important in fiber development (Bao, et al. 2011). Olsen et al. pointed out that many mutations in domestication-related flowering time genes were represented among more than one crop species (Olsen and Wendel 2013). A recent study in sunflower reported the high repeatability of genomic regions underlying adaptive divergence among sunflower species along similar latitudinal gradients (Renaut, et al. 2014), and the genes showing the highest or lowest differentiation were most repeated, suggesting purifying and divergent selection shaping genome variation (Renaut, et al. 2014). Less is known about parallel adaptation in tree species (Holliday, et al. 2016). A recent study in *P. trichocarpa* identified extensive parallelism between loci for local adaption underling phenotypic variation along latitude and altitude transects (Holliday, et al. 2016). The geographical clines in *P. trichocarpa* provide opportunity to explore evidence of parallel adaptation and gene commonalities responding to clinal climate selection.

1.8 Mutational load on plant fitness

Mutations are of great evolutionary significance, providing genetic variation for adaptation and natural selection (Elena, et al. 2007). Mutations can be advantageous, deleterious, or neutral (Loewe and Hill 2010), and the effects of mutations depend on interaction with other alleles and the environment (Loewe and Hill 2010). When beneficial mutations arise, positive selection drives them to high frequency and soon fixation, resulting in selective sweeps at the target and linked sites (Chun and Fay 2011). The rate at which a population can adapt and reach fitness optimum largely depends on the mutation rate and the effect of beneficial mutations (Bataillon

and Bailey 2014). Mutations are defined as deleterious if they cause amino acid change leading to a suboptimal phenotype (Mezmouk and Ross-Ibarra 2014). Most deleterious mutations, individually are mildly deleterious, but collectively can affect population fitness (Bataillon and Bailey 2014). Deleterious mutations are constantly eliminated by purifying selection to maintain survival and health of species (Pritchard, et al. 2011). However, the accumulation of deleterious mutations is inevitable, especially when they are slightly damaging and selection strength is weak (Loewe and Hill 2010). The fate of mutations is determined by many factors including mutational effects, evolutionary factors, and population parameters (Whitlock and Bürger 2004; Loewe and Hill 2010).

Deleterious mutations are proposed to be a genetic basis for many complex diseases in humans (Kryukov, et al. 2007). A mutation can be detrimental if it modifies the protein sequence and structure, leading to detectable human-disease phenotype. Mildly deleterious alleles may cause no detectable phenotypic change, but collectively, accumulation of deleterious alleles can lead to severe "mutation load", defined as the reduction in fitness of individuals due to mutations (Kryukov, et al. 2007; Agrawal and Whitlock 2012). Genetic load can further lead to decreased population size and even extinction, especially in small populations, which is termed as "mutational meltdown" (Kimura, et al. 1963; Whitlock, et al. 2003; Whitlock and Bürger 2004). Deleterious recessive alleles are also the major cause for "inbreeding depression", where inbreeding exposed recessive deleterious alleles in the homozygous state lead to offspring of reduced survival and fertility (Charlesworth and Willis 2009). Damaging alleles can rise to high frequency through random drift or hitchhiking with beneficial mutations (Whitlock, et al. 2003; Chun and Fay 2011). Demographical processes such as bottlenecks and fluctuations in population size also enables deleterious alleles to reach high frequency (Hughes, et al. 2003; Gershoni and Pietrokovski 2014). Thus, the accumulation of deleterious alleles in populations reflects the evolutionary and demographic history of populations.

A large number of methods have been developed to predict deleterious or disease-causing mutations based on the effects of amino acid substitutions (Grantham 1974), effects on protein structure (Chasman and Adams 2001; Sunyaev 2001), and sequence conservation between species (Ng and Henikoff 2001; de Brevern, et al. 2012). Many of these methods were firstly

applied to human genomic variation to characterize the fitness impact of deleterious mutations, and how they evolve across human populations (Chun and Fay 2011; Lohmueller 2014; Gao, et al. 2015). One study examined deleterious mutations across the human genome and detected enrichment of deleterious SNPs within hitchhiking regions (Pritchard, et al. 2011). Another recent study reported no apparent differences in the genetic load and efficacy of selection across human populations (Lohmueller 2014). Comparatively, only a few studies of deleterious alleles have been reported in plant species (Lu, et al. 2006; Gossmann, et al. 2010; Mezmouk and Ross-Ibarra 2014; Renaut and Rieseberg 2015). Mezmouk et al found that deleterious variants were enriched in genes associated with fitness-related traits and hybrid vigor in maize, indicating the functional importance of deleterious alleles (Mezmouk and Ross-Ibarra 2014). Two studies in sunflower and rice both discovered the disproportionally accumulation of deleterious alleles during crop domestication and the preference of deleterious allele in genomic regions of low recombination (Lu, et al. 2006; Renaut, et al. 2014). The accumulation of deleterious mutations affect population fitness, limit the evolutionary potential and capacity of populations to respond novel environments (Hoffmann and Sgrò 2011; Bataillon and Bailey 2014). With limited prior knowledge of deleterious mutations in tree species, further studies are required to understand the fitness consequence of deleterious mutations and their evolutionary trajectory among natural populations.

1.9 Research Objective

The major objective of my study is to dissect the genomic architectures underlying climatic adaptation in *P. trichocarpa*. To accomplish that, I studied phenotypic variation of range-wide poplar samples and the correlation between a set of adaptive traits and geographical and climate variables (Chapter 2). I then used this population to complete a genome-wide association study to identify significant genes associated with adaptive traits (Chapter 3). In chapter 4, I used three selection scans to identify genomic targets of climatic selection. In chapter 5, I explored the distribution and effects of deleterious alleles across the genome and species range. Finally, in chapter 6, I summarized the important findings from the previous chapters.

1.10 Reference cited

Agrawal AF, Whitlock MC. 2012. Mutation Load: The Fitness of Individuals in Populations Where Deleterious Alleles Are Abundant. Annual Review of Ecology, Evolution, and Systematics 43:115-135.

Ågren J, Schemske DW. 2012. Reciprocal transplants demonstrate strong adaptive differentiation of the model organism Arabidopsis thaliana in its native range. New Phytologist 194:1112-1122.

Aitken SN, Bemmels JB. 2016. Time to get moving: assisted gene flow of forest trees. Evolutionary Applications 9:271-290.

Aitken SN, Whitlock MC. 2013. Assisted Gene Flow to Facilitate Local Adaptation to Climate Change. Annual Review of Ecology, Evolution, and Systematics 44:367-388.

Aitken SN, Yeaman S, Holliday JA, Wang TL, Curtis-McLane S. 2008. Adaptation, migration or extirpation: climate change outcomes for tree populations. Evolutionary Applications 1:95-111.

Anderson JT, Panetta AM, Mitchell-Olds T. 2012. Evolutionary and ecological responses to anthropogenic climate change: update on anthropogenic climate change. Plant Physiol 160:1728-1740.

Arendt J, Reznick D. 2008. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? Trends in Ecology & Evolution 23:26-32.

Azeez A, Miskolczi P, Tylewicz S, Bhalerao Rishikesh P. 2014. A Tree Ortholog of APETALA1 Mediates Photoperiodic Control of Seasonal Growth. Current Biology 24:717-724.

Bamshad M, Wooding SP. 2003. Signatures of natural selection in the human genome. Nature Reviews Genetics 4:99-111.

Bao Y, Hu G, Flagel LE, Salmon A, Bezanilla M, Paterson AH, Wang Z, Wendel JF. 2011. Parallel up-regulation of the profilin gene family following independent domestication of diploid and allopolyploid cotton (Gossypium). Proceedings of the National Academy of Sciences 108:21152-21157.

Barreiro LB, Laval G, Quach H, Patin E, Quintana-Murci L. 2008. Natural selection has driven population differentiation in modern humans. Nature Genetics 40:340-345.

Barrett RD, Schluter D. 2008. Adaptation from standing genetic variation. Trends in Ecology & Evolution 23:38-44.

Bataillon T, Bailey SF. 2014. Effects of new mutations on fitness: insights from models and data. Annals of the New York Academy of Sciences 1320:76-92.

Beaumont MA, Balding DJ. 2004. Identifying adaptive genetic divergence among populations from genome scans. Molecular Ecology 13:969-980.

Bertoni G. 2011. Dormancy Cycling in Populus: The Symplasmic Connection. The Plant Cell 23:3-3.

Bohlenius H. 2006. CO/FT Regulatory Module Controls Timing of Flowering and Seasonal Growth Cessation in Trees. Science 312:1040-1043.

Buckling A, Leimu R, Fischer M. 2008. A Meta-Analysis of Local Adaptation in Plants. Plos One 3:e4010.

Caffarra A, Donnelly A. 2010. The ecological significance of phenology in four different tree species: effects of light and temperature on bud burst. International Journal of Biometeorology 55:711-721.

Cannell MGR, Smith RI. 1986. Climatic Warming, Spring Budburst and Forest Damage on Trees. The Journal of Applied Ecology 23:177.

Charlesworth D, Willis JH. 2009. The genetics of inbreeding depression. Nature Reviews Genetics 10:783-796.
Chasman D, Adams RM. 2001. Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation11Edited by F. Cohen. Journal of Molecular Biology 307:683-706.
Chen THH, Howe GT, Bradshaw HD. 2002. Molecular genetic analysis of dormancy-related traits in poplars. Weed Science 50:232-240.
Chun S, Fay JC. 2011. Evidence for hitchhiking of deleterious mutations within the human genome. Plos Genetics 7:e1002240.
Clark PU, Shakun JD, Baker PA, Bartlein PJ, Brewer S, Brook E, Carlson AE, Cheng H, Kaufman DS, Liu Z, et al. 2012. Global climate evolution during the last deglaciation. Proceedings of the National Academy of Sciences 109:E1134-E1142.
Conte GL, Arnegard ME, Peichel CL, Schluter D. 2012. The probability of genetic parallelism and convergence in natural populations. Proceedings of the Royal Society B: Biological Sciences 279:5039-5047.
Cooke JEK, Eriksson ME, Junttila O. 2012. The dynamic nature of bud dormancy in trees: environmental control and molecular mechanisms. Plant, Cell & Environment 35:1707-1728.
Davis MB, Shaw RG. 2001. Range shifts and adaptive responses to Quaternary climate change. Science 292:673-679.
de Brevern AG, Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. 2012. Predicting the Functional Effect of Amino Acid Substitutions and Indels. Plos One 7:e46688.
Dewitt TJ, Sih A, Wilson DS. 1998. Costs and limits of phenotypic plasticity. Trends in Ecology & Evolution 13:77-81.
DS D. 1990. *Populus trichocarpa* Torr. & Gray, Black Cottonwood. In: Burns RM, BH H, editors. Silvics of North America Vol. 2. Hardwoods. Agriculture Handbook 654. US Department of Agriculture, Forest Service: Washington, DC.
Ehrenreich IM, Purugganan MD. 2006. The molecular genetic basis of plant adaptation. American Journal of Botany 93:953-962.
Elena SF, Wilke CO, Ofria C, Lenski RE. 2007. Effects of Population Size and Mutation Rate on the Evolution of Mutational Robustness. Evolution 61:666-674.
Ellis B, Jansson S, Strauss SH, Tuskan GA. 2010. Why and How *Populus* Became a "Model Tree". In: Jansson Sea, editor. Genetics and Genomics of *Populus*, Plant Genetics and Genomics: Crops and Models 8: Springer Science+Business Media, LLC.
Ellstrand NC, Elam DR. 1993. Population Genetic Consequences of Small Population Size: Implications for Plant Conservation. Annual Review of Ecology and Systematics 24:217-242.
Elmer KR, Meyer A. 2011. Adaptation in the age of ecological genomics: insights from parallelism and convergence. Trends in Ecology & Evolution 26:298-306.
Ffrench-Constant RH. 1994. The molecular and population genetics of cyclodiene insecticide resistance. Insect Biochem Mol Biol 24:335-345.
Frewen BE, Chen THH, Howe GT, Davis J, Rohde A, Boerjan W, Bradshaw HD. 2000. Quantitative trait loci and candidate gene mapping of bud set and bud flush in Populus. Genetics 154:837-845.
Furnier GR, Stine M, Mohn CA, Clyde MA. 1991. Geographic patterns of variation in allozymes and height growth in white spruce. Canadian Journal of Forest Research 21:707-712.
Galloway LF, Fenster CB. 2000. Population Differentiation in an Annual Legume: Local Adaptation. Evolution 54:1173-1181.

Gao Z, Waggoner D, Stephens M, Ober C, Przeworski M. 2015. An Estimate of the Average Number of Recessive Lethal Mutations Carried by Humans. Genetics 199:1243-1254.
Garcia-Gil MR, Mikkonen M, Savolainen O. 2003. Nucleotide diversity at two phytochrome loci along a latitudinal cline in Pinus sylvestris. Molecular Ecology 12:1195-1206.
Gershoni M, Pietrokovski S. 2014. Reduced selection and accumulation of deleterious mutations in genes exclusively expressed in men. Nature Communications 5.
Gonzalez P, Neilson RP, Lenihan JM, Drapek RJ. 2010. Global patterns in the vulnerability of ecosystems to vegetation shifts due to climate change. Global Ecology and Biogeography 19:755-768.
Gornall JL, Guy RD. 2007. Geographic variation in ecophysiological traits of black cottonwood (*Populus trichocarpa*)This article is one of a selection of papers published in the Special Issue on Poplar Research in Canada. Canadian Journal of Botany 85:1202-1213.
Gossmann TI, Song BH, Windsor AJ, Mitchell-Olds T, Dixon CJ, Kapralov MV, Filatov DA, Eyre-Walker A. 2010. Genome wide analyses reveal little evidence for adaptive evolution in many plant species. Molecular Biology and Evolution 27:1822-1832.
Grantham R. 1974. Amino Acid Difference Formula to Help Explain Protein Evolution. Science 185:862-864.
Grassein F, Lavorel S, Till-Bottraud I. 2014. The importance of biotic interactions and local adaptation for plant response to environmental changes: field evidence along an elevational gradient. Global Change Biology 20:1452-1460.
Green DS. 2005. Adaptive strategies in seedlings of three co-occurring, ecologically distinct northern coniferous tree species across an elevational gradient. Canadian Journal of Forest Research 35:910-917.
Grimm NB, Chapin FS, Bierwagen B, Gonzalez P, Groffman PM, Luo Y, Melton F, Nadelhoffer K, Pairis A, Raymond PA, et al. 2013. The impacts of climate change on ecosystem structure and function. Frontiers in Ecology and the Environment 11:474-482.
Hamann A, Wang TL. 2006. Potential effects of climate change on ecosystem and tree species distribution in British Columbia. Ecology 87:2773-2786.
Hamrick JL, Godt MJW. 1996. Effects of Life History Traits on Genetic Diversity in Plant Species. Philosophical Transactions of the Royal Society B: Biological Sciences 351:1291-1298.
Hänninen H. 1995. Effects of climatic change on trees from cool and temperate regions: an ecophysiological approach to modelling of bud burst phenology. Canadian Journal of Botany 73:183-199.
Hanninen H, Hakkinen R, Hari P, Koski V. 1990. Timing of growth cessation in relation to climatic adaptation of northern woody plants. Tree Physiology 6:29-39.
Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP. 2006. A single amino acid mutation contributes to adaptive beach mouse color pattern. Science 313:101-104.
Hoffmann AA, Sgrò CM. 2011. Climate change and evolutionary adaptation. Nature 470:479-485.
Holliday JA, Ritland K, Aitken SN. 2010. Widespread, ecologically relevant genetic markers developed from association mapping of climate-related traits in Sitka spruce (Picea sitchensis). New Phytologist 188:501-514.
Holliday JA, Suren H, Aitken SN. 2011. Divergent selection and heterogeneous migration rates across the range of Sitka spruce (Picea sitchensis). Proceedings of the Royal Society B: Biological Sciences 279:1675-1683.

Holliday JA, Zhou L, Bawa R, Zhang M, Oubida RW. 2016. Evidence for extensive parallelism but divergent genomic architecture of adaptation along altitudinal and latitudinal gradients in *Populus trichocarpa*. New Phytologist 209:1240-1251.
Horvath DP, Anderson JV, Chao WS, Foley ME. 2003. Knowing when to grow: signals regulating bud dormancy. Trends in Plant Science 8:534-540.
Howe GT, Aitken SN, Neale DB, Jermstad KD, Wheeler NC, Chen THH. 2003. From genotype to phenotype: unraveling the complexities of cold adaptation in forest trees. Canadian Journal of Botany-Revue Canadienne De Botanique 81:1247-1266.
Hsu CY, Adams JP, Kim H, No K, Ma C, Strauss SH, Drnevich J, Vandervelde L, Ellis JD, Rice BM, et al. 2011. FLOWERING LOCUS T duplication coordinates reproductive and vegetative growth in perennial poplar. Proceedings of the National Academy of Sciences 108:10756-10761.
Hughes AL, Packer B, Welch R, Bergen AW, Chanock SJ, Yeager M. 2003. Widespread purifying selection at polymorphic sites in human protein-coding loci. Proceedings of the National Academy of Sciences 100:15754-15757.
Hurst L, Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A Map of Recent Positive Selection in the Human Genome. PLoS Biol 4:e72.
Ibanez C, Kozarewa I, Johansson M, Ogren E, Rohde A, Eriksson ME. 2010. Circadian clock components regulate entry and affect exit of seasonal dormancy as well as winter hardiness in Populus trees. Plant Physiol 153:1823-1833.
Ingvarson PrK. 2010. Nucleotide Polymorphism, Linkage Disequilibrium and Complex Trait Dissection in Populus. In: al Je, editor. Genetics and Genomics of Populus, Plant Genetics 91 and Genomics: Crops and Models 8: Springer Science+Business Media, LLC 2010.
Jermstad KD, Bassoni DL, Jech KS, Ritchie GA, Wheeler NC, Neale DB. 2003. Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas fir. III. Quantitative trait loci-by-environment interactions. Genetics 165:1489-1506.
Kawecki TJ, Ebert D. 2004. Conceptual issues in local adaptation. Ecology Letters 7:1225-1241.
Keir KR, Bemmels JB, Aitken SN. 2011. Low genetic diversity, moderate local adaptation, and phylogeographic insights in Cornus nuttallii (Cornaceae). American Journal of Botany 98:1327-1336.
Kimura M, Maruyama T, Crow JF. 1963. The Mutation Load in Small Populations. Genetics 48:1303-1312.
Koonin EV, Wolf YI. 2010. Constraints and plasticity in genome and molecular-phenome evolution. Nature Reviews Genetics 11:487-498.
Kryukov GV, Pennacchio LA, Sunyaev SR. 2007. Most Rare Missense Alleles Are Deleterious in Humans: Implications for Complex Disease and Association Studies. The American Journal of Human Genetics 80:727-739.
Lander ES, Botstein D. 1989. Mapping Mendelian Factors Underlying Quantitative Traits Using Rflp Linkage Maps. Genetics 121:185-199.
Leinonen I. 1997. Changing Environmental Effects on Frost Hardiness of Scots Pine During Dehardening. Annals of Botany 79:133-137.
Levsen ND, Tiffin P, Olson MS. 2012. Pleistocene Speciation in the Genus Populus (Salicaceae). Systematic Biology 61:401-412.
Littell JS, McKenzie D, Peterson DL, Westerling AL. 2009. Climate and wildfire area burned in western U.S. ecoprovinces, 1916–2003. Ecological Applications 19:1003-1021.
Loarie SR, Duffy PB, Hamilton H, Asner GP, Field CB, Ackerly DD. 2009. The velocity of climate change. Nature 462:1052-1055.

Loewe L, Hill WG. 2010. The population genetics of mutations: good, bad and indifferent. Philosophical Transactions of the Royal Society B: Biological Sciences 365:1153-1167.
Lohmueller KE. 2014. The distribution of deleterious genetic variation in human populations. Current Opinion in Genetics & Development 29:139-146.
Lu J, Tang T, Tang H, Huang J, Shi S, Wu CI. 2006. The accumulation of deleterious mutations in rice genomes: a hypothesis on the cost of domestication. Trends Genet 22:126-131.
McKown AD, Guy RD, Klapste J, Geraldes A, Friedmann M, Cronk QC, El-Kassaby YA, Mansfield SD, Douglas CJ. 2014. Geographical and environmental gradients shape phenotypic trait variation and genetic structure in *Populus trichocarpa*. New Phytologist 201:1263-1276.
Mezmouk S, Ross-Ibarra J. 2014. The pattern and distribution of deleterious mutations in maize. G3 (Bethesda) 4:163-171.
Neale DB, Savolainen O. 2004. Association genetics of complex traits in conifers. Trends in Plant Science 9:325-330.
Ng PC, Henikoff S. 2001. Predicting Deleterious Amino Acid Substitutions. Genome Research 11:863-874.
Norgaard Nielsen CC, Rasmussen HN. 2009. Frost hardening and dehardening in Abies procera and other conifers under differing temperature regimes and warm-spell treatments. Forestry 82:43-59.
Notivol E, García-Gil MR, Alía R, Savolainen O. 2007. Genetic variation of growth rhythm traits in the limits of a latitudinal cline in Scots pine. Canadian Journal of Forest Research 37:540-551.
Olsen JE, Junttila O, Nilsen J, Eriksson ME, Martinussen I, Olsson O, Sandberg G, Moritz T. 1997. Ectopic expression of oat phytochrome A in hybrid aspen changes critical daylength for growth and prevents cold acclimatization. The Plant Journal 12:1339-1350.
Olsen KM, Wendel JF. 2013. Crop plants as models for understanding plant adaptation and diversification. Front Plant Sci 4.
Ord TJ, Summers TC. 2015. Repeated evolution and the impact of evolutionary history on adaptation. BMC Evolutionary Biology 15.
Parmesan C, Hanley ME. 2015. Plants and climate change: complexities and surprises. Ann Bot 116:849-864.
Parry ML, Rosenzweig C, Iglesias A, Livermore M, Fischer G. 2004. Effects of climate change on global food production under SRES emissions and socio-economic scenarios. Global Environmental Change 14:53-67.
Partanen J, Koski V, Hanninen H. 1998. Effects of photoperiod and temperature on the timing of bud burst in Norway spruce (Picea abies). Tree Physiology 18:811-816.
Porth I, Klápště J, Skyba O, Lai BSK, Geraldes A, Muchero W, Tuskan GA, Douglas CJ, El-Kassaby YA, Mansfield SD. 2013. Populus trichocarpa cell wall chemistry and ultrastructure trait variation, genetic control and genetic correlations. New Phytologist 197:777-790.
Pritchard JK, Chun S, Fay JC. 2011. Evidence for Hitchhiking of Deleterious Mutations within the Human Genome. Plos Genetics 7:e1002240.
Pritchard JK, Przeworski M. 2001. Linkage Disequilibrium in Humans: Models and Data. The American Journal of Human Genetics 69:1-14.
Ralph P, Coop G. 2010. Parallel Adaptation: One or Many Waves of Advance of an Advantageous Allele? Genetics 186:647-668.

Ramos A, Perez-Solis E, Ibanez C, Casado R, Collada C, Gomez L, Aragoncillo C, Allona I. 2005. Winter disruption of the circadian clock in chestnut. Proceedings of the National Academy of Sciences 102:7037-7042.
Reed TE, Waples RS, Schindler DE, Hard JJ, Kinnison MT. 2010. Phenotypic plasticity and population viability: the importance of environmental predictability. Proceedings of the Royal Society B: Biological Sciences 277:3391-3400.
Renaut S, Owens GL, Rieseberg LH. 2014. Shared selective pressure and local genomic landscape lead to repeatable patterns of genomic divergence in sunflowers. Molecular Ecology 23:311-324.
Renaut S, Rieseberg LH. 2015. The Accumulation of Deleterious Mutations as a Consequence of Domestication and Improvement in Sunflowers and Other Compositae Crops. Molecular Biology and Evolution 32:2273-2283.
Repo T, Hanninen H, Kellomaki S. 1996. The effects of long-term elevation of air temperature and CO on the frost hardiness of Scots pine. Plant, Cell and Environment 19:209-216.
Rinne PLH, Welling A, Vahala J, Ripel L, Ruonala R, Kangasjärvi J, van der Schoot C. 2011. Chilling of Dormant Buds HyperinducesFLOWERING LOCUS Tand Recruits GA-Inducible 1,3-β-Glucanases to Reopen Signal Conduits and Release Dormancy in Populus. The Plant Cell 23:130-146.
Rohde A, Storme V, Jorge V, Gaudet M, Vitacolonna N, Fabbrini F, Ruttink T, Zaina G, Marron N, Dillen S, et al. 2011. Bud set in poplar - genetic dissection of a complex trait in natural and hybrid populations. New Phytologist 189:106-121.
Rosenblum EB, Hoekstra HE, Nachman MW. 2004. Adaptive reptile color variation and the evolution of the Mc1r gene. Evolution 58:1794-1808.
Ruttink T, Arend M, Morreel K, Storme V, Rombauts S, Fromm J, Bhalerao RP, Boerjan W, Rohde A. 2007. A Molecular Timetable for Apical Bud Formation and Dormancy Induction in Poplar. The Plant Cell Online 19:2370-2390.
Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, et al. 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature 419:832-837.
Savolainen O, Pyhäjärvi T, Knürr T. 2007. Gene Flow and Local Adaptation in Trees. Annual Review of Ecology, Evolution, and Systematics 38:595-619.
Saxe H, Cannell MGR, Johnsen B, Ryan MG, Vourlitis G. 2001. Tree and forest functioning in response to global warming. New Phytologist 149:369-399.
Saxe H, Ellsworth DS, Heath J. 1998. Tree and forest functioning in an enriched CO2 atmosphere. New Phytologist 139:395-436.
Shafer ABA, Cullingham CI, CÔTÉ SD, Coltman DW. 2010. Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. Molecular Ecology 19:4589-4621.
Shindo C, Aranzana MJ, Lister C, Baxter C, Nicholls C, Nordborg M, Dean C. 2005. Role of FRIGIDA and FLOWERING LOCUS C in determining variation in flowering time of Arabidopsis. Plant Physiol 138:1163-1173.
Slavov GT, DiFazio SP, Martin J, Schackwitz W, Muchero W, Rodgers-Melnick E, Lipphardt MF, Pennacchio CP, Hellsten U, Pennacchio LA, et al. 2012. Genome resequencing reveals multiscale geographic structure and extensive linkage disequilibrium in the forest tree *Populus trichocarpa*. New Phytologist 196:713-725.

Slavov GT, Leonardi S, Adams WT, Strauss SH, DiFazio SP. 2010. Population substructure in continuous and fragmented stands of *Populus trichocarpa*. Heredity 105:348-357.
Sork VL. 2016. Gene flow and natural selection shape spatial patterns of genes in tree populations: implications for evolutionary processes and applications. Evolutionary Applications 9:291-310.
St Clair JB, Howe GT. 2007. Genetic maladaptation of coastal Douglas-fir seedlings to future climates. Global Change Biology 13:1441-1454.
Stayton CT. 2008. Is convergence surprising? An examination of the frequency of convergence in simulated datasets. Journal of Theoretical Biology 252:1-14.
Stern DL. 2013. The genetic causes of convergent evolution. Nature Reviews Genetics 14:751-764.
Stern DL, Orgogozo V. 2008. The Loci of Evolution: How Predictable Is Genetic Evolution? Evolution 62:2155-2177.
Sturm M, Racine C, Tape K. 2001. Climate change. Increasing shrub abundance in the Arctic. Nature 411:546-547.
Sultan SE. 2003. Phenotypic plasticity in plants: a case study in ecological development. Evol Dev 5:25-33.
Sunyaev S. 2001. Prediction of deleterious human alleles. Hum Mol Genet 10:591-597.
Theron E, Hawkins K, Bermingham E, Ricklefs RE, Mundy NI. 2001. The molecular basis of an avian plumage polymorphism in the wild: a melanocortin-1-receptor point mutation is perfectly associated with the melanic plumage morph of the bananaquit, Coereba flaveola. Curr Biol 11:550-557.
Thomashow MF. 1999. PLANT COLD ACCLIMATION: Freezing Tolerance Genes and Regulatory Mechanisms. Annual review of plant physiology and plant molecular biology 50:571-599.
Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, et al. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313:1596-1604.
Ungerer MC, Halldorsdottir SS, Modliszewski JL, Mackay TF, Purugganan MD. 2002. Quantitative trait loci for inflorescence development in Arabidopsis thaliana. Genetics 160:1133-1151.
Vitasse Y, Lenz A, Körner C. 2014. The interaction between freezing tolerance and phenology in temperate deciduous trees. Front Plant Sci 5.
Vitti JJ, Grossman SR, Sabeti PC. 2013. Detecting Natural Selection in Genomic Data. Annual review of genetics 47:97-120.
Weeks AR, Sgro CM, Young AG, Frankham R, Mitchell NJ, Miller KA, Byrne M, Coates DJ, Eldridge MDB, Sunnucks P, et al. 2011. Assessing the benefits and risks of translocations in changing environments: a genetic perspective. Evolutionary Applications 4:709-725.
Weiser CJ. 1970. Cold Resistance and Injury in Woody Plants: Knowledge of hardy plant adaptations to freezing stress may help us to reduce winter damage. Science 169:1269-1278.
Whitlock MC, Bürger R. 2004. Fixation of New Mutations in Small Populations. In: Ferrière R, Dieckmann U, Couvet D, editors. Evolutionary Conservation Biology. Cambridge, UK: Cambridge University Press. p. 155-170.
Whitlock MC, Griswold CK, Peters AD. 2003. Compensating for the meltdown: The critical effective size of a population with deleterious and compensatory mutations. Annales Zoologici Fennici 40:169-183.

Wullschleger SD, Weston DJ, DiFazio SP, Tuskan GA. 2012. Revisiting the sequencing of the first tree genome: *Populus trichocarpa*. Tree Physiology 33:357-364.
Yang RC, Yeh FC, Yanchuk AD. 1996. A comparison of isozyme and quantitative genetic variation in Pinus contorta ssp. latifolia by FST. Genetics 142:1045-1052.
Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, et al. 2005. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics 38:203-208.
Zavaleta ES. 2005. Shrub establishment under experimental global changes in a California grassland. Plant Ecology 184:53-63.

2. Clinal variation of climate-related traits

Abstract

Local adaptation to climate allows plants to cope with heterogeneous environments both temporally and spatially, and understanding the patterns and processes that contribute to local adaptation impinges on our understanding of the potential for future adaptative changes under anthropogenic climate change. *P. trichocarpa,* a temperate tree species native to western North America with an extensive geographic distribution, has emerged as a model system to understand the ecological genomics of local adaptation. This work investigated the distribution of phenotypic traits across both latitudinal and altitudinal transects in *P. trichcarpa*, and how geography and climate shapes the relevant phenotypes, which include vegetative, cold hardiness, growth, and coppice regeneration. We detected significant climate correlation with geography, temperature, precipitation, and frost-free period, suggesting these variables are the primary drivers of observed phenotypic variation. We also demonstrated effect of genotype, environment, and genotype-environment interaction on phenotypic variation in replicated common gardens.

2.1 Introduction

Forest trees comprise more than 80% of terrestrial biomass and hold an ecologically significant role in many terrestrial ecosystems (Aitken, et al. 2008). In addition, trees are economically important in many areas, supplying food, biofuels, timber, and pulp. However, ongoing climate change is intensifying climate-related stress on tree species worldwide by altering the structure, composition, and dynamics of current forest ecosystems (Xie, et al. 2015). Though increasing atmospheric CO_2 and temperature may have a positive impact on average forest biomass and productivity, forests are limited by water and nutrient availability (Boisvenue and Running 2006; van Mantgem and Stephenson 2007). Furthermore, studies suggest that warmer winter temperature is leading to disrupted seasonal timing of growth events. For example, flowering and bud flush have advanced 2 days per decade in North America between 1959 and 1993 (Cleland, et al. 2007; Khanduri, et al. 2008). A meta-investigation on 677 plant and animal species showed that increased temperature led to advanced spring phenology events such as budburst and bird migration (Parmesan and Yohe 2003). Temperate tree species, which transit between periods of growth and dormancy corresponding to seasonal environmental conditions, are especially sensitive and vulnerable to a changing climate, and evidence of mortality attributed to climate-driven abiotic and biotic stress is increasing (van Mantgem and Stephenson 2007; Way and Oren 2010).

Temperate trees alternate between active growth and vegetative dormancy by tracking the day length and temperature (Shim, et al. 2014). In response to short days in late summer, growth cessation and dormancy is induced. The transition to endodormancy is then initiated by endogenous repressing growth factors, exposure to low temperature, and closed symplastic communication (Bertoni 2011; Cooke, et al. 2012). Arrested growth capability is restored with winter chilling temperature, and subsequent heat sum accumulation promotes bud break and a resumption of growth (Horvath, et al. 2003; Shim, et al. 2014). This progression is a protective mechanism against the cold temperatures of winter, which are incompatible with growth. The development of cold acclimation is initiated at the onset of autumn and further developed during the winter (Weiser 1970). Warmer winter temperature can cause trees to leaf-out early or acquire incomplete cold hardiness, which expose trees to early or late spring frost damage. The timing of spring bud break is an adaptive process that balances the need to avoid frost damage while

maximizing growing days (Li, et al. 2010). Experimental work showed that some tree species that are quick in responding to chilling and have quick leaf development may suffer more heavily from frost damage as fluctuating winter temperatures associated with climate change become more pronounced (Linkosalo, et al. 2000; Polgar and Primack 2011). Thus, the proper timing of bud phenology and cold acclimation is of adaptive significance in coordinating plant growth with seasonal climate, and ensuring winter survival (Anderson, et al. 2010).

Plant ecophysiological characteristics such as growth, canopy structure, nutrient allocation and phenological traits such as flowering, timing of flushing and bud set are likely to influence fitness and adaptation to local conditions (Ackerly, et al. 2000). Thus, these traits are of considerable adaptive significance, and ensure individual fitness in a particular environmental niche (Kremer, et al. 2012). Tree species naturally occupy extensive geographical ranges, and concomitant spatially varying selection factors such as climate, photoperiod and water availability often lead to genetic and phenological divergence over time (Savolainen, et al. 2007). Previous common garden experiments showed that wide-ranging populations differ in their timing of seasonal growth cycles and their ability to withstand cold temperatures as a result of climate-associated natural selection (St. Clair 2006). Such clinal variation of these adaptive traits represents local adaptation to environmental gradients (Hurme, et al. 1997; Vihera-Aarnio, et al. 2005). While phenotypic plasticity may play a role, spatial phenotypic differentiation is usually dependent on underlying adaptive genetic variation (Savolainen, et al. 2007), which has been documented in many tree species (e.g., *Picea abies* (Oleksyn, et al. 1998), *Quercus coccifera* (Baquedano, et al. 2008), *Betula pendula* (Kelly, et al. 2003), *Quercus petraea* (Ducousso, et al. 1996), *Populus balsamifera* (Soolanayakanahally, et al. 2009)).

P. trichocarpa is a fast growing hardwood tree species native to the temperate zone of western North America (Gornall and Guy 2007), and occupies a widespread geographical distribution from southern Alaska to California. Extensive variation in adaptive traits has been documented across the species range for traits such as phenology, biomass and growth, wood biochemistry, and lignin content, suggesting these traits are subject to spatially varying selection (Gornall and Guy 2007; Gailing, et al. 2009; McKown, et al. 2014). Thus, natural populations of *P. trichocarpa* represent a genetic resource for exploring adaptive differentiation (Eckert and Dyer

2012; McKown, et al. 2014). Despite extensive literature focusing on phenotypic adaptation across latitudinal gradients, little attention has been drawn on dissecting altitudinal phenotypic differentiation in this species (Klepsatel, et al. 2014; Halbritter, et al. 2015). Altitude should mirror latitude in driving adaptive evolution in tree species (Thomas 2011). However, the latitudinal clines reflects differences in temperature and photoperiod, while altitude is correlated with temperature, frost, atmospheric pressure and solar radiation (Körner 2007; Klepsatel, et al. 2014). Hence, we expect similar phenotypic changes corresponding to environmental selection along latitudinal and altitudinal gradients, though some of the details may be different. To better understand the role of geography and climate on tree growth and phenology, we collected *P. trichocarpa* clones throughout most of its species range and studied the pattern of adaptive traits, including timing of bud set and flush, height, diameter and cold hardiness, regeneration height, and branch number within samples separately across latitudinal and altitudinal samples. We evaluated the relationship between phenotypic traits and climate variables for both latitudinal and altitudinal samples to further illustrate how climate factors affect seasonal growth, and to determine if plants respond differently to climate variables across these two transect types. The replication of genotypes in two common gardens also allowed us to assess the contribution of environment, genotype, and genotype-environment interaction (G x E) to phenotypic variation, which should help understand the stability of tree response to different environments.

2.2 Materials and Methods

Plant material and sampling location

Black cottonwood branch cuttings of unstructured natural populations were collected from 182 provenances in 2010. The provenances span from southeast Alaska to northern California across 20 degrees of latitude and cover most of the species' range (Figure 2.1). The geographical locations (latitude, longitude and elevation) were recorded for each sampling location. The sampled cuttings were first rooted and cultivated in a greenhouse and subsequently planted in May 2011 in two outdoor common gardens located in Virginia and British Columbia. Each of four replicates was planted in a randomized complete block design with four blocks. Climate variables were estimated for each sampling site with software ClimateWNA (Wang, et al. 2012) based on averages for the period 1961-1990. Annual climate variables included mean annual temperature (MAT; °C), mean warmest monthly temperature (MWMT; °C), mean coldest

monthly temperature (MCMT; °C), temperature difference between MWMT and MCMT (TD; °C), mean annual precipitation (MAP; mm), mean annual summer (May to September) precipitation (MSP; mm), annual heat moisture index (AHM), summer heat moisture index (SHM). Derived variables included chilling days below 0 °C (DD_0), growing degree days above 5°C (DD5), frost-free period (FFD) and Hargreaves reference evaporation (Eref; mm).

Phenotyping of climate-related traits

Phenotypic data collected were timing of bud set and bud flush, height and diameter, and cold hardiness. Height and diameter were measured in March 2014. Bud set stage was scored on a weekly basis beginning July 1st until most trees formed terminal bud covered with stipules in year 2012 and 2015 (Frewen, et al. 2000). The bud stages were assessed following the method described in paper (Rohde, et al. 2002) and the days elapsed for trees to reach fully developed bud since January 1st of the year was used as bud set date. Bud flush was defined as the date on which the first fully unfolded leaf emerged (Frewen, et al. 2000). We scored bud flush stages every week starting from observing first flushed tree in spring 2013 until all trees finished bud flush. The timing of bud flush was expressed as the number of days elapsed since January 1st. Cold hardiness was measured for the cambial meristem of lateral branches of all clones with a quantitative assay in 2012. When tissue samples are frozen at temperature beyond their limit, the release of electrolytes from dead cells can increase the electrical conductivity of the solution. Therefore, the rise of electrical conductivity provides proxy measurement for cold injury of the samples. Based on this principle, we first sectioned lateral shoots evenly into small discs of 2 mm width and placed 3 discs of each clone into 0.1mL distilled water with a miniscule amount of silver iodide to aid in ice nucleation. Preliminary tests suggested temperature treatments of -8°C, -14°C and -20°C would provide maximum variation between samples. Samples were placed in a Tenney programmable freezer in which the temperature was lowered from a starting point of 4°C at a rate 4°C/hr until the desired temperature was reached, at which point the samples were held at that temperature for 1hr before being transferred to a refrigerator at 4°C to thaw. After thawing, 2mL distilled water was added to each well of the plate and the plates were shaken for 1 hour on an orbital shaker at 120 rpm before the electrolytic conductivity of the solution was measured. Control samples were maintained at 4 °C throughout the experiment. Conductivity of the solution was measured when plates reach room temperature. Finally, frozen samples and

control samples were heat-killed in 95 °C water bath and were shaken 1 hour followed by a second conductivity measurement. The cold injury index was calculated as follows (Hannerz, et al. 1999):

$$I_t = \frac{100(R_t - R_0)}{1 - R_0}$$

where $R_t = L_t/L_k$, $R_0 = L_0/L_d$, L_t is the solution conductance of sample frozen at temperature t, L_t is the conductance heat-killed of sample treated with temperature t, L_0 is the conductance of unfrozen control group, and L_d is the conductance of the heat-killed sample that are unfrozen. All trees in Virgina garden were chopped down in May 2014 and we then recorded the height of main branch and the number of regenerated branches after regeneration in March 2015.

Estimation of phenotypic BLUP

We estimated clonal best linear unbiased predictors (BLUP) of each trait with a linear mixed model using the *lmer* function in *lmerTest* package in R (http://www.r- project.org/). The model is as follows:

$$y_{ijk} = \mu + b_i + c_j + \varepsilon_{ijk}$$

where y_{ijk} is the phenotype of the kth replicate of jth clone in ith block, μ is the grand mean, b_i is the effect of ith block, c_j is the effect of jth clone, and ε_{ijk} is the random error following $N(0, \sigma_\varepsilon^2 I)$ where σ_ε^2 is variance and I is the identity matrix. Both block and clone effect were fitted as random effects and the BLUPs of each clone were calculated for each trait.

Correlation and regression analysis

To better understand the environmental drivers of the adaptive trait variation, we regressed our phenotypic BLUPs on the above climate and geographical variables. With 451 genotyped poplar clones, we separated the total samples into latitudinal group spanning most of the species range (285 genotypes) and altitudinal group combining samples collected from two transects – Coquihalla Highway (49.75°N, 121.00°W; 63 genotypes) and Highway 99 (50.08°N, 123.06°W; 103 genotypes) in British Columbia. The elevational difference between low and high sampling sites was about 1000m, which corresponds to MAT difference of 6-7 °C for both transects. We first regressed phenotypic BLUPs on latitude/altitude separately for two groups of samples because latitude/altitude are the major geographic factors driving local adaptation. Since some

tree accessions from southern California within latitudinal transect were collected from high elevation, we estimated the phenotypic variation across distance to the species range center (49.36°N, -124.36°W) for each clone as a proxy for latitude. The distance to range center was calculated as the earth circle distance to species center. Pearson's correlation between phenotypic BLUPs, geographic variables (latitude, longitude, elevation) and climate variables were further calculated and compared within latitude and altitude transects. The significance of correlation coefficient was tested with the R package *Hmisc* and derived P values were adjusted with a Bonferroni correction for multiple comparisons.

Estimation of $G \times E$ effect

To determine the influence of environment and genotype on phenotypic traits, we utilized plant height and days of bud set and bud flush collected from two common gardens at different locations. We referred the source location of clones as their genotype. The climate parameters of the two common gardens are extracted with ClimateWNA. The gardens differ in mean annual temperature by 5 °C (Virginia garden: 13.8°C, British Columbia garden: 8.7°C), number of frost-free days by 60 days (Virginia garden: 214, British Columbia garden: 274), mean annual precipitation by 356 mm (Virginia garden: 1166 mm, British Columbia garden: 1522 mm). These differences allow us to determine the effect of genotype, environment, and $G \times E$ interactions on plant phenotypes. Statistical analyses are performed with *lmerTest* R package with model as follows:

$$y_{ijmk} = \mu + e_i + g_j + (eg)_{ij} + b_{im} + \varepsilon_{ijmk}$$

Where y_{ijmk} is the kth replicate of jth clone in mth block in ith environment (common garden), e_i is the effect of ith environment, g_j is the effect of jth genotype (clone), $(eg)_{ij}$ is the interaction between ith environment and jth genotype, b_{im} is the effect of mth block in ith environment, and ε_{ijmk} is the random error following $N(0, \sigma_\varepsilon^2 I)$ where σ_ε^2 is variance and I is the identity matrix. Environmental and interaction effect were fitted as fixed effects and block effect as a random effect nested within environment. To have a reasonable number of samples across both environment, we combined all samples to estimate the genotype-environment interaction.

2.3 Results

Clinal variation of phenotypic traits

We observed strong clinal variation between bud set timing and latitude in both the Virginia ($P<0.001$, $R^2=0.22$) and British Columbia gardens ($P<0.001$, $R^2=0.28$) (Figure 2.2). Latitude had a strong positive relationship with the days to bud break in the Virginia garden ($P<0.001$, $R^2=0.30$), but a non-significant correlation with bud flush days in British Columbia garden ($P>0.05$) (Figure 2.2). There was also a significant negative correlation between plant height and the distances to species center in both common gardens (Virgina garden: $P< 0.001$, $R^2=0.19$; British Columbia garden: $P<0.001$, $R^2=0.45$) (Figure 2.2). Plant diameter significantly decreased as distance to the range center increased (Virgina garden: $P<0.05$, $R^2=0.029$; Figure 2.2). Cold injury index values were much higher in trees of southern origin than northern clones (Virgina garden: $P< 0.001$, $R^2=0.18$; Figure 2.2). Regeneration height significantly decreased as the distance to range center increased (Virgina garden: $P< 0.001$, $R^2=0.062$) and regenerated branch number showed similar decreasing pattern as distance to range center increased (Virgina garden: $P<0.001$, $R^2=0.082$) (Figure 2.2).

Samples across the altitude transect showed similar clinal variation in phenotypic traits but relationships were weaker. Timing of bud set decreased with increasing elevation in both common gardens (Virgina garden: $P< 0.001$, $R^2=0.16$; British Columbia garden: $P<0.05$, $R^2=0.032$) (Figure 2.3). Bud flush timing for high elevations was much later than for low elevations in the Virginia garden (Virgina garden: $P< 0.001$, $R^2=0.15$; British Columbia garden: $P>0.05$, $R^2=0.017$) (Figure 2.3). Plant height (Virgina garden: $P< 0.001$, $R^2=0.14$; British Columbia garden: $P< 0.001$, $R^2=0.11$) and diameter (Virgina garden: $P< 0.001$, $R^2=0.14$) decreases significantly as elevation increases (Figure 2.3). Cold injury showed a slight but significant decrease as elevation increased (Virgina garden: $P< 0.05$, $R^2=0.036$; Figure 2.3). Samples originating from high elevations also displayed lower regeneration height (Virgina garden: $P< 0.001$, $R^2=0.018$; Figure 2.3) and branch number (Virgina garden: $P< 0.001$, $R^2=0.15$; Figure 2.3) than those from lower elevations (Figure 2.3).

Correlation between phenotypes, geography, and climate variables

Most phenotypic traits showed significant co-variation with climate and geographical variables across both transects (Figure 2.4; Figure 2.5). For latitudinal samples, the timing of bud flush and bud set had a strong negative relationship ($r= -0.66$, $P<0.001$; Table 2.1). Timing of bud set was

positively correlated with plant height (r=0.59, P<0.001; Table 2.1) and diameter (r=0.53, P<0.001; Table 2.1), while timing of bud flush had a negative relationship with height (r=-0.56, P<0.001; Table 2.1) and diameter (r=-0.46, P<0.001; Table 2.1). Height, on the other hand, had a very high correlation with diameter (r=0.87, P<0.001; Table 2.1). Pairwise correlations among phenotypic traits for altitudinal samples in general had the same correlation direction but different strength as compared with latitudinal samples (Table 2.2).

We also observed strong correlations for phenotypic traits with climate variables along the latitudinal transect (Table 2.3, Figure 2.4). There was a notable correlation between timing of bud set with most climate variables, including mean annual temperature (MAT) (r=0.49, P<0.001), mean warmest month temperature (MWMT) (r=0.42, P<0.001), mean coldest month temperature (MCMT) (r=0.45, P<0.001), degree days above 5 °C (DD5) (r=0.50, P<0.001), frost-free period (FFP) (r=0.41, P<0.001), temperature difference (TD) (r=-0.32, P<0.001) and degree days above 0 °C (DD_0) (r=-0.38, P<0.001) (Table 2.3). Bud flush, on the other hand showed negative relationships with MAT (r=-0.56, P<0.001), MWMT (r=-0.47, P<0.001), MCMT (r=-0.53, P<0.001), SHM (r=-0.41, P<0.001) and FFP (r=-0.47, P<0.001), Eref (r=-0.53, P<0.001) and positive relationship with TD (r=0.40, P<0.001), DD_0 (r=0.45, P<0.001) (Table 2.3). Height and diameter both showed positive relationship with MAT, MWMT, MCMT, DD5, FFP, Eref, and negative relationship with TD, MAP, DD_0 (Table 2.3). Cold injury showed a positive relationship with MAT (r=0.38, P<0.001), MCMT (r=0.42, P<0.001), DD5 (r=0.36, P<0.001), FFP (r=0.43, P<0.001) and negative relationship with TD (r=-0.41, P<0.001), DD_0 (r=-0.34, P<0.001). MAT, MCMT, TD, DD_0, DD_5 and Eref are most strongly correlated with phenology and physiology traits along latitude samples. Similarly, across altitude, bud phenology and physiological traits were most correlated with temperature-related variables (MAT, MWMT, MCMT, TD) (Table 2.4; Figure 2.5). Most phenotypic traits in the altitudinal transect displayed less strong relationship with moisture-related climate variables (AHM, SHM, Eref) compared with latitudinal samples (Table 2.4). MAT, MWMT, MCMT, TD, DD_0, DD_5 and FFP were most strongly correlated with height and diameter, while AHM/SHM had weak correlations with these traits (Table 2.4; Figure 2.5).

Climate variables related to temperature (MAT, MWT, MCMT, TD), moisture (MSP, AHM, SHM, Eref) and degree-days (DD_0, DD5, FFP) exhibited significant correlations with latitude (Table 2.5). The strongest correlations with latitude were MAT (r=-0.86, P<0.001), DD5 (r=-0.85, P<0.001) and Eref (r=-0.90, P<0.001) (Table 2.5). MAP and MSP co-varied with each other and with heat moisture indices (AHM, SHM), but not strongly with other climate variables (Table 2.5). Latitude showed a negative correlation with longitude (r =-0.63, P<0.001). In contrast, the strongest correlations across altitude were MAT (r=-0.98, P<0.001), MWMT (r=-0.91, P<0.001), MCMT (r=-0.90, P<0.001), DD5 (r=-0.97, P<0.001), DD_0 (r=0.96, P<0.001), FFP (r=-0.95, P<0.001) and Eref (r=-0.84, P<0.001) (Table 2.6). Precipitation and moisture variables were not significantly related with elevation (Table 2.6).

Influence of environment and $G\times E$ interaction

The model output showed that both genotype and environment explained a large proportion of variation in phenotypic traits (Table 2.7, Figure 2.6). Both genotype (P<0.001 for each model) and environment (P<0.005 for each model) contributed significantly to variation in height, timing of bud set, and bud flush. We also found that the $G\times E$ interaction was highly significant for plant height (P<0.001), bud set (P<0.001), and bud flush (P<0.001).

2.4 Discussion

The climatically heterogeneous range of *P. trichocarapa* necessitates local populations adapt to their respective environments, which is reflected by the geographic distribution of phenotypic variation and its correlation to climate (Ingvarsson and Street 2011; Eckert and Dyer 2012; McKown, et al. 2014). Environmental variables driving divergent selection along geographical clines may include photoperiod, temperature, precipitation and soil conditions. Our study has confirmed previous findings that there is a dramatic phenotypic gradient present among natural populations along the latitudinal range of *P. trichocarapa* (Ducousso, et al. 1996; Kelly, et al. 2003; Baquedano, et al. 2008; McKown, et al. 2014), and we have further demonstrated that parallel clines related to altitudinal adaptation also drive phenotypic diversity. This strong clinal phenotypic variation among natural accessions of *P. trichocarpa* along altitudinal and latitudinal gradients suggests strong divergent selection related to local climate shapes phenotypic variation, with temperature-related variables exhibiting the strongest correlations with adaptive traits.

Adaptive trait variation reflect adaptation to temperature regimes

Previous common garden experiments have shown strong phenotypic variation along geographical clines, resulting from spatial divergence in environments (Howe, et al. 2003; Savolainen, et al. 2007; Rohde, et al. 2011). Provenances from high latitude or altitude showed later bud flush and early bud set, thus having a smaller plant size (Zhang, et al. 2004; Rohde, et al. 2011). High latitude samples usually require longer critical day length to form bud thus have earlier growth cessation and bud set (Hurme, et al. 1997; Rohde, et al. 2011). Spring bud break requires a certain amount of fall chilling days before accumulating degree-days (average daily temperature between 0 °C and 5 °C) and sufficient heat-sum (Campbell and Sugano 1975; Hunter and Lechowicz 1992; Li, et al. 2010). The late bud flush of northern clones may be due to an unsatisfied chilling requirement. Growth of high elevation provenances was most limited by temperature – they set bud early and have low heat-sum requirements for bud flush, which allow them to take full advantage of short growing season as well as to avoid cold temperatures (Li, et al. 2010; Basler and Korner 2014).

Our common garden experiments revealed substantial variation in adaptive phenotypes. For example, northern clones from Alaska formed closed buds in late June and broke bud around April, while samples originated from Oregon set bud in late August and flushed in early March in the Virginia common garden. Samples along Coquihalla Highway at 1000m elevation set bud in mid-July, about 3-4 weeks earlier than those from sea level, but broke bud in late April, about 2-3 weeks later than those at sea level. With shorter growing seasons, most northern clones or clones from high elevation origin displayed smaller height and diameter than the clones near the central species range or low elevation. This result is most likely due to their unsatisfied growth period after transfer to non-local environment. Clones from high elevations showing smaller tree dimensions may be also due to the strong effect of radiation. A few southern clones were of smaller size due to their high-altitudinal sampling locations along the Sierra Nevada Mountains. The significantly lower cold injury index among northern samples indicated their deeper cold acclimation and higher freezing tolerance than southern clones. Finally, regeneration ability is an important fitness measure for clonality, which benefits plant survival and resources competition and acquisition (Martínková and Klimešová 2016). However, less is known about the factors

affecting plant regeneration ability. The clinal pattern of regeneration ability may reflect the adaptive differentiation across populations in stress response.

Timing of bud phenology is an important adaptation response to the seasonal changes of temperature and photoperiod, and, therefore, is subject to strong selection of geographical and climatic gradients. Plant height and diameter, on the other hand, are greatly influenced by temperature and precipitation. The development and release of cold acclimation are most sensitive to the availability and extent of freezing temperatures. Such relationships between spatially varying climate and adaptive traits give additional insights into how trees respond to different climate variables. The strong correlation between phenotypes results from their tight link with climate variables and evolutionary advantage of their interplay. For example, freezing resistance may be associated with bud flush timing because early flushing requires higher tolerance to freezing temperature than late flushing trees. The timing of bud phenology determines the lengths of growing season, and is therefore significantly correlated with plant height and diameter. The strong inter-correlation between adaptive traits and their correlations with climate variables may also result from the partially overlapping genetic mechanisms underlying the different local adaption processes (Gornall and Guy 2007; Olson, et al. 2013).

Genetic by environment interaction

Aside from population differentiation in adaptive traits, the magnitude of trait variation among populations also differed between different common gardens. Trees in the Virginia garden set bud an average of 15 days later than those in British Columbia garden, and release from dormancy was 15 days earlier in BC than in Virginia garden. Average tree height in Virginia garden was 50cm higher than in British Columbia. The clinal variation of different source populations shown in both common gardens clearly revealed their latitude/altitude origin being the key driving force. Our results also pointed to the stability of genetic effects as different genotypes for the most part performed consistently across different sites. However, there were a few clones showing less strong or reversed phenotypic difference between sites, which contribute to the strong observed $G \times E$ interaction. Bud phenology and plant height were associated with climate and latitude, particularly with temperature and photoperiod, which differed greatly between two common garden locations. It is well acknowledged that plant

phenology and physiological traits can be influenced by biotic and abiotic factors including temperature, photoperiod, precipitation, and soil fertilization (Gailing, et al. 2009; McKown, et al. 2014). Our result showed that individuals of the same origin and of same genetic background differ in their relative performance between each site as a function their genetic background. The strong $G \times E$ interaction at population level for these adaptive traits indicated that populations differ in how they react to environments and the genetic mechanisms regulating plant adaptation response can be influence by environments.

2.5 Conclusions

Phenotypic traits among *P. trichocarpa* populations vary as a function of the local climate at their geographic origin. We observed strong geographical divergence in adaptive traits, which provides evidence of local adaptation in this species. These adaptive traits are also strong indicators of how well tree populations perform under local or non-local environments. Global climate change is predicted to threaten temperate tree species and force their migration to more northern habitat or higher altitudes, where novel climate and unsynchronized environmental signals will challenge their survival and fecundity. Understanding the relationship between phenology, geography and climate factors informs predictions for how tree populations will respond to climate change, as well as their adaptability to new environments. However, these predictions may be impeded by limited knowledge of the genomic mechanisms controlling local adaptation, and the interaction between these genomic factors and the environment. Thus, dissection of the genomic architecture underlying climatic adaptation will help predict how tree species respond at population and landscape levels under global climate change.

2.6 Reference cited

Ackerly DD, Dudley SA, Sultan SE, Schmitt J, Coleman JS, Linder CR, Sandquist DR, Geber MA, Evans AS, Dawson TE, et al. 2000. The Evolution of Plant Ecophysiological Traits: Recent Advances and Future Directions. BioScience 50:979.

Aitken SN, Yeaman S, Holliday JA, Wang TL, Curtis-McLane S. 2008. Adaptation, migration or extirpation: climate change outcomes for tree populations. Evolutionary Applications 1:95-111.

Anderson JV, Horvath DP, Chao WS, Foley ME. 2010. Bud Dormancy in Perennial Plants: A Mechanism for Survival. 21:69-90.

Baquedano FJ, Valladares F, Castillo FJ. 2008. Phenotypic plasticity blurs ecotypic divergence in the response of Quercus coccifera and Pinus halepensis to water stress. European Journal of Forest Research 127:495-506.

Basler D, Korner C. 2014. Photoperiod and temperature responses of bud swelling and bud burst in four temperate forest tree species. Tree Physiology 34:377-388.

Bertoni G. 2011. Dormancy Cycling in Populus: The Symplasmic Connection. The Plant Cell 23:3-3.

Boisvenue C, Running SW. 2006. Impacts of climate change on natural forest productivity - evidence since the middle of the 20th century. Global Change Biology 12:862-882.

Campbell RK, Sugano AI. 1975. Phenology of Bud Burst in Douglas-Fir Related to Provenance, Photoperiod, Chilling, and Flushing Temperature. Botanical Gazette 136:290-298.

Cleland E, Chuine I, Menzel A, Mooney H, Schwartz M. 2007. Shifting plant phenology in response to global change. Trends in Ecology & Evolution 22:357-365.

Cooke JEK, Eriksson ME, Junttila O. 2012. The dynamic nature of bud dormancy in trees: environmental control and molecular mechanisms. Plant, Cell & Environment 35:1707-1728.

Ducousso A, Guyon JP, Krémer A. 1996. Latitudinal and altitudinal variation of bud burst in western populations of sessile oak (*Quercus petraea (Matt) Liebl*). Annales des Sciences Forestières 53:775-782.

Eckert AJ, Dyer RJ. 2012. Defining the landscape of adaptive genetic diversity. Molecular Ecology 21:2836-2838.

Frewen BE, Chen THH, Howe GT, Davis J, Rohde A, Boerjan W, Bradshaw HD. 2000. Quantitative trait loci and candidate gene mapping of bud set and bud flush in Populus. Genetics 154:837-845.

Gailing O, Vornam B, Leinemann L, Finkeldey R. 2009. Genetic and genomic approaches to assess adaptive genetic variation in plants: forest trees as a model. Physiol Plant 137:509-519.

Gornall JL, Guy RD. 2007. Geographic variation in ecophysiological traits of black cottonwood (*Populus trichocarpa*) This article is one of a selection of papers published in the Special Issue on Poplar Research in Canada. Canadian Journal of Botany 85:1202-1213.

Halbritter AH, Billeter R, Edwards PJ, Alexander JM. 2015. Local adaptation at range edges: comparing elevation and latitudinal gradients. Journal of Evolutionary Biology 28:1849-1860.

Hannerz M, Aitken SN, King JN, Budge S. 1999. Effects of genetic selection for growth on frost hardiness in western hemlock. Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere 29:509-516.

Horvath DP, Anderson JV, Chao WS, Foley ME. 2003. Knowing when to grow: signals regulating bud dormancy. Trends in Plant Science 8:534-540.

Howe GT, Aitken SN, Neale DB, Jermstad KD, Wheeler NC, Chen THH. 2003. From genotype to phenotype: unraveling the complexities of cold adaptation in forest trees. Canadian Journal of Botany-Revue Canadienne De Botanique 81:1247-1266.

Hunter AF, Lechowicz MJ. 1992. Predicting the Timing of Budburst in Temperate Trees. The Journal of Applied Ecology 29:597.
Hurme P, Repo T, Savolainen O, Pääkkönen T. 1997. Climatic adaptation of bud set and frost hardiness in Scots pine (*Pinus sylvestris*). Canadian Journal of Forest Research 27:716-723.
Ingvarsson PK, Street NR. 2011. Association genetics of complex traits in plants. New Phytologist 189:909-922.
Kelly CK, Chase MW, de Bruijn A, Fay MF, Woodward FI. 2003. Temperature-based population segregation in birch. Ecology Letters 6:87-89.
Khanduri VP, Sharma CM, Singh SP. 2008. The effects of climate change on plant phenology. The Environmentalist 28:143-147.
Klepsatel P, Gáliková M, Huber CD, Flatt T. 2014. Similarities and Differences in Altitudinal Versus Latitudinal Variation for Morphological Traits Indrosophila Melanogaster. Evolution 68:1385-1398.
Körner C. 2007. The use of 'altitude' in ecological research. Trends in Ecology & Evolution 22:569-574.
Kremer A, Ronce O, Robledo-Arnuncio JJ, Guillaume F, Bohrer G, Nathan R, Bridle JR, Gomulkiewicz R, Klein EK, Ritland K, et al. 2012. Long-distance gene flow and adaptation of forest trees to rapid climate change. Ecology Letters 15:378-392.
Li H, Wang X, Hamann A. 2010. Genetic adaptation of aspen (*Populus tremuloides*) populations to spring risk environments: a novel remote sensing approach. Canadian Journal of Forest Research 40:2082-2090.
Linkosalo T, Carter TR, Hakkinen R, Hari P. 2000. Predicting spring phenology and frost damage risk of Betula spp. under climatic warming: a comparison of two models. Tree Physiology 20:1175-1182.
Martínková J, Klimešová J. 2016. Enforced Clonality Confers a Fitness Advantage. Front Plant Sci 7.
McKown AD, Guy RD, Klapste J, Geraldes A, Friedmann M, Cronk QC, El-Kassaby YA, Mansfield SD, Douglas CJ. 2014. Geographical and environmental gradients shape phenotypic trait variation and genetic structure in *Populus trichocarpa*. New Phytologist 201:1263-1276.
Oleksyn J, Modrzynski J, Tjoelker MG, Z.ytkowiak R, Reich PB, Karolewski P. 1998. Growth and physiology of Picea abies populations from elevational transects: common garden evidence for altitudinal ecotypes and cold adaptation. Functional Ecology 12:573-590.
Olson MS, Levsen N, Soolanayakanahally RY, Guy RD, Schroeder WR, Keller SR, Tiffin P. 2013. The adaptive potential of *Populus balsamiferaL.* to phenology requirements in a warmer global climate. Molecular Ecology 22:1214-1230.
Parmesan C, Yohe G. 2003. A globally coherent fingerprint of climate change impacts across natural systems. Nature 421:37-42.
Polgar CA, Primack RB. 2011. Leaf-out phenology of temperate woody plants: from trees to ecosystems. New Phytologist 191:926-941.
Rohde A, Bastien C, Boerjan W. 2011. Temperature signals contribute to the timing of photoperiodic growth cessation and bud set in poplar. Tree Physiology 31:472-482.
Rohde A, Prinsen E, De Rycke R, Engler G, Van Montagu M, Boerjan W. 2002. PtABI3 impinges on the growth and differentiation of embryonic leaves during bud set in poplar. (vol 14, pg 1885, 2002). Plant Cell 14:2975-2975.
Savolainen O, Pyhäjärvi T, Knürr T. 2007. Gene Flow and Local Adaptation in Trees. Annual Review of Ecology, Evolution, and Systematics 38:595-619.

Shim D, Ko J-H, Kim W-C, Wang Q, Keathley DE, Han K-H. 2014. A molecular framework for seasonal growth-dormancy regulation in perennial plants. Horticulture Research 1:14059.
Soolanayakanahally RY, Guy RD, Silim SN, Drewes EC, Schroeder WR. 2009. Enhanced assimilation rate and water use efficiency with latitude through increased photosynthetic capacity and internal conductance in balsam poplar (*Populus balsamiferaL.*). Plant, Cell & Environment 32:1821-1832.
St. Clair JB. 2006. Genetic variation in fall cold hardiness in coastal Douglas-fir in western Oregon and Washington. Canadian Journal of Botany 84:1110-1121.
Thomas SC. 2011. Genetic vs. phenotypic responses of trees to altitude. Tree Physiology 31:1161-1163.
van Mantgem PJ, Stephenson NL. 2007. Apparent climatically induced increase of tree mortality rates in a temperate forest. Ecology Letters 10:909-916.
Vihera-Aarnio A, Hakkinen R, Partanen J, Luomajoki A, Koski V. 2005. Effects of seed origin and sowing time on timing of height growth cessation of Betula pendula seedlings. Tree Physiology 25:101-108.
Wang T, Hamann A, Spittlehouse DL, Murdock TQ. 2012. ClimateWNA—High-Resolution Spatial Climate Data for Western North America. Journal of Applied Meteorology and Climatology 51:16-29.
Way DA, Oren R. 2010. Differential responses to changes in growth temperature between trees from different functional groups and biomes: a review and synthesis of data. Tree Physiology 30:669-688.
Weiser CJ. 1970. Cold Resistance and Injury in Woody Plants: Knowledge of hardy plant adaptations to freezing stress may help us to reduce winter damage. Science 169:1269-1278.
Xie Y, Wang X, Silander JA. 2015. Deciduous forest responses to temperature, precipitation, and drought imply complex climate change impacts. Proceedings of the National Academy of Sciences 112:13585-13590.
Zhang X, Friedl MA, Schaaf CB, Strahler AH. 2004. Climate controls on vegetation phenological patterns in northern mid- and high latitudes inferred from MODIS data. Global Change Biology 10:1133-1145.

Table 2.1. Correlation coefficient between phenotypes across latitudinal samples.

	Height	Bud set	Bud flush	Diameter	Cold injury	Regeneration height	Regenerated branch number
Height	1.00	0.59	-0.56	0.87	0.22	0.81	0.56
Bud set	0.59	1.00	-0.66	0.53	0.40	0.62	0.28
Bud flush	-0.56	-0.66	1.00	-0.46	-0.44	-0.57	-0.23
Diameter	0.87	0.53	-0.46	1.00	0.19	0.77	0.58
Cold injury	0.22	0.40	-0.44	0.19	1.00	0.26	0.09
Regeneration height	0.81	0.62	-0.57	0.77	0.26	1.00	0.56
Regenerated branch number	0.56	0.28	-0.23	0.58	0.09	0.56	1.00

Table 2.2. Correlation coefficient between phenotypes across altitudinal samples.

	Height	Bud set	Bud flush	Diameter	Cold injury	Regeneration height	Regenerated branch number
Height	1.00	0.59	-0.41	0.85	0.26	0.79	0.62
Bud set	0.59	1.00	-0.50	0.63	0.21	0.63	0.44
Bud flush	-0.41	-0.50	1.00	-0.37	-0.24	-0.49	-0.26
Diameter	0.85	0.63	-0.37	1.00	0.25	0.75	0.65
Cold injury	0.26	0.21	-0.24	0.25	1.00	0.21	0.19
Regeneration height	0.79	0.63	-0.49	0.75	0.21	1.00	0.64
Regenerated branch number	0.62	0.44	-0.26	0.65	0.19	0.64	1.00

Table 2.3. Correlation coefficient between phenotypes and climate across latitudinal samples.

	MAT	MWMT	MCMT	TD	MAP	MSP	AHM	SHM	DD_0	DD5	FFP	Eref
Height	0.41	0.31	0.39	-0.32	0.03	-0.12	0.10	0.15	-0.34	0.40	0.37	0.31
Bud set	0.49	0.42	0.45	-0.32	-0.05	-0.24	0.20	0.32	-0.38	0.50	0.41	0.46
Bud flush	-0.56	-0.47	-0.53	0.40	0.08	0.31	-0.24	-0.41	0.45	-0.57	-0.47	-0.53
Diameter	0.27	0.21	0.24	-0.18	0.03	-0.08	0.05	0.04	-0.20	0.27	0.25	0.19
Cold injury	0.38	0.20	0.42	-0.41	0.05	-0.13	0.07	0.27	-0.34	0.36	0.43	0.28
Regeneration height	0.35	0.29	0.32	-0.23	0.05	-0.08	0.01	0.04	-0.29	0.34	0.31	0.26
Regenerated branch number	0.21	0.15	0.20	-0.16	0.05	-0.08	-0.04	-0.01	-0.20	0.19	0.22	0.10

Table 2.4. Correlation coefficient between phenotypes and across altitudinal samples.

	MAT	MWMT	MCMT	TD	MAP	MSP	AHM	SHM	DD_0	DD5	FFP	Eref
Height	0.37	0.34	0.36	-0.22	0.07	0.00	0.06	0.06	-0.36	0.37	0.36	0.25
Bud set	0.44	0.39	0.41	-0.25	0.06	-0.01	0.07	0.04	-0.41	0.43	0.43	0.27
Bud flush	-0.41	-0.33	-0.41	0.30	-0.31	-0.23	0.17	0.18	0.39	-0.39	-0.42	-0.22
Diameter	0.38	0.35	0.37	-0.22	0.08	-0.01	0.06	0.08	-0.37	0.38	0.37	0.26
Cold injury	0.22	0.16	0.21	-0.16	0.15	0.11	-0.08	-0.07	-0.22	0.20	0.23	0.16
Regeneration height	0.46	0.42	0.44	-0.27	0.09	0.00	0.06	0.05	-0.42	0.46	0.45	0.26
Regenerated branch number	0.41	0.37	0.43	-0.29	-0.06	-0.19	0.20	0.24	-0.40	0.40	0.41	0.23

Table 2.5. Correlation coefficient between geographic variables and climate across latitudinal samples.

	Latitude	Longitude	Elevation	MAT	MWMT	MCMT	TD	MAP	MSP	AHM	SHM	DD_0	DD5	FFP	Eref
Latitude	1.00	-0.63	-0.19	-0.86	-0.74	-0.82	0.61	-0.01	0.46	-0.27	-0.73	0.73	-0.85	-0.65	-0.90
Longitude	-0.63	1.00	0.24	0.47	0.59	0.31	-0.06	0.06	-0.23	0.09	0.28	-0.33	0.52	0.25	0.59
Elevation	-0.19	0.24	1.00	-0.23	0.04	-0.28	0.38	-0.23	-0.23	0.21	0.26	0.34	-0.13	-0.43	0.20
MAT	-0.86	0.47	-0.23	1.00	0.83	0.95	-0.72	0.21	-0.24	0.12	0.54	-0.91	0.97	0.86	0.83
MWMT	-0.74	0.59	0.04	0.83	1.00	0.63	-0.22	0.06	-0.32	0.26	0.56	-0.63	0.90	0.52	0.90
MCMT	-0.82	0.31	-0.28	0.95	0.63	1.00	-0.90	0.27	-0.18	0.02	0.47	-0.95	0.85	0.88	0.70
TD	0.61	-0.06	0.38	-0.72	-0.22	-0.90	1.00	-0.30	0.05	0.12	-0.28	0.84	-0.55	-0.81	-0.36
MAP	-0.01	0.06	-0.23	0.21	0.06	0.27	-0.30	1.00	0.82	-0.83	-0.51	-0.39	0.10	0.30	-0.11
MSP	0.46	-0.23	-0.23	-0.24	-0.32	-0.18	0.05	0.82	1.00	-0.74	-0.77	0.06	-0.32	-0.06	-0.51
AHM	-0.27	0.09	0.21	0.12	0.26	0.02	0.12	-0.83	-0.74	1.00	0.74	0.15	0.26	-0.01	0.41
SHM	-0.73	0.28	0.26	0.54	0.56	0.47	-0.28	-0.51	-0.77	0.74	1.00	-0.35	0.61	0.32	0.77
DD_0	0.73	-0.33	0.34	-0.91	-0.63	-0.95	0.84	-0.39	0.06	0.15	-0.35	1.00	-0.77	-0.77	-0.64
DD5	-0.85	0.52	-0.13	0.97	0.90	0.85	-0.55	0.10	-0.32	0.26	0.61	-0.77	1.00	0.81	0.88
FFP	-0.65	0.25	-0.43	0.86	0.52	0.88	-0.81	0.30	-0.06	-0.01	0.32	-0.77	0.81	1.00	0.49
Eref	-0.90	0.59	0.20	0.83	0.90	0.70	-0.36	-0.11	-0.51	0.41	0.77	-0.64	0.88	0.49	1.00

Table 2.6. Correlation coefficient between geographic variables and climate across altitudinal samples.

	Latitude	Longitude	Elevation	MAT	MWMT	MCMT	TD	MAP	MSP	AHM	SHM	DD_0	DD5	FFP	Eref
Latitude	1.00	-0.10	0.51	-0.57	-0.43	-0.51	0.36	-0.73	-0.68	0.58	0.59	0.54	-0.55	-0.61	-0.43
Longitude	-0.10	1.00	0.52	-0.53	-0.41	-0.60	0.51	0.50	0.66	-0.66	-0.61	0.58	-0.46	-0.52	-0.25
Elevation	0.51	0.52	1.00	-0.98	-0.91	-0.90	0.49	-0.18	0.07	-0.19	-0.31	0.96	-0.97	-0.95	-0.84
MAT	-0.57	-0.53	-0.98	1.00	0.91	0.93	-0.54	0.23	-0.02	0.11	0.23	-0.97	0.98	0.98	0.81
MWMT	-0.43	-0.41	-0.91	0.91	1.00	0.71	-0.15	0.15	-0.04	0.21	0.36	-0.82	0.97	0.83	0.92
MCMT	-0.51	-0.60	-0.90	0.93	0.71	1.00	-0.80	0.22	-0.07	0.07	0.17	-0.97	0.85	0.97	0.59
TD	0.36	0.51	0.49	-0.54	-0.15	-0.80	1.00	-0.17	0.06	0.08	0.07	0.67	-0.38	-0.66	-0.04
MAP	-0.73	0.50	-0.18	0.23	0.15	0.22	-0.17	1.00	0.94	-0.91	-0.78	-0.21	0.23	0.28	0.22
MSP	-0.68	0.66	0.07	-0.02	-0.04	-0.07	0.06	0.94	1.00	-0.93	-0.89	0.06	0.00	0.02	0.05
AHM	0.58	-0.66	-0.19	0.11	0.21	0.07	0.08	-0.91	-0.93	1.00	0.93	-0.11	0.13	0.03	0.13
SHM	0.59	-0.61	-0.31	0.23	0.36	0.17	0.07	-0.78	-0.89	0.93	1.00	-0.22	0.25	0.14	0.29
DD_0	0.54	0.58	0.96	-0.97	-0.82	-0.97	0.67	-0.21	0.06	-0.11	-0.22	1.00	-0.92	-0.98	-0.74
DD5	-0.55	-0.46	-0.97	0.98	0.97	0.85	-0.38	0.23	0.00	0.13	0.25	-0.92	1.00	0.94	0.88
FFP	-0.61	-0.52	-0.95	0.98	0.83	0.97	-0.66	0.28	0.02	0.03	0.14	-0.98	0.94	1.00	0.72
Eref	-0.43	-0.25	-0.84	0.81	0.92	0.59	-0.04	0.22	0.05	0.13	0.29	-0.74	0.88	0.72	1.00

Table 2.7. ANOVA results of linear mixed model indicating effect of environment, genotype, genotype-environment interaction on phenotypic variation (Type III with Satterthwaite approximation for degrees of freedom).

	Height		**Bud set**		**Bud flush**	
Model	F ratio	P	F ratio	P	F ratio	P
Environment	18.38	0.004	1466.68	<2.20e-16	1622.09	1.18e-14
Genotype	8.87	<2.20e-16	3.87	<2.20e-16	3.86	<2.20e-16
Genotype×Environment	7.92	<2.20e-16	4.70	<2.20e-16	6.74	<2.20e-16

Figure 2.1. Sample origin of 451 clones collected across *P. trichocarpa* species range (Green shaded area is the species range. Red dots indicate the sampling locations).

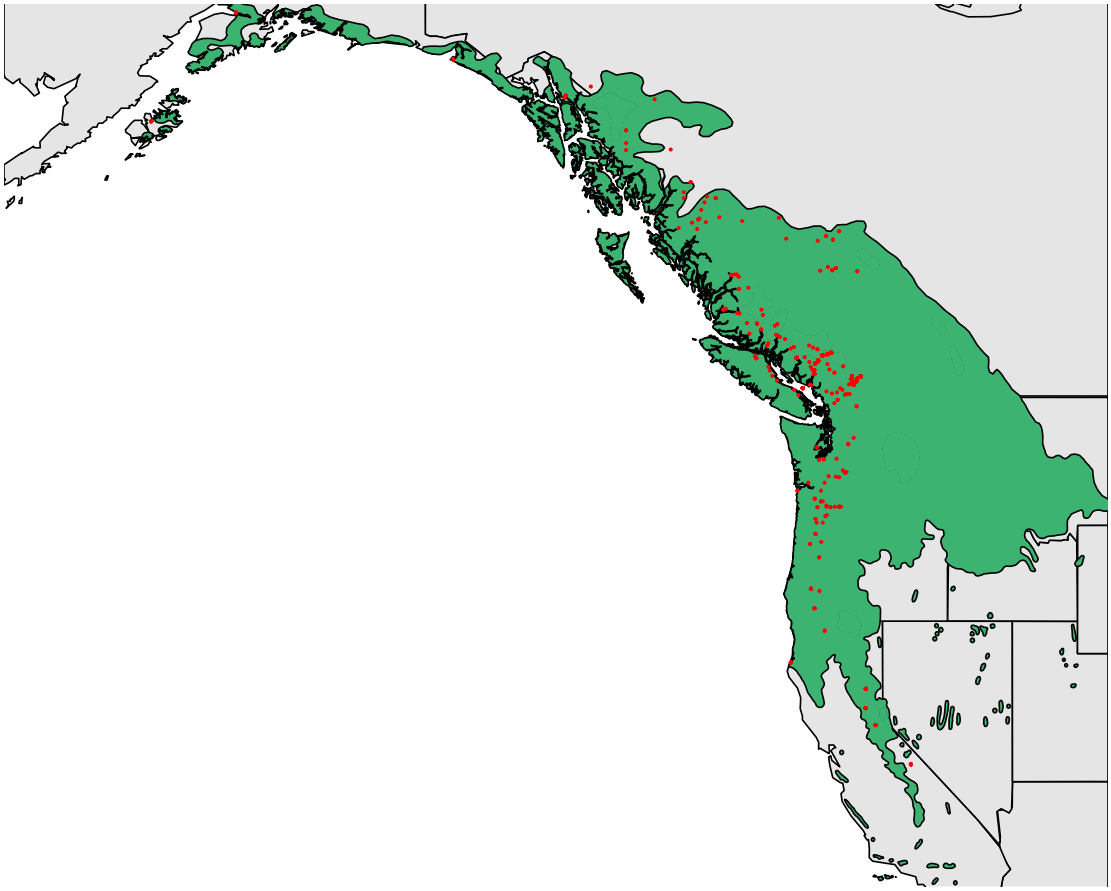


Figure 2.2. Phenotype BLUP in relation to latitude across latitudinal transect. Phenotype traits include: timing of bud set in VA garden (A), timing of bud set in BC garden (B), timing of bud flush in VA garden (C), timing of bud flush in BC garden (D), plant height in VA garden (E), plant height in BC garden (F), diameter in VA garden (G), cold injury in VA garden (H), regeneration height in VA garden (I), and regenerated branch number in VA garden (J) (VA garden refers to the common garden in Virginia and BC garden is the common garden in British Columbia).

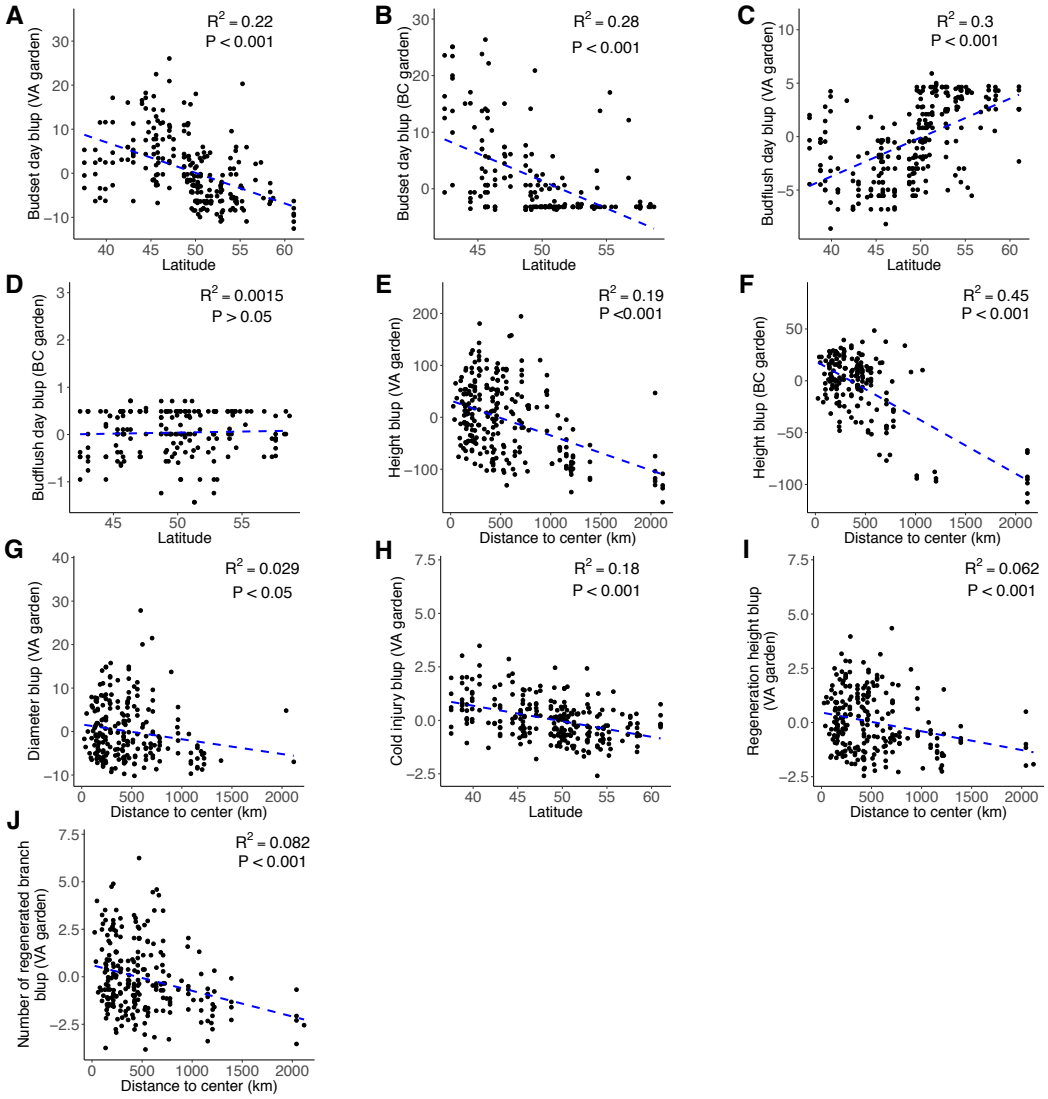


Figure 2.3. Phenotype BLUP in relation to elevation across altitudinal transect. Phenotype traits include: timing of bud set in VA garden (A), timing of bud set in BC garden (B), timing of bud flush in VA garden (C), timing of bud flush in BC garden (D), plant height in VA garden (E), plant height in BC garden (F), diameter in VA garden (G), cold injury in VA garden (H), regeneration height in VA garden (I), and regenerated branch number in VA garden (J) (VA garden refers to the common garden in Virginia and BC garden is the common garden in British Columbia).

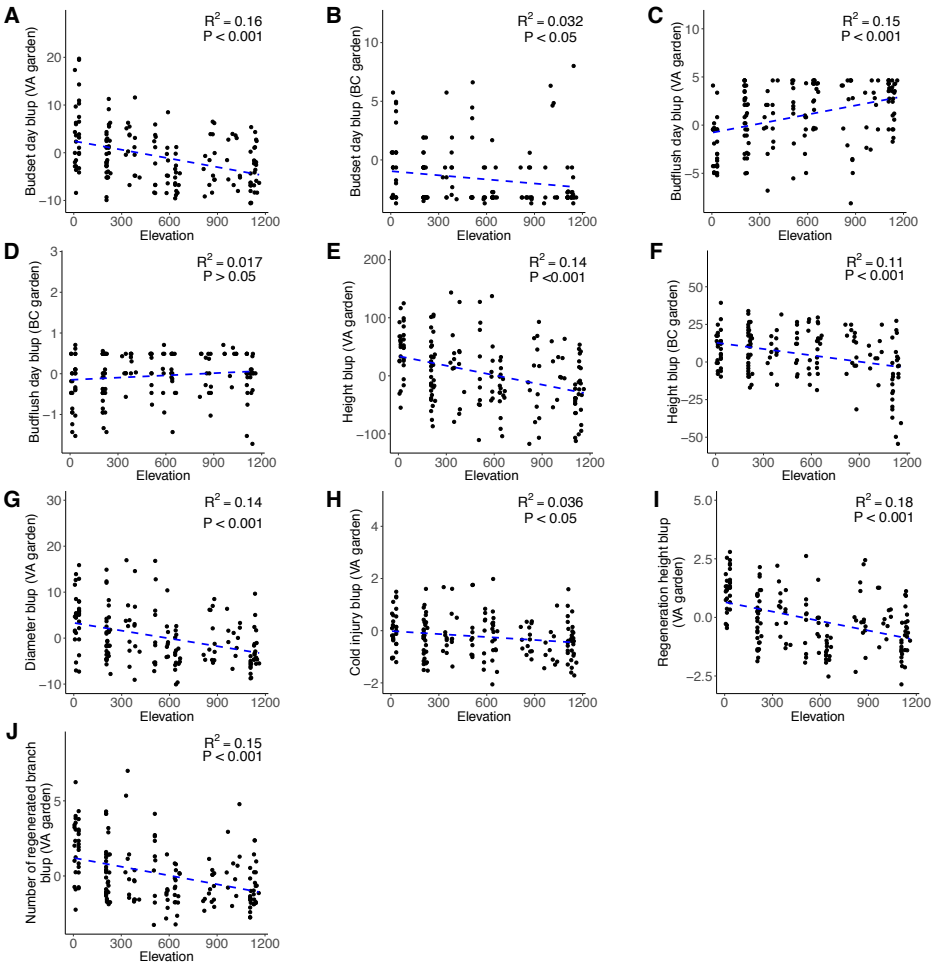


Figure 2.4. Heatmap of correlations between phenotype, geography, and climate across latitude transect (Pearson's correlation coefficients of 1 is indicated by color red and -1 indicated by blue. Phenotype traits include timing of bud set and bud flush, height, diameter, cold injury, regeneration height and branch number. Geography: latitude, longitude, and elevation. Climate includes MAT, MWMT, MCMT, TD, MAP, MSP, AHM, SHM, DD_0, DD5, FFP, Eref).

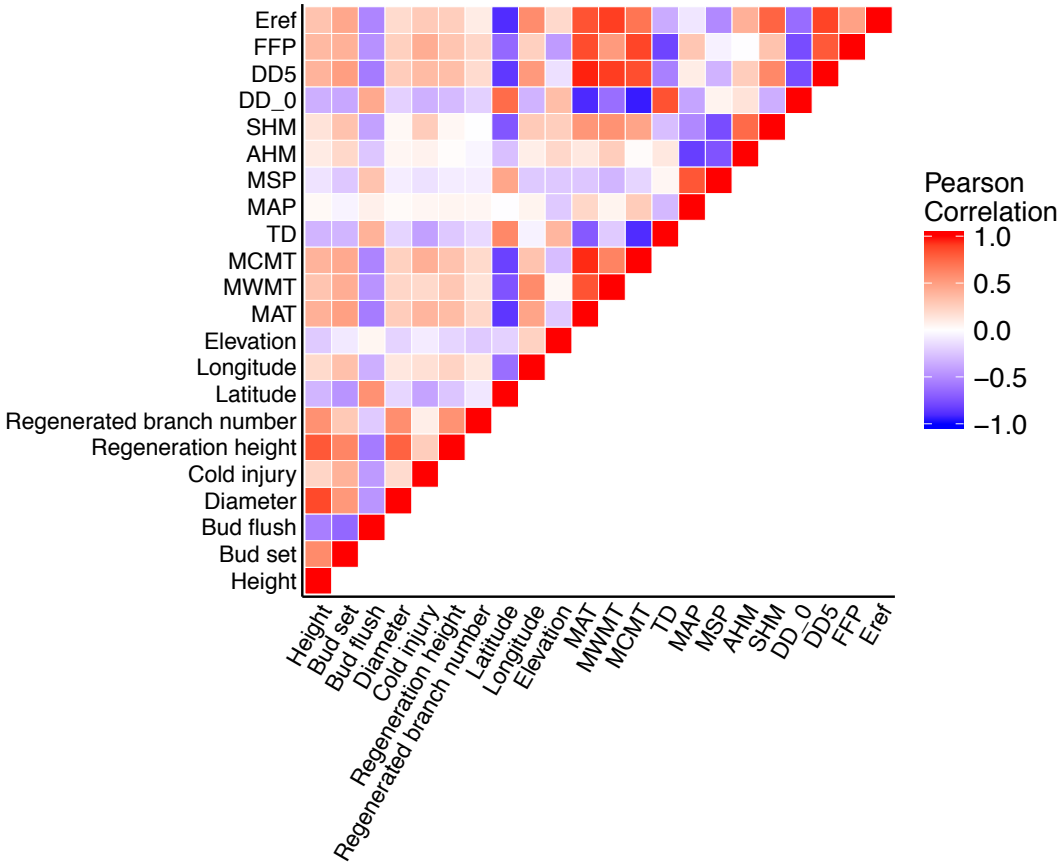


Figure 2.5. Heatmap of correlations between phenotype, geography, and climate across altitude transect (Pearson's correlation coefficients of 1 is indicated by color red and -1 indicated by blue. Phenotype traits include timing of bud set and bud flush, height, diameter, cold injury, regeneration height and branch number. Geography: latitude, longitude, elevation. Climate includes MAT, MWMT, MCMT, TD, MAP, MSP, AHM, SHM, DD_0, DD5, FFP, Eref).

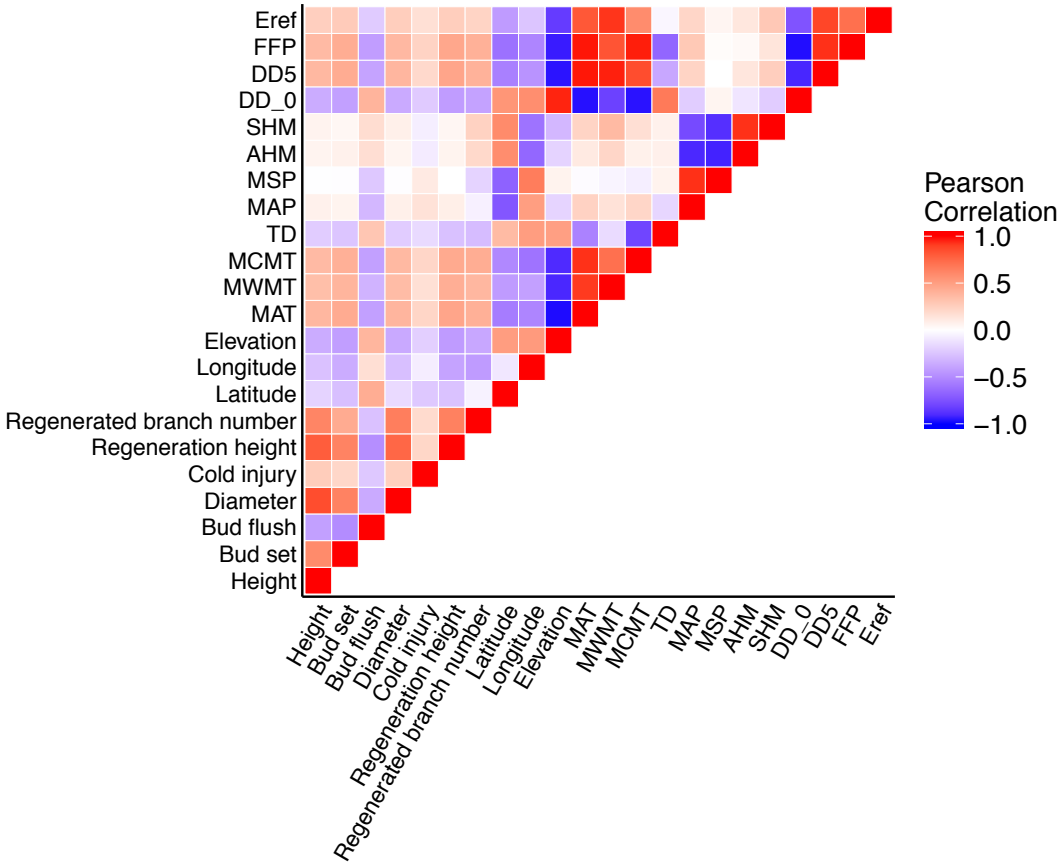
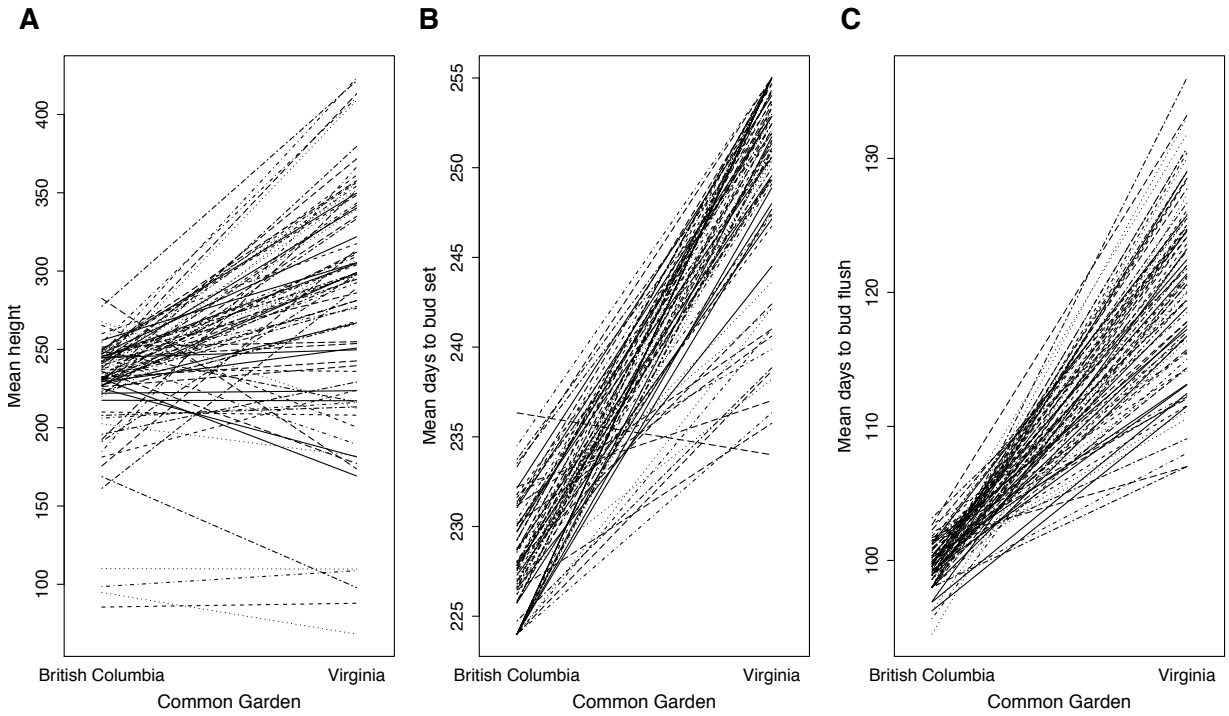


Figure 2.6. Interaction plot of the impact of environment and genotype on height (A), days to bud set (B), and bud flush (C).



3. Genome wide association mapping of climate-related traits

Abstract

Genome-wide association mapping has been an alternative to bi-parental QTL mapping. It has been successfully applied in dissecting many complex traits. We investigated genomic architecture underlying the timing of bud phenology, cold hardiness, and growth traits in *P. trichocarpa* using genome-wide association mapping coupled with sliding window analysis in a sample of 451 natural poplar accessions. The number of associations suggests these traits are highly polygenic: we detected 65 genes associated with timing of bud set, 109 genes for timing of bud flush, 203 genes associated with height, 14 genes for diameter, 28 genes for regeneration height, and 6 genes for regenerated branch number. The sliding window approach enabled us to identify genomic regions with concentrated association signals. The annotation of GWAS results uncovered genes that were previously confirmed to be involved in local adaption process as well as candidate genes of novel functions. Furthermore, we found significant overlap for adaptive genes across latitude and altitudinal transects, which suggests genetic parallelism underlying climatic adaptation.

3.1 Introduction

Local adaptation is critically important for the growth, reproduction, and survival of many organisms in the face of rapid climatic change (Jump and Penuelas 2005). Divergent climate favors individuals best adapted to local conditions, resulting in phenotypic and genetic differentiation across environmental gradients (Savolainen, et al. 2007). Across their wide spatial distributions, many tree species harbor a large amount of both neutral and adaptive standing genetic variation, and disentangling the relationship between this genomic variation and adaptive traits is a central goal of modern tree genomics (Ingvarsson, et al. 2008). While tree species are difficult taxa to work with in some regards, they provide unique insight into the mechanistic basis of physiological characteristics such as perenniality, seasonal growth, and dormancy cycling that cannot be uncovered using conventional model plants (Rohde and Bhalerao 2007). Previous studies have revealed the clinal variation along environmental gradients for locally adaptive traits, including annual growth and dormancy, growth, and cold hardiness (Ducousso, et al. 1996; Howe, et al. 2003; Hall, et al. 2007; Savolainen, et al. 2007). As such, many tree species have undergone adaptive population differentiation despite high levels of gene flow among populations (Howe, et al. 2003; Savolainen, et al. 2007).

Significant progress has been made in recent years to understand the genomic underpinnings of adaptive traits in both annual and perennial plant species (Lin 2000; Valverde, et al. 2004; Kobayashi and Weigel 2007; Rohde, et al. 2007; Ruttink, et al. 2007; Allona, et al. 2008; Hsu, et al. 2011). The adaptation process in temperate woody trees is a composite of complex genetics, physiology and developmental processes (Howe, et al. 2003; Neale and Savolainen 2004; Rohde, Storme, et al. 2011). Functional studies have identified many genes that appear to be related to bud phenology and cold acclimation. Phytochrome and circadian clock players have been shown through transgenic manipulation to regulate photoperiod-dependent growth cessation. For example, overexpressing oat PHYA inactivates cold acclimation in hybrid aspen trees in SD conditions (Olsen, Junttila, Nilsen, et al. 1997; Bohlenius, et al. 2006). Genes involved in ethylene, abscisic acid and gibberellins signal transduction pathways were also found to be directly or indirectly involved in short day dormancy induction and bud break (Olsen, Junttila and Moritz 1997; Chen, et al. 2002; Ruttink, et al. 2007; Cooke, et al. 2012). Candidate genes previously identified in freezing temperature sensing and cold acclimation include ICE (inducer

of CBF expression), CBF transcription factors, and LEA (Late embryogenesis-abundant) proteins (Thomashow 1999; Browse and Xin 2001). However, little is known about how natural genetic variation corresponds to phenotypic diversity for cold hardiness (Mackay, et al. 2009; Ingvarsson and Street 2011).

Association mapping is an attractive approach in dissecting the genomic basis underlying phenotypic variation of complex traits. With the availability of next-generation sequencing methods and genomic resources, association studies based on either candidate gene or unbiased re-sequencing have been employed in many tree species to characterize the genes controlling complex traits, including wood formation and quality, as well as dormancy-related traits (Neale and Savolainen 2004; Ingvarsson, et al. 2008; Evans, et al. 2014; McKown, Klápště, et al. 2014). Association genetics provides higher resolution compared with QTL mapping and makes it possible to simultaneously screen functional sites associated with many traits of interest (Zhao, et al. 2011). Sequencing of the *P. trichocarpa* genome makes association mapping an efficient method for elucidating the genetic basis of complex traits in this species. Trees species are relatively unstructured with high gene flow and have abundant genetic and phenotypic variation, which makes them good models for population genomics studies (Neale and Savolainen 2004). In addition, undomesticated natural populations of poplar have low linkage disequilibrium (LD), such that the markers associated with phenotypes are likely to either be the actual functional loci or tightly linked with the causative genes (Neale and Savolainen 2004). Advances in our knowledge of the genomic architecture underlying climatic adaptation could shed light on the fate of tree populations' adaptation to climate change, and facilitate breeding programs for climatic adaptation (Savolainen, et al. 2013; Hamberger, et al. 2015).

One aspect of tree species ranges that has not been leveraged is the existence of parallel clines, which provide natural experiments that may serve to mitigate false positives in GWA studies, as well as inform our understanding of the evolutionary constraints on adaptation. Previous studies have shown parallel evolution of phenotypic traits across independent lineages inhabiting similar environments has been shaped by similar selective constraints, but also by demographic forces including drift, population size, gene flow and mutation rate (Renaut, et al. 2014; Holliday, et al. 2016; Conte, et al. 2012; Gompel and Prud'homme 2009). Many examples of convergent

phenotypes have been documented either through similar or different paths of molecular evolution (Gompel and Prud'homme 2009; Losos 2011). A recent study in *P. trichocarpa* found a high rate of shared loci under divergent selection across latitudinal and altitudinal clines (Holliday, et al. 2016). These natural populations of *P. trichocarpa* have presumably experienced similar environmental selection and climate adaptation, which resulted in shared genomic divergence (Holliday, et al. 2016). This suggests that we should expect a high degree of shared loci underlying adaptation-related traits along the same latitudinal and altitudinal transects. The main focus of present study is to dissect the genetic mechanism underlying timing of bud phenology, height, diameter, regeneration ability, and cold hardiness in *P. trichocarpa*. We genotyped natural accessions of *P. trichocarpa* that originated from across most of the species range, and employed sliding window analysis based on traditional genome-wide association analysis (GWAS) to identify key loci associated with adaptive traits. We also compared GWAS results from both latitudinal and altitudinal samples to assess the extent of overlap among adaptive loci across these similar climatic gradients.

3.2 Materials and Methods

Genotyping and SNP calling

We sequenced 454 poplar *P. trichocarpa* clones collected from across most of the species range. Genomic DNA was extracted from young leaves following Qiagen DNeasy Plant Mini Kit protocol (Qiagen, Inc, Valencia, CA, USA). The Agilent 'SureSelect' platform (Agilent Technologies, Santa Clara, California, USA) was used for sequence capture targeting only the exons and regulatory regions of 39,739 poplar genes. Oligonucleotide baits targeting regulatory regions, and intergenic control regions were designed using SureSelect eArray software based on the black cottonwood annotation and assembly (version 3.0 on Phytozome (http://phytozome.jgi.doe.gov)). Library preparation and target enrichment followed the Agilent SureSelectXT protocol (Version B, April 2012), which was detailed described in (Zhou and Holliday 2012; Holliday, et al. 2016). Briefly, poplar genomic DNA was sheered on a Covaris S220 instrument at the Virgina Bioinformatics Institute (VBI) into fragments of ~200bp, followed by end repair, 3'-end adenylation, adaptor ligation and amplification. Target regions were then retrieved by hybridization to the baits, which were then purified, amplified and

indexed for sequencing. The captured libraries were then sequenced on Illumina HiSeq 2000 System in 2x100 paired end format with 16 samples per lane at VBI.

Sequence preprocessing was conducted with a custom pipeline for quality control, de-multiplexing, and trimming adapter sequence. Reads that passed the quality score (Q-score) threshold were re-paired and aligned to reference *P. trichocarpa* genome with Burrows-Wheeler Aligner (BWA) software, using the *sampe* function (Li and Durbin 2009). SNPs were subsequently called with the SAMtools software mpileup function (Li, et al. 2009). Missing genotypes on chromosome 1 to chromosome 19 were imputed using BEAGLE software (Browning and Browning 2016). The detailed sequencing and SNP calling procedure are described in (Zhou and Holliday 2012; Holliday, et al. 2016). Finally, we obtained 1,379,703 SNPs SNPs for chromosome 1 to chromosome 19 with minor allele frequency (MAF) > 0.05 and 116,398 SNPs on other scaffolds. Finally, we excluded 3 possible hybrids for further analysis.

Population structure estimation

To estimate population stratification, we excluded SNPs that failed Hardy-Weinberg Equilibrium tests with software PLINK (Purcell, et al. 2007) and obtained 125,245 SNPs. We performed principal component analysis on these 125,245 SNPs with the *prcomp* function in R (R Core Team 2015) and selected first two principal components (PC) to use as covariates for population structure (P matrix) in association tests according to the broken-stick rule (Jackson 1993) implemented in R package 'vegan'. The first two principal components explained 3.4% and 1.5% of genotype variation respectively (Figure 3.1). The kinship matrix was estimated with TASSEL version 5 (Bradbury, et al. 2007) to control for familial relatedness and negative values between individuals were set to 0.

Association analysis of climate-related traits

Inclusion of kinship matrix generated a uniform distribution of p-value (Figure S3.1) for most traits, which indicates over-correction of population structure and reduced power in detecting true positives. This is mostly likely due to the dense SNP coverage and linkage between SNPs and the over-correction for confounding effect of population structure may lead to excessive false negatives (McKown, Klápště, et al. 2014). Thus, we proceeded with GWAS on GLM

model using P matrix across 451 clones alone as control for population stratification. To assess the effect of population structure control on reducing false positives, we estimated the empirical false discovery rate (FDR) with 1400 putatively neutral SNPs (intergenic SNPs). We defined the empirical FDR as the number of significant intergenic SNPs (false positives) divided by the number of total outlier SNPs by assuming the intergenic SNPs are all neutral and non-functional. The empirical FDRs range from 0 to 0.0013. Though we cannot rule of the possibility that intergenic SNPs may play a role in adaptation or be influenced by hitchhiking at adjacent genic SNPs, the random placement of these intergenic SNPs across 19 chromosomes makes it less likely to be targets of selection. Finally, we separated 451 clones into two major groups of samples across altitude cline (166 individuals) and laitutde cline (285 individuals) and performed association analysis with GLM procedure in TASSEL. The population stratification was not obvious among either latitudinal nor altitudinal samples (Figure S3.2). The first two principal components estimated separately for these two transects were used to control for population structure.

The poplar collections in two common gardens were phenotyped for the timing of bud set and bud flush, height, diameter, cold hardiness, and regenerated height and branch number. Clonal BLUPs (Best linear unbiased predictor) were estimated for each trait (as described in the previous chapter). Association models were fitted using GLM procedure with software TASSEL (Bradbury, et al. 2007). GLM (generalized linear model) model was performed in with model as following:

$$y = S\alpha + Q\nu + \varepsilon$$

where y is a vector of phenotypic observations, α is a vector of SNP effects, ν is a covariate of population structure effects (P matrix), ε is residuals following $N(0, \sigma_\varepsilon^2 I)$. S and Q are incidence matrices of 1's and 0's relating y to α, ν, respectively.

The p-values generated from the model were adjusted for multiple testing using Benjamini & Hochberg false discovery rate (FDR) (Benjamini and Hochberg 1995). And we defined FDR adjusted p-values as adjusted P.

SNPs uncovered by GLM model across total samples were classified as significantly associated where adjusted P ≤ 0.01 and were annotated to their respective closest gene using the *P.*

trichocarpa genome annotation version 3.0 on Phytozome (http://phytozome.jgi.doe.gov). Significant SNPs were further assessed with sliding window analysis to identify genomic regions with excessive proportion of significant SNPs. Since drift or population structure cause diffuse genomic divergence among populations, we expected true association signals to be more strong and concentrated than random association signals (i.e., those due to population structure). Thus we estimated the number of significant SNPs and total numbers of SNPs within each 10kb genomic window with 5kb slide across 19 chromosomes. We only included the genomic windows contained at least 1 significant SNP due to zero-inflated distribution of the number of significant SNPs within each slide. For each trait, we assumed the number of significant SNPs per sliding window (N_+) repsents sampling without replacement from a pool of significant SNPs following hypergeometric distribution $N_+ \sim Hyper(N, K, n)$ where N is the total number of SNPs across all windows for each trait, K is the total number of SNPs per window, n is the total number of significant SNPs across windows. As N is extrmely large in our case, N_+ has an approximately binomial distribution $N_+ \sim Bin(K, p)$ where p equals to n/N (Rivals, et al. 2006). We then selected genomic windows where the proportion of significant SNPs exceeded the 99% upper bound of the estimated proportion of significant SNPs (p) based on the bionomial approximation (Figure 3.2). These significant genomic regions were annotated with genes found in the intervals. We further annotated candidate genes to their 'biological process' GO categorizations (Gene Ontology 2015).

In addition to GWAS using the total sample, we partitioned that sample into latitudinal and altitudinal components to test for parallelism across these two clines. Due to the small population size of altitudinal samples, which limits the power of GWAS, we set an adjusted $P \leq 0.1$ as cutoff to expand the power in detecting candidate SNP associations across latitude and altitude transects. We first identified common genes containing SNP outliers shared between latitudinal and altitudinal transects for each trait. We used the *dhyper* function in R to assess the significance level for the number of overlapping loci, which was calculated as the probability of obtaining the observed number or more under the null hypothesis that there is no enrichment of overlap.

Cumulative R^2 of significant SNPs

In order to assess the total phenotypic variation explained by all trait-associated SNPs, we calculated a '*cumulative* R^2' as the difference in adjusted R^2 between a full model and reduced model (Ingvarsson, et al. 2008; McKown, et al. 2014). Due to the colinearity between markers, the full model was determined as the model comprising a subset of significant SNP markes and population structure explaining the greatest phenotypic variation. The best model was selected using BIC (Bayesian information criterion) based on forward stepwise regression implemented in R package *'leaps'* (Lumley and Millar 2009). The reduced model contains only population structure. The '*cumulative* R^2' provides an estimate for the phenotypic variance explained by significant SNP markers. We repeated this method for estimating phenotypic variance explained by significant markers detected with the sliding window approach as well.

3.3 Results

Phenotype-genotype associations across total samples

With adjusted $P \leq 0.01$, we detected 649 significant associations for bud set, 1120 for bud flush, 2266 for height, 159 for diameter, 10 for cold hardiness, 384 for regeneration height, and 33 SNPs associated with regenerated branch number (Table 3.1). The percentage of variation explained (PVE) by individual SNPs range from 3.8 to 14.8%. Cumulatively, significant associated SNPs explained from 21.2 to 42.7% of total phenotypic variation (Table 3.1). We identified total 316 unique genomic regions containing a number of significant SNPs that exceeded the one-sided 99% upper bound of the estimated genomic average of significant SNPs per window (Figure 3.2). Cold hardiness was the only trait for which no outlier windows were found. Significant windows comprised approximately 12-33% the total number of significant SNPs (Table 3.2), and these SNPs explained slightly less total phenotypic variance (11.2%-35.1%) than the full set of significant SNPs.

Genes associated with bud phenology

We identified 65 unique genes within 70 regions that had strong associations with bud set (Table 3.3) and 109 genes within 105 regions associated with bud flush (Table 3.4). Annotations of these phenology-associated suggest a variety of biological processes, including stress response, cell cycling, protein methylation, lipid metabolic process and hormone regulation. Several candidate genes associated with bud set have been previously characterized as important players

in photoperiod-dependent bud dormancy pathway. These included, Potri.008G166600 (CRY2; CRYPTOCHROME 2) and Potri.006G161400 (PAP2; PHYTOCHROME-ASSOCIATED PROTEIN 2). 65 genes contained 160 significant SNPs were identified and the PVEs of individual SNPs are between 5.6% and 10.0% of total variation in the timing of bud set timing (Table 3.2; Figure 3.3; Figure S3.3).

There were 109 genes associated with bud flush identified, including Potri.001G132600 (GRF2; growth-regulating factor 2), Potri.012G101900 (HSD2; HYDROXYSTEROID DEHYDROGENASE 2), Potri.012G102000 (HSD4; HYDROXYSTEROID DEHYDROGENASE 4), and Potri.008G131700 (RGA1; REPRESSOR OF GA1). There were 315 significant SNPs within candidate genes captured by the sliding window analysis. These significant SNPs showed strong association signals with timing of bud flush and had high PVEs, ranging from 4.9% to 8.6% (Table 3.2; Figure 3.3; Figure S3.3). There were 13 genes exhibiting significant association with both the timing of bud flush and bud set. One example is a SNP within Potri.002G013100 (LDL1; LYSINE SPECIFIC HISTONE DEMETHYLASE LIKE 1), which showed little linkage with neighboring SNPs and alone explained 5.1% total variation of bud flush and 5.0% of bud set (Figure 3.4). Genotypes possessing homozygous minor allele for this SNP exhibited later bud set and earlier bud flush compared with other genotypes (Figure 3.4).

Genes associated with height and diameter

We detected 203 unique genes within 199 genomic regions that were associated with height (Table 3.5), and 14 genes within 17 regions associated with diameter (Table 3.6; Figure 3.5). Candidate associated with height included Potri.002G089000 (PHYA; PHYTOCHROME A), Potri.008G131700 (RGA1), Potri.013G149200 (MYB14), and Potri.004G019400 (BGLU46; BETA GLUCOSIDASE 46) (Table 3.5). Genes associated with height were enriched for pathways related to organ development, cellular metabolic and biosynthesis process, stress response and carbohydrate metabolic process. 13 out of 14 genes associated with diameter also had association with height. Among them, Potri.003G185700 (WAKs; WALL-ASSOCIATED KINASE FAMILY PROTEIN) and Potri.006G160700 (XTH26; XYLOGLUCAN HYDROLASE 26), which are involved in carbohydrate metabolic process, contained SNPs that

individually explained more than 14% of total variation in height and more than 8% of the total variation of diameter (Figure S3.4).

Genes associated with regeneration

We detected 28 unique genes within 31 genomic regions associated with regeneration height (Table 3.7), and 6 genes within 5 genomic intervals associated with regenerated branch number (Table 3.8). The significant SNPs within genes captured by sliding window analysis together explained 35.1% of total variation in regeneration height (97 SNPs) and 11.2% of total variation in regenerated branch number (4 SNPs), compared with 41.4% of total variation for regeneration height (384 SNPs) and 31.5% for branch number (33 SNPs) explained by all significant SNP associations (Table 3.1; Table 3.2). The candidate genes encompassed transporters, protein kinase, and proteins involved in catalytic activity and nucleic acid binding. For example, genes associated with regeneration height was Potri.001G075700 (embryo defective 1025; EMB1025), and genes associated with regenerated branch number Potri.001G142600 (BOU; A BOUT DE SOUFFLE) and Potri.013G056200 (EMBRYO SAC DEVELOPMENT ARREST 30). SNPs within associated genes explained a large proportion of phenotypic variation in regeneration height and branch number (Figure S3.5). There was no overlapping regions associated with both regeneration height and regeneration branch number.

Genes associated with multiple traits

A few genomic regions uncovered with sliding window analysis showed association with multiple traits (Table 3.9). One example is region around Potri.015G136400 (CBF4; C-REPEAT-BINDING FACTOR 4) and Potri.015G136600 (PERL-LIKE FAMILY PROTEIN), which were in high linkage and showed significant association with timing of bud set, bud flush, and height. One significant nonsynonymous SNP within Potri.015G136400 explained 8.4 % of total variation in bud set, 6.3 % for bud flush, and 6.7% for height (Figure 3.6). The SNP of largest allelic effect was within the 5'UTR of Potri.015G136600, and explained 10% total variation for bud set, 6.5% for bud flush, and 7.4% for height (Figure 3.6). Interestingly, these two SNPs showed similar overdominant allelic effects, where the heterozygous genotype showed later bud set, earlier bud flush and larger plant height than two homozygous genotypes (Figure 3.6). Another example is Potri.008G131700 (RGA1), which was associated with both height and

timing of bud flush. Two significant nonsynonymous SNPs within this gene explained 4.2% and 4.8% of total variation in height and 5.7% and 5.9% of variation in bud flush, respectively (Figure 3.7). Individuals homozygous for minor alleles of these SNPs had greater height and flushed earlier than other genotypes (Figure 3.7).

Overlapping phenotype-genotype associations between spatial transects

The latitudinal transect had more associations than the altitudinal transects (adjusted P≤ 0.1), and substantial overlap was observed between latitude and altitude for most traits (Figure S3.6). For all traits except regeneration branch number, significant hypergeometric test statistics indicated non-random overlap between SNP associations for altitude and latitude. There were 29 overlapping genes associated with bud flush (Hypergeometric P=0.02; Table 3.10) and 70 overlapping genes for bud set (Hypergeometric P=1.2e-7; Table 3.11) (Figure S3.6). These candidate genes are involved in biological process such as disease resistance, hormone signal transduction, lipid metabolic process and chromatin methylation. Interesting genes included Potri.001G132600 (SFR6; SENSITIVE TO FREEZING 6) and Potri.001G011900 (GSL10; GLUCAN SYNTHASE-LIKE 10) associated with bud set, as well as Potri.006G164000 (CDC27) associated with bud flush. There were 58 genes associated plant height (Hypergeometric P=0.002; Table 3.12) and 8 genes associated with diameter (Hypergeometric P=0.007; Table 3.13) that were shared between the transects (Figure S3.6), and their annotations suggest roles in hormone signaling. Sixteen genes associated with regeneration height (Hypergeometric P=0.008; Table 3.14) and 1 gene with regenerated branch number were shared studies (Hypergeometric P=0.3; Figure S3.6). These overlapping genes are mostly involved in disease resistance, lipid transport and biosynthesis, and carbohydrate metabolism. For example, Potri.007G093400 encodes an ethylene-induced calmodulin-binding protein involved in leaf senescence, defense response to cold, and pathogen resistance. The only common gene associated with regenerated branch number across both geographical transects is Potri.019G070500, which belongs to TIR-NBS-LRR class disease resistance protein family.

3.4 Discussion

In this study, we combined sliding window analysis with traditional GWAS and uncovered numerous loci with annotations strongly suggestive of involvement in timing of bud phenology,

height, diameter, regeneration height, and branch number. While our sliding window approach had advantages in terms of separating true from false positives, this method to some extent limits our ability to identify the specific genes and alleles driving associations. Despite the limited power of GWAS for our altitudinal samples, we detected a significant number of overlapping loci shared between the two transects for each trait, suggesting parallel adaptation at genotypic level recapitulates parallelism at the phenotypic level for climate adaption traits.

Population structure and gene flow

Despite significant intraspecific phenotypic variation, tree species usually exhibit low population differentiation across most loci genome-wide (Howe, et al. 2003; Savolainen, et al. 2007; Holliday, et al. 2010). Our extensive sampling of the *P. trichocarpa* range supports this general trend. Natural accessions were clustered into two major latitudinal groups that divided from the species center and several distinct subgroups in the range peripheries (Figure 3.1). We identified candidate genes exhibiting strong statistical associations with climate-related traits after controlling for population structure, which suggests that selection is sufficient at these loci to antagonize the homogenizing effects of gene flow and facilitate local adaptation (Holliday, et al. 2010). The low empirical FDR estimates also suggested that the spurious associations were well controlled. At the same time, adjusting population structure may lead to false negatives in species such as *P. trichocarpa*, for which climatic selection covaries with population stratification (Holliday, et al. 2010).

Genes underlying bud phenlogy

A large number of genes contained significant SNP associated with timing of bud set and bud flush, suggesting the complex genetic nature of phenology traits. These included many genes with functions related to flowering and dormancy. For example, Potri.008G166600 encodes CRY2 protein involved in light signal entraining to circadian system and floral initiation in Arabidopsis (Mockler, et al. 2003; Valverde, et al. 2004). With evidence of a shared photoperiod-dependent mechanism between flowering and dormancy, it is likely some genes in flowering pathways may play important roles in determining bud dormancy in *P. trichocarpa* (Rohde and Bhalerao 2007; Horvath 2009). Potri.006G161400 (PAP2) was reported to integrate auxin signaling pathway and light response pathway (Liscum and Reed 2002), but its implication

was not characterized in *Populus*. Associations with bud flush included genes related to light perception, hormone signaling, and growth promoting genes. The Arabidopsis homolog of Potri.001G132600 (GRF2) encodes a transcription activator involved in leaf development, and is directly activated by floral identity factors, suggesting this gene is possibly related with bud structure and development. Potri.012G101900 (HSD2) and Potri.012G102000 (HSD4) encode two hydroxysteroid dehydrogenase, the overexpression of which were reported to have increased growth and stress tolerance, and reduced seed dormancy in Arabidopsis (Li, et al. 2007). Most SNP associations within associated genes had relatively small PVE, suggesting the evolution of variation in bud phenology involves numerous loci with small effects (Rockman 2012). This complex genetic architecture for bud phenology may also be attributed to the complexity of dormancy cycling, which requires the perception of environmental signals including daylength and temperature, inner clock synchronization and hormone regulation, morphological change, as well as organ development.

Genes underlying biomass and regeneration

Many strong candidates associated with height or diameter had functions related to light sensing, carbonhydrate metabolism, and secondary cell wall growth. For instance, Potri.002G089000 (PHYA) encodes red/far-red photoreceptor that controls tree growth cessation and bud dormancy by interacting with CO/FT regulon in trees (Bohlenius 2006; Rohde, Bastien, et al. 2011). Potri.004G019400 (BGLU46), Potri.003G185700 (WAKs) and Potri.006G160700 (XTH26) are involved in carbohydrate metabolic and secondary growth. Genes controlling embryo development were found to be associated with regeneration. For example, Potri.001G075700 (EMB1025), associated with regeneration height, is essential for embryo growth and development during seed dormancy (Cushing, et al. 2005).

Genes underlying multiple traits

A few genes were associated with multiple traits. Though it is unclear whether this can be ascribed to pleiotropy or simply correlations between different traits, these identified loci are potentially functional genes important for environmental response in *P. trichocarpa*. Potri.002G013100 (LDL1) is associated with timing of bud set and bud flush. This gene was reported previously to function in the autonomous pathways of flowering in Arabidopsis and

promote floral initiation through H3K4 methylation of FLC (FLOWERING TIME LOCUS C) and FWA (FLOWERING WAGENINGEN) (Abou-Elwafa, et al. 2010). In addition, this gene controls primary seed dormancy by regulating ABA signaling-related genes in Arabidopsis (Zhao, et al. 2015). With partly conserved pathways between flowering and dormancy pathways, Potri.002G013100 may be a novel player in determining the timing of transition to dormancy and growth resumption in the following year. Associated with both plant height and bud phenology, Potri.008G131700 (RGA1) belongs to the DELLA regulatory family and represses gibberellin-indcued vegetative growth and floral initiation. GA synthesis and metabolism is important for tree growth, morphology and physiological responses to adverse conditions (Zawaski, et al. 2011; Zhang, et al. 2014). Previous study identified RGA's role in the inhibiting response to drought and short days in *Populus* tree (Zhang, et al. 2014). Potri.015G136400 (CBF4) and Potri.015G136600 (Perl-like family protein) were in high linkage and associated with bud phenology and height. Potri.015G136400 (CBF4) belongs to the DREB family and is responsive to drought stress and abscisic acid treatment in Arabidopsis, but not to low temperature (Haake 2002). However, the expression of CBFs in birch tree was strongly induced during freezing treatment along with CBF target genes, indicating CBF possibly functions in cold acclimation and winter hardiness in birch tree (Welling and Palva 2008). Potri.015G136400 encodes a Per1-like family protein of unknown biological function, and may have shown associations due to close physical linkage with Potri.015G136400. These potential pleitropic loci functions in pathways controlling multiple traits, including hormone signaling, stress response, and development. Though the underlying phenotypic traits are statistically correlated, the shared associations among these genes may partly explain the strong genetic correlations between traits (Oubida, et al. 2015).

E*valuation of sliding window analyses*

The sliding window approach we employed was able to capture genomic regions that harbored more concentrated association signals than the genomic background. The candidate genes uncovered also contained significant SNP associations of large phenotypic effects for corresponding traits, suggesting the power of sliding window technique in detecting genes important to climate-related traits. Though there were fewer variants within outlier bins, collectively they explained a moderate proportion of phenotypic variation, which is about 35.7%-

85.4% of the cumulative PVE explained by all significant variants. With less stringent control for population structure, traditional GWAS approach may fail to distinguish true positives from the spurious associations caused by population stratification (Cardon and Palmer 2003). Sliding window analysis is more robust to false positive discoveries by focusing on genomic clusters of strong associations exceeding a noisy background. In addition, genome-wide positive selection can increase local linkage disequilibrium, which may lead to non-functional SNPs showing strong association with phenotypes. While traditional GWAS is vulnerable to these indirect SNP associations, sliding window analysis is independent of single SNP association, and therefore is less likely to uncover isolated indirect association signals.

Adaptive genes conserved across geographic transects

Previous studies have detected evidence of parallel adaptation across biotic and abiotic gradients in many species (Conte, et al. 2012; Renaut, et al. 2014; Soria-Carrasco, et al. 2014; Holliday, et al. 2016). However, parallel adaptation has been less explored in plant species (Holliday, et al. 2016). Two recent studies in plants have revealed shared genetic divergence in response to similar local climatic selection constraints along latitude and altitude transects, indicating parallel adaptation (Renaut, et al. 2014; Holliday, et al. 2016). We found a significant number of overlapping gene associations with bud phenology, height, diameter, and regeneration height that are conserved for both latitudinal and altitudinal transects. This level of overlap was very unlikely to have occurred by chance, and is likely due to shared genetic mechanism underlying climate adaptation. Though the insufficient sample size limited the power of GWAS across altitudinal samples, we detected a number of interesting genes with phenotypic association patterns in both transects. Several of these overlapping genes have function suggestive of roles in seasonal dormancy and cold acclimation. These include Potri.006G164000 (CDC27), which is required for cell division and differentiation in meristems; and Potri.001G132600 (SFR6), which acts downstream of CBF pathway and inputs environmental signal such as freezing temperature to plant circadian clock (Knight, et al. 2008). In addition, recent evidence suggests that short-day induced callose deposition seals cell wall plasmodesmata in the shoot apical meristem and disconnects the transport of growth regulators to the dormant bud and thus to maintain bud in dormant state (van der Schoot and Rinne 2011). In light of this finding, it is interesting that Potri.001G011900 was among the genes that overlapped between latitude and altitude, as it

encodes a protein that is essential for 1,3-beta glucan synthesis and the formation of callose deposition in cell wall. Taken together, the constellation of genes that were associated with adaptive traits across both latitude and altitude suggest functional constraints on local adaptation.

Heritability of adaptive trait

Recent common garden studies reported low to moderate estimate of heritability H^2 for a number of traits in *P. trichocarpa* (McKown, Guy, et al. 2014; Oubida, et al. 2015). The estimates vary across studies and years, however, bud phenology and growth related traits tend to have relatively higher H^2 (McKown, Guy, et al. 2014; Oubida, et al. 2015). The broad-sense heritability (H^2) for bud set is between 0.58-0.87, bud break between 0.4-0.82, height between 0.44-0.75 and diameter 36% (McKown, Guy, et al. 2014; Oubida, et al. 2015). Based on the PVE collectively explained by our significant SNP associations, at least 35.6% of heritability can be explained by trait-associated variants (adjusted $P \leq 0.01$) for bud phenology and height. While we thus captured a substantial proportion of the segregating variation responsible for the genetic component of trait variation, the proportion of unexplained heritability leads to the question of "missing heritability (Manolio, et al. 2009). One explanation for the missing heritability is that some casual variants may be rare and therefore difficult to capture in GWAS studies with limited sample sizes. Additionally, incomplete LD between casual variants and genotyped SNPs, epistasis, and genetic-environment interaction may also lead to missing heritability (Manolio, et al. 2009; Yang, et al. 2010; Zuk, et al. 2012).

3.5 Conclusions

Climate adaptation in temperate tress is a complex physiological process involving the synchronization of seasonal development to environmental conditions. Our data revealed the high level of phenotypic variation and little population structure in *P. trichocarpa*, which makes this species an ideal system for GWAS studies. By employing the sliding window approach to scan for genomic regions enriched with significant variants underlying bud phenology, biomass, and regeneration related traits, we detected many genes, some of which were previously characterized while others were novel. One limitation of this approach is that it has less power in detecting scattered association signals where LD is low. Finally, by comparing two independent GWAS scans performed on latitude and altitude transects, a number of overlapping loci were

detected, indicating shared genetic mechanism controlling climate adaptation under similar selection constraints along parallel geographical clines.

3.6 Reference cited

Abou-Elwafa SF, Buttner B, Chia T, Schulze-Buxloh G, Hohmann U, Mutasa-Gottgens E, Jung C, Muller AE. 2010. Conservation and divergence of autonomous pathway genes in the flowering regulatory network of Beta vulgaris. Journal of Experimental Botany 62:3359-3374.
Allona I, Ramos A, Ibanez C, Contreras A, Casado R, Aragoncillo C. 2008. Molecular control of winter dormancy establishment in trees. Spanish Journal of Agricultural Research 6:201-210.
Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B-Statistical Methodology 57:289–300.
Bohlenius H. 2006. CO/FT Regulatory Module Controls Timing of Flowering and Seasonal Growth Cessation in Trees. Science 312:1040-1043.
Bohlenius H, Huang T, Charbonnel-Campaa L, Brunner AM, Jansson S, Strauss SH, Nilsson O. 2006. CO/FT regulatory module controls timing of flowering and seasonal growth cessation in trees. Science 312:1040-1043.
Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-2635.
Browning Brian L, Browning Sharon R. 2016. Genotype Imputation with Millions of Reference Samples. The American Journal of Human Genetics 98:116-126.
Browse J, Xin Z. 2001. Temperature sensing and cold acclimation. Current Opinion in Plant Biology 4:241-246.
Cardon LR, Palmer LJ. 2003. Population stratification and spurious allelic association. The Lancet 361:598-604.
Chen THH, Howe GT, Bradshaw HD. 2002. Molecular genetic analysis of dormancy-related traits in poplars. Weed Science 50:232-240.
Conte GL, Arnegard ME, Peichel CL, Schluter D. 2012. The probability of genetic parallelism and convergence in natural populations. Proceedings of the Royal Society B: Biological Sciences 279:5039-5047.
Cooke JE, Eriksson ME, Junttila O. 2012. The dynamic nature of bud dormancy in trees: environmental control and molecular mechanisms. Plant, Cell & Environment 35:1707-1728.
Cushing DA, Forsthoefel NR, Gestaut DR, Vernon DM. 2005. Arabidopsis emb175 and other ppr knockout mutants reveal essential roles for pentatricopeptide repeat (PPR) proteins in plant embryogenesis. Planta 221:424-436.
Ducousso A, Guyon JP, Krémer A. 1996. Latitudinal and altitudinal variation of bud burst in western populations of sessile oak (Quercus petraea (Matt) Liebl). Annales des Sciences Forestières 53:775-782.
Evans LM, Slavov GT, Rodgers-Melnick E, Martin J, Ranjan P, Muchero W, Brunner AM, Schackwitz W, Gunter L, Chen JG, et al. 2014. Population genomics of *Populus trichocarpa* identifies signatures of selection and adaptive trait associations. Nature Genetics 46:1089-1096.
Gene Ontology C. 2015. Gene Ontology Consortium: going forward. Nucleic Acids Res 43:D1049-1056.
Gompel N, Prud'homme B. 2009. The causes of repeated genetic evolution. Developmental Biology 332:36-47.
Haake V. 2002. Transcription Factor CBF4 Is a Regulator of Drought Adaptation in Arabidopsis. Plant Physiology 130:639-648.

Hall D, Luquez V, Garcia VM, St Onge KR, Jansson S, Ingvarsson PK. 2007. Adaptive population differentiation in phenology across a latitudinal gradient in European aspen (*Populus tremula, L.*): a comparison of neutral markers, candidate genes and phenotypic traits. Evolution 61:2849-2860.
Hamberger B, Porth I, Klápště J, McKown AD, La Mantia J, Guy RD, Ingvarsson PK, Hamelin R, Mansfield SD, Ehlting J, et al. 2015. Evolutionary Quantitative Genomics of *Populus trichocarpa*. Plos One 10:e0142864.
Holliday JA, Ritland K, Aitken SN. 2010. Widespread, ecologically relevant genetic markers developed from association mapping of climate-related traits in Sitka spruce (Picea sitchensis). New Phytologist 188:501-514.
Holliday JA, Zhou L, Bawa R, Zhang M, Oubida RW. 2016. Evidence for extensive parallelism but divergent genomic architecture of adaptation along altitudinal and latitudinal gradients in *Populus trichocarpa*. New Phytologist 209:1240-1251.
Horvath D. 2009. Common mechanisms regulate flowering and dormancy. Plant Science 177:523-531.
Howe GT, Aitken SN, Neale DB, Jermstad KD, Wheeler NC, Chen THH. 2003. From genotype to phenotype: unraveling the complexities of cold adaptation in forest trees. Canadian Journal of Botany-Revue Canadienne De Botanique 81:1247-1266.
Hsu CY, Adams JP, Kim H, No K, Ma C, Strauss SH, Drnevich J, Vandervelde L, Ellis JD, Rice BM, et al. 2011. FLOWERING LOCUS T duplication coordinates reproductive and vegetative growth in perennial poplar. Proc Natl Acad Sci U S A 108:10756-10761.
Ingvarsson PK, Garcia MV, Luquez V, Hall D, Jansson S. 2008. Nucleotide Polymorphism and Phenotypic Associations Within and Around the phytochrome B2 Locus in European Aspen (*Populus tremula*, Salicaceae). Genetics 178:2217-2226.
Ingvarsson PK, Street NR. 2011. Association genetics of complex traits in plants. New Phytologist 189:909-922.
Jackson DA. 1993. Stopping Rules in Principal Components Analysis: A Comparison of Heuristical and Statistical Approaches. Ecology 74:2204.
Jump AS, Penuelas J. 2005. Running to stand still: adaptation and the response of plants to rapid climate change. Ecology Letters 8:1010-1020.
Knight H, Thomson AJ, McWatters HG. 2008. Sensitive to freezing6 integrates cellular and environmental inputs to the plant circadian clock. Plant Physiol 148:293-303.
Kobayashi Y, Weigel D. 2007. Move on up, it's time for change - mobile signals controlling photoperiod-dependent flowering. Genes & Development 21:2371-2384.
Li F, Asami T, Wu X, Tsang EWT, Cutler AJ. 2007. A Putative Hydroxysteroid Dehydrogenase Involved in Regulating Plant Growth and Development. Plant Physiology 145:87-97.
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-1760.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-2079.
Lin C. 2000. Photoreceptors and regulation of flowering time. Plant Physiology 123:39-50.
Liscum E, Reed JW. 2002. Genetics of Aux/IAA and ARF action in plant growth and development. Plant Molecular Biology 49:387-400.
Losos JB. 2011. Convergence, Adaptation, and Constraint. Evolution 65:1827-1840.
Lumley T, Millar A. 2009. leaps: Regression Subset Selection. R package version 2.9.

Mackay TF, Stone EA, Ayroles JF. 2009. The genetics of quantitative traits: challenges and prospects. Nature Reviews Genetics 10:565-577.
Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, et al. 2009. Finding the missing heritability of complex diseases. Nature 461:747-753.
McKown AD, Guy RD, Klapste J, Geraldes A, Friedmann M, Cronk QC, El-Kassaby YA, Mansfield SD, Douglas CJ. 2014. Geographical and environmental gradients shape phenotypic trait variation and genetic structure in *Populus trichocarpa*. New Phytologist 201:1263-1276.
McKown AD, Klápště J, Guy RD, Geraldes A, Porth I, Hannemann J, Friedmann M, Muchero W, Tuskan GA, Ehlting J, et al. 2014. Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of *Populus trichocarpa*. New Phytologist 203:535-553.
Mockler T, Yang H, Yu X, Parikh D, Cheng YC, Dolan S, Lin C. 2003. Regulation of photoperiodic flowering by Arabidopsis photoreceptors. Proceedings of the National Academy of Sciences of the United States of America 100:2140-2145.
Neale DB, Savolainen O. 2004. Association genetics of complex traits in conifers. Trends in Plant Science 9:325-330.
Olsen JE, Junttila O, Moritz T. 1997. Long-day induced bud break in Salix pentandra is associated with transiently elevated levels of GA(1) and gradual increase in indole-3-acetic acid. Plant and Cell Physiology 38:536-540.
Olsen JE, Junttila O, Nilsen J, Eriksson ME, Martinussen I, Olsson O, Sandberg G, Moritz T. 1997. Ectopic expression of oat phytochrome A in hybrid aspen changes critical daylength for growth and prevents cold acclimatization. Plant Journal 12:1339-1350.
Oubida RW, Gantulga D, Zhang M, Zhou L, Bawa R, Holliday JA. 2015. Partitioning of multivariate phenotypes using regression trees reveals complex patterns of adaptation to climate across the range of black cottonwood (*Populus trichocarpa*). Front Plant Sci 6.
Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics 81:559-575.
R Core Team. 2015. R: A language and environment for statistical computing: R Foundation for Statistical Computing.
Renaut S, Owens GL, Rieseberg LH. 2014. Shared selective pressure and local genomic landscape lead to repeatable patterns of genomic divergence in sunflowers. Molecular Ecology 23:311-324.
Rivals I, Personnaz L, Taing L, Potier MC. 2006. Enrichment or depletion of a GO category within a class of genes: which test? Bioinformatics 23:401-407.
Rockman MV. 2012. The Qtn Program and the Alleles That Matter for Evolution: All That's Gold Does Not Glitter. Evolution 66:1-17.
Rohde A, Bastien C, Boerjan W. 2011. Temperature signals contribute to the timing of photoperiodic growth cessation and bud set in poplar. Tree Physiology 31:472-482.
Rohde A, Bhalerao RP. 2007. Plant dormancy in the perennial context. Trends in Plant Science 12:217-223.
Rohde A, Ruttink T, Hostyn V, Sterck L, Van Driessche K, Boerjan W. 2007. Gene expression during the induction, maintenance, and release of dormancy in apical buds of poplar. Journal of Experimental Botany 58:4047-4060.

Rohde A, Storme V, Jorge V, Gaudet M, Vitacolonna N, Fabbrini F, Ruttink T, Zaina G, Marron N, Dillen S, et al. 2011. Bud set in poplar--genetic dissection of a complex trait in natural and hybrid populations. The New phytologist 189:106-121.
Ruttink T, Arend M, Morreel K, Storme V, Rombauts S, Fromm J, Bhalerao RP, Boerjan W, Rohde A. 2007. A molecular timetable for apical bud formation and dormancy induction in poplar. Plant Cell 19:2370-2390.
Savolainen O, Lascoux M, Merilä J. 2013. Ecological genomics of local adaptation. Nature Reviews Genetics 14:807-820.
Savolainen O, Pyhäjärvi T, Knürr T. 2007. Gene Flow and Local Adaptation in Trees. Annual Review of Ecology, Evolution, and Systematics 38:595-619.
Soria-Carrasco V, Gompert Z, Comeault AA, Farkas TE, Parchman TL, Johnston JS, Buerkle CA, Feder JL, Bast J, Schwander T, et al. 2014. Stick Insect Genomes Reveal Natural Selection's Role in Parallel Speciation. Science 344:738-742.
Thomashow MF. 1999. PLANT COLD ACCLIMATION: Freezing Tolerance Genes and Regulatory Mechanisms. Annual review of plant physiology and plant molecular biology 50:571-599.
Valverde F, Mouradov A, Soppe W, Ravenscroft D, Samach A, Coupland G. 2004. Photoreceptor regulation of CONSTANS protein in photoperiodic flowering. Science 303:1003-1006.
van der Schoot C, Rinne PLH. 2011. Dormancy cycling at the shoot apical meristem: Transitioning between self-organization and self-arrest. Plant Science 180:120-131.
Welling A, Palva ET. 2008. Involvement of CBF Transcription Factors in Winter Hardiness in Birch. Plant Physiology 147:1199-1211.
Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, et al. 2010. Common SNPs explain a large proportion of the heritability for human height. Nature Genetics 42:565-569.
Zawaski C, Kadmiel M, Pickens J, Ma C, Strauss S, Busov V. 2011. Repression of gibberellin biosynthesis or signaling produces striking alterations in poplar growth, morphology, and flowering. Planta 234:1285-1298.
Zhang J, Zawaski C, Busov VB. 2014. Roles of Gibberellin Catabolism and Signaling in Growth and Physiological Response to Drought and Short-Day Photoperiods in Populus Trees. Plos One 9:e86217.
Zhao K, Tung C-W, Eizenga GC, Wright MH, Ali ML, Price AH, Norton GJ, Islam MR, Reynolds A, Mezey J, et al. 2011. Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. Nature Communications 2:467.
Zhao M, Yang S, Liu X, Wu K. 2015. Arabidopsis histone demethylases LDL1 and LDL2 control primary seed dormancy by regulating DELAY OF GERMINATION 1 and ABA signaling-related genes. Front Plant Sci 6.
Zhou L, Holliday JA. 2012. Targeted enrichment of the black cottonwood (*Populus trichocarpa*) gene space using sequence capture. Bmc Genomics 13:703.
Zuk O, Hechter E, Sunyaev SR, Lander ES. 2012. The mystery of missing heritability: Genetic interactions create phantom heritability. Proceedings of the National Academy of Sciences 109:1193-1198.

Table 3.1. Significant SNP associations from GWAS analysis of total samples.

Trait	N[a]	Phenotypic variance (%)[b]	Cumulative phenotypic variance (%)[c]
Bud set (VA)	617	4.75-10.30	38.68
Bud flush (VA)	1111	3.98-9.28	42.70
Height (VA)	1143	3.81-10.33	39.68
Diameter (VA)	159	5.47-11.20	38.37
Cold hardiness (VA)	10	5.55-7.29	21.21
Regeneration height (VA)	384	4.87-9.86	41.39
Regenerated branch number (VA)	33	5.32-8.92	31.46
Bud set (BC)	32	7.13-9.75	38.65
Bud flush (BC)	9	7.83-9.06	32.01
Height (BC)	1124	5.63-14.81	23.87

a Number of significant associations (FDR adjusted $P \leq 0.01$).
b Range of percentage trait variance (marker R^2) explained by genetic effects of associated SNPs (output from TASSEL).
c Phenotypic variance collectively explained by significantly associated SNPs (estimated as 'cumulative R^2').

Table 3.2. Significant SNP associations from sliding window analysis of total samples.

Trait	N[a]	Phenotypic variance (%)[b]	Cumulative phenotypic variance (%)[c]
Bud set	160	5.56-10.0	30.83
Bud flush	315	4.91-8.56	34.14
Height	341	4.45-8.10	29.84
Diameter	53	5.47-11.20	30.78
Regeneration height	97	5.02-9.86	35.13
Regenerated branch number	4	5.40-6.91	11.22

a Number of significant associations within outlier genomic bins.
b Range of percentage trait variance (marker R^2) explained by genetic effects of associated SNPs (output from TASSEL).
c Phenotypic variance collectively explained by associated SNPs (estimated as 'cumulative R^2').

Table 3.3. Genes identified by sliding window analysis across 451 samples associated with timing of bud set.

Gene model	Trait	AT homolog	Annotated description
Potri.001G146200	Bud set	AT4G18020.2	PRR2 (CheY-like two-component responsive regulator family protein)
Potri.001G244600	Bud set	AT2G29210.1	splicing factor PWI domain-containing protein
Potri.001G244700	Bud set		
Potri.002G007200	Bud set	AT5G24360.1	IRE1-1 (inositol requiring 1-1)
Potri.002G007300	Bud set	AT5G24318.1	O-Glycosyl hydrolases family 17 protein
Potri.002G007400	Bud set	AT1G42540.1	GLR3.3 (glutamate receptor 3.3)
Potri.002G012800	Bud set	AT1G20510.1	OPCL1
Potri.002G012900	Bud set	AT1G47670.1	Transmembrane amino acid transporter family protein
Potri.002G013000	Bud set	AT1G20460.1	
Potri.002G013100	Bud set	AT1G62830.1	LDL1 (LSD1-like 1)
Potri.002G068100	Bud set		
Potri.002G068200	Bud set	AT1G43640.1	TLP5 (tubby like protein 5)
Potri.002G077200	Bud set	AT4G21470.1	FHY (FMN hydrolase)
Potri.002G081300	Bud set	AT1G21730.1	P-loop containing nucleoside triphosphate hydrolases superfamily protein
Potri.002G081400	Bud set	AT1G44130.1	Eukaryotic aspartyl protease family protein
Potri.003G035900	Bud set	AT3G13730.1	CYP90D1
Potri.003G155300	Bud set	AT1G30900.1	VSR6 (VACUOLAR SORTING RECEPTOR 6)
Potri.004G011500	Bud set		
Potri.004G011600	Bud set		
Potri.004G127400	Bud set	AT5G39570.1	
Potri.004G220100	Bud set	AT1G35520.1	ARF15 (auxin response factor 15)
Potri.005G036500	Bud set	AT4G16710.1	glycosyltransferase family protein 28
Potri.005G036600	Bud set	AT1G54820.1	Protein kinase superfamily protein
Potri.005G036700	Bud set	AT3G05210.1	UVR7
Potri.006G029000	Bud set	AT5G58290.1	RPT3 (regulatory particle triple-A ATPase 3)
Potri.006G029100	Bud set	AT5G58290.1	RPT3 (regulatory particle triple-A ATPase 3)
Potri.006G029200	Bud set	AT3G10480.2	NAC050 (NAC domain containing protein 50)
Potri.006G029300	Bud set	AT3G10480.2	ANAC050,NAC050
Potri.006G156100	Bud set	AT4G28910.2	NINJA (novel interactor of JAZ)
Potri.006G160400	Bud set	AT4G29000.1	Tesmin/TSO1-like CXC domain-containing protein

Potri.006G160500	Bud set	AT4G28990.2	RNA-binding protein-related
Potri.006G160600	Bud set	AT1G67280.1	Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily protein
Potri.006G160700	Bud set	AT4G28850.1	XTH26 (xyloglucan hydrolase 26)
Potri.006G161300	Bud set	AT5G66680.1	DGL1
Potri.006G161400	Bud set	AT4G29080.1	PAP2 (phytochrome associated protein 2)
Potri.006G161500	Bud set	ATCG00270.1	photosystem II reaction center protein D
Potri.006G170900	Bud set	AT5G52450.1	MATE efflux family protein
Potri.006G171000	Bud set	AT5G17050.1	UGT78D2 (UDP-glucosyl transferase 78D2)
Potri.006G171100	Bud set	AT5G17050.1	UGT78D2 (UDP-glucosyl transferase 78D2)
Potri.006G171200	Bud set	AT5G17050.1	UGT78D2 (UDP-glucosyl transferase 78D2)
Potri.006G244000	Bud set	AT3G15605.4	nucleic acid binding
Potri.006G244100	Bud set	AT5G25590.1	Protein of unknown function (DUF630 and DUF632)
Potri.008G166600	Bud set	AT1G04400.2	CRY2 (cryptochrome 2)
Potri.008G185500	Bud set	AT3G26430.1	GDSL-like Lipase/Acylhydrolase superfamily protein
Potri.011G037300	Bud set	AT4G21390.1	B120
Potri.013G132200	Bud set	AT4G03520.1	ATHM2
Potri.014G159400	Bud set		
Potri.014G184500	Bud set	AT2G01050.1	zinc ion binding;nucleic acid binding
Potri.015G119100	Bud set	AT1G31200.1	PP2-A9 (phloem protein 2-A9)
Potri.015G119200	Bud set	AT5G04690.1	Ankyrin repeat family protein
Potri.015G119300	Bud set	AT1G10180.1	
Potri.015G136400	Bud set	AT5G51990.1	CBF4 (C-repeat-binding factor 4)
Potri.015G136500	Bud set	AT1G06010.1	
Potri.015G136600	Bud set	AT1G16560.2	Per1-like family protein
Potri.015G136700	Bud set	AT4G25450.1	NAP8 (non-intrinsic ABC protein 8)
Potri.015G136800	Bud set		
Potri.015G136900	Bud set	AT1G12580.1	PEPKR1 (phosphoenolpyruvate carboxylase-related kinase 1)
Potri.015G142200	Bud set	AT5G62070.1	IQD23
Potri.015G142300	Bud set		
Potri.015G142400	Bud set		
Potri.015G142500	Bud set	AT3G22942.1	AGG2 (G-protein gamma subunit 2)
Potri.016G103600	Bud set	AT2G37920.1	emb1513
Potri.016G103700	Bud set	AT3G09280.1	
Potri.017G076800	Bud set	AT5G62680.1	Major facilitator superfamily protein
Potri.018G008600	Bud set		

Table 3.4. Genes identified by sliding window analysis across 451 samples associated with timing of bud flush.

Gene model	Trait	AT homolog	Annotated description
Potri.001G132400	Bud flush	AT4G19070.1	Putative membrane lipoprotein
Potri.001G132500	Bud flush	AT4G27890.1	HSP20-like chaperones superfamily protein
Potri.001G132600	Bud flush	AT4G37740.1	GRF2 (growth-regulating factor 2)
Potri.001G132700	Bud flush	AT4G19045.1	Mob1/phocein family protein
Potri.001G320300	Bud flush	AT2G31970.1	RAD50 (DNA repair-recombination protein)
Potri.001G331100	Bud flush	AT3G01900.1	CYP94B2
Potri.001G331200	Bud flush	AT3G27120.1	P-loop containing nucleoside triphosphate hydrolases superfamily protein
Potri.001G395200	Bud flush	AT2G20562.1	
Potri.002G013000	Bud flush	AT1G20460.1	
Potri.002G013100	Bud flush	AT1G62830.1	LDL1 (LSD1-like 1)
Potri.002G030200	Bud flush	AT1G19720.1	Pentatricopeptide repeat (PPR-like) superfamily protein
Potri.002G030300	Bud flush		
Potri.002G030400	Bud flush	AT1G19715.3	Mannose-binding lectin superfamily protein
Potri.002G068100	Bud flush		
Potri.002G068200	Bud flush	AT1G43640.1	TLP5 (tubby like protein 5)
Potri.002G192900	Bud flush		
Potri.002G193000	Bud flush		
Potri.004G124000	Bud flush	AT3G28730.1	HMG (high mobility group)
Potri.004G163200	Bud flush	AT2G21270.1	UFD1 (ubiquitin fusion degradation 1)
Potri.004G163300	Bud flush	AT4G34710.2	ADC2 (arginine decarboxylase 2)
Potri.004G215100	Bud flush	AT5G57620.1	MYB36 (myb domain protein 36)
Potri.005G000600	Bud flush	AT3G05940.1	Protein of unknown function (DUF300)
Potri.005G000700	Bud flush	AT5G26731.1	
Potri.005G036500	Bud flush	AT4G16710.1	glycosyltransferase family protein 28
Potri.005G036600	Bud flush	AT1G54820.1	Protein kinase superfamily protein
Potri.005G036700	Bud flush	AT3G05210.1	UVR7
Potri.005G049100	Bud flush	AT4G32640.2	Sec23/Sec24 protein transport family protein
Potri.005G049200	Bud flush	AT3G23750.1	Leucine-rich repeat protein kinase family protein
Potri.005G049300	Bud flush	AT5G28050.1	Cytidine/deoxycytidylate deaminase family protein
Potri.005G077300	Bud flush	AT1G78080.1	RAP2.4 (related to AP2 4)
Potri.005G236200	Bud flush	AT1G19870.1	IQ-domain 32

Potri.005G236300	Bud flush	AT1G19860.1	Zinc finger C-x8-C-x5-C-x3-H type family protein
Potri.005G236400	Bud flush	AT5G39590.1	TLD-domain containing nucleolar protein
Potri.006G047500	Bud flush	AT4G20990.1	ACA4 (alpha carbonic anhydrase 4)
Potri.006G047600	Bud flush	AT3G57290.1	EIF3E (eukaryotic translation initiation factor 3E)
Potri.006G047700	Bud flush	AT2G41520.1	Heat shock protein DnaJ with tetratricopeptide repeat
Potri.006G081800	Bud flush	AT5G24710.1	Transducin/WD40 repeat-like superfamily protein
Potri.006G081900	Bud flush	AT2G28780.1	
Potri.006G082000	Bud flush	AT2G28780.1	
Potri.006G082100	Bud flush	AT2G37480.1	
Potri.006G171200	Bud flush	AT5G17050.1	UGT78D2
Potri.006G179300	Bud flush	AT2G18660.1	PNP-A
Potri.006G244100	Bud flush	AT5G25590.1	Protein of unknown function (DUF630 and DUF632)
Potri.006G244200	Bud flush	AT2G25735.1	
Potri.006G244300	Bud flush	AT4G32690.1	GLB3 (hemoglobin 3)
Potri.007G011700	Bud flush	AT5G65530.1	
Potri.007G011800	Bud flush	AT5G65540.1	
Potri.007G067600	Bud flush	AT1G79180.1	MYB63 (myb domain protein 63)
Potri.008G110800	Bud flush	AT1G68740.1	PHO1
Potri.008G110900	Bud flush	AT1G68730.1	Zim17-type zinc finger protein
Potri.008G111000	Bud flush		
Potri.008G131600	Bud flush	AT1G70780.1	
Potri.008G131700	Bud flush	AT2G01570.1	RGA1 (repressor of GA)
Potri.009G035000	Bud flush		
Potri.010G079700	Bud flush		
Potri.010G079900	Bud flush		
Potri.010G080000	Bud flush		
Potri.010G080100	Bud flush		
Potri.010G080400	Bud flush		
Potri.010G154800	Bud flush		
Potri.010G154900	Bud flush		
Potri.010G155000	Bud flush		
Potri.011G072300	Bud flush	AT1G29750.2	RKF1 (receptor-like kinase in flowers 1)
Potri.011G072400	Bud flush	AT2G34480.1	Ribosomal protein L18ae/LX family protein
Potri.011G147600	Bud flush	AT1G33170.1	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
Potri.012G028100	Bud flush		
Potri.012G101800	Bud flush	AT5G50610.1	

Potri.012G101900	Bud flush	AT5G50590.1	HSD4 (hydroxysteroid dehydrogenase 4)
Potri.012G102000	Bud flush	AT3G47350.1	HSD2 (hydroxysteroid dehydrogenase 2)
Potri.013G013000	Bud flush	AT1G56230.1	Protein of unknown function (DUF1399)
Potri.013G013100	Bud flush	AT5G27920.1	F-box family protein
Potri.013G109600	Bud flush	AT1G35530.2	DEAD/DEAH box RNA helicase family protein
Potri.013G149200	Bud flush	AT2G31180.1	MYB14 (myb domain protein 14)
Potri.014G049000	Bud flush		
Potri.014G084900	Bud flush	AT2G45910.1	U-box domain-containing protein kinase family protein
Potri.014G159400	Bud flush		
Potri.014G179500	Bud flush	AT2G44525.1	Protein of unknown function (DUF498/DUF598)
Potri.015G022200	Bud flush		
Potri.015G022300	Bud flush	AT5G53210.1	SPCH
Potri.015G136400	Bud flush	AT5G51990.1	CBF4 (C-repeat-binding factor 4)
Potri.015G136500	Bud flush	AT1G06010.1	
Potri.015G136600	Bud flush	AT1G16560.2	Per1-like family protein
Potri.015G136700	Bud flush	AT4G25450.1	NAP8 (non-intrinsic ABC protein 8)
Potri.015G136800	Bud flush		
Potri.015G136900	Bud flush	AT1G12580.1	PEPKR1 (phosphoenolpyruvate carboxylase-related kinase 1)
Potri.016G047500	Bud flush	AT1G20970.1	
Potri.016G047600	Bud flush	AT5G13020.1	
Potri.016G088200	Bud flush	AT3G09740.1	SYP71 (syntaxin of plants 71)
Potri.016G088300	Bud flush	AT1G61110.1	NAC025 (NAC domain containing protein 25)
Potri.016G088500	Bud flush	AT5G03150.1	JKD (C2H2-like zinc finger protein)
Potri.016G088700	Bud flush	AT5G03170.1	FLA11 (FASCICLIN-like arabinogalactan-protein 11)
Potri.016G088800	Bud flush	AT3G53390.1	Transducin/WD40 repeat-like superfamily protein
Potri.016G088900	Bud flush	AT3G53400.1	
Potri.016G089000	Bud flush	AT3G09770.1	RING/U-box superfamily protein
Potri.016G089300	Bud flush	AT3G58750.1	CSY2 (citrate synthase 2)
Potri.016G089400	Bud flush		
Potri.016G091100	Bud flush	AT2G37040.1	PAL1 (PHE ammonia lyase 1)
Potri.017G015400	Bud flush	AT3G14460.1	LRR and NB-ARC domains-containing disease resistance protein
Potri.017G015700	Bud flush	AT1G03820.1	
Potri.017G015800	Bud flush	AT2G20570.1	GPRI1 (GBF pro-rich region-interacting factor 1)
Potri.017G074700	Bud flush	AT3G28460.1	methyltransferases

Potri.017G094200	Bud flush	AT3G02120.1	hydroxyproline-rich glycoprotein family protein
Potri.017G094300	Bud flush	AT5G39200.1	
Potri.017G130600	Bud flush		
Potri.017G136300	Bud flush		
Potri.018G008500	Bud flush		
Potri.018G008600	Bud flush		
Potri.018G008700	Bud flush		
Potri.019G046700	Bud flush		

Table 3.5. Genes identified by sliding window analysis across 451 samples associated with height.

Gene model	Trait	AT homolog	Annotated description
Potri.001G058700	Height	AT1G24265.2	Protein of unknown function (DUF1664)
Potri.001G098000	Height	AT1G64080.1	
Potri.001G098100	Height	AT2G46150.1	LEA (Late embryogenesis abundant protein)
Potri.001G211200	Height	AT1G60890.2	Phosphatidylinositol-4-phosphate 5-kinase family protein
Potri.001G308600	Height	AT1G07650.1	Leucine-rich repeat transmembrane protein kinase
Potri.001G308700	Height	AT1G29660.1	GDSL-like Lipase
Potri.001G395200	Height	AT2G20562.1	
Potri.001G395500	Height	AT5G28780.1	PIF1 helicase
Potri.001G448500	Height	AT1G64700.1	
Potri.001G448600	Height		
Potri.002G066100	Height	AT1G76010.1	Alba DNA/RNA-binding protein
Potri.002G068100	Height		
Potri.002G068200	Height	AT1G43640.1	TLP5 (tubby like protein 5)
Potri.002G077200	Height	AT4G21470.1	FHY (FMN hydrolase)
Potri.002G088700	Height	AT4G08460.2	Protein of unknown function (DUF1644)
Potri.002G088800	Height		
Potri.002G088900	Height	AT4G08500.1	MEKK1 (MAPK/ERK kinase kinase 1)
Potri.002G089000	Height	AT1G09570.1	PHYA (phytochrome A)
Potri.002G089100	Height	AT1G36990.1	
Potri.002G102900	Height	AT4G24660.1	HB22 (homeobox protein 22)
Potri.002G107400	Height	AT1G27595.1	
Potri.002G194000	Height		
Potri.003G063300	Height	AT3G13990.2	Kinase-related protein of unknown function (DUF1296)
Potri.003G167200	Height	AT5G50780.1	Histidine kinase family protein
Potri.003G167300	Height		
Potri.003G185700	Height	AT1G16260.2	Wall-associated kinase family protein
Potri.003G193300	Height	AT5G07010.1	ST2A (sulfotransferase 2A)
Potri.004G011500	Height		
Potri.004G011600	Height		
Potri.005G036500	Height	AT4G16710.1	glycosyltransferase family protein 28
Potri.005G036600	Height	AT1G54820.1	Protein kinase superfamily protein
Potri.005G036700	Height	AT3G05210.1	UVR7
Potri.006G068100	Height	AT1G02630.1	Nucleoside transporter family protein

Potri.006G068200	Height	AT5G20190.1	Tetratricopeptide repeat (TPR)-like superfamily protein
Potri.006G081800	Height	AT5G24710.1	Transducin/WD40 repeat-like superfamily protein
Potri.006G081900	Height	AT2G28780.1	
Potri.006G082000	Height	AT2G28780.1	
Potri.006G111800	Height	AT5G01960.1	RING/U-box superfamily protein
Potri.006G156100	Height	AT4G28910.2	NINJA (novel interactor of JAZ)
Potri.006G160600	Height	AT1G67280.1	Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily protein
Potri.006G160700	Height	AT4G28850.1	XTH26 (xyloglucan hydrolase 26)
Potri.006G170900	Height	AT5G52450.1	MATE efflux family protein
Potri.006G171000	Height	AT5G17050.1	UGT78D2 (UDP-glucosyl transferase 78D2)
Potri.006G171100	Height	AT5G17050.1	UGT78D2 (UDP-glucosyl transferase 78D2)
Potri.006G171200	Height	AT5G17050.1	UGT78D2 (UDP-glucosyl transferase 78D2)
Potri.006G244000	Height	AT3G15605.4	nucleic acid binding
Potri.006G244100	Height	AT5G25590.1	Protein of unknown function (DUF630 and DUF632)
Potri.006G244200	Height	AT2G25735.1	
Potri.006G244300	Height	AT4G32690.1	GLB3 (hemoglobin 3)
Potri.006G248000	Height	AT1G52820.1	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
Potri.006G248100	Height	AT4G32551.2	LUG (LisH dimerisation motif protein)
Potri.006G254000	Height	AT5G25360.2	
Potri.007G011700	Height	AT5G65530.1	Protein kinase superfamily protein
Potri.007G011800	Height	AT5G65540.1	
Potri.007G020800	Height	AT5G17760.1	P-loop containing nucleoside triphosphate hydrolases superfamily protein
Potri.007G076700	Height	AT2G21790.1	RNR1 (ribonucleotide reductase 1)
Potri.008G025600	Height	AT5G03680.1	PTL (Duplicated homeodomain-like superfamily protein)
Potri.008G110800	Height	AT1G68740.1	PHO1
Potri.008G110900	Height	AT1G68730.1	Zim17-type zinc finger protein
Potri.008G111000	Height		
Potri.008G111100	Height	AT1G68720.1	TADA (tRNA arginine adenosine deaminase)
Potri.008G131600	Height	AT1G70780.1	
Potri.008G131700	Height	AT2G01570.1	RGA1 (repressor of GA)
Potri.008G169300	Height	AT3G23410.1	FAO3 (fatty alcohol oxidase 3)
Potri.008G169400	Height	AT3G23430.1	PHO1 (phosphate 1)
Potri.008G185500	Height	AT3G26430.1	GDSL-like Lipase/Acylhydrolase superfamily protein

Potri.008G185600	Height	AT1G67850.2	Protein of unknown function (DUF707)
Potri.009G035000	Height		
Potri.010G011900	Height		
Potri.010G012000	Height		
Potri.010G056900	Height		
Potri.010G079500	Height		
Potri.010G079700	Height		
Potri.010G080000	Height		
Potri.011G006500	Height		
Potri.011G006600	Height		
Potri.011G006700	Height	AT1G62250.1	
Potri.011G006800	Height	AT1G11860.2	Glycine cleavage T-protein family
Potri.011G092800	Height	AT3G13120.1	Ribosomal protein S10p/S20e family protein
Potri.013G109600	Height	AT1G35530.2	DEAD/DEAH box RNA helicase family protein
Potri.014G166600	Height		
Potri.014G166700	Height		
Potri.014G166800	Height		
Potri.015G022200	Height		
Potri.015G022300	Height	AT5G53210.1	SPCH
Potri.015G046800	Height	AT3G18400.1	NAC058 (NAC domain containing protein 58)
Potri.015G072800	Height	AT3G48050.2	BAH domain protein
Potri.015G136400	Height	AT5G51990.1	CBF4 (C-repeat-binding factor 4)
Potri.015G136500	Height	AT1G06010.1	
Potri.015G136600	Height	AT1G16560.2	Per1-like family protein
Potri.015G136700	Height	AT4G25450.1	NAP8 (non-intrinsic ABC protein 8)
Potri.015G136800	Height		
Potri.015G136900	Height	AT1G12580.1	PEPKR1 (phosphoenolpyruvate carboxylase-related kinase 1)
Potri.016G065200	Height	AT5G06560.1	Protein of unknown function (DUF593)
Potri.016G065300	Height	AT4G37980.1	ELI3 (elicitor-activated gene 3)
Potri.016G119300	Height	AT2G38280.2	FAC1
Potri.017G041800	Height	AT4G33650.1	DRP3A (dynamin-related protein 3A)
Potri.017G041900	Height	AT5G54010.1	UDP-Glycosyltransferase superfamily protein
Potri.017G042000	Height		
Potri.017G073000	Height	AT3G55250.1	
Potri.017G074400	Height	AT3G28450.1	Leucine-rich repeat protein kinase family protein
Potri.018G011400	Height		
Potri.018G011500	Height		

Potri.018G011600	Height		
Potri.018G011700	Height		
Potri.018G012000	Height		
Potri.018G066900	Height		
Potri.018G092100	Height		
Potri.018G092200	Height		
Potri.018G092300	Height		
Potri.001G122000	Height	AT4G12550.1	AIR1 (Auxin-Induced in Root cultures 1)
Potri.001G122100	Height	AT3G22142.1	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
Potri.001G122200	Height	AT5G47570.1	
Potri.001G122300	Height		
Potri.001G126400	Height	AT5G51300.1	splicing factor-related
Potri.001G191100	Height	AT1G51980.1	Insulinase (Peptidase family M16) protein
Potri.001G192800	Height	AT1G52565.1	
Potri.001G194700	Height		
Potri.001G218800	Height	AT3G44350.2	NAC061 (NAC domain containing protein 61)
Potri.001G218900	Height		
Potri.001G387300	Height	AT1G78650.1	POLD3
Potri.001G387400	Height	AT1G53400.1	Ubiquitin domain-containing protein
Potri.001G390100	Height	AT3G14880.2	
Potri.002G066600	Height	AT2G21770.1	CESA9 (cellulose synthase A9)
Potri.002G090200	Height	AT1G36370.1	SHM7 (serine hydroxymethyltransferase 7)
Potri.002G090300	Height	AT1G22030.1	
Potri.002G104100	Height	AT4G24610.1	
Potri.002G104200	Height	AT5G49710.3	
Potri.002G150200	Height	AT2G45550.1	CYP76C4
Potri.002G227300	Height	AT2G32540.1	CSLB04 (cellulose synthase-like B4)
Potri.002G232500	Height	AT2G43060.1	IBH1 (ILI1 binding bHLH 1)
Potri.002G232600	Height	AT2G31660.1	URM9
Potri.002G241800	Height	AT2G32700.6	LUH (LEUNIG_homolog)
Potri.002G241900	Height	AT3G24800.1	PRT1 (proteolysis 1)
Potri.002G242000	Height	AT5G24090.1	CHIA (chitinase A)
Potri.003G076200	Height	AT5G47510.1	Sec14p-like phosphatidylinositol transfer family protein
Potri.003G077500	Height	AT4G16650.1	O-fucosyltransferase family protein
Potri.003G077600	Height	AT4G17330.1	G2484-1 (G2484-1 protein)
Potri.003G129600	Height	AT4G23440.1	Disease resistance protein (TIR-NBS class)
Potri.003G129700	Height	AT1G63800.1	UBC5 (ubiquitin-conjugating enzyme 5)

Potri.004G019200	Height		
Potri.004G019300	Height	AT1G61820.1	BGLU46 (beta glucosidase 46)
Potri.004G019400	Height	AT1G61820.1	BGLU46 (beta glucosidase 46)
Potri.004G025100	Height	AT4G23180.1	RLK4 (RECEPTOR-like protein kinase 4)
Potri.004G025200	Height	AT4G23180.1	RLK4 (RECEPTOR-like protein kinase 4)
Potri.004G027100	Height	AT5G54470.1	
Potri.004G027200	Height	AT4G27280.1	
Potri.005G065000	Height		
Potri.005G065100	Height	AT5G20680.3	TBL16
Potri.005G204100	Height	AT5G50890.1	alpha/beta-Hydrolases superfamily protein
Potri.005G204200	Height	AT2G42975.1	
Potri.005G251800	Height	AT4G38350.1	Patched family protein
Potri.006G045100	Height	AT2G41450.1	N-acetyltransferases;N-acetyltransferases
Potri.006G045200	Height	AT3G57190.1	peptide chain release factor, putative
Potri.006G045300	Height	AT3G52590.1	UBQ1 (ubiquitin extension protein 1)
Potri.006G173300	Height	AT2G24720.1	GLR2.2 (glutamate receptor 2.2)
Potri.006G217900	Height	AT5G21280.1	hydroxyproline-rich glycoprotein family protein
Potri.006G218500	Height	AT2G22620.1	Rhamnogalacturonate lyase family protein
Potri.006G218600	Height	AT5G20080.1	FAD/NAD(P)-binding oxidoreductase
Potri.006G218700	Height		
Potri.007G034900	Height	AT3G50380.1	Protein of unknown function (DUF1162)
Potri.007G035000	Height	AT3G50380.1	Protein of unknown function (DUF1162)
Potri.007G051500	Height	AT4G34280.1	transducin family protein / WD-40 repeat family protein
Potri.007G066200	Height		
Potri.007G094900	Height	AT2G22360.1	DNAJ heat shock family protein
Potri.007G095000	Height	AT1G22340.1	UGT85A7
Potri.007G139000	Height		
Potri.007G139100	Height		
Potri.007G139200	Height		
Potri.007G139300	Height		
Potri.007G139400	Height		
Potri.008G157400	Height	AT5G43850.1	RmlC-like cupins superfamily protein
Potri.008G157500	Height	AT4G14710.1	RmlC-like cupins superfamily protein
Potri.008G158600	Height	AT1G04210.1	Leucine-rich repeat protein kinase family protein
Potri.009G120900	Height		
Potri.010G130300	Height		
Potri.011G023400	Height	AT4G22330.1	ATCES1 (Alkaline phytoceramidase)

Potri.011G023500	Height	AT3G17640.1	Leucine-rich repeat (LRR) family protein
Potri.012G003900	Height	AT5G53110.1	RING/U-box superfamily protein
Potri.012G077800	Height	AT3G48040.1	ROP10 (RHO-related protein from plants 10)
Potri.013G005100	Height	AT1G56700.2	Peptidase C15, pyroglutamyl peptidase I-like
Potri.013G005200	Height	AT1G09420.1	G6PD4
Potri.013G005300	Height		
Potri.013G100900	Height	AT2G01050.1	zinc ion binding;nucleic acid binding
Potri.013G149200	Height	AT2G31180.1	MYB14 (myb domain protein 14)
Potri.013G149400	Height	AT1G15030.1	Protein of unknown function (DUF789)
Potri.014G146700	Height	AT5G16740.1	Transmembrane amino acid transporter family protein
Potri.014G152000	Height	AT1G05230.4	HDG2 (homeodomain GLABROUS 2)
Potri.014G152100	Height	AT3G21700.3	SGP2 (Ras-related small GTP-binding family protein)
Potri.014G152200	Height	AT1G05200.1	GLR3 (glutamate receptor 3)
Potri.014G172000	Height	AT5G48310.2	
Potri.014G172100	Height		
Potri.014G172200	Height	AT4G12650.1	Endomembrane protein 70 protein family
Potri.014G172300	Height	AT4G11120.1	translation elongation factor Ts (EF-Ts), putative
Potri.014G172400	Height	AT1G61520.3	LHCA3 (photosystem I light harvesting complex gene 3)
Potri.014G172500	Height	AT5G48330.2	Regulator of chromosome condensation (RCC1) family protein
Potri.014G191400	Height	AT5G48655.1	RING/U-box superfamily protein
Potri.016G017900	Height	AT4G39720.1	VQ motif-containing protein
Potri.016G033500	Height	AT5G05130.1	DNA/RNA helicase protein
Potri.016G033600	Height	AT5G05130.1	DNA/RNA helicase protein
Potri.016G033700	Height	AT5G05130.1	DNA/RNA helicase protein
Potri.016G036600	Height	AT3G56710.1	SIB1 (sigma factor binding protein 1)
Potri.016G036700	Height	AT3G12600.1	NUDT16 (nudix hydrolase homolog 16)
Potri.016G036800	Height	AT3G56960.1	PIP5K4 (phosphatidyl inositol monophosphate 5 kinase 4)
Potri.016G036900	Height	AT5G04140.1	GLU1 (glutamate synthase 1)

Table 3.6. Genes identified by sliding window analysis across 451 samples associated with diameter.

Gene model	Trait	AT homolog	Annotated description
Potri.002G194000	Diameter		
Potri.003G185700	Diameter	AT1G16260.2	Wall-associated kinase family protein
Potri.006G160600	Diameter	AT1G67280.1	Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily protein
Potri.006G160700	Diameter	AT4G28850.1	XTH26 (xyloglucan hydrolase 26)
Potri.006G248000	Diameter	AT1G52820.1	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
Potri.006G248100	Diameter	AT4G32551.2	LUG (LisH dimerisation motif protein)
Potri.011G092800	Diameter	AT3G13120.1	Ribosomal protein S10p/S20e family protein
Potri.018G011400	Diameter		
Potri.018G011500	Diameter		
Potri.018G011600	Diameter		
Potri.018G011700	Diameter		
Potri.018G011800	Diameter		
Potri.018G012000	Diameter		
Potri.018G092200	Diameter		

Table 3.7. Genes identified by sliding window analysis across 451 samples associated with regeneration height.

Gene model	Trait	AT homolog	Annotated description
Potri.001G075600	Regeneration height	AT1G30900.1	VSR6 (VACUOLAR SORTING RECEPTOR 6)
Potri.001G075700	Regeneration height	AT4G20090.1	EMB1025 (embryo defective 1025)
Potri.001G181900	Regeneration height		
Potri.001G182000	Regeneration height	AT1G79900.1	BAC2 (Mitochondrial substrate carrier family protein)
Potri.001G420200	Regeneration height	AT5G54800.1	GPT1 (glucose 6-phosphate/phosphate translocator 1)
Potri.002G194000	Regeneration height		
Potri.003G185700	Regeneration height	AT1G16260.2	Wall-associated kinase family protein
Potri.006G081800	Regeneration height	AT5G24710.1	Transducin/WD40 repeat-like superfamily protein
Potri.006G081900	Regeneration height	AT2G28780.1	

Potri.006G082000	Regeneration height	AT2G28780.1	
Potri.006G082100	Regeneration height	AT2G37480.1	
Potri.006G139400	Regeneration height	AT1G19180.1	JAZ1 (jasmonate-zim-domain protein 1)
Potri.006G141600	Regeneration height	ATCG00270.1	photosystem II reaction center protein D
Potri.009G063400	Regeneration height		
Potri.009G063500	Regeneration height		
Potri.010G078400	Regeneration height		
Potri.010G079600	Regeneration height		
Potri.010G079700	Regeneration height		
Potri.010G080000	Regeneration height		
Potri.010G080200	Regeneration height		
Potri.011G003200	Regeneration height	AT1G50660.1	
Potri.012G121700	Regeneration height	AT5G45030.2	Trypsin family protein
Potri.013G109600	Regeneration height	AT1G35530.2	DEAD/DEAH box RNA helicase family protein
Potri.017G016100	Regeneration height	AT2G20560.1	DNAJ heat shock family protein
Potri.017G016200	Regeneration height	AT5G44180.1	Homeodomain-like transcriptional regulator
Potri.017G054300	Regeneration height	AT2G32720.1	CB5-B
Potri.018G005100	Regeneration height		
Potri.018G005200	Regeneration height		

Table 3.8. Genes identified by sliding window analysis across 451 samples associated with regenerated branch number.

Gene model	Trait	AT homolog	Annotated description
Potri.001G142600	Regenerated branch number	AT5G46800.1	BOU (A BOUT DE SOUFFLE)
Potri.013G056200	Regenerated branch number	AT3G03810.1	EDA30 (EMBRYO SAC DEVELOPMENT ARREST 30)
Potri.013G056300	Regenerated branch number	AT2G37270.2	RPS5B (ribosomal protein 5B)
Potri.016G039100	Regenerated branch number	AT3G50120.1	Plant protein of unknown function (DUF247)
Potri.016G039200	Regenerated branch number	AT3G50150.1	Plant protein of unknown function (DUF247)
Potri.016G039300	Regenerated branch number	AT3G50150.1	Plant protein of unknown function (DUF247)

Table 3.9. Genes identified by sliding window analysis across 451 samples associated with multiple traits.

Gene model	Trait	AT homolog	Annotated description
Potri.001G395200	Bud flush, Height	AT2G20562.1	
Potri.002G013000	Bud set, Bud flush	AT1G20460.1	
Potri.002G013100	Bud set, Bud flush	AT1G62830.1	LDL1 (LSD1-like 1)
Potri.002G068100	Bud set, Bud flush, Height		
Potri.002G068200	Bud set, Bud flush, Height	AT1G43640.1	TLP5 (tubby like protein 5)
Potri.002G194000	Height, Diameter, Regeneration height		
Potri.003G185700	Height, Diameter, Regeneration height	AT1G16260.2	Wall-associated kinase family protein
Potri.005G036500	Bud set, Bud flush, Height	AT4G16710.1	glycosyltransferase family protein 28
Potri.005G036600	Bud set, Bud flush, Height	AT1G54820.1	Protein kinase superfamily protein
Potri.005G036700	Bud set, Bud flush, Height	AT3G05210.1	UVR7
Potri.006G081800	Bud flush, Height, Regeneration height	AT5G24710.1	Transducin/WD40 repeat-like superfamily protein
Potri.006G081900	Bud flush, Height, Regeneration height	AT2G28780.1	
Potri.006G082000	Bud flush, Height, Regeneration height	AT2G28780.1	
Potri.006G082100	Bud flush, Regeneration height	AT2G37480.1	
Potri.006G160600	Height, Diameter	AT1G67280.1	Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily protein
Potri.006G160700	Height, Diameter	AT4G28850.1	XTH26 (xyloglucan hydrolase 26)
Potri.006G171200	Bud set, Bud flush, Height	AT5G17050.1	UGT78D2
Potri.006G244100	Bud set, Bud flush, Height	AT5G25590.1	Protein of unknown function (DUF630 and DUF632)
Potri.006G244200	Bud flush, Height	AT2G25735.1	

Potri.006G244300	Bud flush, Height	AT4G32690.1	GLB3 (hemoglobin 3)
Potri.006G248000	Height, Diameter	AT1G52820.1	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
Potri.006G248100	Height, Diameter	AT4G32551.2	LUG,RON2
Potri.007G011700	Bud flush, Height	AT5G65530.1	
Potri.007G011800	Bud flush, Height	AT5G65540.1	
Potri.008G110800	Bud flush, Height	AT1G68740.1	PHO1
Potri.008G110900	Bud flush, Height	AT1G68730.1	Zim17-type zinc finger protein
Potri.008G111000	Bud flush, Height		
Potri.008G131600	Bud flush, Height	AT1G70780.1	
Potri.008G131700	Bud flush, Height	AT2G01570.1	RGA1 (repressor of GA)
Potri.009G035000	Bud flush, Height		
Potri.009G035000	Bud flush, Height		
Potri.010G079700	Bud flush, Height, Regeneration height		
Potri.010G080000	Bud flush, Height, Regeneration height		
Potri.011G092800	Height, Diameter	AT3G13120.1	Ribosomal protein S10p/S20e family protein
Potri.013G109600	Bud flush, Height, Regeneration height	AT1G35530.2	DEAD/DEAH box RNA helicase family protein
Potri.013G149200	Bud flush, Height	AT2G31180.1	MYB14 (myb domain protein 14)
Potri.014G159400	Bud set, Bud flush		
Potri.015G022200	Bud flush, Height		
Potri.015G022300	Bud flush, Height	AT5G53210.1	SPCH
Potri.015G136400	Bud set, Bud flush, Height	AT5G51990.1	CBF4 (C-repeat-binding factor 4)
Potri.015G136500	Bud set, Bud flush, Height	AT1G06010.1	
Potri.015G136600	Bud set, Bud flush, Height	AT1G16560.2	Per1-like family protein
Potri.015G136700	Bud set, Bud flush, Height	AT4G25450.1	NAP8 (non-intrinsic ABC protein 8)
Potri.015G136800	Bud set, Bud flush, Height		
Potri.015G136900	Bud set, Bud flush, Height	AT1G12580.1	PEPKR1 (phosphoenolpyruvate carboxylase-related kinase 1)

Potri.018G008600	Bud set, Bud flush		
Potri.018G011400	Height, Diameter		
Potri.018G011500	Height, Diameter		
Potri.018G011600	Height, Diameter		
Potri.018G011700	Height, Diameter		
Potri.018G012000	Height, Diameter		
Potri.018G092200	Height, Diameter		

Table 3.10. Overlapping genes associated with timing of bud flush across latitude and altitude transects.

GeneID	Trait	AT homolog	Annotated description
Potri.001G214500	Bud flush	AT1G08470.1	SSL3 (strictosidine synthase-like 3)
Potri.001G256300	Bud flush	AT5G12470.1	Protein of unknown function (DUF3411)
Potri.001G442000	Bud flush	AT3G16030.1	CES101 (lectin protein kinase family protein)
Potri.002G216700	Bud flush	NA	
Potri.005G000100	Bud flush	AT5G26742.2	DEAD box RNA helicase (RH3)
Potri.005G000300	Bud flush	AT3G05940.1	Protein of unknown function (DUF300)
Potri.005G004400	Bud flush	AT5G36930.2	Disease resistance protein (TIR-NBS-LRR class) family
Potri.005G016100	Bud flush	AT1G74190.1	RLP15 (receptor like protein 15)
Potri.010G080000	Bud flush	NA	
Potri.011G052600	Bud flush	AT2G33770.1	PHO2 (phosphate 2)
Potri.012G009200	Bud flush	AT3G05660.1	RLP33 (receptor like protein 33)
Potri.012G042300	Bud flush	AT2G01970.1	Endomembrane protein 70 protein family
Potri.014G159400	Bud flush	NA	
Potri.015G019200	Bud flush	AT5G53140.1	Protein phosphatase 2C family protein
Potri.015G083700	Bud flush	AT1G23280.1	MAK16 protein-related
Potri.015G107400	Bud flush	AT5G45470.1	Protein of unknown function (DUF594)
Potri.015G113800	Bud flush	AT5G45470.1	Protein of unknown function (DUF594)
Potri.016G086500	Bud flush	AT1G07940.1	GTP binding Elongation factor Tu family protein
Potri.017G036300	Bud flush	AT3G59350.1	Protein kinase superfamily protein
Potri.018G070900	Bud flush		
Potri.019G006000	Bud flush		
Potri.019G115900	Bud flush		
Potri.001G240300	Bud flush	AT3G55440.1	TPI (triosephosphate isomerase)
Potri.001G443200	Bud flush	AT2G35630.1	MOR1 (ARM repeat superfamily protein)
Potri.004G117500	Bud flush	AT1G13680.1	PLC-like phosphodiesterases superfamily protein

Potri.004G173400	Bud flush	AT4G34950.1	Major facilitator superfamily protein
Potri.005G157300	Bud flush	AT4G24570.1	DIC2 (dicarboxylate carrier 2)
Potri.006G145600	Bud flush	AT5G35460.1	
Potri.006G164000	Bud flush	AT2G20000.1	CDC27 (CDC27 family protein)

Table 3.11. Overlapping genes associated with timing of bud set across latitude and altitude transects.

GeneID	Trait	AT homolog	Annotated description
Potri.001G276900	Bud set	AT5G19420.1	Regulator of chromosome condensation (RCC1) family protein
Potri.001G365400	Bud set	AT3G26300.1	CYP71B34
Potri.002G086800	Bud set	AT2G41705.1	camphor resistance CrcB family protein
Potri.003G021700	Bud set	AT4G22970.1	ESP (homolog of separase)
Potri.003G044000	Bud set		
Potri.003G183200	Bud set	AT1G15520.1	PDR12 (pleiotropic drug resistance 12)
Potri.004G040300	Bud set	AT4G04920.2	SFR6 (sensitive to freezing 6)
Potri.004G127200	Bud set	AT1G26690.1	GOLD family protein
Potri.004G175800	Bud set	AT1G68010.1	HPR (hydroxypyruvate reductase)
Potri.004G217800	Bud set	AT1G22930.1	T-complex protein 11
Potri.005G000300	Bud set	AT3G05940.1	Protein of unknown function (DUF300)
Potri.005G066600	Bud set	AT5G63950.1	CHR24 (chromatin remodeling 24)
Potri.005G246700	Bud set	AT1G76110.1	HMG (high mobility group) box protein with ARID/BRIGHT DNA-binding domain
Potri.006G221900	Bud set	AT5G21430.1	Chaperone DnaJ-domain superfamily protein
Potri.006G271300	Bud set	AT4G30360.1	CNGC17 (cyclic nucleotide-gated channel 17)
Potri.006G276300	Bud set	AT4G31360.1	selenium binding
Potri.007G044600	Bud set	AT5G67060.1	HEC1
Potri.007G047800	Bud set	AT4G39220.1	RER1A (Rer1 family protein)
Potri.008G035800	Bud set	AT3G54840.1	RABF1 (Ras-related small GTP-binding family protein)
Potri.008G163300	Bud set	AT1G04300.1	TRAF-like superfamily protein
Potri.010G080200	Bud set		
Potri.010G120900	Bud set		
Potri.010G145000	Bud set		
Potri.011G013200	Bud set	AT5G36930.2	Disease resistance protein (TIR-NBS-LRR class) family
Potri.011G013700	Bud set	AT5G36930.2	Disease resistance protein (TIR-NBS-LRR class) family
Potri.011G036500	Bud set	AT1G65790.1	RK1 (receptor kinase 1)
Potri.011G053900	Bud set	AT2G33800.1	Ribosomal protein S5 family protein
Potri.011G060200	Bud set	AT1G28530.1	
Potri.011G082800	Bud set	AT5G46180.1	DELTA-OAT
Potri.012G061000	Bud set	AT3G18480.1	CASP
Potri.014G001900	Bud set	AT3G14470.1	NB-ARC domain-containing disease resistance protein
Potri.014G009400	Bud set	AT3G14470.1	NB-ARC domain-containing disease resistance protein

Potri.014G024800	Bud set	AT3G57870.1	SCE1 (sumo conjugation enzyme 1)
Potri.014G055400	Bud set	AT3G17380.1	TRAF-like family protein
Potri.017G010500	Bud set	AT3G49190.1	O-acyltransferase (WSD1-like) family protein
Potri.017G103500	Bud set	AT5G17680.1	disease resistance protein (TIR-NBS-LRR class), putative
Potri.017G137700	Bud set		
Potri.018G008600	Bud set		
Potri.018G011800	Bud set		
Potri.018G069000	Bud set		
Potri.018G074500	Bud set		
Potri.018G088800	Bud set		
Potri.018G144600	Bud set		
Potri.019G022800	Bud set		
Potri.019G030500	Bud set		
Potri.019G059200	Bud set		
Potri.019G070600	Bud set		
Potri.019G095900	Bud set		
Potri.019G122900	Bud set		
Potri.001G011900	Bud set	AT3G07160.1	ATGSL10,CALS9,gsl10
Potri.001G110600	Bud set	AT1G21280.1	
Potri.001G347400	Bud set	AT3G27770.1	
Potri.002G135100	Bud set	AT3G60190.1	DL1E (DYNAMIN-like 1E)
Potri.003G189400	Bud set	AT3G54360.1	zinc ion binding
Potri.004G126400	Bud set	AT1G19260.1	TTF-type zinc finger protein with HAT dimerisation domain
Potri.004G129800	Bud set	AT2G34590.1	Transketolase family protein
Potri.004G172400	Bud set	AT2G04500.1	Cysteine/Histidine-rich C1 domain family protein
Potri.005G007600	Bud set	AT1G58190.2	RLP9 (receptor like protein 9)
Potri.005G036300	Bud set	AT5G19090.2	Heavy metal transport/detoxification superfamily protein
Potri.006G092600	Bud set	AT5G02010.1	ROPGEF7 (RHO guanyl-nucleotide exchange factor 7)
Potri.008G129900	Bud set	AT1G14990.1	
Potri.008G143000	Bud set	AT1G65810.1	P-loop containing nucleoside triphosphate hydrolases superfamily protein
Potri.012G088000	Bud set	AT3G48210.1	
Potri.012G090200	Bud set	AT4G24770.1	RBP31
Potri.012G106500	Bud set	AT5G13980.2	Glycosyl hydrolase family 38 protein
Potri.012G122400	Bud set	AT3G14470.1	NB-ARC domain-containing disease resistance protein
Potri.014G176200	Bud set		
Potri.015G025300	Bud set	AT2G34930.1	disease resistance family protein / LRR family protein

Potri.015G090600	Bud set	AT3G25100.1	CDC45 (cell division cycle 45)
Potri.019G035400	Bud set		

Table 3.12. Overlapping genes associated with height across latitude and altitude transects.

GeneID	Trait	AT homolog	Annotated description
Potri.001G004200	Height	AT1G55325.2	GCT
Potri.001G058700	Height	AT1G24265.2	Protein of unknown function (DUF1664)
Potri.001G089500	Height	AT5G42070.1	
Potri.001G214500	Height	AT1G08470.1	SSL3 (strictosidine synthase-like 3)
Potri.001G391400	Height	AT1G31860.1	HISN2 (histidine biosynthesis bifunctional protein)
Potri.002G151800	Height	AT2G45670.1	calcineurin B subunit-related
Potri.003G026100	Height	AT3G14840.2	Leucine-rich repeat transmembrane protein kinase
Potri.003G185700	Height	AT1G16260.2	Wall-associated kinase family protein
Potri.004G231400	Height	AT1G79620.1	Leucine-rich repeat protein kinase family protein
Potri.004G231500	Height	AT5G49760.1	Leucine-rich repeat protein kinase family protein
Potri.005G041100	Height	AT4G27220.1	NB-ARC domain-containing disease resistance protein
Potri.005G137700	Height	AT3G50370.1	
Potri.006G275800	Height	AT3G14470.1	NB-ARC domain-containing disease resistance protein
Potri.006G276400	Height	AT5G19940.1	Plastid-lipid associated protein PAP / fibrillin family protein
Potri.006G276600	Height	AT5G24810.1	ABC1 family protein
Potri.007G042400	Height	AT4G37030.1	
Potri.008G078800	Height	AT5G16290.1	VAT1 (VALINE-TOLERANT 1)
Potri.009G124500	Height		
Potri.010G042700	Height		
Potri.010G080000	Height		
Potri.010G246800	Height	AT5G17770.1	CBR1 (cytochrome B5 reductase 1)
Potri.011G075400	Height	AT1G29740.1	Leucine-rich repeat transmembrane protein kinase
Potri.013G048100	Height	AT3G04380.2	SUVR4
Potri.014G024700	Height	AT5G47820.1	FRA1
Potri.014G081100	Height	AT4G00660.1	RH8 (RNAhelicase-like 8)
Potri.014G128200	Height	AT4G02390.1	PARP1
Potri.014G177200	Height	AT3G22400.1	LOX5
Potri.017G024300	Height	AT3G60470.1	Plant protein of unknown function (DUF247)

Potri.018G081200	Height		
Potri.019G110600	Height		
Potri.001G385600	Height	AT1G53430.1	Leucine-rich repeat transmembrane protein kinase
Potri.001G457200	Height	AT5G65950.1	
Potri.001G463300	Height	AT2G34790.1	EDA28 (FAD-binding Berberine family protein)
Potri.002G025800	Height	AT4G31500.1	CYP83B1
Potri.002G207200	Height		
Potri.002G245000	Height	AT3G07180.1	GPI transamidase component PIG-S-related
Potri.004G192500	Height	AT1G16260.1	Wall-associated kinase family protein
Potri.005G012100	Height	AT1G58190.2	RLP9 (receptor like protein 9)
Potri.006G045400	Height	AT3G10160.1	DFC (DHFS-FPGS homolog C)
Potri.006G222500	Height	AT5G12080.3	MSL10 (mechanosensitive channel of small conductance-like 10)
Potri.007G040100	Height	AT2G15790.1	CYP40
Potri.007G144500	Height		
Potri.008G103400	Height	AT5G49510.2	PFD3 (prefoldin 3)
Potri.008G189000	Height	AT1G23420.1	INO
Potri.009G111500	Height		
Potri.010G245700	Height	AT5G04890.1	RTM2
Potri.011G035000	Height	AT1G65790.1	RK1 (receptor kinase 1)
Potri.012G005700	Height	AT3G11010.1	RLP34 (receptor like protein 34)
Potri.012G009900	Height	AT1G45616.1	RLP6 (receptor like protein 6)
Potri.013G109600	Height	AT1G35530.2	DEAD/DEAH box RNA helicase family protein
Potri.013G110000	Height	AT1G35530.2	DEAD/DEAH box RNA helicase family protein
Potri.014G154600	Height	AT1G05120.1	Helicase protein with RING/U-box domain
Potri.015G065200	Height		
Potri.015G107500	Height	AT5G45470.1	Protein of unknown function (DUF594)
Potri.015G119900	Height		
Potri.015G132800	Height	AT4G39490.1	CYP96A10
Potri.018G146300	Height		
Potri.019G117300	Height		

Table 3.13. Overlapping genes associated with diameter across latitude and altitude transects.

GeneID	Trait	AT homolog	Annotated description
Potri.003G185700	Diameter	AT1G16260.2	Wall-associated kinase family protein
Potri.005G031600	Diameter	AT4G12010.1	Disease resistance protein (TIR-NBS-LRR class) family
Potri.006G275800	Diameter	AT3G14470.1	NB-ARC domain-containing disease resistance protein
Potri.010G080200	Diameter		
Potri.011G083800	Diameter	AT3G20860.1	NEK5 (NIMA-related kinase 5)
Potri.018G012000	Diameter		
Potri.018G141900	Diameter		
Potri.019G110600	Diameter		

Table 3.14. Overlapping genes associated with regeneration height across latitude and altitude transects.

GeneID	Trait	AT homolog	Annotated description
Potri.001G037700	Regeneration height	AT5G57330.1	Galactose mutarotase-like superfamily protein
Potri.001G433100	Regeneration height		
Potri.002G017000	Regeneration height	AT1G76040.2	CPK29 (calcium-dependent protein kinase 29)
Potri.002G151800	Regeneration height	AT2G45670.1	calcineurin B subunit-related
Potri.003G060100	Regeneration height	AT1G80230.1	Rubredoxin-like superfamily protein
Potri.005G000300	Regeneration height	AT3G05940.1	Protein of unknown function (DUF300)
Potri.005G000500	Regeneration height	AT5G26742.2	DEAD box RNA helicase
Potri.006G087400	Regeneration height	AT3G09320.1	DHHC-type zinc finger family protein
Potri.007G093400	Regeneration height	AT2G22300.2	SR1 (signal responsive 1)
Potri.010G080200	Regeneration height		
Potri.012G027400	Regeneration height	AT3G11080.1	AtRLP35,RLP35
Potri.012G122600	Regeneration height	AT3G14470.1	NB-ARC domain-containing disease resistance protein
Potri.014G103800	Regeneration	AT1G01180.1	S-adenosyl-L-methionine-dependent

	height		methyltransferases superfamily protein
Potri.016G088900	Regeneration height	AT3G53400.1	
Potri.018G031000	Regeneration height		
Potri.019G097500	Regeneration height		

Figure 3.1. Population structure of 451 samples.

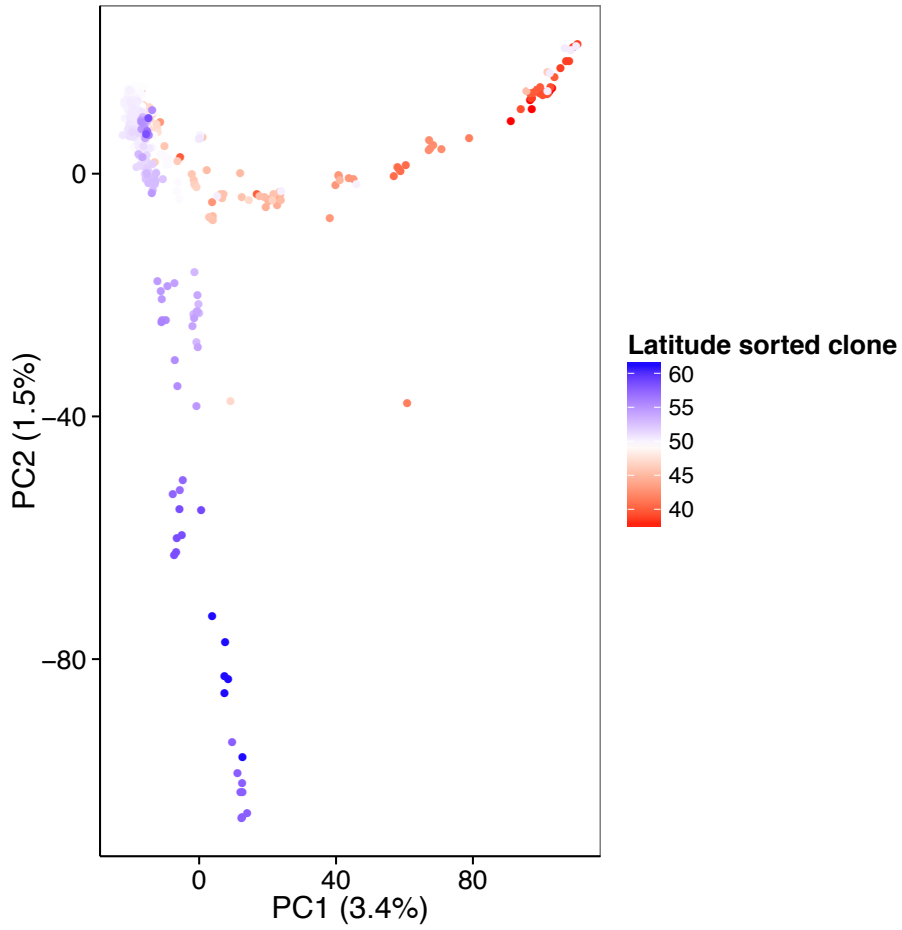


Figure 3.2. Sliding window analysis of significant number of SNPs associated with plant height (A), timing of bud flush (B), timing of bud set (C), diameter (D), regeneration height (E), and regenerated branch number (F) (The red line is the 99% confidence upper bound of estimated genomic average number of significant SNPs per window).

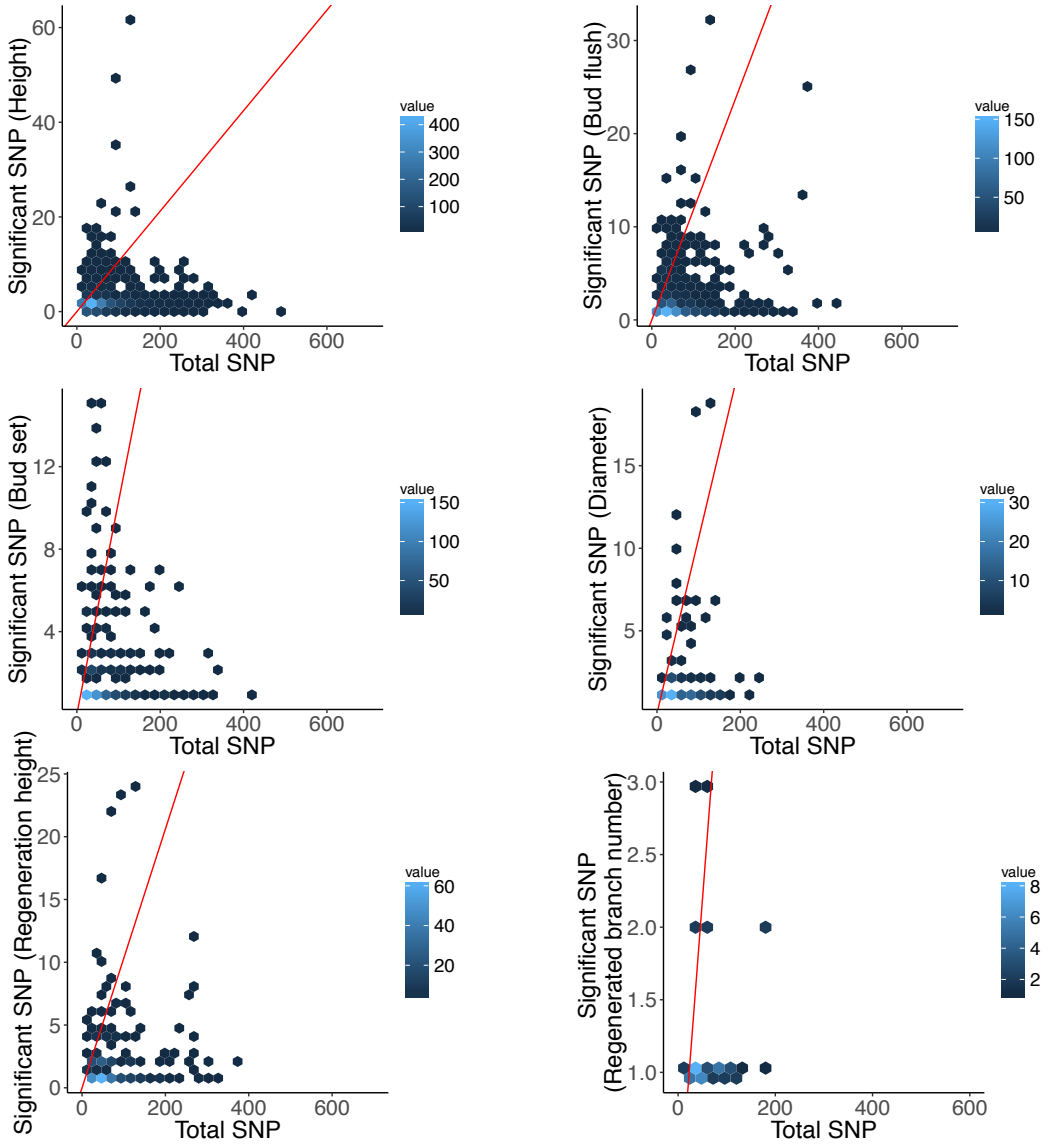


Figure 3.3. Scatterplot of significant SNPs associated with timing of bud set and bud flush (The dark red lines indicate candidate gene models within the sliding window enriched with significant SNPs).

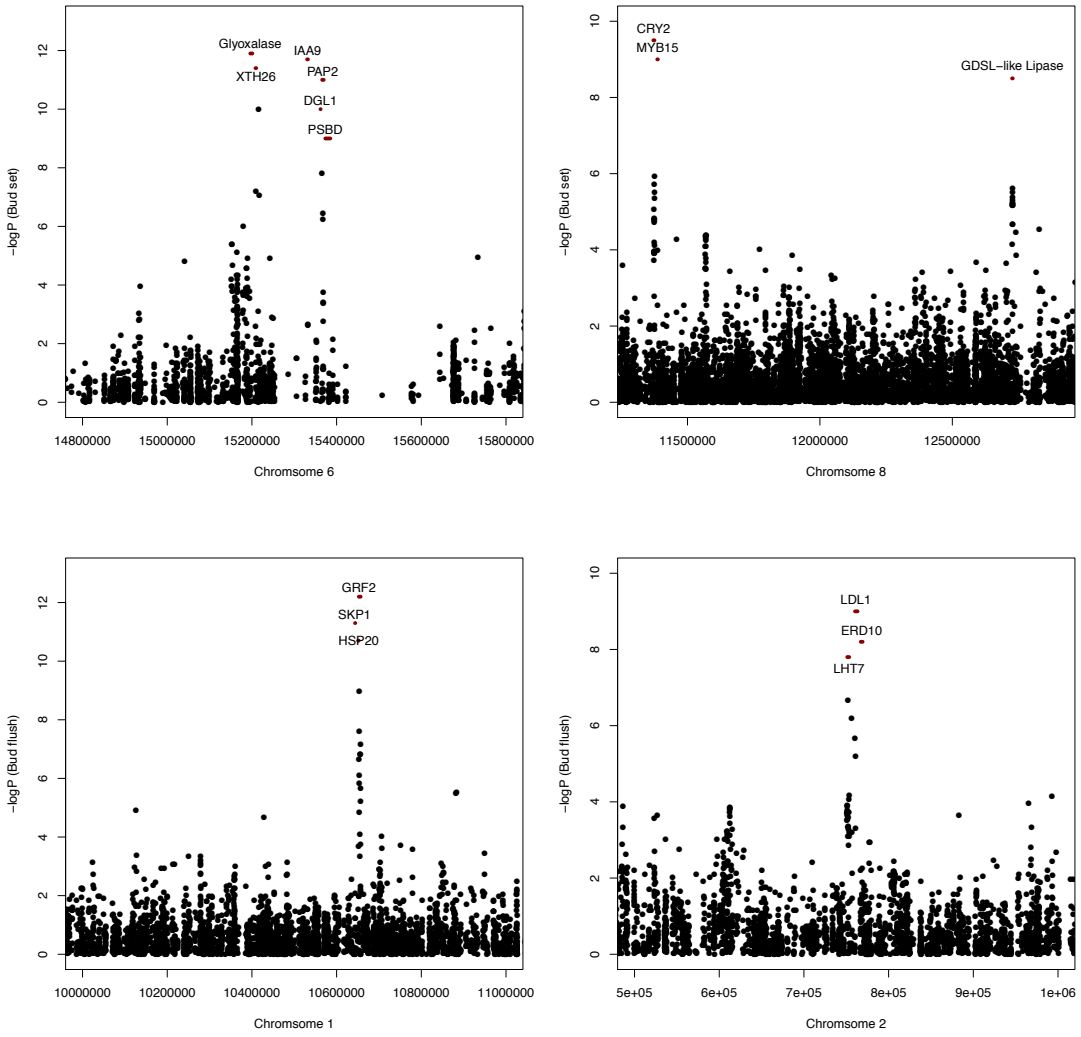


Figure 3.4. Structure of LDL1 gene (Potri.002G013100) containing SNPs associated with timing of bud flush and bud set (The heatmap indicates pairwise linkage between SNPs. The boxplots depict phenotypic effects of SNPs on the days to bud flush and bud set. The Gene structure is represented in exon (red boxes)-intron (lines)-UTR (grey boxes) patterns. The sign * indicates SNPs of significant associations).

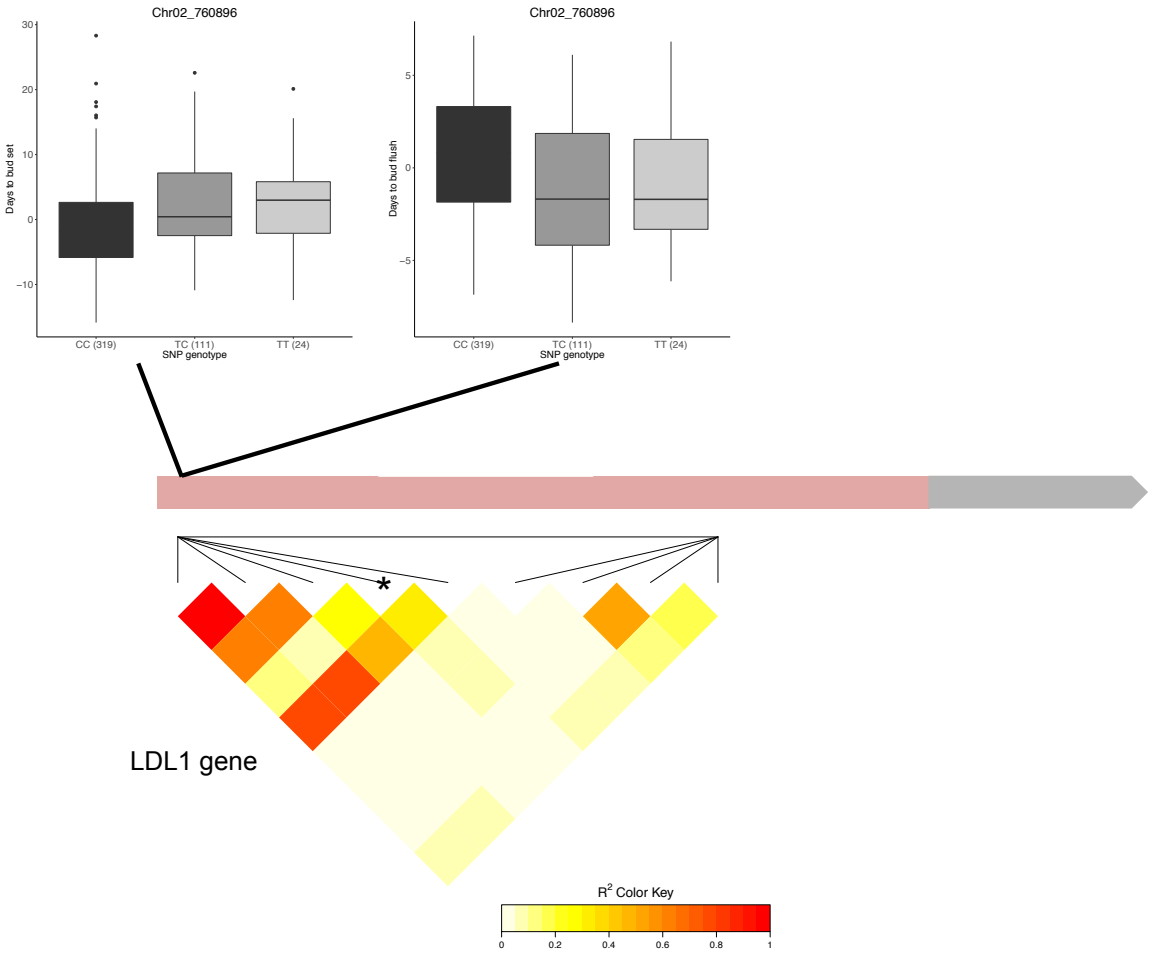


Figure 3.5. Scatterplot of significant SNPs associated with height and diameter (The dark red lines indicate candidate gene models within the sliding window enriched with significant SNPs).

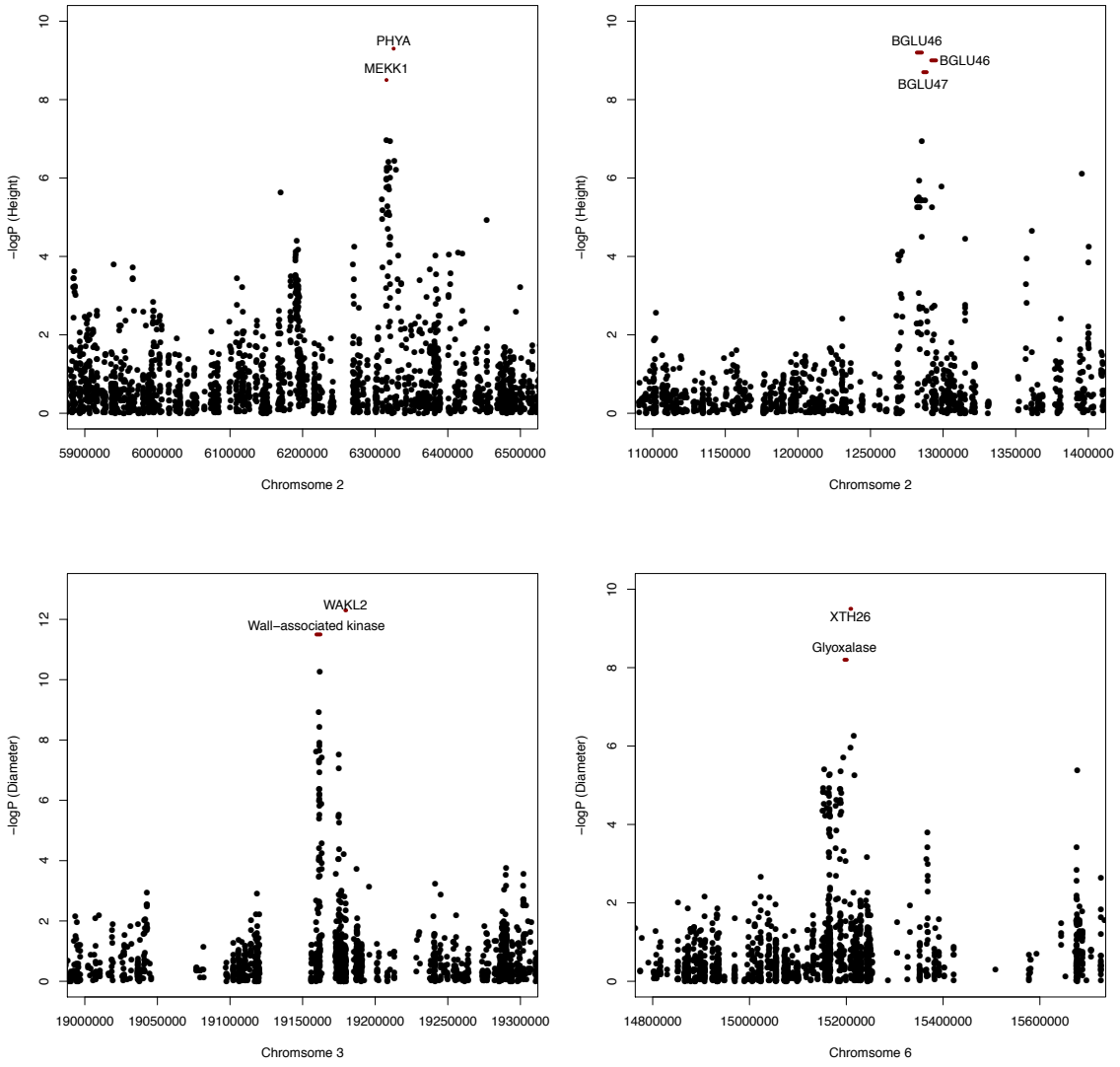


Figure 3.6. Structure of a gene block including CBF4 (Potri.015G136400) and Perl-like family protein (Potri.015G136600) containing SNPs associated with timing of bud flush and bud set (The heatmap indicates pairwise linkage between SNPs. The boxplots depict phenotypic effects of SNPs on height and the days to bud set. The Gene structure is represented in exon (red boxes)-intron (lines)-UTR (grey boxes) patterns. The sign * indicates SNPs of significant associations).

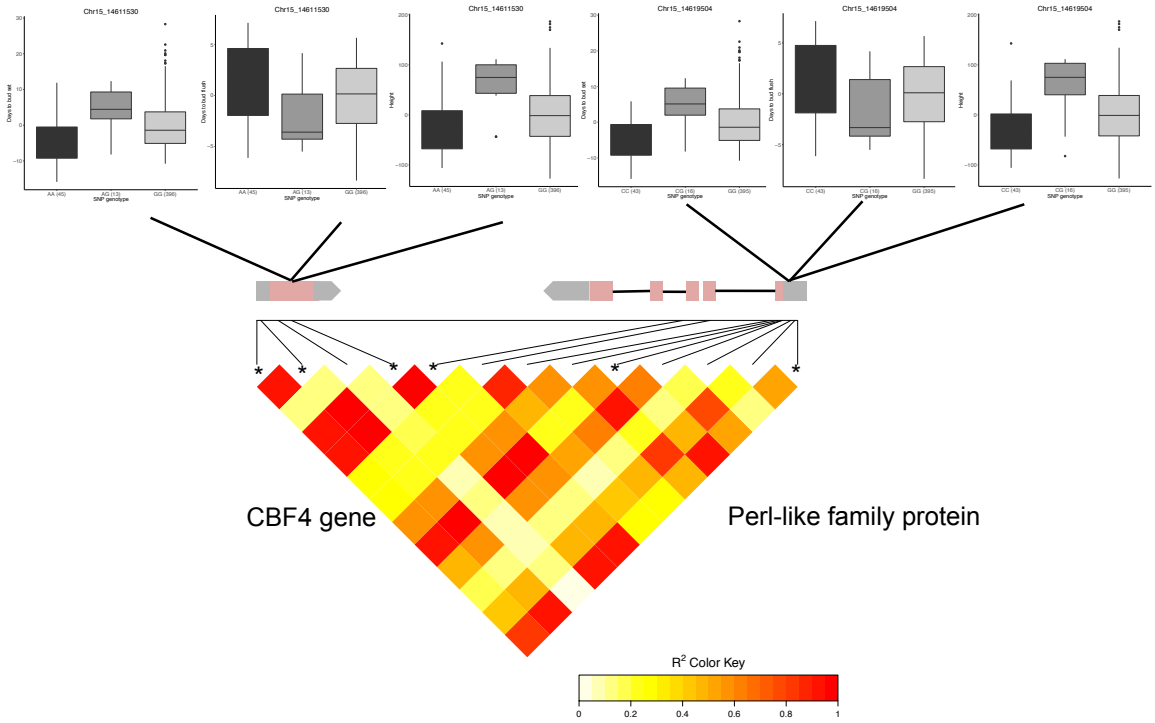


Figure 3.7. Structure of gene RGA1 (Potri.008G131700) containing SNPs associated with timing of bud flush and height (The heatmap indicates pairwise linkage between SNPs. The boxplots depict phenotypic effects of SNPs on height and the days to bud flush. The Gene structure is represented in exon (red boxes)-intron (lines)-UTR (grey boxes) patterns. The sign * indicates SNPs of significant associations).

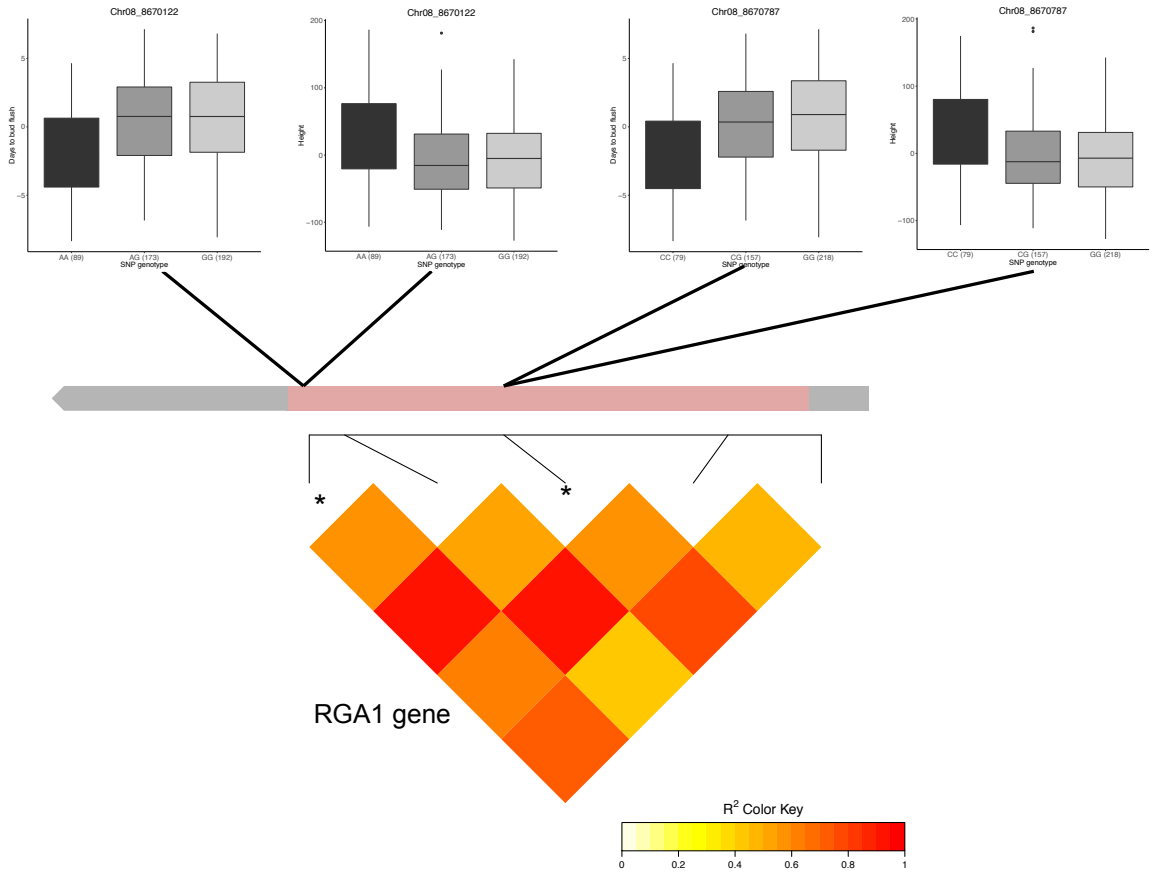


Figure S3.1. Quantile-quantile (QQ) plot of GWAS on height in VA garden (A), height in BC garden (B), timing of bud set in VA garden (C), timing of bud set in BC garden (D), timing of bud flush in VA garden (E), timing of bud flush in BC garden (F), cold injury in VA garden (G), diameter in VA garden (H), regeneration height in VA garden (I), regenerated branch number in VA garden (J) with P+K matrix correcting for population structure (VA garden is the common garden in Virginia and BC garden is the common garden in British Columbia).

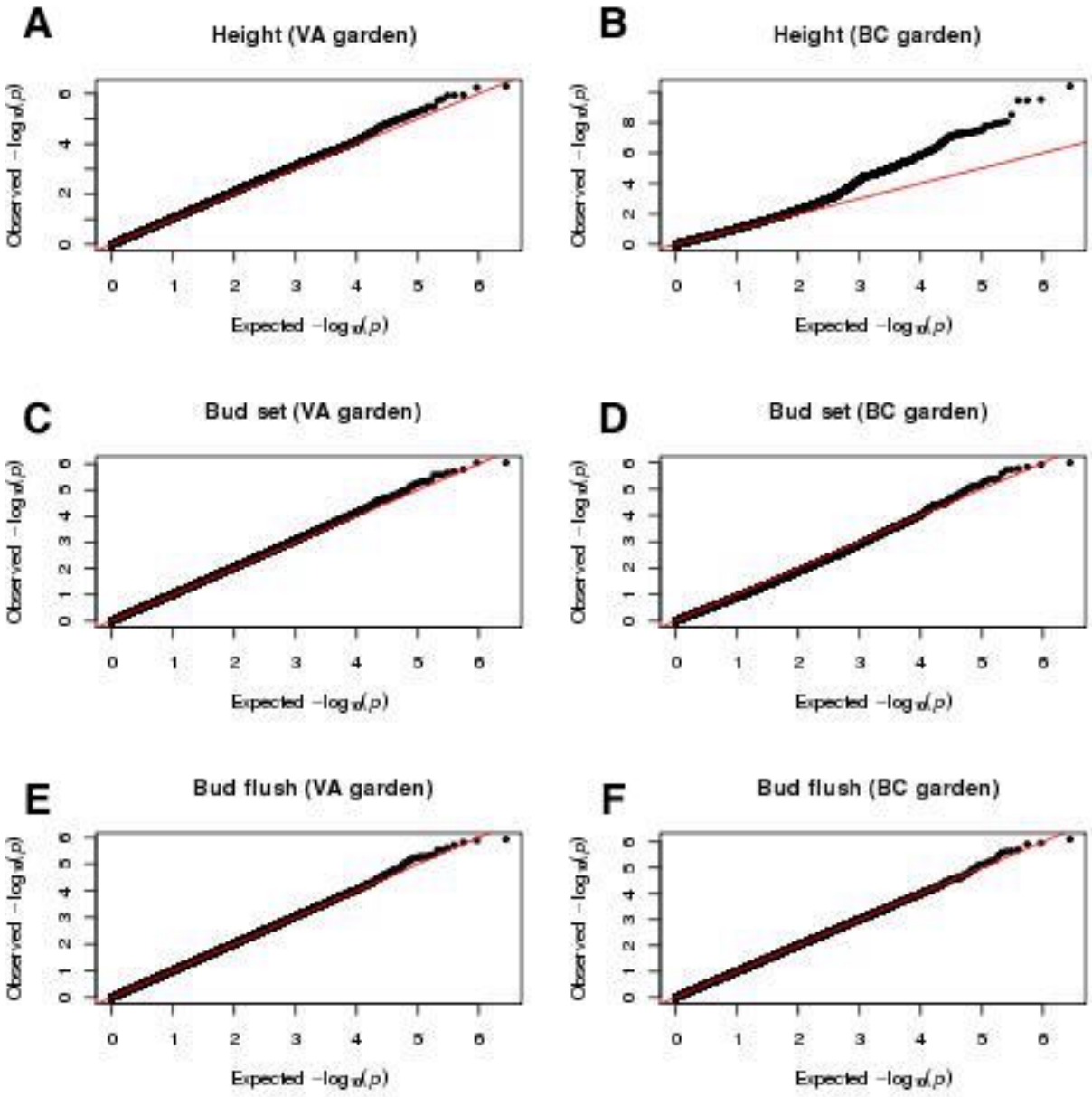


Figure S3.1. (Continued)

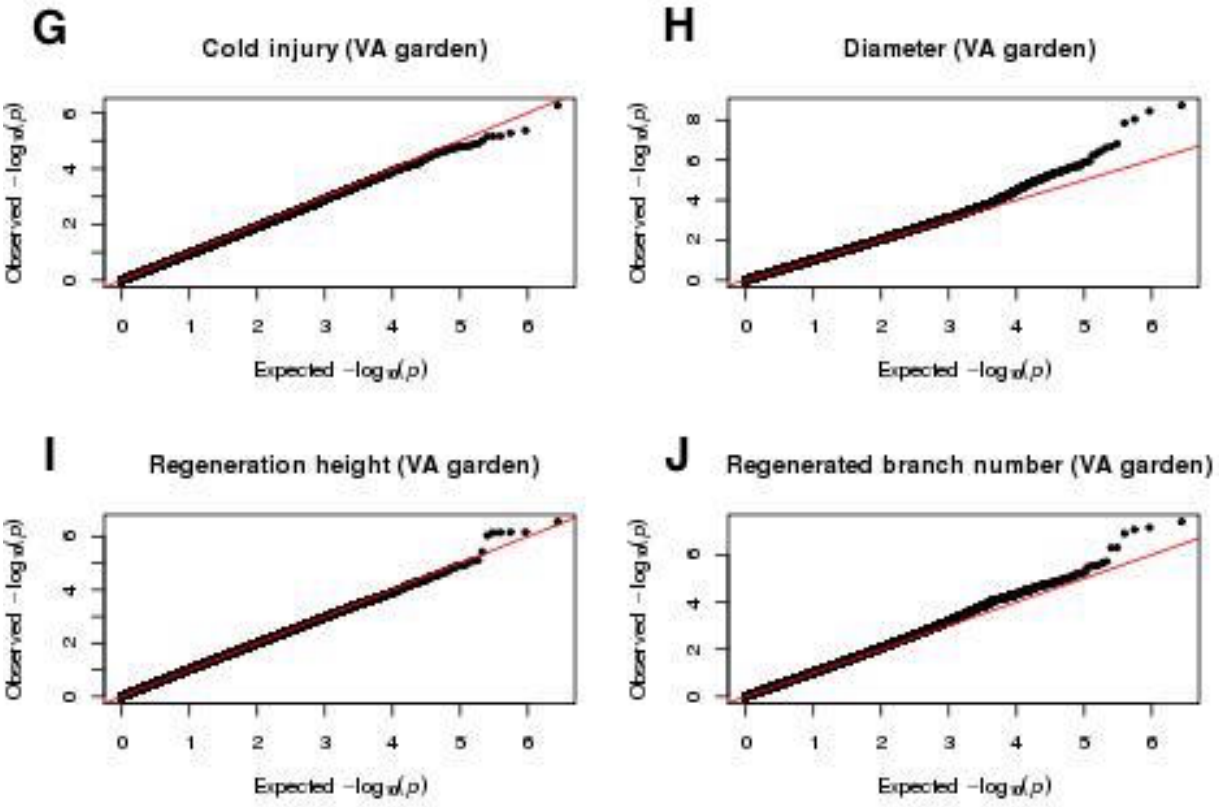


Figure S3.2. Population structure of latitudinal samples (A) and altitudinal samples (B).

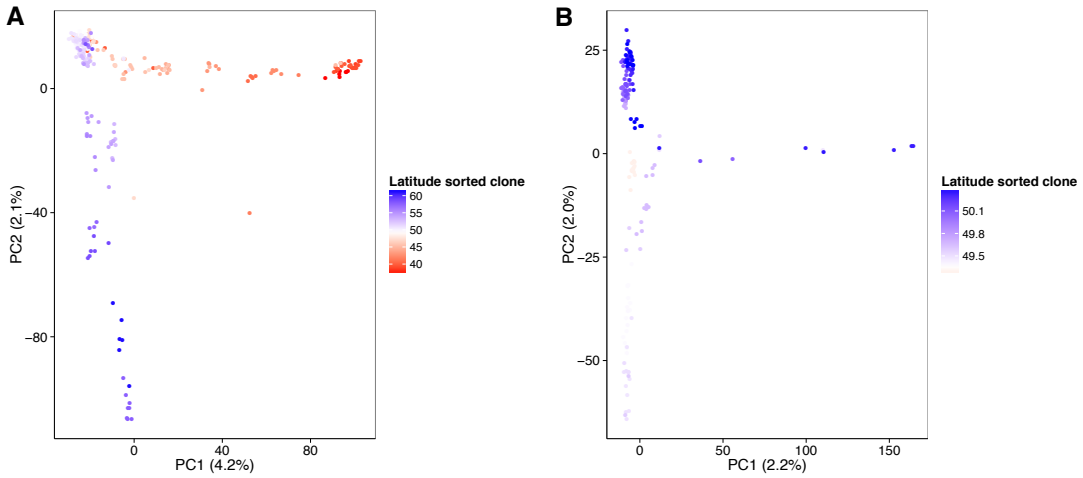


Figure S3.3. Phenotypic effects of SNPs associated with timing of bud set and bud flush.

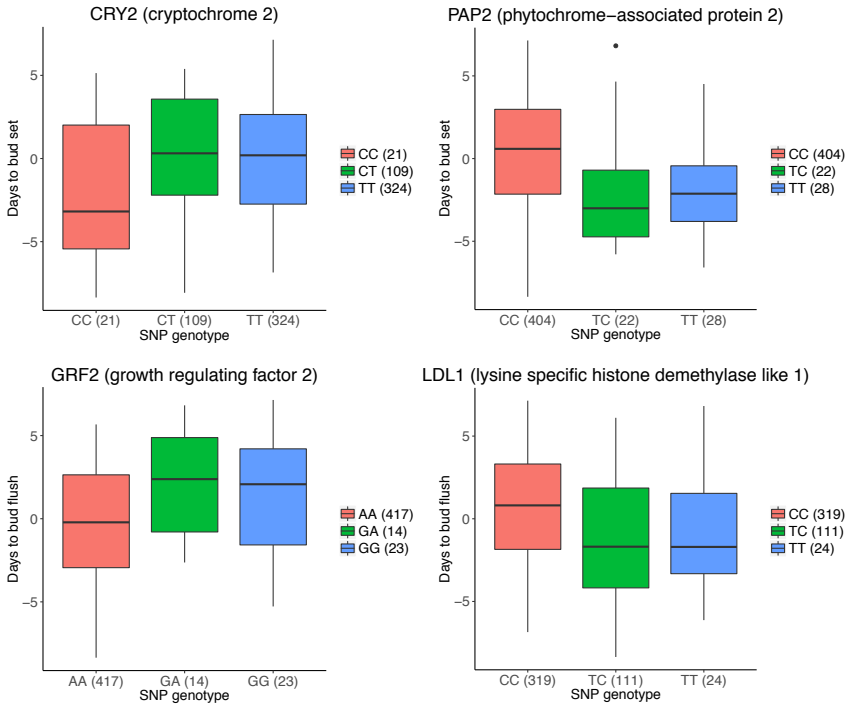


Figure S3.4. Phenotypic effects of SNPs associated with height and diameter.

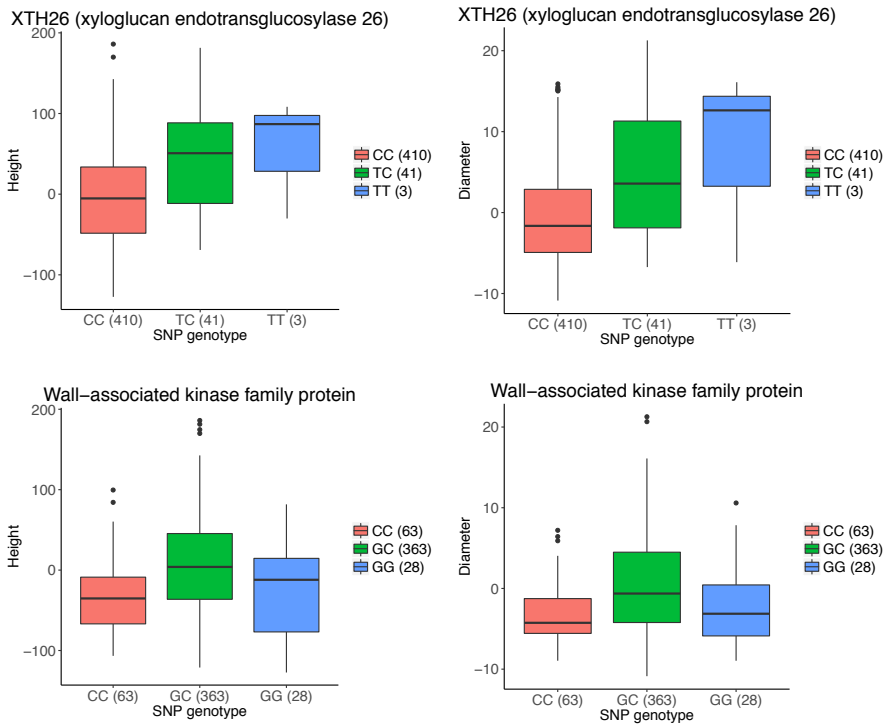


Figure S3.5. Phenotypic effects of SNPs associated with regeneration height and branch number.

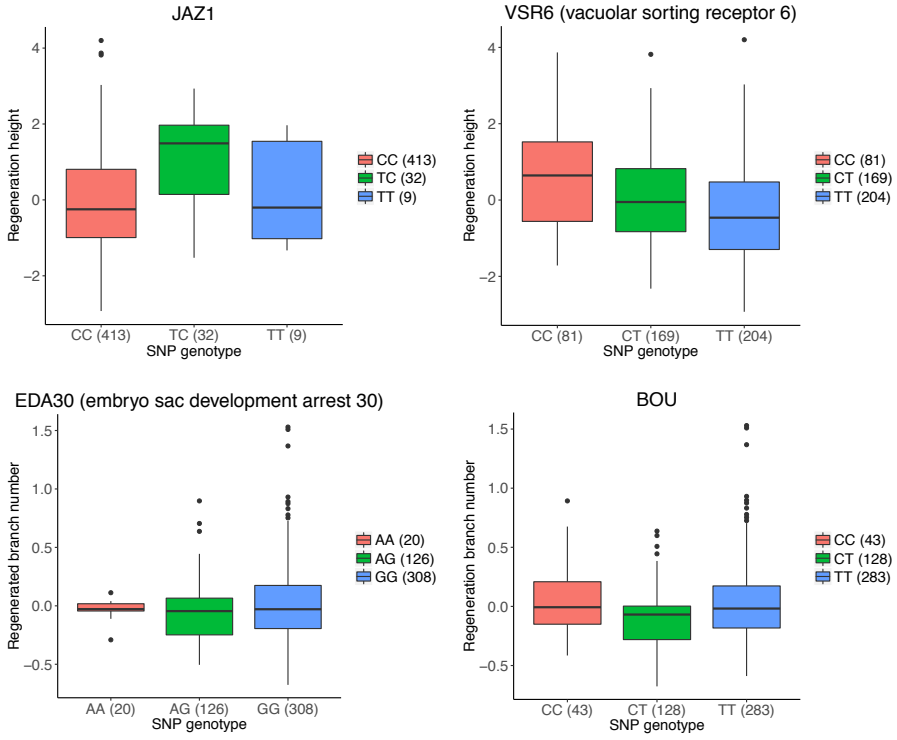
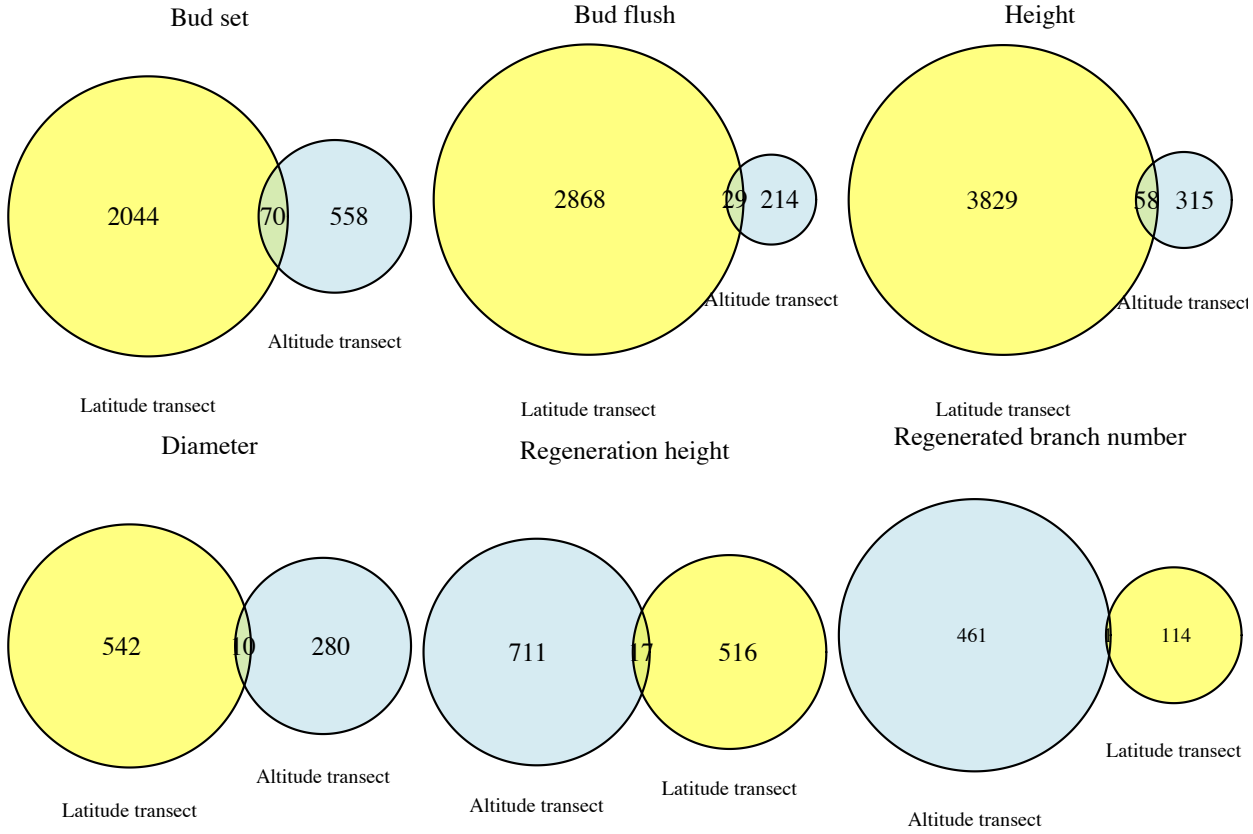


Figure S3.6. Venn diagram of the number of candidate genes from GWAS on latitudinal and altitudinal samples for phenotypic traits including timing of bud set (A), timing of bud flush (B), height (C), diameter (D), regeneration height (E), regenerated branch number (F) (yellow for samples across latitude and blue for altitude).



4. Selection scans for adaptive genomic divergence

Abstract

Identifying the genomic targets of climate-related selection is a key step in understanding the underlying mechanism of local adaptation, as well as for developing strategies for assisted gene flow and conservation under climate change. We used natural populations of *P. trichocarpa* sampled across altitude and latitude coupled with sliding window selection scans to search for imprints of natural selection. We employed three tests of selection – F_{ST}, iHS, and BayEnv – and assessed the degree of overlap in candidate regions between transects. Up to 14% of outlier regions were overlapped in general between altitude and latitude, while genomic regions associated with climate variables showed moderate level of overlap (up to ~7%). This level of overlap among genomic targets of selection is very unlikely to have occurred by chance, which implies parallel adaptation to the same selective agents. Genes within these overlapping regions were involved in the photosynthesis, photoperiodism, response to environmental stimuli, and disease resistance. Additionally, candidate regions identified from selection scans were enriched for genotype-phenotype associations with adaptive traits. These findings provided insights into the role of climatic selection in shaping genetic and phenotypic evolution.

4.1 Introduction

Natural selection is the key evolutionary force that molds species' phenotypes to better suit their local environmental conditions (Hall, et al. 2007; Le Corre and Kremer 2012). Diverse environments result in heterogeneous physiological challenges for plant populations, which in turn lead to morphological and physiological differentiation (Feng, et al. 2015). At genetic level, positive selection favors the spread of beneficial genetic variation (Sabeti 2006; Oleksyk, et al. 2009), and leaves distinct patterns on genome: background selection removes deleterious mutations and eliminates variation at linked sites, while positive selection involves genetic hitchhiking whereby neutral variation is depleted surrounding selected sites (Kaplan, et al. 1989; Charlesworth, et al. 1993; Bamshad and Wooding 2003). Many methods have been devised to capture the distinct genomic signatures of selection (Sabeti 2006; Evans, et al. 2007; Oleksyk, et al. 2009; Wollstein and Stephan 2015). These generally fall into three categories: tests for hitchhiking, tests for excessive population differentiation, and correlations between genotype and environment. The integrated haplotype score (iHS), proposed by Voight *et al*. (2006), leverages the theoretical expectation that haplotype decay around beneficial derived allele takes longer than for ancestral alleles. Directional selection may also operate differently for each locus across populations, and can cause a high degree of genetic differentiation for selected sites compared with neutral loci (Kim 2013). F_{ST} is the most widely applied statistic to quantify allele frequency differentiation between populations, with outlier values being indicative of spatially divergent selection (Wright 1950). Methods based on genotype-environment association models include Bayenv, Latent Factor Mixed Models (LFMMs) (de Villemereuil, et al. 2014), and spatial ancestry analysis (Spanelborowski, et al. 1983; Yang, et al. 2012). Bayenv is a method developed to determine potential loci involved in local adaptation using allele frequency correlations with environmental variables (Coop, et al. 2010; Blair, et al. 2014), and uses a Bayesian model-based framework to first estimate the empirical null covariance of allele frequencies across populations, and then to test for individual SNPs that have correlations between allele frequency and ecological variables deviating from this null model. Such outlier loci may be under selection driven by those environmental factors or correlated selection pressures (Coop, et al. 2010). This method has been successfully applied to a variety of systems, including plants (Eckert, et al. 2010; Hancock, et al. 2011; Fang, et al. 2012; Pyhajarvi, et al. 2013).

Previous studies of environmental adaptation have employed varying geographic transects, including latitude, longitude, and altitude, which reflects environmental heterogeneity for environmental variables such as temperature, precipitation, aridity, and biotic interactions (Richardson, et al. 2014; Adrion, et al. 2015). Some of these studies have revealed convergent or parallel evolution of phenotypes in response to similar selection pressure across populations or species groups (Renaut, et al. 2014; Messer, et al. 2015). This parallelism has been shown to extend to the molecular level – common genes or polymorphisms corresponding to similar selection pressures have been identified in a number of cases (Renaut, et al. 2014; Messer, et al. 2015). By surveying published genetic studies, Conte *et al*. (2012) estimated the probability of gene reuse during repeated phenotypic evolution among different lineages to be 0.32-0.55. However, phenotypic convergence can be achieved through differed genetic paths (Arendt and Reznick 2008; Holliday, et al. 2016). For example, a study of beach mice along two distant coasts suggested completely different genetic mutations result in similar adaptive coloration (Steiner, et al. 2009). Thus, the probability of parallel genetic basis between populations or species depends on the sparsity of geographic distribution, standing genetic variation for the trait, as well as precision of selective constraints (Hermisson 2005; Ralph and Coop 2010; Stern 2013; Messer, et al. 2015).

While several examples of parallel genetic adaptation exist for animals, less is known for plants in general and trees in particular. Range shifts for tree species associated with past climatic change have resulted in steep latitudinal and altitudinal clines that provide natural experiments to test the role of parallel environmental selection in shaping the genomic basis for adaptation (Oubida, et al. 2015; Holliday, et al. 2016). While environmental selection across latitude is well studied in trees, altitude represents a much steeper gradient and differs in possible climate drivers such as atmospheric pressure and solar radiation (Körner 2007; Klepsatel, et al. 2014). Using an F_{ST} based approach, a recent study reported a high level of parallel genetic adaptation across latitudinal and altitudinal clines in the model tree *Populus trichocarpa* (Holliday, et al. 2016). In the current study, we sought to extend this work to understand the role of parallelism in generating replicated phenotypic clines in *P. trichocarpa* by synergizing genotype-phenotype associations with additional tests of selection that directly consider environmental variables.

4.2 Materials and Methods

Genotyping and population assignment

Sequencing, alignment, SNP calling, and filtering procedures were described in detail in Chapter III. We generated 1,496,101 SNP of MAF above 0.05 for 451 individuals, which were used for F_{ST} and Bayenv analyses. We grouped the 451 clones into 80 populations (67 latitudinal groups and 13 altitudinal groups) according to their sampling location (Figure 4.1). These groupings were used for both F_{ST} and Bayenv analyses.

Genome wide selection scans

We used two methods to detect positive selection. First, we estimated the integrated haplotype score (iHS) across the genome, which evaluates the extent of linkage disequilibrium for haplotypes carrying the ancestral allele compared with haplotypes carrying the derived allele (Voight, et al. 2006). Standardized iHS scores were calculated as the log of the integrated haplotype homozygosity (iHH_A) for the ancestral allele divided by integrated Haplotype homozygosity for derived allele (iHH_D) (Voight, et al. 2006). To mitigate stochastic variation caused by low frequency alleles, we normalized raw iHS scores within allele frequency bins (Evans, et al. 2014). To detect evidence of divergent selection, we estimated population differentiation (F_{ST}) across the 67 latitudinal and 13 altitudinal populations separately, as well as for the 80 populations in total with the R package *'hierfstat'* (Goudet 2014). F_{ST} quantifies differences in allele frequency between populations to infer positive selection on specific loci (Sabeti, et al. 2006). Signatures of selection can be recognized when targeted loci and adjacent SNPs show excessive F_{ST} values (Smith and Haigh 1974). Both the normalized iHS scores and F_{ST} estimates were averaged across SNPs within each 10kb sliding window with a 5kb slide. We defined the top 1% of the empirical iHS and F_{ST} distributions as outlier regions that are potential targets of selection.

Genotype-environmental association analysis

A third scan for targets of selection was performed with Bayenv2 software, with the goal of identifying loci that show unusual correlation between allele frequencies and environmental variables. We ran Bayenv2 on both raw climate variables (mean annual temperature (MAT),

mean coldest month temperature (MCMT), mean annual precipitation (MAP), annual heat-to-moisture index (AHM), and reference atmospheric evaporative demand (Eref) as well as principal components of 21 variables (MAT, MWMT, MCMT, TD, MAP, MSP, AHM, SHM, DD_0, DD5, DD_18, DD18, NFFD, bFFP, eFFP, FFP, PAS, EMT, EXT, Eref and CMD). We incorporated the first three principal components for the latter, as they explained over 90% of total variation across all annual climate variables. PC1 among latitudinal groups was strongly correlated with temperature and frost period, PC2 was correlated with precipitation and aridity, and PC3 was a combination of precipitation and heat period (Table S4.1). For altitudinal groups, PC1 was correlated with temperature, degree-days and frost period, PC2 was correlated with precipitation and moisture, while PC3 was correlated with aridity and continentality (Table S4.1). Across total sample, PC1 was correlated with temperature and frost period, while PC2 and PC3 together partitioned precipitation, moisture, and aridity (Table S4.1).

To estimate the neutral allelic covariance matrix for Bayenv2, we used 1500 intergenic SNPs as a neutral panel. We then tested for covariance between population-specific allele frequencies and the eight climatic variables with 100,000 iterations of Bayenv2 for latitudinal groups, altitudinal and total 80 groups. We completed three replicate runs and averaged the results across runs to reduce the run-to-run variability and achieve more robust estimation (Blair, et al. 2014). Bayesian factor (BF) is the ratio of posterior probability of a SNP under the alternative hypothesis of adaptation divided by the posterior probability of the data under the null hypothesis of pure random drift (Blair, et al. 2014). We averaged the Bayesian Factors and posterior spearman correlation estimate across SNPs for each 10kb overlapping windows with 5kb slide throughout the genome. We classified regions in the top 1% Bayesian factor associated with each climate variables and PCs as candidate selection regions for each grouping, resulting 8 independent Bayenv scans for latitude groups, altitude groups, and the total 80 range wide groups.

Identification of parallel adaptive divergence

To quantify genetic parallelism across latitude and altitude transects, we first assessed the pairwise overlap between regions of divergence detected within each selection scan (F_{ST} and Bayenv) and within each climate variables (MAT, MCMT, MAP, AHM, Eref, PC1, PC2, PC3)

across latitude and altitude. Then we considered all selection scans performed on these two transects in general by classifying the top 1% of genomic regions for at least two scans as regions for each transect (F_{ST} and 8 Bayenv scans for each transect). We then calculated the overlap between the outlier regions and genes from each transect. We further annotated these candidate regions using Gene Ontology (GO) enrichment analysis (Gene Ontology 2015) and compared the overrepresented biological processes. The significance level of overlap was estimated with the *dhyper* function in R as the probability of obtaining the observed number of overlapping regions or genes under null hypothesis that overlap occurs by pure chance.

Signal of association within candidate selection regions

To assess the functional importance of outliers, we also investigated genomic overlap between outlier regions from selection scans with regions enriched for genotype-phenotype associations detected from the sliding window analysis of GWAS (Chapter III). To test whether candidate selection regions show signs of phenotypic association, we averaged $-\log_{10}(P)$ within each region. To further determine if selection outliers show greater phenotypic association than expected by chance, we permuted 10,000 times on the average $-\log_{10}(P)$ of 10kb bins throughout the genome and compared the distribution of $-\log_{10}(P)$ across selection regions versus the genomic average with a Wilcoxon signed rank test.

4.3 Results

Genomic regions showing signs of hitchhiking

Using a sliding window approach, we have partitioned the genome into 78,811 10kb-windows. Among these regions, 513 genomic regions showed average iHS>2. The hitchhiking regions were scattered across chromosome 1 to 18 except chromosome 9. A total of 641 genes fell within these regions, including genes for pathogen resistance (LRR and NB-ARC disease resistance proteins, WIN1), genes responsible for stress response (AREB3, CIPK23, and CIPK4); and photoperiod genes (PPR3, COP1, and REF) (Table 4.1). No GO term was statistically enriched among these regions. The GO terms in top 1% of fold enrichment were stress response to copper ion (GO:1990169), regulation of mitotic recombination (GO:0000019), detoxification of copper ion (GO:0010273), miRNA transport (GO:1990428), cellular defense response (GO:0006968).

Genomic divergence across populations

Sliding window analyses of F_{ST} across altitude identified 692 windows that exceeded the empirical 99th percentile of the F_{ST} distribution (mean F_{ST}>0.080; range: 0.081-0.24). These windows contained 632 genes. Across latitude, the 692 outlier windows (mean F_{ST}>0.18; range: 0.18-0.73) contained 566 genes. Between these two transects, 8.8% of genomic regions and 11.1% genes were shared (P<2e-16 based on hypergeometric test for both) (Figure S4.2). The selection outliers overlap between transects are enriched in biological processes such as photosynthesis (GO:0015979), light reaction (GO:0019684), aromatic compound biosynthetic process (GO:0019438) and generation of precursor metabolites and energy (GO:0006091) (FDR adjusted P≤0.05). The overlapping photosynthesis genes included photosynthetic electron transfer proteins, photosystem reaction proteins, and ribosomal structural proteins (Table 4.2). These photosynthesis-related genes were concentrated on chromosome 13. There were also a few stomata related genes found in both transects, including SPCH (SPEECHLESS) and KAN4 (KANADI 4) (Table 4.2).

Genetic-environment associations

Bayenv identified a number of genomic regions showing strongly differentiated allele frequency along gradient of 5 climate variables and 3 climate PCs (Figure S4.3) separately for latitude and altitude transect. We found 43 shared outlier regions (6.21% of all regions) associated with mean annual temperature (MAT) (P<2e-16 based on hypergeometric test) and 7.04% shared genes (P<2e-16 based on hypergeometric test) across both transects. The 46 MAT-associated genes identified in both transects were involved in response to hormone, leaf development and photosynthesis (Table 4.3). Among outlier regions associated with MCMT, 37 regions (6.07%) containing 27 genes (7.94%) were shared across latitude and altitude transect (Region: P= P<2e-16; Gene: P=3.9e-11 based on hypergeometric test). And 92.60% of genes associated with MCMT were same as genes associated with MAT. For mean annual precipitation, 26 out of 538 genomic regions identified across altitude (P=1.5e-12 based on hypergeometric test) were found in overlap with selection regions identified across latitude (P<2e-16 based on hypergeometric test). 45 genes associated with MAP had less overlap with temperature related climate variables (with MAT: 28.89%; with MCMT: 24.44%). There were 48 candidate regions (P<2e-16 based

on hypergeometric test) shared between altitude and latitude, which contained 55 genes, of which ~73% were also associated with MAT or MCMT, and ~20% associated with MAP. Regions associated with Eref had less overlap in associated genomic regions (2.31%; P=0.0019) and genes (2.60 %; P>0.05) between latitude and altitude compared with other climate variables.

Genomic regions showed less degree of overlap with climate principal components. For example, only 1.16% regions associated with PC1 across altitude were found among top 1% PC1 associated regions across latitude (P>0.05). Similarly, regions associated with PC2 and PC3 also have less overlap between latitude and altitude (PC2: 0.43%, P>0.05; PC3: 1.59%; P>0.05). Genes associated with climate principal components also had no overlap with genes associated with MAT, MCMT, MAP, AHM, and Eref. In general, the climate-associated genes shared across both transect included hormone regulators, genes response to cold and heat stimuli, and genes related to lipid metabolism.

Parallel selection signals across scans

Here, we considered the top 1% of outlier regions in at least in two selection scans (including F_{ST} and 8 climate Bayenv scans) as candidate selection regions for latitude and altitude respectively. In total, we identified 3307 putative selection regions containing 2991 genes across latitude, and 3017 candidate genes within 3534 regions across altitude (Figure 4.2; Table 4.4). In general, there were substantial overlapping selection outliers between two transects: 424 regions (corresponding to 12.82% of latitude and 14.05 % of altitude; P<2e-16) and 412 genes (13.66% of latitude and 13.77% of altitude; P <2e-16) were shared. To examine parallel imprints of selection captured with multiple scans, we screened the genome for regions with elevated F_{ST} and strong association with climate variables separately for the two transects. One striking observation is a 50 kb region located on chromosome 13, which was associated with MAT, AHM and F_{ST} (Figure 4.3). This region contained several genes including: Potri.013G139300 and Potri.013G139600, which encodes chloroplast NDHF dehydrogenases; Potri.013G139700, encodes protein located on chloroplast inner envelope membrane; and several genes encoding ribosomal proteins. Another example is a 20kb region on chromosome 2, which contained Potri.002G067700 (Heat shock protein 20; HSP20) and Potri.002G068200 (TLP5; TUBBY LIKE PROTEIN 5), which is responsive to fungus, ABA, and osmic stress (Bao, et al. 2014)

(Figure S4.4). This region is associated with heat moisture index and had an average F_{ST} above the 99th percentile in both transects (Figure S4.4).

GO enrichment analysis identified overrepresented pathways among parallel outliers, which included metabolic process (GO:0008152), photosynthesis (GO:0015979), response to chitin (GO:0010200), response to chemical stimulus (GO:0042221), and post-embryonic development (GO:0009791) (FDR≤0.05). We further compared the relative enrichment of biological process among candidate genes identified across latitude and altitude on a selected set of GO terms (Figure 4.4). A few pathways were relatively enriched such as response to salt (GO:1902074), response to desiccation (GO:0009269), regulation of root development (GO:2000280), meristem determinacy (GO:0010022) and photosynthesis (GO:0015979). Several flowering genes were detected across both transects: EMF1 (EMBRYONIC FLOWER), COL9 (CONSTANS-LIKE 9), PPR5 (PSEUDO-RESPONSE REGULATOR 5), and AGL16 (AGAMOUS-LIKE 16) (Table 4.4).

Selection outliers also differ in enriched pathways across latitude and altitude (Figure 4.4). For example, genes responsive to UV were shared across both transects, while gene related to pigment accumulation under UV were enriched for altitude (Figure 4.4). Among these, PAP18 (PURPLE ACID PHOSPHATASE 18), ANL2 (ANTHOCYANINLESS 2), FSD2 (FE SUPEROXIDE DISMUTASE 2), were identified across altitude, while F3H (FLAVANONE 3-HYDROXYLASE) was identified for latitude. Genes related to photoperiodism and osmotic stress were more overrepresented across latitude. For example, photoperiodic genes across latitude included CUL4 (CULLIN4), SPA1 (SUPPRESSOR OF PHYA), and CRY2 (CRYPTOCHROME 2).

Overlap among selection scans

To study the convergence of selection signals captured with different methods, we focused on selection scans on total 80 populations and classified the top 1% regions from iHS and F_{ST} and Bayenv scans as candidate selection regions. Most of these candidate regions occurred uniquely among selection scans with little pairwise overlap (Figure 4.5). Less than 5% of iHS regions were in common with outliers from either F_{ST} or Bayenv ($P>0.5$ for both hypergeometric tests).

However, there was more overlap between candidate regions for F_{ST} and Bayenv (Regions: 32.08% of F_{ST} and 6.31% of Bayenv outliers were in common; P<2e-16 based on hypergeometric test) (Figure 4.5). Genes spanning selection regions for these metrics were also enriched in several biological pathways (Figure 4.6), particularly response to environmental stimuli such as photoperiod, salt, and heat, while genes within hitchhiking regions were relatively enriched cellular water homeostasis, recombination and transport and stress response to metal ion.

Relationship between outlier regions and GWAS

Climatic selections preferably target genomic regions containing functional sites associated with advantageous phenotypes. To relate selection regions with regions related to adaptive traits, we first performed GWAS analysis with bud phenology and tree height across range wide samples. We averaged the $-\log_{10}(P)$ as an index of association strength for SNPs within each 10kb sliding window with a 5kb slide. We found that candidate selection regions identified from the F_{ST} scan had stronger phenotypic associations with bud phenology and height compared with the genomic average (P<2e-16 based on Wilcoxon signed rank test for all traits) (Figure 4.7; Figure S4.5; Figure S4.6). Outlier Bayenv regions were also significantly enriched for associations, though the $-\log_{10}(P)$ distribution is not clearly shifted to the right (P<7-e5 for all traits) (Figure 4.7; Figure S4.5; Figure S4.6). On the other hand, regions identified with iHS only showed slight and not significant enrichment for strong association with bud set (P=0.052) (Figure 4.7). To test whether these selection candidate regions are functionally related to adaptive traits, we performed sliding window analysis based on GWAS results and identified regions with significant proportion of SNP-phenotype associations for height, timing of bud set and bud flush. We found a small number of selection regions containing strong phenotypic association signals with adaptive traits (Figure 4.8; Figure S4.7; Figure S4.8). Among 15 hitchhiking regions associated with timing of bud flush, a 15kb region on chromosome 11 had a large number of high iHS scores (|iHS| >2) and was associated with bud flush (Figure 4.9). The strong selection signal occurred near Potri.011G021700 (COP1; CONSTITUTIVE PHOTOMORPHOGENIC 1), which is involved in photomorphogenesis and entrainment of circadian clock. COP1 and SPA1 (SUPPRESSOR OF PHYA 1) form ubiquitin ligase complex, which degrades CO (CONSTANS) and represses flowering in Arabidopsis (Hajdu, et al. 2015). Another example is the Bayenv

outlier regions associated with climate variable PC1 and PC2, which contained strong association signals with timing of bud set and bud flush (Figure 4.10). One gene within this region, Potri.005G065700, encodes a Copine (calcium-dependent phospholipid-binding protein) family protein. This gene is involved in Ca^{2+} dependent stimulus-responses and was found required for pollen fertility by regulating Ca^{2+} and phospholipid signaling pathways in Rice (Yang, et al. 2008). A second gene (Potri.005G065800; EMB2780; EMBRYO DEFECTIVE 2780) is a DNA-directed DNA polymerase that functions in DNA replication and mitotic cell cycle.

We also detected candidate selection regions that showed no strong phenotypic association but contained genes potentially involved in photoperiodism and dormancy pathways. For example, a 10kb sliding window on chromosome 14 had high F_{ST} and was associated with MAT, MCMT, and AHM (Figure 4.11). COL9 (Potri.014G170600; CONSTANS-LIKE 9) was in this region, which delays flowering in Arabidopsis by down-regulating CO (CONSTANS) and FT (FLOWERING LOCUS T), the key regulators of flowering and dormancy pathways in *Populus* (Cheng and Wang 2005; Bohlenius 2006; Hsu, et al. 2011). Another gene in this region, Potri.014G170800, was reported be related to meristem vegetative to reproductive phase transition (Doyle and Amasino 2009).

4.4 Discussion

We used three genome-wide scans across two natural transects of *P. trichocarpa* to identify regions of the genome that have been targeted by selection. We compared the putative selection-targeted genes in our study with one recent study revealing targets of selection in *P. trichocarapa* (Evans, et al. 2014). We found 5.1% ~ 10.3% candidate genes of exeme iHS, F_{ST}, Bayesian factors associated with PC1, PC2 in this study were also within candidate regions detected in our study. We also found substantial overlap in regions with signatures of positive or divergent selection between two geographical clines, indicating that parallel adaptation likely occurred during colonization of the two transects. At the same time, many regions did not overlap between transects, which suggests selection pressures also differ to some extent between latitude and altitude. Many selection signals overlapped with genotype-phenotype associations, and it is therefore likely that the candidate regions were under selective pressure due to their functional

importance in phenotypic adaption. There was little convergence in candidate selection regions found across selection scans, which indicates that various selection pressures and evolutionary processes resulted in a complex genome landscape of selection imprints.

Parallel and divergent selection

Parallel adaptation has been a hot topic in evolutionary biology for many years (Losos 2011), as it provides a window into the repeatability of evolution. The natural latitude and altitude transects we studied represent similar ecological clines varying in the type and amount of precipitation, mean and seasonal temperatures, and yearly timing of growing seasons. One major shared selection agent across latitude and altitude transect is the local differences in mean annual temperature. Northern clones and clones from high altitude are usually adapted to lower mean annual temperature and colder winters. Consequently, we expected the genomes of local populations that experience similar temperature regimes to respond in parallel, to the extent that genetic constraints exist for the loci of adaptation. To detect these signatures, we performed sliding window analysis using F_{ST} and Bayenv scans. For F_{ST}, we detected a significant level of overlap among outlier regions and the composite genes, which is unlikely to occur by chance. This suggests that local selection targets similar standing genetic variations and drives adaptive allele frequency shifts across spatial groups. The Bayenv scan pointed out the many targeted genomic regions and genes for specific selection agents across the two transects. Interestingly, there was significant overlap between candidate selection regions associated with temperature and moisture related variables, while there was less overlap with Eref and climate principal components. With temperature and water availability being two major factors explaining environmental variation, it is likely that same genetic variants were favored in response to similar selective constraint. The parallel frequency shifts we observed across similar environmental clines confirmed the role of climatic selection in driving parallel adaptation. As these two natural transect share common ancestral populations, genomic context and standing genetic variants may limit options of adaptation pathways, and thus partially explain the level of overlap we observed (Losos 2011; Stern 2013).

We also observed a large number of discordant selection outliers between latitude and altitude. This non-overlap is likely generated by unique selective constraints to each transect favoring

different genetic variants. For example, atmospheric pressure and daily temperature difference tend to influence plant physiology and metabolism in altitudinal gradients, while day length and solar angle of incidence vary dramatically across latitudinal gradients (Altshuler 2006). Our GO enrichment analysis revealed that photoperiodic genes were among selection outliers in both transect, but were more enriched across latitude, suggesting photoperiod being a major force driving population differentiation. Similarly, genes responsive to UV light were found in both clines, but UV-induced pigment accumulation process was overrepresented across altitude, suggesting solar radiation being a more pronounced selection constraint for this transect. Altitude also reflects a steeper temperature cline than latitude (Altshuler 2006; Holliday, et al. 2016), which may also play a role in shaping the genomic targets of selection. Moreover, historical distributions and complex demographical processes specific to local populations can also be responsible for the lack of concordance between two transects (Renaut, et al. 2014). Future analysis can explore how genomic features and population parameters limit parallelism.

Functional role of selection regions

Many candidate regions contain genes that were plausibly related to adaptation to biotic and abiotic conditions. Genes involved in pathways related to photoperiodic response, dormancy, resistance to herbivores and pathogens, and response and tolerance to environmental stimuli were especially enriched, which is consistent with previous studies (Evans, et al. 2014; Holliday, et al. 2016). Heat shock proteins (HSP) are mostly chaperones that respond to stressful conditions (not just heat) and affect protein folding and conformation (Wang, et al. 2008). HSPs displayed association with MAT and MCMT for both transects. One gene (Potri.010G217200; low temperature and salt responsive family protein) was found associated with MAP across both transects. The Arabidopsis ortholog of this gene belongs to a low temperature and salt responsive gene family and is involved in cold and salinity response. Parallel selection outliers were also enriched for genes in the photosynthesis pathway, photoperiodism regulation, meristem determinacy and response to salt. Interesting genes included HT1 (HIGH LEAF TEMPERATURE 1; Potri.012G080000), which controls red-light induced leaf stomatal opening (Matrosova, et al. 2015); COL9 (CONSTANS-LIKE 9; Potri.014G170600), which belongs to the CO (CONSTANS) gene family and can delay flowering by down-regulating expression CO and FT (Cheng and Wang 2005); HOS15 (HIGH EXPRESSION OF OSMOTICALLY

RESPONSIVE GENES; Potri.005G144400), which is involved in histone deacetylation in response to abiotic stress such as salt and cold, and can modulate flowering timing through regulation of FT and SOC expression (Zhu, et al. 2008); and AGL16 (AGAMOUS-LIKE 16; Potri.005G150500), which acts as part of the reproductive pathway in Arabidopsis. These selection outlier genes provide reliable candidate targets for climatic selection. Functional studies is further required confirm their role in local adaptation in tree species.

Convergence and discordance between selection scanning methods

Among 492 regions identified as under positive selection, only 7 regions showed extreme differentiated allele frequencies across population and 8 regions showed correlation with climatic variables detected by Bayenv. The lack of congruence between iHS and the other two methods was also observed in previous studies in *Populus* (Evans, et al. 2014). iHS (Integrated Haplotype Score) is a statistics developed to detect genetic imprints of recent positive selection at a locus (Voight, et al. 2006). It searches for regions of high linkage disequilibrium surrounding a positively selected allele comparing with local background alleles (Sabeti, et al. 2002; Voight, et al. 2006). Regions of extreme iHS score (|iHS>2|) can occur when certain alleles are favored and adjacent alleles are subjected to hitchhiking (Voight, et al. 2006). This method is powerful for detecting recent selection events where selected alleles swept to intermediate frequencies, but fails to capture ancient positive selection or beneficial alleles that have reached fixation (Sabeti, et al. 2002; Voight, et al. 2006). Another reason for a lack of overlap between the iHS results and F_{ST}/Bayenv is that many adaptive events in natural populations occur by polygenic adaptation rather than hard selective sweeps (Pritchard and Di Rienzo 2010). In this context, a dramatic adaptive response could result from slight frequency shifts across many loci, which is unlikely to leave selection signatures standing out above neutral background variants (Pritchard and Di Rienzo 2010). iHS may fail to capture signatures of selection in such cases.

There was more substantial concordance between F_{ST} and Bayenv than either with iHS scan. F_{ST} detects genetic adaptive divergence manifested in allele frequency differentiation at a locus across populations (Jakobsson, et al. 2012), while Bayenv detects associations between climate and allelic frequencies across populations (Gunther and Coop 2013). Thus, with climate diverging across geographic clines, adaptive loci can differentiate across environmental gradients

and spatial groups, resulting in signals of local adaptation captured in both methods. The large number of non-overlapping selection regions between these tests may reflect the different assumptions of the two methods. Bayenv accounted for the problem of sample size difference and the shared history of populations which can lead to false positives and negatives by including estimates of background covariance using a set of control markers (Gunther and Coop 2013). While F_{ST} is considerably dependent on the allele-frequency distribution and genetic diversity within populations (Charlesworth 1998; Jakobsson, et al. 2012). Another explanation for the discrepancies between these methods is that genes displaying high levels of allele frequency differentiation across populations may be targets of ecological variables such as soil nutrients, biotic competition, or herbivory that are unrelated to gradients in temperature or precipitation (Alberto, et al. 2013). Future studies would benefit from measuring additional environmental variables beyond climate, which should capture additional locally-adaptive variants.

4.5 Conclusions

Climate is one major abiotic constraint directing adaptive change in morphology, phenology, and physiology of plants. Here we detected candidate selection regions with three genome-wide selection scans and demonstrated a substantial overlap between selected regions across latitude and altitude. Additionally, we found evidence of parallel adaptation across these gradients, which suggested some genetic constraint on the loci of adaptation. Our study also integrated association with adaptive traits and demonstrated that these selection targets are potentially important for climatic adaptation.

4.6 Reference cited

Adrion JR, Hahn MW, Cooper BS. 2015. Revisiting classic clines in Drosophila melanogaster in the age of genomics. Trends in Genetics 31:434-444.

Alberto FJ, Derory J, Boury C, Frigerio JM, Zimmermann NE, Kremer A. 2013. Imprints of Natural Selection Along Environmental Gradients in Phenology-Related Genes of Quercus petraea. Genetics 195:495-512.

Altshuler DL. 2006. Adaptations to life at high elevation: An introduction to the symposium. Integrative and Comparative Biology 46:3-4.

Arendt J, Reznick D. 2008. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? Trends in Ecology & Evolution 23:26-32.

Bamshad M, Wooding SP. 2003. Signatures of natural selection in the human genome. Nature Reviews Genetics 4:99-111.

Bao Y, Song W-M, Jin Y-L, Jiang C-M, Yang Y, Li B, Huang W-J, Liu H, Zhang H-X. 2014. Characterization of Arabidopsis Tubby-like proteins and redundant function of AtTLP3 and AtTLP9 in plant response to ABA and osmotic stress. Plant Molecular Biology 86:471-483.

Blair LM, Granka JM, Feldman MW. 2014. On the stability of the Bayenv method in assessing human SNP-environment associations. Human Genomics 8:1.

Bohlenius H. 2006. CO/FT Regulatory Module Controls Timing of Flowering and Seasonal Growth Cessation in Trees. Science 312:1040-1043.

Charlesworth B. 1998. Measures of divergence between populations and the effect of forces that reduce variability. Molecular Biology and Evolution 15:538-543.

Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics 134:1289-1303.

Cheng X-F, Wang Z-Y. 2005. Overexpression of COL9, a CONSTANS-LIKE gene, delays flowering by reducing expression of CO and FT in Arabidopsis thaliana. The Plant Journal 43:758-768.

Conte GL, Arnegard ME, Peichel CL, Schluter D. 2012. The probability of genetic parallelism and convergence in natural populations. Proceedings of the Royal Society B: Biological Sciences 279:5039-5047.

Coop G, Witonsky D, Di Rienzo A, Pritchard JK. 2010. Using Environmental Correlations to Identify Loci Underlying Local Adaptation. Genetics 185:1411-1423.

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

Doyle MR, Amasino RM. 2009. A Single Amino Acid Change in the Enhancer of Zeste Ortholog CURLY LEAF Results in Vernalization-Independent, Rapid Flowering in Arabidopsis. Plant Physiology 151:1688-1697.

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

Evans LM, Slavov GT, Rodgers-Melnick E, Martin J, Ranjan P, Muchero W, Brunner AM, Schackwitz W, Gunter L, Chen JG, et al. 2014. Population genomics of *Populus trichocarpa* identifies signatures of selection and adaptive trait associations. Nature Genetics 46:1089-1096.

Evans SN, Shvets Y, Slatkin M. 2007. Non-equilibrium theory of the allele frequency spectrum. Theor Popul Biol 71:109-119.

Fang Z, Pyhajarvi T, Weber AL, Dawe RK, Glaubitz JC, Gonzalez JdJS, Ross-Ibarra C, Doebley J, Morrell PL, Ross-Ibarra J. 2012. Megabase-Scale Inversion Polymorphism in the Wild Ancestor of Maize. Genetics 191:883-894.
Fay JC, Wu CI. 2000. Hitchhiking under positive Darwinian selection. Genetics 155:1405-1413.
Feng X-J, Jiang G-F, Fan Z. 2015. Identification of outliers in a genomic scan for selection along environmental gradients in the bamboo locust, Ceracris kiangsu. Scientific Reports 5:13758.
Gene Ontology C. 2015. Gene Ontology Consortium: going forward. Nucleic Acids Res 43:D1049-1056.
Goudet J. 2014. hierfstat: Estimation and tests of hierarchical F-statistics: R package version 0.04-14.
Gunther T, Coop G. 2013. Robust Identification of Local Adaptation from Allele Frequencies. Genetics 195:205-220.
Hajdu A, Ádám É, Sheerin DJ, Dobos O, Bernula P, Hiltbrunner A, Kozma-Bognár L, Nagy F. 2015. High-level expression and phosphorylation of phytochrome B modulates flowering time in Arabidopsis. The Plant Journal 83:794-805.
Hall D, Luquez V, Garcia VM, St Onge KR, Jansson S, Ingvarsson PK. 2007. Adaptive population differentiation in phenology across a latitudinal gradient in European aspen (*Populus tremula*, L.): a comparison of neutral markers, candidate genes and phenotypic traits. Evolution 61:2849-2860.
Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, Toomajian C, Roux F, Bergelson J. 2011. Adaptation to Climate Across the Arabidopsis thaliana Genome. Science 334:83-86.
Hermisson J. 2005. Soft Sweeps: Molecular Population Genetics of Adaptation From Standing Genetic Variation. Genetics 169:2335-2352.
Holliday JA, Zhou L, Bawa R, Zhang M, Oubida RW. 2016. Evidence for extensive parallelism but divergent genomic architecture of adaptation along altitudinal and latitudinal gradients in *Populus trichocarpa*. New Phytologist 209:1240-1251.
Hsu CY, Adams JP, Kim H, No K, Ma C, Strauss SH, Drnevich J, Vandervelde L, Ellis JD, Rice BM, et al. 2011. FLOWERING LOCUS T duplication coordinates reproductive and vegetative growth in perennial poplar. Proc Natl Acad Sci U S A 108:10756-10761.
Jakobsson M, Edge MD, Rosenberg NA. 2012. The Relationship Between FST and the Frequency of the Most Frequent Allele. Genetics 193:515-528.
Kaplan NL, Hudson RR, Langley CH. 1989. The "hitchhiking effect" revisited. Genetics 123:887-899.
Kim Y. 2013. Stochastic patterns of polymorphism after a selective sweep over a subdivided population. Genetics Research 95:57-67.
Klepsatel P, Gáliková M, Huber CD, Flatt T. 2014. Similarities and Differences in Altitudinal Versus Latitudinal Variation for Morphological Traits Indrosophila Melanogaster. Evolution 68:1385-1398.
Körner C. 2007. The use of ‘altitude’ in ecological research. Trends in Ecology & Evolution 22:569-574.
Le Corre V, Kremer A. 2012. The genetic differentiation at quantitative trait loci under local adaptation. Molecular Ecology 21:1548-1566.
Losos JB. 2011. Convergence, Adaptation, and Constraint. Evolution 65:1827-1840.
Matrosova A, Bogireddi H, Mateo-Peñas A, Hashimoto-Sugimoto M, Iba K, Schroeder JI, Israelsson-Nordström M. 2015. The HT1 protein kinase is essential for red light-induced

stomatal opening and genetically interacts with OST1 in red light and CO2-induced stomatal movement responses. New Phytologist 208:1126-1137.
Messer PW, Ralph PL, Coop G. 2015. Convergent Evolution During Local Adaptation to Patchy Landscapes. Plos Genetics 11:e1005630.
Nielsen R, Hellmann I, Hubisz M, Bustamante C, Clark AG. 2007. Recent and ongoing selection in the human genome. Nature Reviews Genetics 8:857-868.
Oleksyk TK, Smith MW, O'Brien SJ. 2009. Genome-wide scans for footprints of natural selection. Philosophical Transactions of the Royal Society B: Biological Sciences 365:185-205.
Oubida RW, Gantulga D, Zhang M, Zhou L, Bawa R, Holliday JA. 2015. Partitioning of multivariate phenotypes using regression trees reveals complex patterns of adaptation to climate across the range of black cottonwood (*Populus trichocarpa*). Front Plant Sci 6.
Pritchard JK, Di Rienzo A. 2010. Adoption – not by sooty alone. Nature Reviews Genetics 11:665-667.
Pyhajarvi T, Hufford MB, Mezmouk S, Ross-Ibarra J. 2013. Complex Patterns of Local Adaptation in Teosinte. Genome Biology and Evolution 5:1594-1609.
Ralph P, Coop G. 2010. Parallel Adaptation: One or Many Waves of Advance of an Advantageous Allele? Genetics 186:647-668.
Renaut S, Owens GL, Rieseberg LH. 2014. Shared selective pressure and local genomic landscape lead to repeatable patterns of genomic divergence in sunflowers. Molecular Ecology 23:311-324.
Richardson JL, Urban MC, Bolnick DI, Skelly DK. 2014. Microgeographic adaptation and the spatial scale of evolution. Trends in Ecology & Evolution 29:165-176.
Sabeti PC. 2006. Positive Natural Selection in the Human Lineage. Science 312:1614-1620.
Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, et al. 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature 419:832-837.
Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES. 2006. Positive natural selection in the human lineage. Science 312:1614-1620.
Smith JM, Haigh J. 1974. The hitch-hiking effect of a favourable gene. Genet Res 23:23-35.
Spanelborowski K, Petterborg LJ, Reiter RJ. 1983. Morphological and Morphometric Changes in the Ovaries of White-Footed Mice (Peromyscus-Leucopus) Following Exposure to Long or Short Photoperiod. Anatomical Record 205:13-19.
Steiner CC, Rompler H, Boettger LM, Schoneberg T, Hoekstra HE. 2009. The genetic basis of phenotypic convergence in beach mice: similar pigment patterns but different genes. Molecular Biology and Evolution 26:35-45.
Stern DL. 2013. The genetic causes of convergent evolution. Nature Reviews Genetics 14:751-764.
Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. PLoS Biol 4:e72.
Wang Y, Zhang WZ, Song LF, Zou JJ, Su Z, Wu WH. 2008. Transcriptome Analyses Show Changes in Gene Expression to Accompany Pollen Germination and Tube Growth in Arabidopsis. Plant Physiology 148:1201-1211.
Wollstein A, Stephan W. 2015. Inferring positive selection in humans from genomic data. Investigative Genetics 6:5.
Wright S. 1950. Genetical structure of populations. Nature 166:247-249.

Yang W-Q, Lai Y, Li M-N, Xu W-Y, Xue Y-B. 2008. A Novel C2-Domain Phospholipid-Binding Protein, OsPBP1, Is Required for Pollen Fertility in Rice. Molecular Plant 1:770-785.
Yang W-Y, Novembre J, Eskin E, Halperin E. 2012. A model-based approach for analysis of spatial structure in genetic data. Nature Genetics 44:725-731.
Zhu J, Jeong JC, Zhu Y, Sokolchik I, Miyazaki S, Zhu JK, Hasegawa PM, Bohnert HJ, Shi H, Yun DJ, et al. 2008. Involvement of Arabidopsis HOS15 in histone deacetylation and cold tolerance. Proceedings of the National Academy of Sciences 105:4945-4950.

Table 4.1. Genes within genomic regions in top 1% of empirical distribution of iHS.

Gene model	AT homolog	Annotated description
Potri.001G043500	AT1G80880.1	Tetratricopeptide repeat (TPR)-like superfamily protein
Potri.001G043600	AT5G15350.1	early nodulin-like protein 17
Potri.001G043700	AT1G80870.1	Protein kinase superfamily protein
Potri.001G082800	AT4G19550.1	zinc ion binding;transcription regulators
Potri.001G082900	AT1G56140.1	Leucine-rich repeat transmembrane protein kinase
Potri.001G180100	AT1G79920.1	Heat shock protein 70 (Hsp 70) family protein
Potri.001G180400	AT1G79960.1	ovate family protein 14
Potri.001G193100	AT1G15370.1	SNARE-like superfamily protein
Potri.001G193200	AT5G20150.1	SPX domain gene 1
Potri.001G193300	AT3G45890.1	Protein of unknown function, DUF647
Potri.001G197900	AT3G13920.1	eukaryotic translation initiation factor 4A1
Potri.001G199200	AT5G20490.1	Myosin family protein with Dil domain
Potri.001G199500	AT1G51730.1	Ubiquitin-conjugating enzyme family protein
Potri.001G220600	AT2G41060.2	RNA-binding (RRM/RBD/RNP motifs) family protein
Potri.001G220700	AT3G56850.1	ABA-responsive element binding protein 3
Potri.001G229300	AT2G44430.1	DNA-binding bromodomain-containing protein
Potri.001G246400	AT5G59010.1	Protein kinase protein with tetratricopeptide repeat domain
Potri.001G246500	AT2G29530.1	Tim10/DDP family zinc finger protein
Potri.001G246600	AT2G29540.2	RNApolymerase 14 kDa subunit
Potri.001G246700	AT5G05830.1	RING/FYVE/PHD zinc finger superfamily protein
Potri.001G261300	AT3G14460.1	LRR and NB-ARC domains-containing disease resistance protein
Potri.001G261400	AT3G14470.1	NB-ARC domain-containing disease resistance protein
Potri.001G261500	AT3G14470.1	NB-ARC domain-containing disease resistance protein
Potri.001G294600	AT3G19610.1	Plant protein of unknown function (DUF936)
Potri.001G294700	AT3G19620.1	Glycosyl hydrolase family protein
Potri.001G305200	AT5G47260.1	ATP binding;GTP binding;nucleotide binding;nucleoside-triphosphatases
Potri.001G307500	AT5G05320.1	FAD/NAD(P)-binding oxidoreductase family protein
Potri.001G307600	AT5G09850.1	Transcription elongation factor (TFIIS) family protein
Potri.001G327600	AT5G14100.1	non-intrinsic ABC protein 14
Potri.001G327700	AT1G30890.1	Integral membrane HRF1 family protein
Potri.001G334700	AT4G31940.1	cytochrome P450, family 82, subfamily C, polypeptide 4
Potri.001G335000	AT3G27260.1	global transcription factor group E8
Potri.001G335100	AT5G14280.1	DNA-binding storekeeper protein-related
Potri.001G349100	AT3G27870.1	ATPase E1-E2 type family protein / haloacid dehalogenase-like hydrolase family protein

Potri.001G356500	AT5G55230.1	microtubule-associated proteins 65-1
Potri.001G357800	AT1G30460.1	cleavage and polyadenylation specificity factor 30
Potri.001G397600	AT3G15220.1	Protein kinase superfamily protein
Potri.001G410800	AT4G27290.1	S-locus lectin protein kinase family protein
Potri.001G419800	AT4G27220.1	NB-ARC domain-containing disease resistance protein
Potri.001G426600	AT4G27220.1	NB-ARC domain-containing disease resistance protein
Potri.001G437300	AT1G78340.1	glutathione S-transferase TAU 22
Potri.001G437400	AT1G78380.1	glutathione S-transferase TAU 19
Potri.001G463200	AT3G26710.1	cofactor assembly of complex C
Potri.001G463300	AT2G34790.1	FAD-binding Berberine family protein
Potri.002G057800	AT2G38020.1	vacuoleless1 (VCL1)
Potri.002G075800	AT2G34300.1	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
Potri.002G075900	AT1G21270.1	wall-associated kinase 2
Potri.002G079900	AT1G21680.1	DPP6 N-terminal domain-like protein
Potri.002G080000	AT1G44110.1	Cyclin A1;1
Potri.002G080300	AT1G77405.1	Pentatricopeptide repeat (PPR) superfamily protein
Potri.002G080400	AT1G15960.1	NRAMP metal ion transporter 6
Potri.002G080500	AT1G80830.1	natural resistance-associated macrophage protein 1
Potri.002G092500	AT1G22150.1	sulfate transporter 1;3
Potri.002G092600	AT1G78010.1	tRNA modification GTPase, putative
Potri.002G094100	AT1G78070.1	Transducin/WD40 repeat-like superfamily protein
Potri.002G108600	AT5G64460.4	Phosphoglycerate mutase family protein
Potri.002G108700	AT3G28670.1	oxidoreductase, zinc-binding dehydrogenase family protein
Potri.002G144500	AT1G49920.1	MuDR family transposase
Potri.002G150000	AT2G45540.1	WD-40 repeat family protein / beige-related
Potri.003G001000	AT1G59750.4	auxin response factor 1
Potri.003G014200	AT5G36930.2	Disease resistance protein (TIR-NBS-LRR class) family
Potri.003G027600	AT2G25470.1	receptor like protein 21
Potri.003G043900	AT3G13930.1	Dihydrolipoamide acetyltransferase, long form protein
Potri.003G045700	AT4G19210.1	RNAse l inhibitor protein 2
Potri.003G045900	AT3G13900.1	ATPase E1-E2 type family protein / haloacid dehalogenase-like hydrolase family protein
Potri.003G051000	AT3G16230.2	Predicted eukaryotic LigT
Potri.003G064100	AT1G17440.2	Transcription initiation factor TFIID subunit A
Potri.003G069300	AT3G14130.1	Aldolase-type TIM barrel family protein
Potri.003G108100	AT1G76520.1	Auxin efflux carrier family protein
Potri.003G110800	AT4G00150.1	GRAS family transcription factor
Potri.003G134100	AT3G18440.1	aluminum-activated malate transporter 9
Potri.003G172100	AT2G23520.1	Pyridoxal phosphate (PLP)-dependent transferases

		superfamily protein
Potri.003G172200	AT3G04280.2	response regulator 22
Potri.003G173200	AT1G67370.1	DNA-binding HORMA family protein
Potri.003G176600	AT3G23940.1	dehydratase family
Potri.003G204600	AT5G13490.2	ADP/ATP carrier 2
Potri.003G204700	AT1G21200.1	sequence-specific DNA binding transcription factors
Potri.003G204800	AT1G63440.1	heavy metal atpase 5
Potri.003G216200	AT1G55210.2	Disease resistance-responsive (dirigent-like protein) family protein
Potri.003G218000	AT5G17230.1	PHYTOENE SYNTHASE
Potri.003G218100	AT1G55280.1	Lipase/lipooxygenase, PLAT/LH2 family protein
Potri.004G068300	AT3G47570.1	Leucine-rich repeat protein kinase family protein
Potri.004G068500	AT5G14940.1	Major facilitator superfamily protein
Potri.004G072400	AT5G24060.1	Pentatricopeptide repeat (PPR) superfamily protein
Potri.004G072500	AT5G52810.1	NAD(P)-binding Rossmann-fold superfamily protein
Potri.004G072800	AT5G37930.1	Protein with RING/U-box and TRAF-like domains
Potri.004G072900	AT5G52780.1	Protein of unknown function (DUF3464)
Potri.004G090400	AT5G03900.2	Iron-sulphur cluster biosynthesis family protein
Potri.004G090500	AT2G36230.1	Aldolase-type TIM barrel family protein
Potri.004G101100	AT3G02280.1	Flavodoxin family protein
Potri.004G101200	AT1G67140.1	HEAT repeat-containing protein
Potri.004G169100	AT4G38710.2	glycine-rich protein
Potri.004G169200	AT2G16630.1	Pollen Ole e 1 allergen and extensin family protein
Potri.004G169300	AT4G34880.1	Amidase family protein
Potri.004G201300	AT2G27280.1	Coiled-coil domain-containing protein 55 (DUF2040)
Potri.004G201400	AT2G27450.1	nitrilase-like protein 1
Potri.004G201500	AT5G22390.1	Protein of unknown function (DUF3049)
Potri.005G051200	AT4G34670.1	Ribosomal protein S3Ae
Potri.005G051300	AT3G04830.1	Protein prenylyltransferase superfamily protein
Potri.005G076700	AT4G39740.1	Thioredoxin superfamily protein
Potri.005G086900	AT1G53050.1	Protein kinase superfamily protein
Potri.005G095800	AT1G80600.1	HOPW1-1-interacting 1
Potri.005G099500	AT2G17650.1	AMP-dependent synthetase and ligase family protein
Potri.005G099600	AT4G35260.1	isocitrate dehydrogenase 1
Potri.005G100000	AT2G17620.1	Cyclin B2;1
Potri.005G100100	AT2G17580.1	Polynucleotide adenylyltransferase family protein
Potri.005G134200	AT5G60200.1	TARGET OF MONOPTEROS 6
Potri.005G151200	AT1G11290.1	Pentatricopeptide repeat (PPR) superfamily protein
Potri.005G153100	AT4G24480.1	Protein kinase superfamily protein
Potri.005G153200	AT1G09700.1	dsRNA-binding domain-like superfamily protein
Potri.005G153600	AT4G24510.1	HXXXD-type acyl-transferase family protein

Potri.005G153800	AT4G24520.1	P450 reductase 1
Potri.005G154000	AT1G58200.2	MSCS-like 3
Potri.005G154200	AT2G25540.1	cellulose synthase 10
Potri.005G155600	AT1G09730.1	Cysteine proteinases superfamily protein
Potri.005G161900	AT4G09010.1	ascorbate peroxidase 4
Potri.005G162000	AT1G11870.3	Seryl-tRNA synthetase
Potri.005G162100	AT1G22450.1	cytochrome C oxidase 6B
Potri.005G191400	AT1G43670.1	Inositol monophosphatase family protein
Potri.005G191500	AT1G21170.1	Exocyst complex component SEC5
Potri.005G191600	AT1G76860.1	Small nuclear ribonucleoprotein family protein
Potri.005G224900	AT3G54670.3	Structural maintenance of chromosomes (SMC) family protein
Potri.005G225000	AT1G06210.1	ENTH/VHS/GAT family protein
Potri.005G225100	AT2G31140.1	Peptidase S24/S26A/S26B/S26C family protein
Potri.005G225200	AT1G06190.1	Rho termination factor
Potri.006G003500	AT5G47760.1	2-phosphoglycolate phosphatase 2
Potri.006G003600	AT1G60870.1	maternal effect embryo arrest 9
Potri.006G003700	AT4G26590.1	oligopeptide transporter 5
Potri.006G010000	AT1G07250.1	UDP-glucosyl transferase 71C4
Potri.006G010100	AT4G22350.2	Ubiquitin C-terminal hydrolases superfamily protein
Potri.006G019400	AT4G16380.1	Heavy metal transport/detoxification superfamily protein
Potri.006G020000	AT4G16380.1	Heavy metal transport/detoxification superfamily protein
Potri.006G022100	AT4G16660.1	heat shock protein 70 (Hsp 70) family protein
Potri.006G022200	AT2G46950.1	cytochrome P450, family 709, subfamily B, polypeptide 2
Potri.006G022300	AT1G22400.1	UDP-Glycosyltransferase superfamily protein
Potri.006G062800	AT1G30270.1	CBL-interacting protein kinase 23
Potri.006G062900	AT3G18670.1	Ankyrin repeat family protein
Potri.006G063000	AT1G15690.1	Inorganic H pyrophosphatase family protein
Potri.006G066800	AT3G15880.1	WUS-interacting protein 2
Potri.006G067200	AT5G20230.1	blue-copper-binding protein
Potri.006G075400	AT2G19260.1	RING/FYVE/PHD zinc finger superfamily protein
Potri.006G075500	AT4G29940.1	pathogenesis related homeodomain protein A
Potri.006G086300	AT2G29050.1	RHOMBOID-like 1
Potri.006G086400	AT3G09300.1	OSBP(oxysterol binding protein)-related protein 3B
Potri.006G095500	AT3G54110.1	plant uncoupling mitochondrial protein 1
Potri.006G095600	AT5G01160.1	RING/U-box superfamily protein
Potri.006G127100	AT5G03280.1	NRAMP metal ion transporter family protein
Potri.006G145700	AT5G20490.1	Myosin family protein with Dil domain
Potri.006G163900	AT5G19680.1	Leucine-rich repeat (LRR) family protein

Potri.006G163900	AT5G19680.1	Leucine-rich repeat (LRR) family protein
Potri.006G174900	AT3G46550.1	Fasciclin-like arabinogalactan family protein
Potri.006G175000	AT2G01050.1	zinc ion binding;nucleic acid binding
Potri.006G174900	AT3G46550.1	Fasciclin-like arabinogalactan family protein
Potri.006G175000	AT2G01050.1	zinc ion binding;nucleic acid binding
Potri.006G197900	AT3G58460.1	RHOMBOID-like protein 15
Potri.006G198000	AT1G10385.1	Vps51/Vps67 family (components of vesicular transport) protein
Potri.006G198100	AT2G38110.1	glycerol-3-phosphate acyltransferase 6
Potri.006G198200	AT3G11930.4	Adenine nucleotide alpha hydrolases-like superfamily protein
Potri.006G240300	AT5G11560.1	catalytics
Potri.006G240300	AT5G11560.1	catalytics
Potri.006G240400	AT5G11550.1	ARM repeat superfamily protein
Potri.006G253100	AT5G11170.1	DEAD/DEAH box RNA helicase family protein
Potri.006G253200	AT3G47650.1	DnaJ/Hsp40 cysteine-rich domain superfamily protein
Potri.006G253300	AT5G16080.1	carboxyesterase 17
Potri.006G253400	AT2G26660.1	SPX domain gene 2
Potri.006G253300	AT5G16080.1	carboxyesterase 17
Potri.006G253400	AT2G26660.1	SPX domain gene 2
Potri.006G254100	AT2G25490.1	EIN3-binding F box protein 1
Potri.006G254200	AT1G30380.1	photosystem I subunit K
Potri.006G277800	AT5G24800.1	basic leucine zipper 9
Potri.006G277900	AT1G29150.1	non-ATPase subunit 9
Potri.007G040100	AT2G15790.1	peptidyl-prolyl cis-trans isomerase / cyclophilin-40 (CYP40) / rotamase
Potri.007G040100	AT2G15790.1	peptidyl-prolyl cis-trans isomerase / cyclophilin-40 (CYP40) / rotamase
Potri.007G061300	AT2G01050.1	zinc ion binding;nucleic acid binding
Potri.007G061300	AT2G01050.1	zinc ion binding;nucleic acid binding
Potri.007G062700	AT2G17220.1	Protein kinase superfamily protein
Potri.007G062700	AT2G17220.1	Protein kinase superfamily protein
Potri.007G070300	AT1G50200.1	Alanyl-tRNA synthetase
Potri.007G074400	AT5G23960.2	terpene synthase 21
Potri.007G102800	AT5G63950.1	chromatin remodeling 24
Potri.007G102900	AT3G22560.1	Acyl-CoA N-acyltransferases (NAT) superfamily protein
Potri.007G103000	AT4G30770.1	Putative membrane lipoprotein
Potri.007G103100	AT2G13440.1	glucose-inhibited division family A protein
Potri.007G102900	AT3G22560.1	Acyl-CoA N-acyltransferases (NAT) superfamily protein
Potri.007G103000	AT4G30770.1	Putative membrane lipoprotein

Potri.007G103100	AT2G13440.1	glucose-inhibited division family A protein
Potri.007G103200	AT5G10830.1	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
Potri.007G103300	AT4G22530.1	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
Potri.008G005000	AT1G78900.2	vacuolar ATP synthase subunit A
Potri.008G005100	AT5G65720.1	nitrogen fixation S (NIFS)-like 1
Potri.008G005200	AT1G50430.1	Ergosterol biosynthesis ERG4/ERG24 family
Potri.008G005300	AT3G17380.1	TRAF-like family protein
Potri.008G005700	AT3G17380.1	TRAF-like family protein
Potri.008G023500	AT3G18670.1	Ankyrin repeat family protein
Potri.008G033900	AT3G47570.1	Leucine-rich repeat protein kinase family protein
Potri.008G034000	AT3G47570.1	Leucine-rich repeat protein kinase family protein
Potri.008G075900	AT3G03300.3	dicer-like 2
Potri.008G150900	AT3G05960.1	sugar transporter 6
Potri.008G151000	AT5G26330.1	Cupredoxin superfamily protein
Potri.008G151100	AT5G26340.1	Major facilitator superfamily protein
Potri.008G160200	AT4G14580.1	CBL-interacting protein kinase 4
Potri.008G160300	AT3G23020.1	Tetratricopeptide repeat (TPR)-like superfamily protein
Potri.008G178400	AT1G24340.1	FAD/NAD(P)-binding oxidoreductase family protein
Potri.008G178500	AT1G24350.3	Acid phosphatase/vanadium-dependent haloperoxidase-related protein
Potri.008G179800	AT3G26000.1	Ribonuclease inhibitor
Potri.008G185500	AT3G26430.1	GDSL-like Lipase/Acylhydrolase superfamily protein
Potri.008G188500	AT1G70520.1	cysteine-rich RLK (RECEPTOR-like protein kinase) 2
Potri.008G188600	AT1G23360.1	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
Potri.008G188700	AT1G23380.1	KNOTTED1-like homeobox gene 6
Potri.010G178100	AT5G48380.1	BAK1-interacting receptor-like kinase 1
Potri.010G178200	AT5G16290.1	VALINE-TOLERANT 1
Potri.010G187700	AT5G05570.2	transducin family protein / WD-40 repeat family protein
Potri.010G190700	AT5G05740.2	ethylene-dependent gravitropism-deficient and yellow-green-like 2
Potri.010G190800	AT1G04050.1	homolog of SU(var)3-9 1
Potri.010G190700	AT5G05740.2	ethylene-dependent gravitropism-deficient and yellow-green-like 2
Potri.010G190800	AT1G04050.1	homolog of SU(var)3-9 1
Potri.010G202000	AT3G12380.1	actin-related protein 5
Potri.010G202100	AT5G13840.2	FIZZY-related 3
Potri.010G202200	AT5G13830.1	FtsJ-like methyltransferase family protein
Potri.010G215200	AT5G60100.3	pseudo-response regulator 3

Potri.010G215900	AT5G13510.1	Ribosomal protein L10 family protein
Potri.010G216000	AT3G45740.1	hydrolase family protein / HAD-superfamily protein
Potri.011G003400	AT4G04720.1	calcium-dependent protein kinase 21
Potri.011G003500	AT1G64570.1	Homeodomain-like superfamily protein
Potri.011G015400	AT1G27170.1	transmembrane receptors;ATP binding
Potri.011G019500	AT3G54070.1	Ankyrin repeat family protein
Potri.011G019600	AT3G18670.1	Ankyrin repeat family protein
Potri.011G021300	AT4G13820.1	Leucine-rich repeat (LRR) family protein
Potri.011G021700	AT2G32950.1	Transducin/WD40 repeat-like superfamily protein
Potri.011G021800	AT4G22260.1	Alternative oxidase family protein
Potri.011G036600	AT1G11340.1	S-locus lectin protein kinase family protein
Potri.011G038500	AT1G11330.1	S-locus lectin protein kinase family protein
Potri.011G068900	AT3G58700.1	Ribosomal L5P family protein
Potri.011G069200	AT3G58700.1	Ribosomal L5P family protein
Potri.011G085600	AT5G55580.1	Mitochondrial transcription termination factor family protein
Potri.011G085900	AT5G55580.1	Mitochondrial transcription termination factor family protein
Potri.011G086500	AT5G35700.1	fimbrin-like protein 2
Potri.011G096500	AT3G15050.1	IQ-domain 10
Potri.011G096600	AT3G15030.3	TCP family transcription factor 4
Potri.011G101800	AT1G53520.1	Chalcone-flavanone isomerase family protein
Potri.011G101900	AT1G78530.1	Protein kinase superfamily protein
Potri.011G102000	AT1G78530.1	Protein kinase superfamily protein
Potri.011G161300	AT1G30760.1	FAD-binding Berberine family protein
Potri.011G161400	AT1G30760.1	FAD-binding Berberine family protein
Potri.011G161500	AT2G34790.1	FAD-binding Berberine family protein
Potri.011G161600	AT5G44400.1	FAD-binding Berberine family protein
Potri.011G161700	AT5G44360.1	FAD-binding Berberine family protein
Potri.011G161800	AT5G44360.1	FAD-binding Berberine family protein
Potri.011G161900	AT2G34780.1	maternal effect embryo arrest 22
Potri.012G057400	AT1G18490.1	Protein of unknown function (DUF1637)
Potri.012G060000	AT3G18524.1	MUTS homolog 2
Potri.012G074000	AT1G74850.1	plastid transcriptionally active 2
Potri.012G074200	AT1G74850.1	plastid transcriptionally active 2
Potri.012G075600	AT1G74920.1	aldehyde dehydrogenase 10A8
Potri.012G077700	AT3G48050.1	BAH domain ;TFIIS helical bundle-like domain
Potri.012G077800	AT3G48040.1	RHO-related protein from plants 10
Potri.012G105800	AT5G23720.1	dual specificity protein phosphatase family protein
Potri.012G105900	AT5G50920.1	CLPC homologue 1
Potri.012G115100	AT5G65260.1	RNA-binding (RRM/RBD/RNP motifs) family protein
Potri.012G123500	AT3G14470.1	NB-ARC domain-containing disease resistance protein

Potri.013G038300	AT3G22780.1	Tesmin/TSO1-like CXC domain-containing protein
Potri.013G038400	AT3G22910.1	ATPase E1-E2 type family protein / haloacid dehalogenase-like hydrolase family protein
Potri.013G065800	AT5G17860.1	calcium exchanger 7
Potri.013G065900	AT5G17850.1	Sodium/calcium exchanger family protein
Potri.013G086800	AT3G02830.1	zinc finger protein 1
Potri.013G087000	AT5G16530.1	Auxin efflux carrier family protein
Potri.013G094400	AT1G33470.2	RNA-binding (RRM/RBD/RNP motifs) family protein
Potri.013G096400	AT2G19130.1	S-locus lectin protein kinase family protein
Potri.013G096500	AT5G60900.1	receptor-like protein kinase 1
Potri.013G096400	AT2G19130.1	S-locus lectin protein kinase family protein
Potri.013G096500	AT5G60900.1	receptor-like protein kinase 1
Potri.013G096600	AT1G71350.1	eukaryotic translation initiation factor SUI1 family protein
Potri.013G099500	AT4G19710.2	aspartate kinase-homoserine dehydrogenase ii
Potri.013G099700	AT1G31480.1	shoot gravitropism 2 (SGR2)
Potri.013G132000	AT5G56170.1	LORELEI-LIKE-GPI-ANCHORED PROTEIN 1
Potri.013G132100	AT1G03687.1	DTW domain-containing protein
Potri.013G132200	AT4G03520.1	Thioredoxin superfamily protein
Potri.013G132400	AT1G47710.1	Serine protease inhibitor (SERPIN) family protein
Potri.013G132500	AT2G20710.1	Tetratricopeptide repeat (TPR)-like superfamily protein
Potri.013G133700	AT1G03670.1	ankyrin repeat family protein
Potri.014G001400	AT1G78955.1	camelliol C synthase 1
Potri.014G012000	AT3G14470.1	NB-ARC domain-containing disease resistance protein
Potri.014G054600	AT4G13970.1	zinc ion binding
Potri.014G054700	AT3G60460.1	myb-like HTH transcriptional regulator family protein
Potri.014G074500	AT4G00230.1	xylem serine peptidase 1
Potri.014G074600	AT1G01900.1	subtilase family protein
Potri.014G106400	AT2G46820.2	photosystem I P subunit
Potri.014G106500	AT5G54160.1	O-methyltransferase 1
Potri.014G183200	AT3G46940.1	DUTP-PYROPHOSPHATASE-LIKE 1
Potri.014G183600	AT2G30920.1	coenzyme Q 3
Potri.014G183500	AT2G16230.1	O-Glycosyl hydrolases family 17 protein
Potri.014G183700	AT2G30920.1	coenzyme Q 3
Potri.014G184200	AT2G16230.1	O-Glycosyl hydrolases family 17 protein
Potri.014G185100	AT2G16230.1	O-Glycosyl hydrolases family 17 protein
Potri.014G185200	AT3G19830.2	Calcium-dependent lipid-binding (CaLB domain) family protein
Potri.014G190200	AT1G74260.1	purine biosynthesis 4
Potri.014G190300	AT5G55160.1	small ubiquitin-like modifier 2
Potri.014G190400	AT5G33406.1	hAT dimerisation domain-containing protein /

		transposase-related
Potri.014G190500	AT3G17450.1	hAT dimerisation domain-containing protein
Potri.014G190600	AT3G07140.1	GPI transamidase component Gpi16 subunit family protein
Potri.014G190800	AT3G07170.1	Sterile alpha motif (SAM) domain-containing protein
Potri.014G191200	AT5G48910.1	Pentatricopeptide repeat (PPR) superfamily protein
Potri.015G012400	AT5G24300.2	Glycogen/starch synthases, ADP-glucose type
Potri.015G012500	AT3G49260.2	IQ-domain 21
Potri.015G016900	AT5G53450.1	OBP3-responsive gene 1
Potri.015G017000	AT4G01500.1	AP2/B3-like transcriptional factor family protein
Potri.015G017100	AT5G53450.1	OBP3-responsive gene 1
Potri.015G037000	AT1G18270.1	ketose-bisphosphate aldolase class-II family protein
Potri.015G047500	AT1G73990.1	signal peptide peptidase
Potri.015G047600	AT1G18485.1	Pentatricopeptide repeat (PPR) superfamily protein
Potri.015G051500	AT1G23710.1	Protein of unknown function (DUF1645)
Potri.015G055600	AT1G18490.1	Protein of unknown function (DUF1637)
Potri.015G055700	AT2G23810.1	tetraspanin8
Potri.015G055800	AT1G74030.1	enolase 1
Potri.015G089000	AT3G48430.1	relative of early flowering 6
Potri.015G089100	AT3G48440.1	Zinc finger C-x8-C-x5-C-x3-H type family protein
Potri.015G103300	AT5G50860.1	Protein kinase superfamily protein
Potri.015G107000	AT1G27060.1	Regulator of chromosome condensation (RCC1) family protein
Potri.015G117100	AT5G14345.1	early nodulin-like protein 21
Potri.015G117200	AT5G45110.1	NPR1-like protein 3
Potri.015G119400	AT1G03670.1	ankyrin repeat family protein
Potri.015G119600	AT3G18670.1	Ankyrin repeat family protein
Potri.015G119700	AT5G04690.1	Ankyrin repeat family protein
Potri.016G011400	AT1G53440.1	Leucine-rich repeat transmembrane protein kinase
Potri.016G016600	AT3G21790.1	UDP-Glycosyltransferase superfamily protein
Potri.016G016700	AT3G21750.1	UDP-glucosyl transferase 71B1
Potri.016G016800	AT3G21760.1	UDP-Glycosyltransferase superfamily protein
Potri.016G024900	AT3G56950.2	small and basic intrinsic protein 2;1
Potri.016G025000	AT3G56940.1	dicarboxylate diiron protein, putative (Crd1)
Potri.016G025100	AT3G56930.1	DHHC-type zinc finger family protein
Potri.016G025200	AT5G05070.1	DHHC-type zinc finger family protein
Potri.016G025500	AT3G10740.1	alpha-L-arabinofuranosidase 1
Potri.016G045000	AT5G06740.1	Concanavalin A-like lectin protein kinase family protein
Potri.016G045100	AT1G71840.1	transducin family protein / WD-40 repeat family protein
Potri.016G045200	AT3G12100.1	Cation efflux family protein

Potri.016G045300	AT2G42070.1	nudix hydrolase homolog 23
Potri.016G045400	AT2G42080.1	Chaperone DnaJ-domain superfamily protein
Potri.016G062100	AT3G58490.1	Phosphatidic acid phosphatase (PAP2) family protein
Potri.016G062400	AT5G06610.1	Protein of unknown function (DUF620)
Potri.016G062500	AT1G55310.1	SC35-like splicing factor 33
Potri.016G095700	AT5G03730.2	Protein kinase superfamily protein
Potri.016G120500	AT1G45616.1	receptor like protein 6
Potri.016G120600	AT1G45616.1	receptor like protein 6
Potri.017G006500	AT2G43890.1	Pectin lyase-like superfamily protein
Potri.017G013500	AT5G44240.2	aminophospholipid ATPase 2
Potri.017G034400	AT1G61190.1	LRR and NB-ARC domains-containing disease resistance protein
Potri.017G034500	AT1G66980.1	suppressor of npr1-1 constitutive 4
Potri.017G034600	AT1G18390.2	Protein kinase superfamily protein
Potri.017G057700	AT4G13780.1	methionine--tRNA ligase, putative / methionyl-tRNA synthetase, putative / MetRS, putative
Potri.017G062200	AT5G41130.1	Esterase/lipase/thioesterase family protein
Potri.017G069600	AT5G40870.1	uridine kinase/uracil phosphoribosyltransferase 1
Potri.017G078300	AT2G01420.2	Auxin efflux carrier family protein
Potri.017G078400	AT5G15110.1	Pectate lyase family protein
Potri.017G088700	AT3G29200.1	chorismate mutase 1
Potri.017G089600	AT5G15400.1	U-box domain-containing protein
Potri.017G116400	AT5G16150.1	plastidic GLC translocator

Table 4.2. Overlapping genes within genomic regions in top 1% of empirical distribution of F_{ST} across latitude and altitude transect.

Gene model	AT homolog	Annotated description
Potri.001G324300	AT5G64670.1	Ribosomal protein L18e/L15 superfamily protein
Potri.001G324400	AT5G41130.1	Esterase/lipase/thioesterase family protein
Potri.001G410400	AT5G54510.1	DFL1 (DWARF IN LIGHT 1)
Potri.002G068200	AT1G43640.1	TLP5 (tubby like protein 5)
Potri.003G079300	AT2G44940.1	Integrase-type DNA-binding superfamily protein
Potri.005G060000	AT1G77120.1	ADH1 (alcohol dehydrogenase 1)
Potri.006G156200	AT2G20020.1	RNA-binding CRS1 / YhbY (CRM) domain-containing protein
Potri.006G164600	AT4G26080.1	ABI1 (ABA INSENSITIVE 1)
Potri.011G107800	AT3G14860.2	NHL domain-containing protein
Potri.012G035400	AT1G64110.3	P-loop containing nucleoside triphosphate hydrolases superfamily protein
Potri.013G136500	ATCG00830.1	ribosomal protein L2
Potri.013G136600	ATCG00810.1	ribosomal protein L22
Potri.013G136700	ATCG00720.1	photosynthetic electron transfer B
Potri.013G136800	ATCG00730.1	photosynthetic electron transfer D
Potri.013G136900	ATCG00710.1	photosystem II reaction center protein H
Potri.013G137000	ATCG00700.1	photosystem II reaction center protein N
Potri.013G137100	ATCG00680.1	photosystem II reaction center protein B
Potri.013G139300	ATCG01010.1	NADH-Ubiquinone oxidoreductase (complex I), chain 5 protein
Potri.013G139400	ATCG01040.1	Cytochrome C assembly protein
Potri.013G139600	ATCG01110.1	NAD(P)H dehydrogenase subunit H
Potri.013G139700	ATCG01130.1	Ycf1 protein
Potri.013G140300	ATCG00810.1	ribosomal protein L22
Potri.013G140400	ATCG00800.1	structural constituent of ribosome
Potri.013G140500	ATCG00790.1	ribosomal protein L16
Potri.013G140600	ATCG00780.1	ribosomal protein L14
Potri.013G140700	ATCG00770.1	ribosomal protein S8
Potri.013G140800	ATCG00740.1	RNA polymerase subunit alpha
Potri.013G142100	ATCG00220.1	photosystem II reaction center protein M
Potri.013G142200	ATCG00190.1	RNA polymerase subunit beta
Potri.013G142300	ATCG00180.1	DNA-directed RNA polymerase family protein
Potri.013G142400	ATCG00170.1	DNA-directed RNA polymerase family protein
Potri.013G142600	ATCG00160.1	ribosomal protein S2
Potri.013G142700	ATCG00140.1	ATP synthase subunit C family protein
Potri.013G142800	ATCG00120.1	ATP synthase subunit alpha
Potri.014G037200	AT5G42630.1	KAN4 (Homeodomain-like superfamily protein)

Potri.014G179300	AT5G46570.1	BSK2 (BRASSINOSTEROID SIGNALING KINASE 2)
Potri.015G022300	AT5G53210.1	SPCH (SPEECHLESS)
Potri.016G013400	AT4G22360.1	SWIB complex BAF60b domain-containing protein
Potri.016G083400	AT5G09360.1	LAC14 (laccase 14)
Potri.016G088500	AT5G03150.1	JKD (C2H2-like zinc finger protein)
Potri.016G089000	AT3G09770.1	RING/U-box superfamily protein
Potri.017G010000	AT3G26040.1	HXXXD-type acyl-transferase family protein

Table 4.3. Overlapping genes within genomic regions in top 1% of Bayesian Factor associated with MAT across latitude and altitude transects.

Gene model	AT homolog	Annotated description
Potri.001G046100	AT1G15950.1	CCR1 (cinnamoyl coa reductase 1)
Potri.001G302900	AT2G15490.3	UGT73B4 (UDP glycosyltransferase 73B4)
Potri.001G359900	AT5G55390.1	EDM2 (ENHANCED DOWNY MILDEW 2)
Potri.002G107400	AT1G27595.1	TANG1
Potri.003G057400	AT1G15550.1	GA3OX1 (gibberellin 3-oxidase 1)
Potri.004G093800	AT5G39080.1	HXXXD-type acyl-transferase family protein
Potri.006G156200	AT2G20020.1	CAF1 (CRM domain-containing protein)
Potri.011G018200	AT5G35810.1	Ankyrin repeat family protein
Potri.012G045900	AT3G17940.1	Galactose mutarotase-like superfamily protein
Potri.013G136500	ATCG00830.1	ribosomal protein L2
Potri.013G136900	ATCG00710.1	photosystem II reaction center protein H
Potri.013G139600	ATCG01110.1	NAD(P)H dehydrogenase subunit H
Potri.013G140300	ATCG00810.1	ribosomal protein L22
Potri.013G140400	ATCG00800.1	structural constituent of ribosome
Potri.013G140500	ATCG00790.1	ribosomal protein L16
Potri.013G140600	ATCG00780.1	ribosomal protein L14
Potri.013G140700	ATCG00770.1	ribosomal protein S8
Potri.013G140800	ATCG00740.1	RNA polymerase subunit alpha
Potri.013G142300	ATCG00180.1	DNA-directed RNA polymerase family protein
Potri.013G142400	ATCG00170.1	DNA-directed RNA polymerase family protein
Potri.013G142600	ATCG00160.1	ribosomal protein S2
Potri.013G142700	ATCG00140.1	ATP synthase subunit C family protein
Potri.013G142800	ATCG00120.1	ATP synthase subunit alpha
Potri.014G171900	AT4G12640.1	RNA recognition motif (RRM)-containing protein
Potri.015G008300	AT1G55580.1	LAS (LATERAL SUPPRESSOR)
Potri.017G107600	AT1G13570.1	F-box/RNI-like superfamily protein
Potri.017G107800	AT1G33390.1	FAS4 (FASCIATED STEM 4)
Potri.017G107900	AT1G33390.1	FAS4 (FASCIATED STEM 4)

Table 4.4. Overlapping candidate genes under selection across latitude and altitude transects (Candidate genes were defined as top 1% outliers in at least two selection scans).

Gene model	AT homolog	Annotated description
Potri.001G015400	AT3G45140.1	LOX2 (lipoxygenase 20)
Potri.001G046100	AT1G15950.1	CCR1 (cinnamoyl coa reductase 1)
Potri.001G060500	AT5G13140.1	Pollen Ole e 1 allergen and extensin family protein
Potri.001G060600	AT5G14070.1	ROXY2 (Thioredoxin superfamily protein)
Potri.001G088000	AT5G63610.1	CDKE (cyclin-dependent kinase E)
Potri.001G116000	AT1G12400.1	Nucleotide excision repair, TFIIH, subunit TTDA
Potri.001G124200	AT3G47420.1	PS3 (phosphate starvation-induced gene 3)
Potri.001G169100	AT3G13960.1	GRF5 (growth-regulating factor 5)
Potri.001G201000	AT1G72960.1	Root hair defective 3 GTP-binding protein (RHD3)
Potri.001G302900	AT2G15490.3	UGT73B4 (UDP-glycosyltransferase 73B4)
Potri.001G321800	AT3G23790.1	AAE16 (AMP-dependent synthetase and ligase family protein)
Potri.001G340700	AT1G59970.1	Matrixin family protein
Potri.001G344000	AT5G40530.1	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
Potri.001G359300	AT3G20800.1	Cell differentiation, Rcd1-like protein
Potri.001G359400	AT1G04950.1	TAF6 (TATA BOX ASSOCIATED FACTOR II 59)
Potri.001G359900	AT5G55390.1	EDM2 (ENHANCED DOWNY MILDEW 2)
Potri.001G360500	AT5G59240.1	Ribosomal protein S8e family protein
Potri.001G383800	AT3G42170.1	BED zinc finger ;hAT family dimerisation domain
Potri.001G395200	AT2G20562.1	
Potri.001G395700	AT5G39080.1	HXXXD-type acyl-transferase family protein
Potri.001G398500	AT5G53890.1	PSKR2 (phytosylfokine-alpha receptor 2)
Potri.001G409300	AT4G21380.1	RK3 (receptor kinase 3)
Potri.001G436600	AT1G78380.1	GSTU19 (glutathione S-transferase TAU 19)
Potri.001G453800	AT5G47610.1	RING/U-box superfamily protein
Potri.001G460600	AT1G30650.1	WRKY14 (WRKY DNA-binding protein 14)
Potri.001G464700	AT5G44440.1	FAD-binding Berberine family protein
Potri.001G467800	AT1G55810.3	UKL3 (uridine kinase-like 3)
Potri.001G468100	AT4G26530.1	Aldolase superfamily protein
Potri.002G025400	AT3G26320.1	cytochrome P450, family 71, subfamily B, polypeptide 36
Potri.002G068200	AT1G43640.1	TLP5 (tubby like protein 5)
Potri.002G076800	AT5G54160.1	OMT1 (O-methyltransferase 1)
Potri.002G083500	AT1G44575.1	Chlorophyll A-B binding family protein
Potri.002G098000	AT1G78240.2	QUA2 (QUASIMODO2)
Potri.002G105300	AT3G56770.1	basic helix-loop-helix (bHLH) DNA-binding superfamily protein

Potri.002G106400	AT1G58210.1	EMB1674 (EMBRYO DEFECTIVE 1674)
Potri.002G106500	AT1G09710.2	Homeodomain-like superfamily protein
Potri.002G114700	AT1G10120.1	basic helix-loop-helix DNA-binding superfamily protein
Potri.002G143200	AT5G07670.1	RNI-like superfamily protein
Potri.002G149300	AT2G45430.1	AHL22 (AT-hook motif nuclear-localized protein 22)
Potri.002G160100	AT1G01690.1	PRD3 (putative recombination initiation defects 3)
Potri.002G245400	AT3G25070.1	RIN4 (RPM1 interacting protein 4)
Potri.002G245500	AT5G48655.1	RING/U-box superfamily protein
Potri.003G042200	AT2G26190.1	calmodulin-binding family protein
Potri.003G052400	AT1G79840.1	GL2 (GLABRA 2)
Potri.003G057400	AT1G15550.1	GA3OX1 (gibberellin 3-oxidase 1)
Potri.003G088800	AT1G32450.1	NRT1.5 (nitrate transporter 1.5)
Potri.003G096100	AT5G46910.1	Transcription factor jumonji (jmj) family protein / zinc finger (C5HC2 type) family protein
Potri.003G108400	AT5G62410.1	SMC2 (structural maintenance of chromosomes 2)
Potri.003G211100	AT5G44640.1	BGLU13 (beta glucosidase 13)
Potri.003G214200	AT5G13000.1	GSL12 (glucan synthase-like 12)
Potri.003G214600	AT5G06730.1	Peroxidase superfamily protein
Potri.003G214700	AT5G06720.1	PA2 (peroxidase 2)
Potri.004G008400	AT4G14305.1	Peroxisomal membrane 22 kDa (Mpv17/PMP22) family protein
Potri.004G018200	AT4G04860.1	DER2.2 (DERLIN-2.2)
Potri.004G051500	AT5G28010.1	Polyketide cyclase/dehydrase and lipid transport superfamily protein
Potri.004G051700	AT4G17500.1	ERF1 (ethylene responsive element binding factor 1)
Potri.004G051800	AT4G18450.1	Integrase-type DNA-binding superfamily protein
Potri.004G051900	AT5G45950.1	GDSL-like Lipase/Acylhydrolase superfamily protein
Potri.004G052300	AT2G34040.1	Apoptosis inhibitory protein 5 (API5)
Potri.004G052400	AT2G29120.1	GLR2.7 (glutamate receptor 2.7)
Potri.004G093800	AT5G39080.1	HXXXD-type acyl-transferase family protein
Potri.004G128100	ATCG01280.1	Chloroplast Ycf2
Potri.004G157100	AT4G39210.1	APL3
Potri.004G183600	AT4G38380.1	MATE efflux family protein
Potri.004G183700	AT2G32730.1	26S proteasome regulatory complex, non-ATPase subcomplex, Rpn2/Psmd1 subunit
Potri.004G209700	AT5G58140.2	PHOT2 (phototropin 2)
Potri.004G209800	AT4G38130.1	HD1 (histone deacetylase 1)
Potri.005G015600	AT1G18610.1	Galactose oxidase/kelch repeat superfamily protein
Potri.005G018800	AT1G56340.2	CRT1 (calreticulin 1)
Potri.005G018900	AT3G62550.1	Adenine nucleotide alpha hydrolases-like superfamily protein
Potri.005G019400	AT1G09190.1	Tetratricopeptide repeat (TPR)-like superfamily protein

Potri.005G019500	AT5G27030.2	TPR3 (TOPLESS-related 3)
Potri.005G060600	AT3G24000.1	Tetratricopeptide repeat (TPR)-like superfamily protein
Potri.005G060700	AT5G48920.1	TED7 (tracheary element differentiation-related 7)
Potri.005G060800	AT1G48880.1	TBL7 (TRICHOME BIREFRINGENCE-LIKE 7)
Potri.005G070400	AT3G05270.2	Plant protein of unknown function (DUF869)
Potri.005G082800	AT4G39640.2	GGT1 (gamma-glutamyl transpeptidase 1)
Potri.005G103800	AT4G35720.1	Arabidopsis protein of unknown function (DUF241)
Potri.005G103900	AT3G51390.1	DHHC-type zinc finger family protein
Potri.005G144400	AT5G67320.1	HOS15 (HIGH EXPRESSION OF OSMOTICALLY RESPONSIVE GENES 15)
Potri.005G150500	AT3G57230.1	AGL16 (AGAMOUS-like 16)
Potri.005G179300	AT1G44170.2	ALDH4 (aldehyde dehydrogenase 3H1)
Potri.005G179400	AT1G11040.1	HSP40/DnaJ peptide-binding protein
Potri.005G189800	AT3G18710.1	PUB29 (plant U-box 29)
Potri.005G207700	AT5G37020.1	ARF8 (auxin response factor 8)
Potri.005G230100	AT1G19640.1	JMT (jasmonic acid carboxyl methyltransferase)
Potri.006G066800	AT3G15880.1	WSIP2 (WUS-interacting protein 2)
Potri.006G109000	AT1G61670.1	Lung seven transmembrane receptor family protein
Potri.006G111800	AT5G01960.1	RING/U-box superfamily protein
Potri.006G126800	AT2G37040.1	PAL1 (PHE ammonia lyase 1)
Potri.006G156100	AT4G28910.2	NINJA (novel interactor of JAZ)
Potri.006G156200	AT2G20020.1	CAF1
Potri.006G162300	AT3G24480.1	Leucine-rich repeat (LRR) family protein
Potri.006G162500	AT5G19590.1	Protein of unknown function, DUF538
Potri.006G204700	AT3G52500.1	Eukaryotic aspartyl protease family protein
Potri.006G240500	AT5G11530.1	EMF1 (embryonic flower 1)
Potri.006G244000	AT3G15605.4	nucleic acid binding
Potri.006G244100	AT5G25590.1	Protein of unknown function (DUF630 and DUF632)
Potri.006G246800	AT5G25520.2	SPOC domain / Transcription elongation factor S-II protein
Potri.006G247000	AT4G32600.1	RING/U-box superfamily protein
Potri.006G248100	AT4G32551.2	LUG
Potri.006G250900	AT4G32420.1	Cyclophilin-like peptidyl-prolyl cis-trans isomerase family protein
Potri.006G255900	AT5G11160.1	APT5
Potri.006G256000	AT4G32160.1	Phox (PX) domain-containing protein
Potri.007G009200	AT5G65700.2	BAM1 (BARELY ANY MERISTEM 1)
Potri.007G017500	AT3G51060.1	SRS1
Potri.007G055300	AT4G37630.2	CYCD5 (CYCLIN D5)
Potri.007G098700	AT5G08740.1	NDC1 (NADH dehydrogenase C1)
Potri.007G098900	AT5G08750.1	RING/FYVE/PHD zinc finger superfamily protein
Potri.008G017100	AT5G04770.1	CAT6 (cationic amino acid transporter 6)

Potri.008G020700	AT5G04740.1	ACT domain-containing protein
Potri.008G020900	AT3G54260.1	TBL36 (TRICHOME BIREFRINGENCE-LIKE 36)
Potri.008G021000	AT2G36290.1	alpha/beta-Hydrolases superfamily protein
Potri.008G030900	AT5G04420.2	Galactose oxidase/kelch repeat superfamily protein
Potri.008G032500	AT5G22930.1	Protein of unknown function (DUF1635)
Potri.008G036500	AT3G10360.1	PUM4 (pumilio 4)
Potri.008G036600	AT2G39120.1	Ubiquitin carboxyl-terminal hydrolase family protein
Potri.008G036700	AT5G04250.1	Cysteine proteinases superfamily protein
Potri.008G056600	AT3G13570.1	SCL30A (SC35-like splicing factor 30A)
Potri.008G069100	AT2G40130.2	Double Clp-N motif-containing P-loop nucleoside triphosphate hydrolases superfamily protein
Potri.008G073400	AT2G40330.1	PYL6 (PYR1-like 6)
Potri.008G089700	AT3G01140.1	MYB106 (myb domain protein 106)
Potri.008G142000	AT1G14600.1	Homeodomain-like superfamily protein
Potri.008G186100	AT1G67900.2	Phototropic-responsive NPH3 family protein
Potri.010G242900	AT3G52060.2	Core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein
Potri.010G243000	AT2G26790.1	Pentatricopeptide repeat (PPR) superfamily protein
Potri.010G246600	AT3G07040.1	RPS3 (RESISTANCE TO PSEUDOMONAS SYRINGAE 3)
Potri.010G246800	AT5G17770.1	CYTOCHROME B5 REDUCTASE 1
Potri.010G251900	AT1G71070.1	Core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein
Potri.011G016000	AT3G18670.1	Ankyrin repeat family protein
Potri.011G016100	AT4G03500.1	Ankyrin repeat family protein
Potri.011G018200	AT5G35810.1	Ankyrin repeat family protein
Potri.011G023400	AT4G22330.1	ATCES1 (Alkaline phytoceramidase)
Potri.011G023500	AT3G17640.1	Leucine-rich repeat (LRR) family protein
Potri.011G028500	AT4G05200.1	CRK25 (cysteine-rich RLK 25)
Potri.011G033300	AT4G21350.1	PUB8 (plant U-box 8)
Potri.011G043500	AT4G21540.1	SPHK1 (sphingosine kinase 1)
Potri.011G053400	AT4G20970.1	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
Potri.011G075100	ATCG00890.1	NADH-Ubiquinone/plastoquinone (complex I) protein
Potri.011G075200	ATCG01240.1	ribosomal protein S7
Potri.011G088000	AT2G26730.1	Leucine-rich repeat protein kinase family protein
Potri.011G097900	AT5G54010.1	UDP-Glycosyltransferase superfamily protein
Potri.011G131100	AT1G50600.1	SCL5 (scarecrow-like 5)
Potri.011G164900	AT5G56020.1	Got1/Sft2-like vescicle transport protein family
Potri.012G024200	AT5G17260.1	NAC086 (NAC domain containing protein 86)
Potri.012G024300	AT2G36170.1	Ubiquitin supergroup;Ribosomal protein L40e
Potri.012G034300	AT5G65770.3	LINC4 (little nuclei4)

Potri.012G034600	AT2G34930.1	disease resistance family protein / LRR family protein
Potri.012G034700	AT2G34790.1	FAD-binding Berberine family protein
Potri.012G045900	AT3G17940.1	Galactose mutarotase-like superfamily protein
Potri.012G052100	AT1G18040.1	CDKD1 (cyclin-dependent kinase D1)
Potri.012G103700	AT3G48770.1	DNA binding;ATP binding
Potri.012G129000	AT5G51570.1	SPFH/Band 7/PHB domain-containing membrane-associated protein family
Potri.013G044100	AT3G04580.2	EIN4
Potri.013G044300	AT1G54650.2	Methyltransferase family protein
Potri.013G056300	AT2G37270.2	RPS5B (ribosomal protein 5B)
Potri.013G056400	AT3G13540.1	MYB5 (myb domain protein 5)
Potri.013G065700	AT3G03710.1	PNP (polyribonucleotide nucleotidyltransferase)
Potri.013G074700	ATCG01110.1	NAD(P)H dehydrogenase subunit H
Potri.013G074800	ATCG01100.1	NADH dehydrogenase family protein
Potri.013G074900	ATCG01100.1	NADH dehydrogenase family protein
Potri.013G075000	ATCG01090.1	NADPH dehydrogenases
Potri.013G116300	AT3G02430.1	Protein of unknown function (DUF679)
Potri.013G116400	AT1G34260.1	FORMS APLOID AND BINUCLEATE CELLS 1A
Potri.013G136200	AT1G03620.1	ELMO/CED-12 family protein
Potri.013G136300	AT4G35160.1	O-methyltransferase family protein
Potri.013G136400	AT4G35160.1	O-methyltransferase family protein
Potri.013G136500	ATCG00830.1	ribosomal protein L2
Potri.013G136600	ATCG00810.1	ribosomal protein L22
Potri.013G136700	ATCG00720.1	photosynthetic electron transfer B
Potri.013G136800	ATCG00730.1	photosynthetic electron transfer D
Potri.013G136900	ATCG00710.1	photosystem II reaction center protein H
Potri.013G137000	ATCG00700.1	photosystem II reaction center protein N
Potri.013G137100	ATCG00680.1	photosystem II reaction center protein B
Potri.013G139300	ATCG01010.1	NADH-Ubiquinone oxidoreductase (complex I), chain 5 protein
Potri.013G139400	ATCG01040.1	Cytochrome C assembly protein
Potri.013G139600	ATCG01110.1	NAD(P)H dehydrogenase subunit H
Potri.013G139700	ATCG01130.1	Ycf1 protein
Potri.013G140300	ATCG00810.1	ribosomal protein L22
Potri.013G140400	ATCG00800.1	structural constituent of ribosome
Potri.013G140500	ATCG00790.1	ribosomal protein L16
Potri.013G140600	ATCG00780.1	ribosomal protein L14
Potri.013G140700	ATCG00770.1	ribosomal protein S8
Potri.013G140800	ATCG00740.1	RNA polymerase subunit alpha
Potri.013G141100	ATCG00500.1	acetyl-CoA carboxylase carboxyl transferase subunit beta
Potri.013G142100	ATCG00220.1	photosystem II reaction center protein M

Potri.013G142200	ATCG00190.1	RNA polymerase subunit beta
Potri.013G142300	ATCG00180.1	DNA-directed RNA polymerase family protein
Potri.013G142400	ATCG00170.1	DNA-directed RNA polymerase family protein
Potri.013G142600	ATCG00160.1	ribosomal protein S2
Potri.013G142700	ATCG00140.1	ATP synthase subunit C family protein
Potri.013G142800	ATCG00120.1	ATP synthase subunit alpha
Potri.013G143000	ATCG00070.1	photosystem II reaction center protein K precursor
Potri.014G010700	AT3G14470.1	NB-ARC domain-containing disease resistance protein
Potri.014G012700	AT1G58470.1	RNA-binding protein 1
Potri.014G027400	AT5G42930.1	alpha/beta-Hydrolases superfamily protein
Potri.014G027500	AT5G42930.1	alpha/beta-Hydrolases superfamily protein
Potri.014G050200	AT1G05490.1	chromatin remodeling 31
Potri.014G103500	AT2G46700.1	CRK3 (CDPK-related kinase 3)
Potri.014G120800	AT3G62420.1	BZIP53 (basic region/leucine zipper motif 53)
Potri.014G135000	AT2G47920.1	Kinase interacting (KIP1-like) family protein
Potri.014G135100	AT2G47930.1	AGP26 (arabinogalactan protein 26)
Potri.014G136100	AT1G03440.1	Leucine-rich repeat (LRR) family protein
Potri.014G170600	AT3G07650.1	COL9 (CONSTANS-like 9)
Potri.014G171900	AT4G12640.1	RNA recognition motif (RRM)-containing protein
Potri.014G191400	AT5G48655.1	RING/U-box superfamily protein
Potri.014G191500	AT5G48660.1	B-cell receptor-associated protein 31-like
Potri.014G191600	AT5G61540.1	N-terminal nucleophile aminohydrolases (Ntn hydrolases) superfamily protein
Potri.015G000300	AT5G24680.1	Peptidase C78, ubiquitin fold modifier-specific peptidase 1/ 2
Potri.015G000400	AT3G49590.2	Autophagy-related protein 13
Potri.015G000500	AT5G24660.1	LSU2 (response to low sulfur 2)
Potri.015G001800	AT4G01575.1	serine protease inhibitor, Kazal-type family protein
Potri.015G002000	AT2G04230.1	FBD, F-box and Leucine Rich Repeat domains containing protein
Potri.015G002100	AT4G03220.1	Protein with RNI-like/FBD-like domains
Potri.015G002200	AT5G24490.1	30S ribosomal protein, putative
Potri.015G002300	AT5G24470.1	PRR5 (PSEUDO-RESPONSE REGULATOR 5)
Potri.015G002800	AT5G24530.1	DMR6
Potri.015G003000	AT5G45140.1	NRPC2 (nuclear RNA polymerase C2)
Potri.015G006000	AT3G53480.1	PDR9 (pleiotropic drug resistance 9)
Potri.015G006200	AT5G53660.1	GRF7 (growth-regulating factor 7)
Potri.015G008300	AT1G55580.1	SCL18 (SCARECROW-LIKE 18)
Potri.015G050800	AT1G73960.1	TBP-associated factor 2
Potri.015G075200	AT1G62400.1	HT1 (high leaf temperature 1)
Potri.015G112200	AT2G44740.1	CYCP4
Potri.015G136600	AT1G16560.2	Per1-like family protein

Potri.015G136700	AT4G25450.1	NAP8
Potri.015G136900	AT1G12580.1	PEPKR1 (phosphoenolpyruvate carboxylase-related kinase 1)
Potri.016G033300	AT5G05140.1	Transcription elongation factor (TFIIS) family protein
Potri.016G036600	AT3G56710.1	SIB1 (sigma factor binding protein 1)
Potri.016G036700	AT3G12600.1	NUDT16 (nudix hydrolase homolog 16)
Potri.016G037400	AT3G56990.1	EDA7 (embryo sac development arrest 7)
Potri.016G037500	AT3G57000.1	nucleolar essential protein-related
Potri.016G053800	AT2G41890.1	curculin-like lectin family protein
Potri.016G083100	AT1G07910.2	RNL (RNAligase)
Potri.016G110800	AT3G09090.1	DEX1 (defective in exine formation protein)
Potri.016G126300	AT3G51740.1	IMK2 (inflorescence meristem receptor-like kinase 2)
Potri.016G126400	AT2G38360.1	PRA1 (prenylated RAB acceptor 1)
Potri.016G126500	AT5G01650.2	Tautomerase/MIF superfamily protein
Potri.016G136200	AT3G10910.1	RING/U-box superfamily protein
Potri.017G000500	AT2G44140.1	Peptidase family C54 protein
Potri.017G000600	AT2G26710.1	CYP734A1 (Cytochrome P450 superfamily protein)
Potri.017G031500	AT1G05910.1	cell division cycle protein 48-related / CDC48-related
Potri.017G042200	AT3G21175.2	ZML1 (ZIM-like 1)
Potri.017G042300	AT1G51600.2	ZML2 (ZIM-LIKE 2)
Potri.017G075400	AT5G40030.1	Protein kinase superfamily protein
Potri.017G090800	AT5G15470.1	GAUT14 (galacturonosyltransferase 14)
Potri.017G101500	AT2G03890.1	GAMMA 7
Potri.017G107600	AT1G13570.1	F-box/RNI-like superfamily protein
Potri.017G107800	AT1G33390.1	FAS4

Table S4.1. PC loadings of 21 annual climate variables across geographical groups.

	Latitudinal groups			Altitudinal groups			Total groups		
	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3
MAT	0.28	-0.04	0.07	-0.26	-0.09	-0.01	0.27	-0.05	0.07
MWMT	0.24	0.13	0.30	-0.23	-0.12	0.27	0.24	0.12	0.28
MCMT	0.26	-0.12	-0.06	-0.25	-0.05	-0.19	0.26	-0.13	-0.03
TD	-0.19	0.23	0.26	0.22	0.00	0.38	-0.19	0.24	0.21
MAP	0.01	-0.38	0.42	0.07	-0.47	-0.05	-0.02	-0.38	0.42
MSP	-0.12	-0.33	0.32	0.13	-0.42	-0.01	-0.13	-0.32	0.29
AHM	0.06	0.38	-0.24	-0.16	0.37	0.05	0.08	0.38	-0.24
SHM	0.17	0.29	-0.02	-0.18	0.35	0.15	0.17	0.30	0.00
DD_0	-0.24	0.15	-0.04	0.26	0.09	0.07	-0.24	0.16	-0.03
DD5	0.27	0.04	0.12	-0.26	-0.10	0.08	0.27	0.03	0.12
DD_18	-0.28	0.05	-0.04	0.26	0.10	0.02	-0.27	0.06	-0.05
DD18	0.23	0.13	0.33	-0.24	-0.04	0.26	0.23	0.13	0.30
NFFD	0.25	-0.17	-0.08	-0.26	-0.07	-0.13	0.25	-0.17	-0.09
bFFP	-0.22	0.19	0.24	0.26	0.08	0.08	-0.22	0.19	0.22
eFFP	0.24	-0.20	-0.12	-0.25	-0.07	-0.19	0.24	-0.20	-0.12
FFP	0.23	-0.20	-0.19	-0.26	-0.08	-0.13	0.23	-0.19	-0.18
PAS	-0.21	-0.14	0.33	0.23	-0.18	0.06	-0.20	-0.12	0.38
EMT	0.25	-0.17	-0.09	-0.26	-0.03	-0.16	0.25	-0.18	-0.08
EXT	0.22	0.20	0.29	-0.13	-0.23	0.49	0.22	0.20	0.32
Eref	0.23	0.20	0.23	-0.15	-0.21	0.47	0.23	0.20	0.27
CMD	0.17	0.33	0.07	-0.11	0.39	0.29	0.17	0.34	0.06

Figure 4.1. Sampling origins for 80 groups across latitude and altitude transects (The yellow dots are 67 latitudinal populations and red triangles are 13 altitudinal populations).

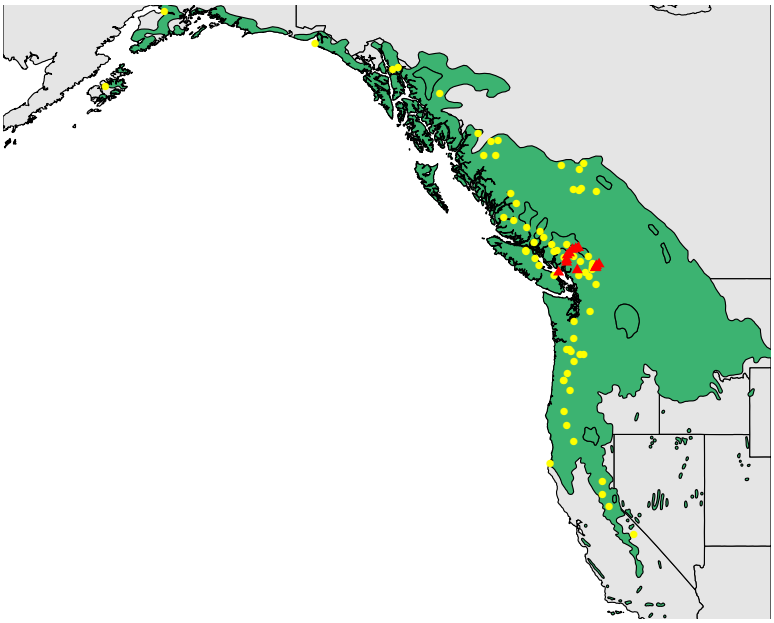


Figure 4.2. Venn diagram of the number of selection regions (A) and genes (B) identified for latitude and altitude transects (The blue circle is for latitude transect and yellow for altitude transect).

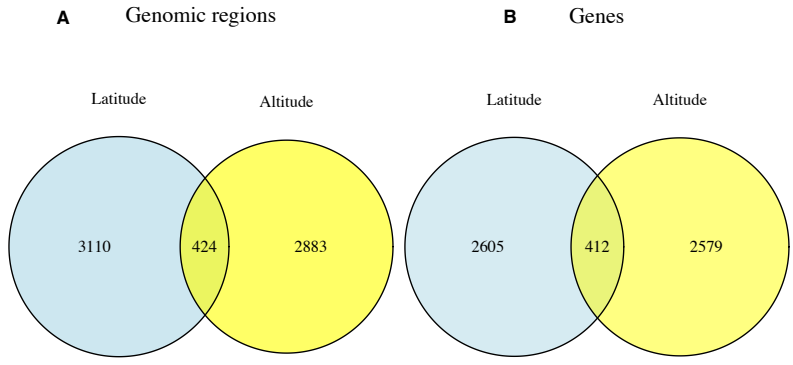


Figure 4.3. Distribution of F_{ST}, Bayesian Factors associated with MAT, and AHM along chromosome 13 (The blue line is loess smoothed line fitted for genomic regions across latitude transect and orange loess smoothed line is fitted for genomic regions across altitude transect).

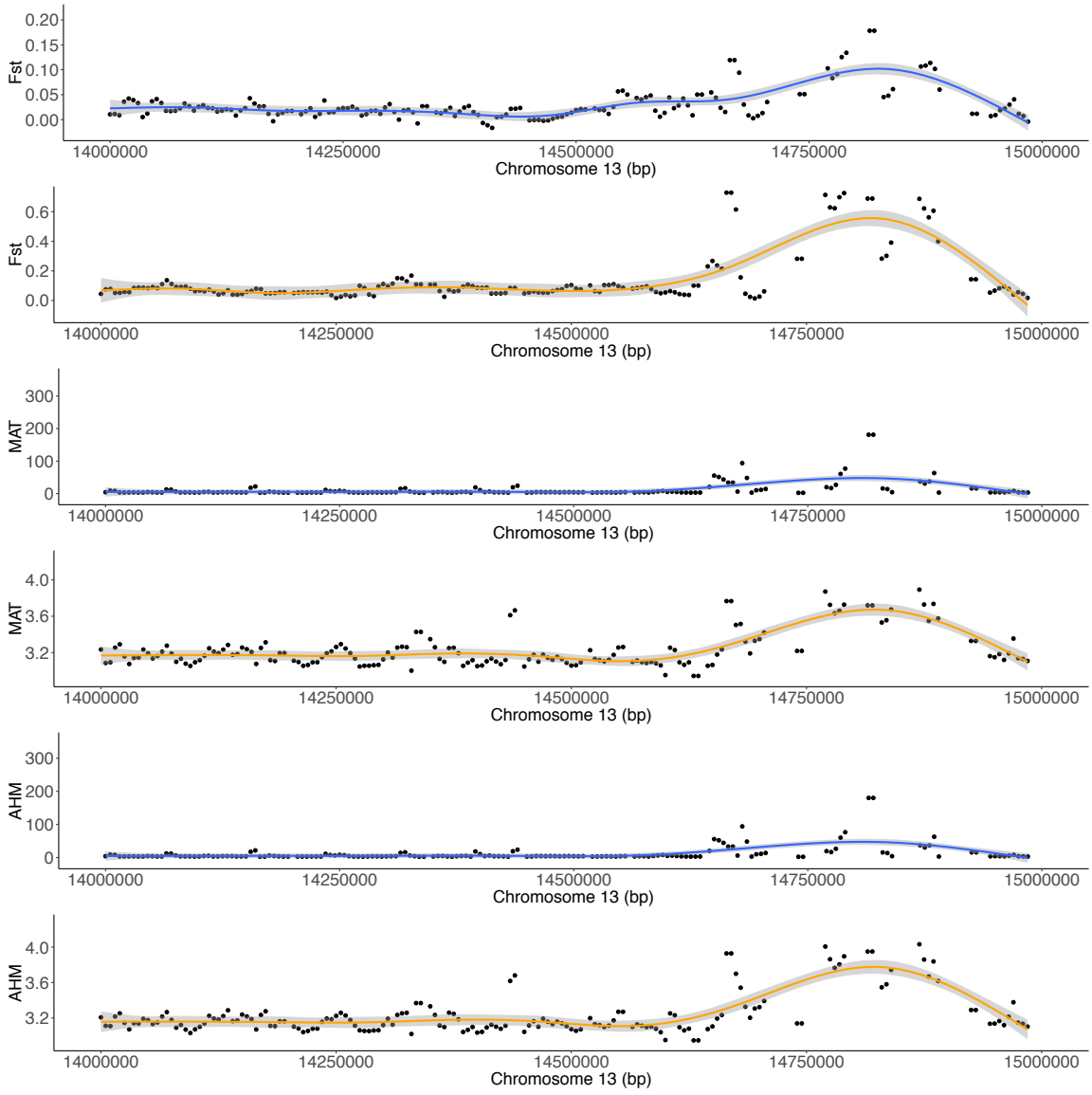


Figure 4.4. Heatmap of biological process enriched across latitude, altitude transects and their overlap (The color scale indicates normalized fold enrichment score for each biological process GO terms).

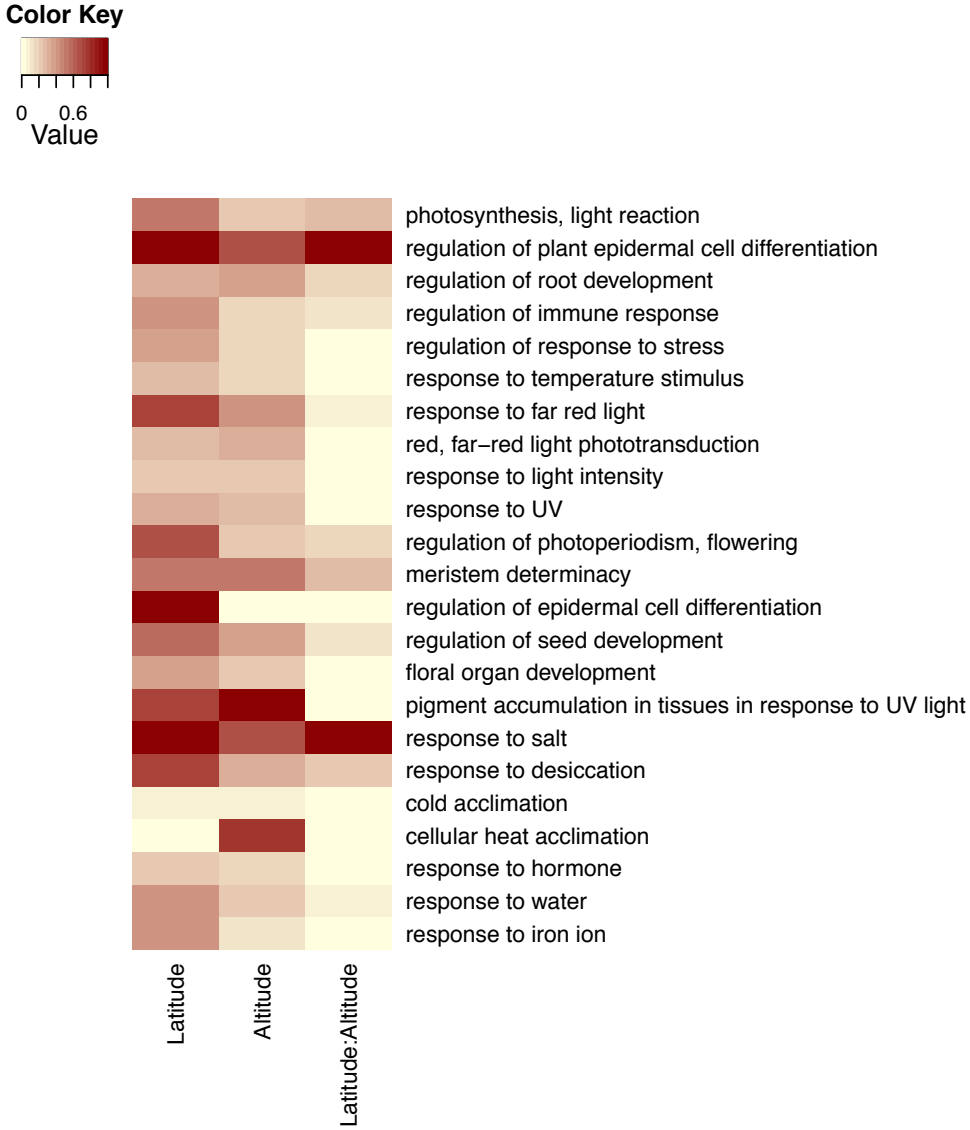


Figure 4.5. Venn diagram of the number of regions identified from three selection scans (The blue circle is for iHS, yellow for F_{ST}, and red for Bayenv).

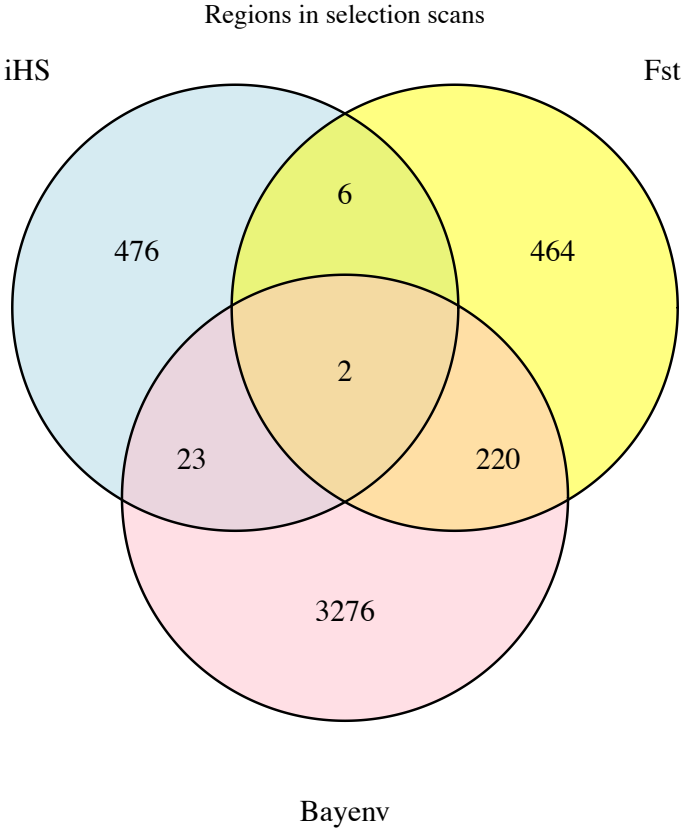


Figure 4.6. Heatmap of biological process for selection outliers identified from three selection scans (The color scale indicates normalized fold enrichment score for each biological process GO terms).

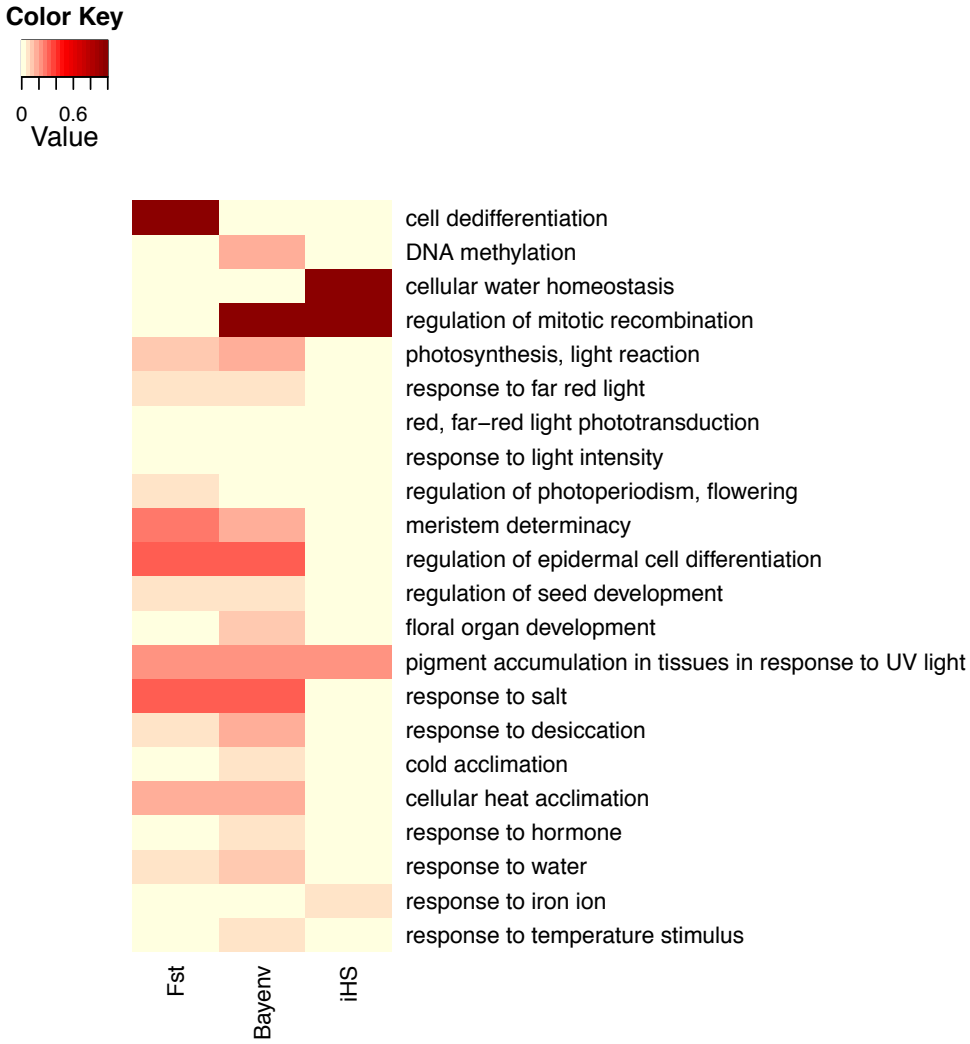


Figure 4.7. The association signal with timing of bud set within candidate selection regions (The distribution of $-\log_{10}(P)$ is from GWAS on timing of bud set in 10-kb windows through the genome (green) and within selection outlier regions (blue) detected with iHS (A), F_{ST} (B), and Bayenv (C). The red lines indicate the mean $-\log_{10}(P)$ of candidate regions).

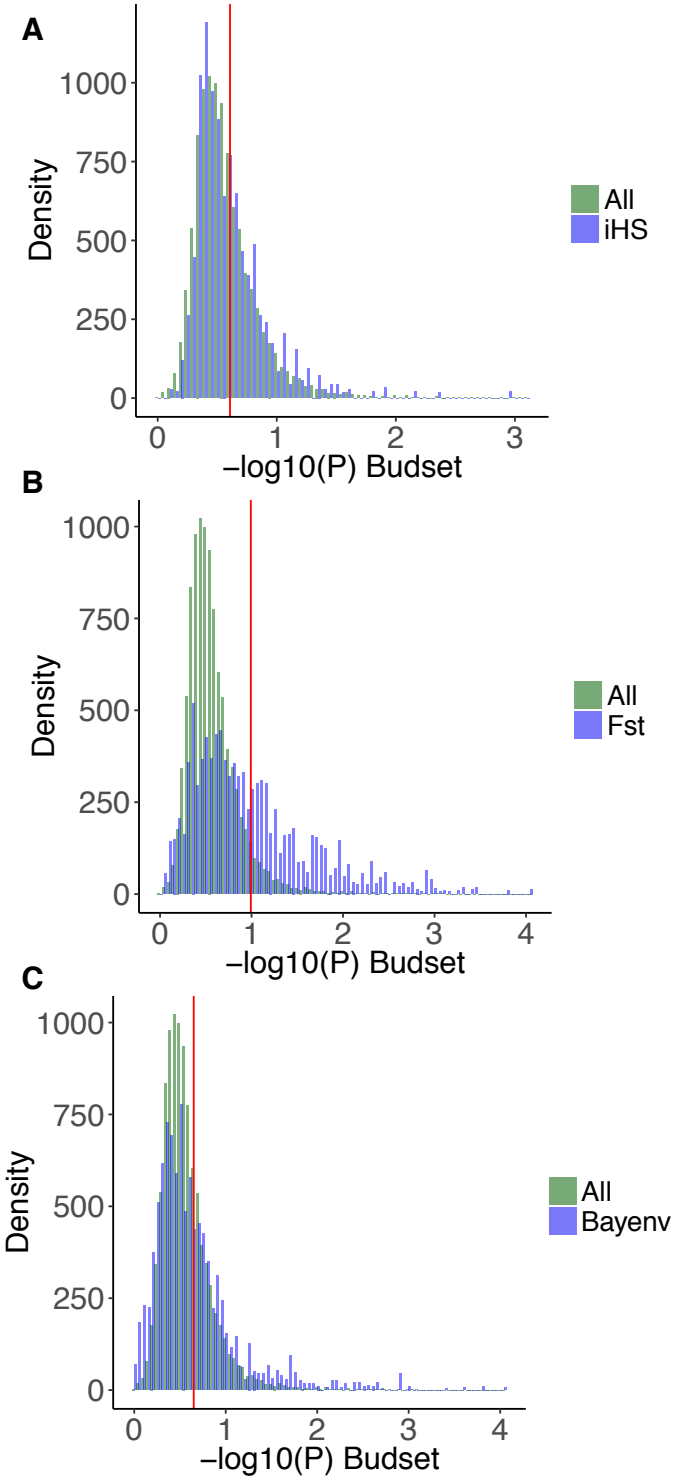


Figure 4.8. Venn diagram of the number of selection outliers and regions associated with timing of bud set from GWAS (The blue circle is for iHS, pink for Bayenv, orchid for GWAS, and yellow for F_{ST}).

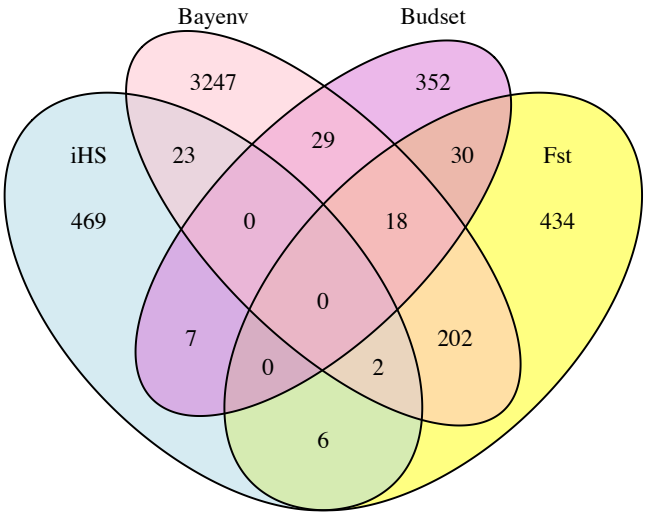


Figure 4.9. A region of chromosome 11 showed abundant number SNP associations with timing of bud flush (A) and high iHS scores (B) (The dark red lines indicate gene models).

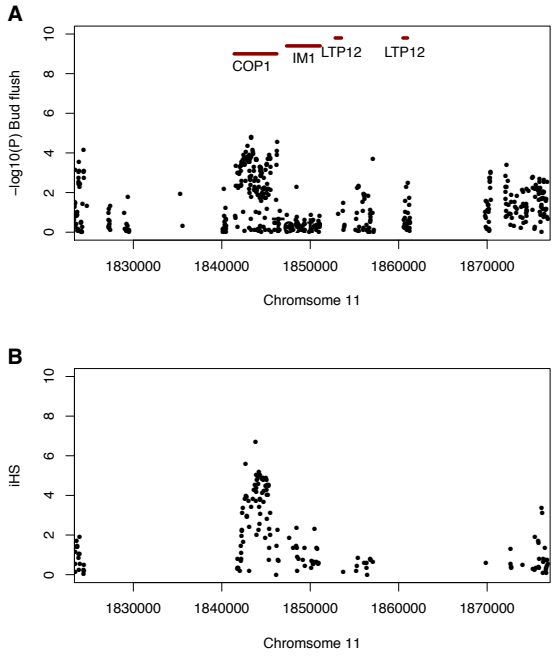


Figure 4.10. A region of chromosome 5 contained SNP associations with timing of bud flush (A), timing of bud set (B), and strong correlation with climate PC1 (C) and PC2 (D) (The dark red lines indicate gene models).

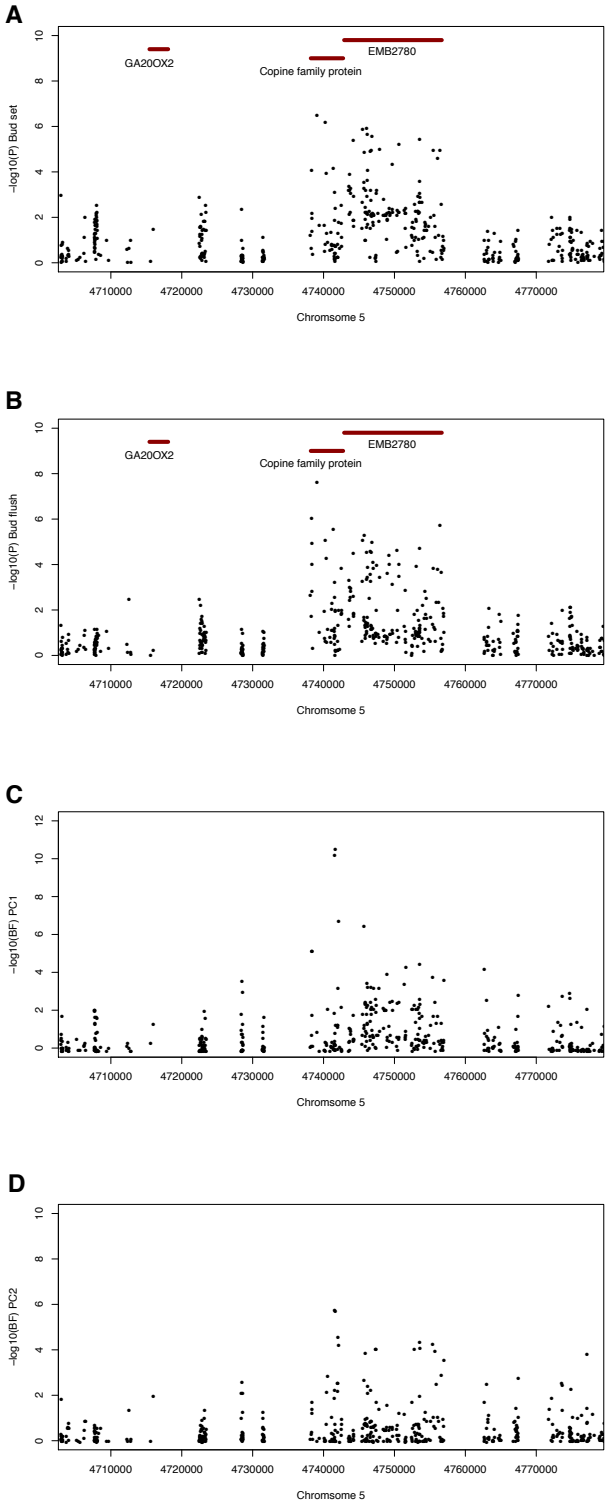


Figure 4.11. A region of chromosome 14 showed elevated F_{ST} (A) and strong correlation with climate variables MAT (B), MCMT (C), and AHM (D) (The dark red lines indicate gene models).

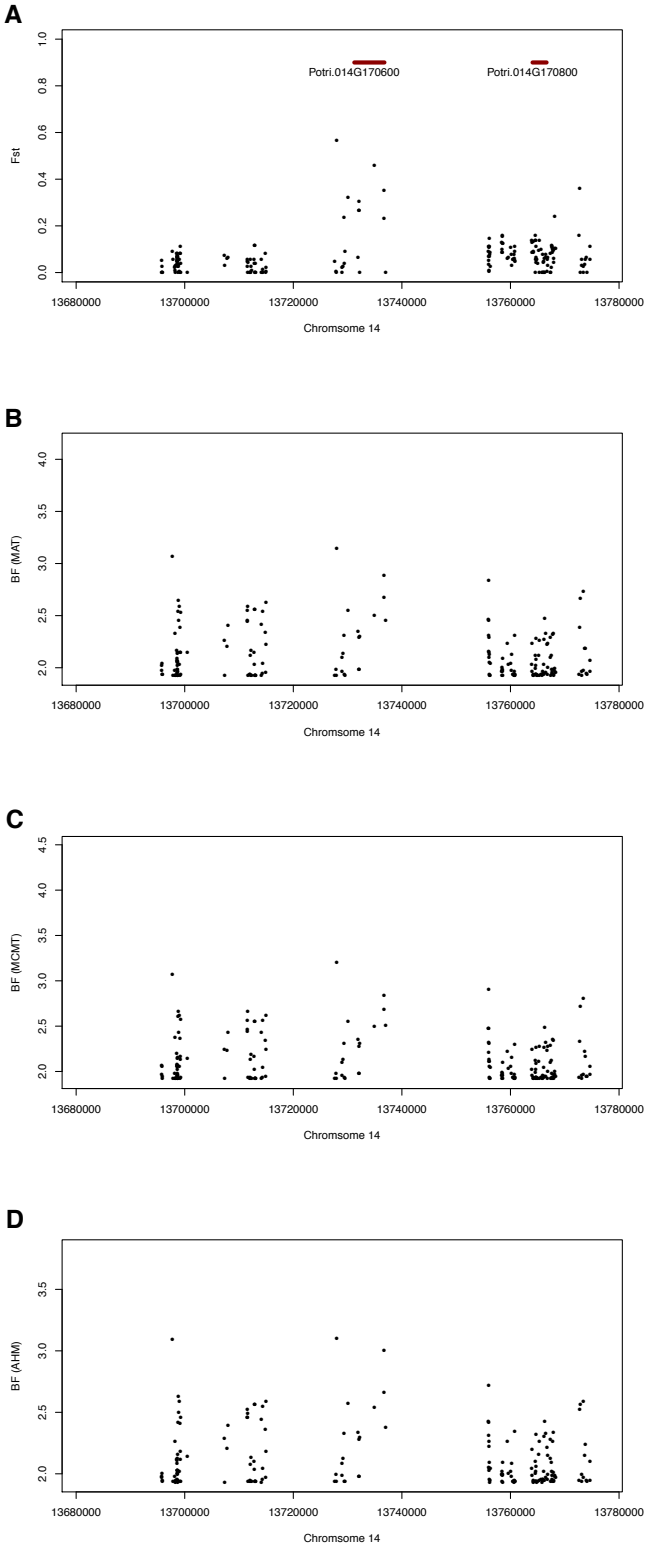


Figure S4.1. Mean annual temperature in relation to latitude across latitude transect (A) and in relation to elevation across altitude transect (B).

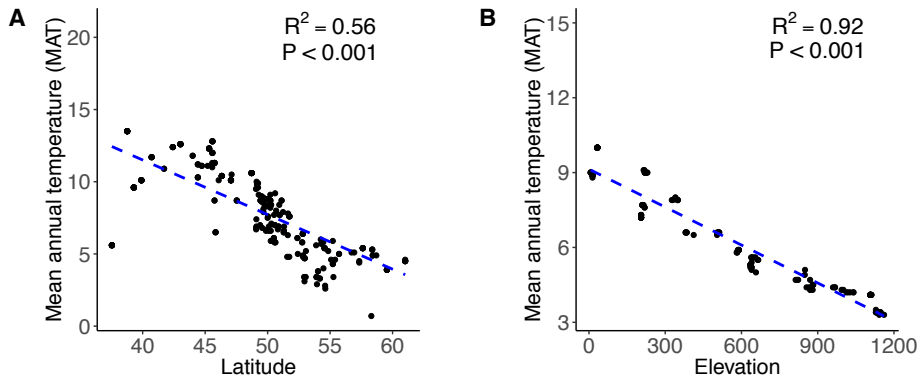


Figure S4.2. Venn diagram of the number of regions in the top 1% F_{ST} across latitude and altitude transects (The yellow circle is for altitude and blue for latitude).

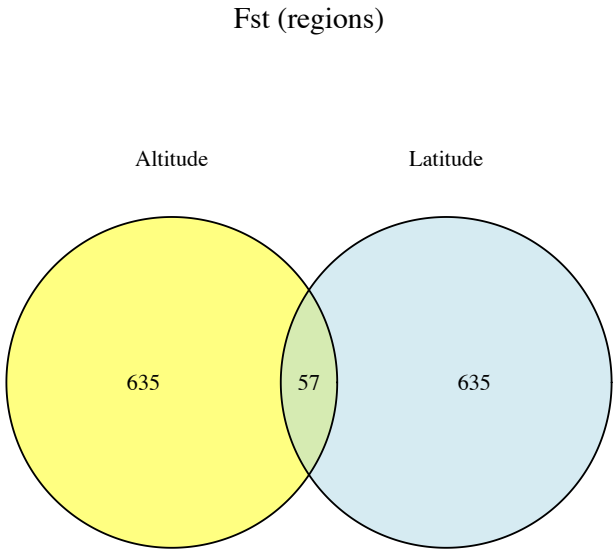


Figure S4.3. Venn diagram of the number of regions in the top 1% Bayesian factor for 8 climate variables including MAT (A), MCMT (B), MAP (C), AHM(D), Eref (E), PC1 (F), PC2 (G), PC3 (H) across latitude and altitude transects (The yellow circle is for altitude and blue for latitude).

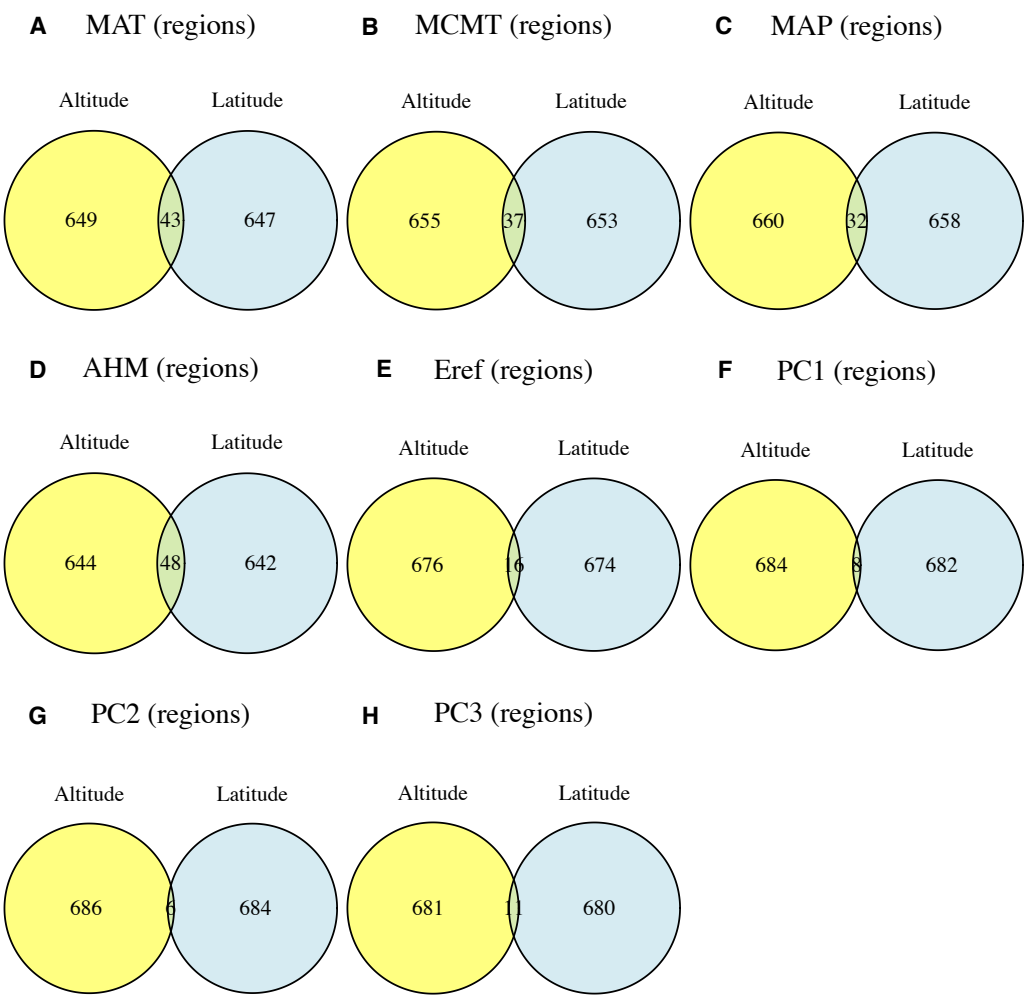


Figure S4.4. Distribution of F_{ST}, Bayesian Factors associated with AHM along chromosome 2 (Blue line is loess smoothed line fitted for genomic regions across latitude transect and orange loess smoothed line is fitted for genomic regions across altitude transect).

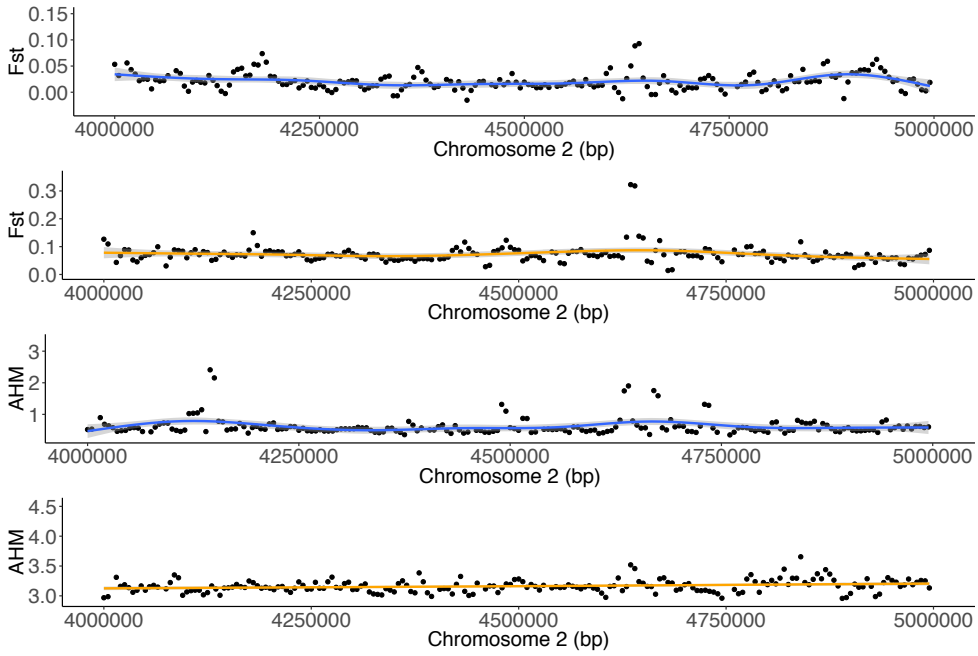


Figure S4.5. The association signal with timing of bud flush within candidate selection regions (The distribution of $-\log_{10}(P)$ is from GWAS on timing of bud flush in 10-kb windows through the genome (green) and within selection outlier regions (blue) detected with iHS (A), F_{ST} (B), and Bayenv (C). The red lines indicate the mean $-\log_{10}(P)$ of candidate regions).

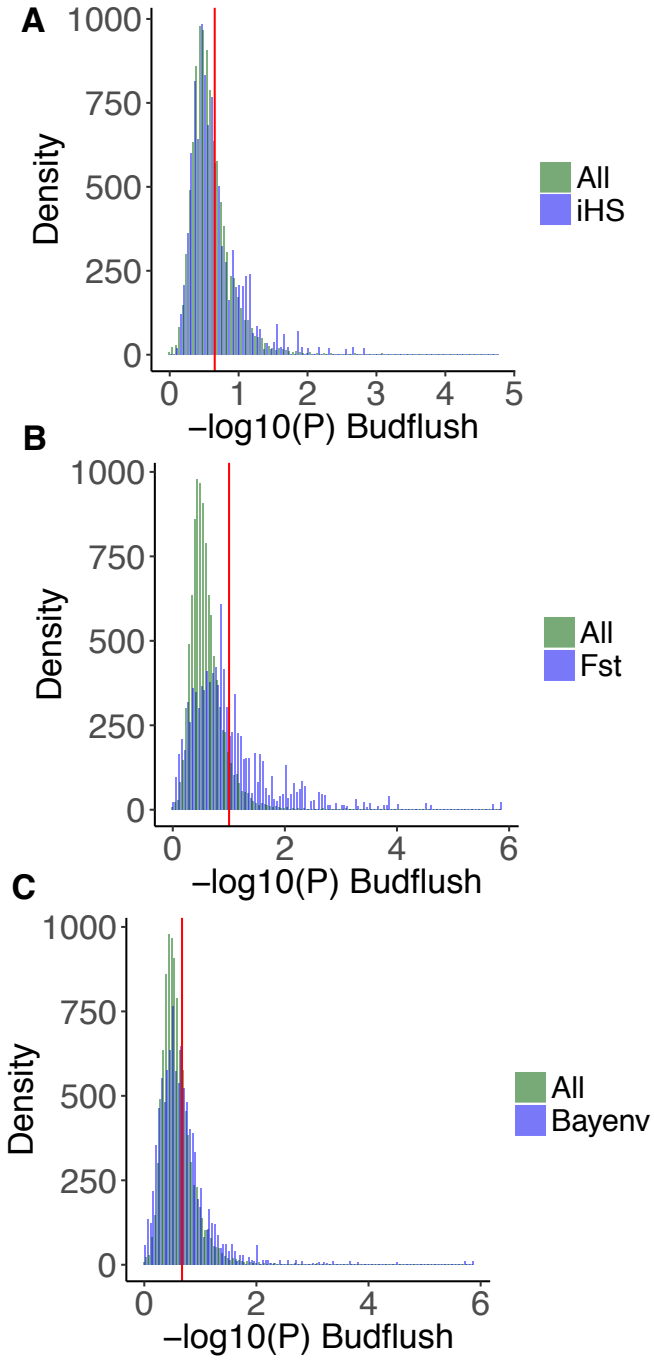


Figure S4.6. The association signal with height within candidate selection regions (The distribution of $-\log_{10}(P)$ is from GWAS on timing of bud flush in 10-kb windows through the genome (green) and within selection outlier regions (blue) detected with iHS (A), F_{ST} (B), and Bayenv (C). The red lines indicate the mean $-\log_{10}(P)$ of candidate regions).

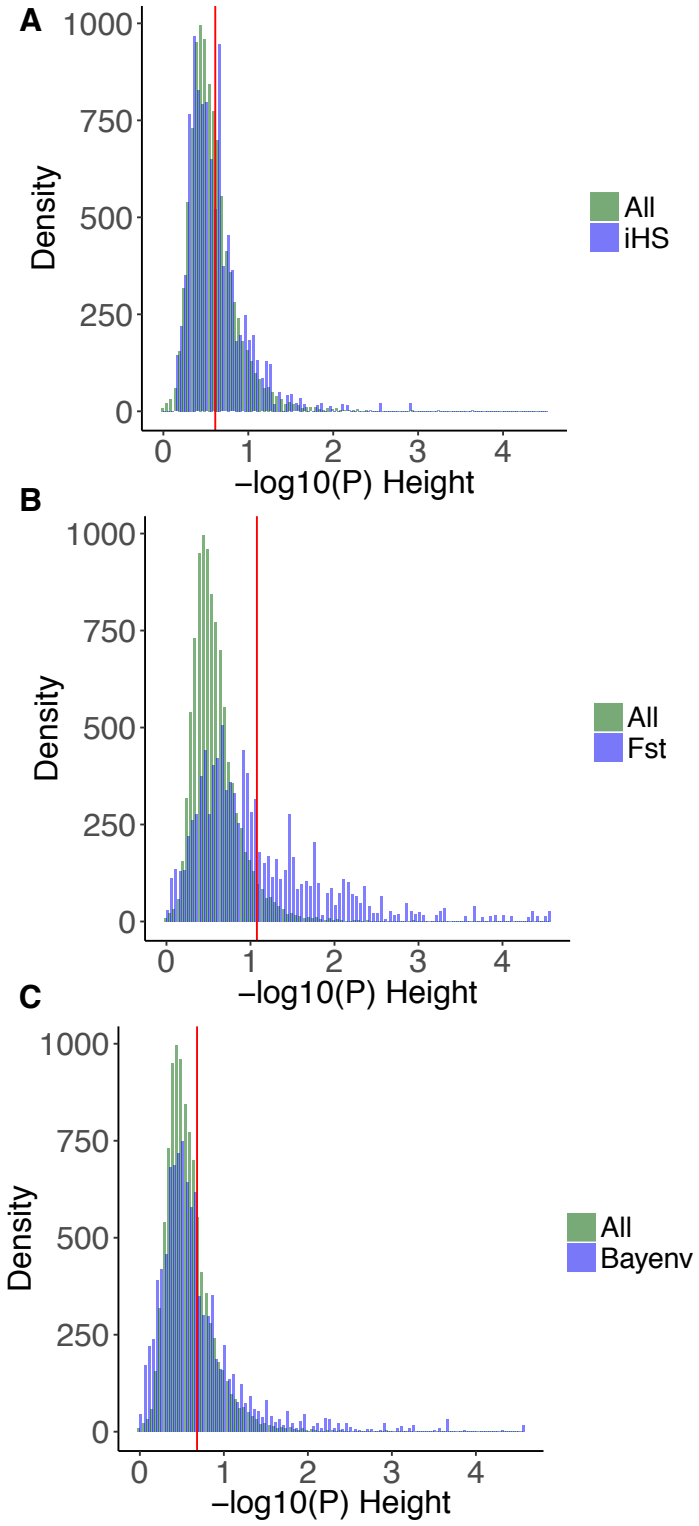


Figure S4.7. Venn diagram of the number of selection outlier and regions associated with timing of bud flush from GWAS (The blue circle is for iHS, pink for Bayenv, orchid for GWAS, and yellow for F_{ST}).

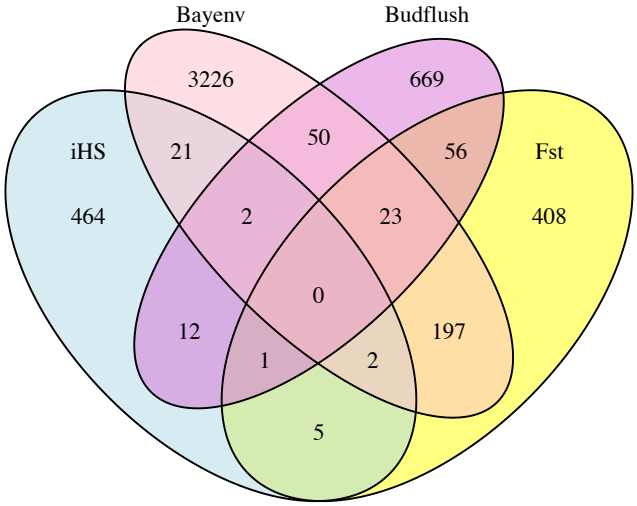
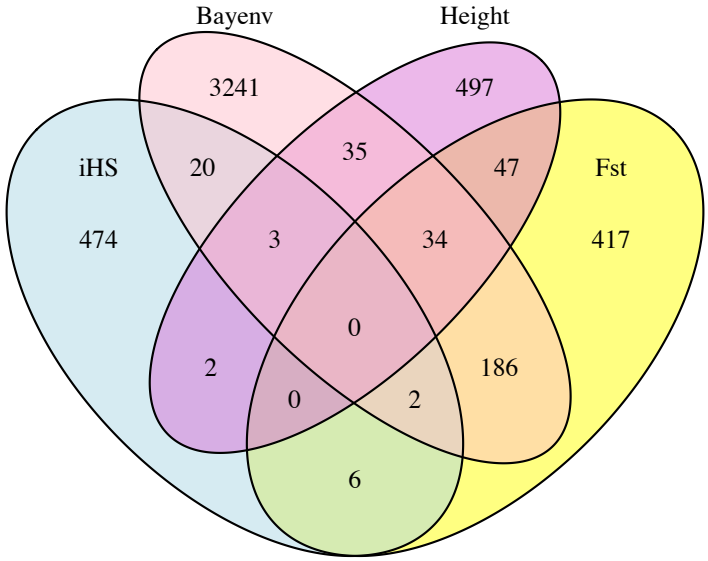


Figure S4.8. Venn diagram of the number of selection outlier and regions associated with plant height from GWAS (The blue circle is for iHS, pink for Bayenv, orchid for GWAS, and yellow for F_{ST}).



5. Recombination rate variation, hitchhiking, and demographic history shape deleterious load in poplar

Abstract

Deleterious alleles are expected to be purged by purifying selection or maintained at low frequency. However, many additional evolutionary forces may shape the pattern of deleterious mutations across the genome and among populations, including selection, hitchhiking, recombination, and demographic history. We used exome capture data to estimate the genome-wide distribution of deleterious alleles across natural populations of the model tree black cottonwood (*Populus trichocarpa*). While deleterious alleles were on average present at low frequency, suggesting purifying selection, they were preferentially enriched both within genomic regions of low-recombination and in regions showing evidence of positive selection. The demographic history of this species also appeared to play a role in the distribution of deleterious alleles among populations, with peripheral populations having higher rates of deleterious homozygosity. This suggests that marginal populations bear more deleterious mutations due in part to less efficient selection arising from smaller effective population sizes, and possibly also due to recent bottlenecks associated with postglacial recolonization. Finally, correlations between deleterious homozygosity and plant growth suggest a significant effect of deleterious load on fitness. Our results show that both genomic context and historical demography play a role in shaping the distribution of deleterious alleles across the genome and range of *P. trichocarpa*.

5.1 Introduction

Mutation is the ultimate source of genetic variation, providing the raw material for selection and local adaptation, but most new mutations are deleterious and reduce fitness (Eyre-Walker and Keightley 2007; Whitlock, et al. 2003). Mutations may be deleterious if they cause an amino acid change that reduces fitness, imposing a genetic load on populations (Mezmouk and Ross-Ibarra 2014). The fate of deleterious alleles depends on the mutation rate, population size and structure, strength of selection, effects on fitness, and other evolutionary forces (Fu, et al. 2014; Whitlock and Bürger 2004). In small populations, the accumulation of deleterious alleles can lead to "mutational meltdown"– a decline in population size and fitness that may lead to local or global extinction (Kimura, et al. 1963; Lande 1994; Loewe and Hill 2010; Lynch, et al. 1995; Whitlock and Bürger 2004; Whitlock, et al. 2003). Fewer mutations can drift to high frequency in large populations, and those that reach fixation may have negligible effect on fitness (Whitlock 2000). At the same time, the accumulation of beneficial mutations may compensate for the fitness consequences of deleterious allele fixation (Lu, et al. 2006). While natural selection may increase the frequency of adaptive mutations and remove deleterious alleles at the same time (Whitlock, et al. 2003), deleterious variants may also hitchhike with advantageous mutation to high frequency if they have intermediate fitness costs (Chun and Fay 2011). Additionally, recombination facilitates the removal of damaging alleles from a population (Charlesworth 2007; Chun and Fay 2011; Felsenstein 1974; Rodgers-Melnick, et al. 2015), and genomic regions of low recombination accumulate more deleterious alleles as the efficiency of purifying selection is reduced (Charlesworth 2007; Chun and Fay 2011).

Deleterious alleles clearly play a role in phenotype change and may affect population fitness (Charlesworth and Willis 2009; Mezmouk and Ross-Ibarra 2014), and detecting deleterious variants that contribute to disease has been a focus of human genetics research for decades (Boyko, et al. 2008; Sunyaev, et al. 2001; Xue, et al. 2012). Nonsynonymous mutations have been implicated in numerous pathologies, from rare diseases such as amyotrophic lateral sclerosis (ALS) (Cirulli, et al. 2015), to more common phenotypes such as triglyceride and cholesterol levels (Lange, et al. 2014; Walter, et al. 2015) that predispose individuals to myocardial infarction (Do, et al. 2015). In outcrossing plant species, deleterious mutations are the predominant cause of inbreeding depression on the one hand, and heterosis on the other

(Charlesworth and Willis 2009; Paige 2010). Inbreeding unmasks deleterious recessive alleles, leading to lower fertility and survival, as well as increased susceptibility of populations to environment change (Charlesworth and Willis 2009; Paige 2010). Recent efforts have employed genome-wide data to understand the genomic distribution of predicted deleterious alleles that do not have a clear relationship to phenotypes. Such studies have been undertaken in several plant species, including Arabidopsis, rice, maize, and sunflower (Cao, et al. 2011; Gossmann, et al. 2010; Gunther and Schmid 2010; Lu, et al. 2006; Mezmouk and Ross-Ibarra 2014; Renaut and Rieseberg 2015). Lu et al (2006) found that artificial selection during rice domestication increased allele frequency of deleterious variants due to effect of 'hitchhiking' (Lu, et al. 2006). Analyses in *Arabidopsis* suggest differential selection on deleterious nonsynonymous SNPs among subpopulations (Gunther and Schmid 2010), where marginal populations had a greater proportion of deleterious mutations (Cao, et al. 2011). In maize, deleterious variants are enriched in genes associated with complex traits and hybrid vigor (Mezmouk and Ross-Ibarra 2014). By contrast, little is known about the maintenance and effects of deleterious alleles in natural populations of trees. Trees are useful model in this sense because they usually have a high outcrossing rate and associated inbreeding depression that may be a function of deleterious load. *Populus trichocarpa* (poplar) is a tree species native to western North America, with a geographic range extending from southeast Alaska to southern California. With a compact and fully sequenced genome, ease of vegetative propagation, and rapid growth, poplar has emerged as the model for tree biology in general and tree genomics in particular (Jansson and Douglas 2007). Previous studies have revealed that poplar populations across the wide latitudinal species range have a high degree of phenotypic adaptation in growth and phenology, as well as adaptive genotypic variation along environmental gradients (Evans, et al. 2014; McKown, et al. 2014; Oubida, et al. 2015). While much is known about the genomic basis for local adaptation in poplar, the constraints imposed by deleterious mutations have not been explored. The main objective of this study is to quantify the distribution of deleterious mutations across the genome and range of poplar, to understand the evolutionary forces that maintain them, and to assess their effects on fitness and adaptability.

5.2 Materials and Methods

Plant material and exome capture

P. trichocarpa cuttings were collected from 182 sites across 15 degrees of latitude ranging from California to southern Alaska. The cuttings were rooted on a mist bench in 2012 and subsequently transferred to common gardens at the Virginia Tech Reynold's Homestead Forestry Research Center (36.63 °N, 80.15°W) and at Campbell River, British Columbia, Canada (50.06 °N, 125.32 °W). Genomic DNA was extracted from young leaves of 449 poplar clones using QiagenDNeasy Plant Mini Kit (Qiagen, Inc, Valencia, CA, USA). Agilent 'SureSelect' technology was used for sequence capture targeting only the exons and regulatory regions of 39,739 poplar genes, and capture was successful (>10X average depth) for 113,374 contiguous regions with an average length of 344bp, for a total of ~39Mb of sequence data (Zhou, et al. 2014). Library preparation followed the Agilent SureSelect^{XT} protocol (Version B, April 2012) (Zhou and Holliday 2012; Holliday, et al. 2016). The prepared libraries were sequenced on Illumina HiSeq platform, with 16 samples per lane. Sequence preprocessing was conducted to remove adapter sequence and low-quality reads prior to alignment with *Burrows-Wheeler Aligner* (BWA) software (Li and Durbin 2009) to version 3.0 of *P. trichocarpa* reference genome (http://phytozome.jgi.doe.gov).

SNPs were called with both the SAMtools *mpileup* function (Li, et al. 2009; Zhou and Holliday 2012; Zhou, et al. 2014) as well as the Genome Analysis ToolKit Unified Genotyper (https://www.broadinstitute.org/gatk/). For *mpileup*, SNPs with quality scores >30 that were located in genomic regions of a minimum length of 60 base-pairs and unanimously covered with a minimum depth of 10X across 75% of individual clones were used. Prior to GATK SNP calling, duplicates were identified and removed using the Picard (https://broadinstitute.github.io/picard/command-line-overview.html) MarkDuplicates and GATK DuplicateReadFilter functions, respectively (DePristo, et al. 2011), and indels realigned using the GATK IndelRealigner function. Candidate SNPs from GATK were considered where their quality score was >50 and missing data <25%. To arrive at the final SNP dataset, we merged the results from GATK and *mpileup* and considered only sites that were called by both methods. In addition to the *P. trichocarpa* samples above, we included two samples from each of the following four outgroup species for the purpose of polarizing variable sites into ancestral and derived states: *P. deltoides*, *P. fremontii*, *P. tremula*, and *P. tremuloides*. SNPs were called for these samples as above, except that the EMIT_ALL_SITES option was used in GATK so as to

acquire genotypes for sites that were invariant in the outgroups but variable in *P. trichocarpa* (i.e., for the purposes of polarizing SNPs called in *P. trichocarpa*).

Identification of deleterious mutation

To identify putatively damaging mutations, we used SIFT4G, a GPU optimized version of SIFT (Sorting Intolerant From Tolerant) software (Vaser, et al. 2016; Ng and Henikoff 2006; Ng and Henikoff 2003; Ng and Henikoff 2001). We chose SIFT because it has similar sensitivity, specificity, and accuracy compared with other available deleterious allele classification software (Choi, et al. 2012), and because SIFT4G is much faster compared both with the original SIFT algorithm and related methods, thus making it appropriate for genomic studies. SIFT queries the NCBI protein database with PSI-BLAST to identify orthologous sequences, and uses these alignments to calculate conservation values at variable sites to predict whether a nonsynonymous mutation is likely to affect protein function. The intuition behind this method is that if a particular allele or amino acid substitution type is infrequently or never seen in a protein alignment with related species, it is more likely to be deleterious. The ancestral allele was determined as the consensus allele among the outgroup species above. To confirm our SIFT results, we also used PROVEAN v.1.1.5 (Choi, et al. 2012) to predict deleterious mutations for nonsynonymous SNPs. PROVEAN is also an alignment-based method that compares sequence similarity of query protein sequence with their protein homologs before and after the introduction of variant. Protein query sequences were blasted against the latest NCBI nr database with homolog search restricted to dicots.

Signatures of selection

To better understand the extent to which deleterious sites are influenced by positive selection, we used two approaches. First, to identify species-wide selective sweeps, we estimated the integrated haplotype score as described in Voight, et al. (2006). Because this method is computationally demanding, we first reduced the SNP set using an MAF cutoff 0.05, which left 322,230 SNPs. The standardized iHS was calculated as the log of the integrated haplotype homozygosity (iHH_A) for the ancestral allele divided by the integrated Haplotype homozygosity for the derived allele (iHH_D) (Voight, et al. 2006). These raw iHS scores were then normalized within allele frequency bins to mitigate stochastic variation due to low frequency alleles (Evans,

et al. 2014). We also estimated F_{ST} using the R package 'hierfstat' (Goudet 2014) across 42 geographical populations, each corresponding to approximately 0.5 degree of latitude to identify regions of the genome that may be locally adaptive. Normalized |iHS| score were averaged for each 100kb overlapping window with a 10kb slide across the 19 chromosomes.

Estimation of recombination

To assess the effect of recombination on deleterious allele accumulation, recombination rates in our sample were estimated with PHASE version 2.1 with default settings and the MR0 model (Stephens, et al. 2001; Stephens and Donnelly 2003; Stephens and Scheet 2005). The centromere region was estimated based on genome complexity and degree of methylation from Slavov et al. (2012). We partitioned the genome into 100kb overlapping window with 10kb slide and assigned recombination rates to each bins. The windows within centromere bounds were assigned as centromeric regions (c=1) and the other windows are assigned as non-centromeric regions (c=0). The relationship between rate of deleterious SNPs, recombination rate, and centromeres was assessed with following logistic mixed effect model, which was fit using the glmer package in R:

$$Logit\left(\frac{DEL_{ij}}{TOL_{ij}+DEL_{ij}}\right)=\mu+\beta_1 r_i+\beta_2 I_i+\beta_3 c_j+\varepsilon_{ij}$$

where DEL_{ij} is the number of deleterious SNPs, TOL_{ij} is the sum of tolerated and synonymous SNPs in i^{th} window on j^{th} chromosome, $\frac{DEL_{ij}}{TOL_{ij}+DEL_{ij}}$ is defined as proportion of deleterious alleles within i^{th} window, r_i is the median recombination rate of i^{th} window, I_i is an indicator variable for i^{th} window in centromere regions, c_j is the chromosome-specific random effect for j^{th} chromosome, and ε_{ij} is individual-specific error for i^{th} window on j^{th} chromosome. To avoid skewed data, we transformed the proportion of deleterious alleles from range [0,1] into (0,1) following procedures proposed by Smithson and Verkuilen (2006).

To identify windows of exceptionally high levels of deleterious SNPs, we used a Poisson distribution to evaluate sliding windows with the highest density of deleterious SNPs. The expectation of the number of deleterious SNPs per window was set as the rounded average deleterious proportion multiplied by 100. Among all windows of at least two deleterious SNPs,

309 windows fell in the top 1% of the P-value distribution within each chromosome. The selected genomic regions were further annotated using GO enrichment analysis with FDR control. Finally, mean normalized iHS between deleterious enriched regions and other genomic intervals regions were compared with a Mann–Whitney U test.

Genome-wide association for plant height

Plant height was used as a surrogate for local adaption because it integrates bud phenology in spring and fall (i.e., timing of growth initiation and cessation) as well as growth rate. Plant heights were collected in 2014. Clonal best linear unbiased predictors (BLUPs) were used for association tests and were calculated using the following model:

$$y = \mu + R\alpha + C\beta + G\gamma + \varepsilon$$

where μ is the phenotype mean; R, C and G are incidence matrices; α, β and γ, are vectors of random effects for row, column, and clone; and ε is residuals following $N(0, \sigma_\varepsilon^2 I)$. Association analysis was performed in TASSEL v5.0 with false discovery rate (FDR) control (*qvalue* package in R; Dabney and Storey 2015). We used principal component analysis to adjust for population structure in the association model (*prcomp* function in R (R Core Team 2015), with the first 2 principal components selected based on broken stick method as predictors (Jackson 1993). We directly tested for genotype-phenotype associations at the SNP level, and compared P-value distributions between deleterious and tolerated SNPs with Mann–Whitney U test.

To gain more power in detecting associated genes, we performed gene-based association test by combining independent P-values within genes into a gene-level P-value following the FDR method of Peng, et al. (2009). Given a set of P-values $\{p_1, p_2, \ldots\ p_m\}$ computed for m hypothesis tests, FDR estimates are calculated as $t_i = \frac{V(p_i)}{F\ (p_i)}$, where $p_1 \leq p_2 \leq \cdots \leq p_m$ are ordered P-values, α is the given threshold for false positive rate, $V(p)$ is the expected proportion of tests giving false positive results with significant level p, and $F(p)$ is the expected proportion of tests yielding P-values less than or equal to p. For each SNP i, $q_i = \min_{j \geq i} t_j$. For each gene, $q_1 \leq q_2 \leq \cdots \leq q_m$ are ordered false discovery rates, and we took q_1 as the gene-level P-value (Benjamini and Hochberg 2000; Pounds and Cheng 2006). With linkage disequilibrium between genic SNPs, FDR correction allows correlation structure between SNPs and controls false

positives at a similar level as other classical methods without losing power (Wojcik, et al. 2015). Without knowing true distribution of gene-level P-values, we bootstrapped 10,000 replicates to create an empirical distribution of gene-level P-values. Any genes with P <0.0001 (empirical P less than 0.01) were considered candidate associated genes. We also tested for enrichment of deleterious genes among associated genes with a Chi-square test for symmetry. Deleterious genes were classified as those containing deleterious SNPs, while tolerated genes were those only containing tolerated SNPs. We also compared the gene level P-value of deleterious genes compared with tolerated genes to test if there is difference in their P-value distributions using a Mann–Whitney U test. Finally, we calculated the ratio of number deleterious alleles versus number of tolerated alleles within each gene and compared this ratio between height associated genes versus non-associated genes.

Distribution of deleterious alleles among populations

To evaluate the effect of deleterious alleles on growth, we used multiple linear regression to quantify the mutual relationship between height and deleterious homozygosity. Because both height and deleterious homozygosity covary with distance from the range center, we adjusted this model using centralized clone distance as a proxy for local adaptation, and further included the two principal components from GWAS as controls for population structure. We thus fit the following multiple linear regression:

$$y = \beta_0 + \beta_1 L + \beta_2 D + PC_1 + PC_2$$

where y is height, L is centralized clone distance, D is deleterious homozygosity, PC_1 and PC_2 are first two principal component score, ε is residuals following $N(0, \sigma_\varepsilon^2 I)$. The coefficient of partial determination ($R^2_{(y,D|L)}$) was calculated as the coefficient of simple determination (R^2) between residuals from a regression of height on covariates (centralized clone distance+principal components) and residuals from deleterious homozygosity on covariates (centralized clone distance+principal components). The coefficient of partial determination explains the correlation between deleterious homozygosity and plant height with local adaptation and population structure adjusted. Similar analyses were performed to assess relationships between deleterious homozygosity and effective population size. For the latter, N_e was estimated according to McEvoy et al (2011). This approach exploits the relationship of between r_{LD}^2 within recombination distance categories and N_e, which can be approximated as $E(r_{LD}^2)=1/(\alpha+ 4N_e c)$,

where α=2 and accounts for impact of mutation and c is the recombination distance in Morgans. We thus calculated N_e using average r_{LD}^2 for each binned time interval (0.005-0.25cM) according to $N_e=1/[(4c) * [(1/r_{LD}^2) - 2]]$.

5.3 Results

Frequency spectrum of deleterious alleles

We used hybrid capture to retrieve and sequence most of the exons from a sample of 449 poplar clones spanning the latitudinal range of the species (Figure 5.1), and predicted deleterious alleles on the basis of these data using both Sorting Intolerant From Tolerant (SIFT) (Ng and Henikoff 2001) and PROVEAN (Choi, et al. 2012) software. As most available deleterious allele prediction algorithms are subject to reference bias (variants are less likely to be classified as deleterious when the reference allele is derived (Simons, et al. 2014)), we only considered sites where the reference allele was ancestral as reliable deleterious alleles (Simons, et al. 2014). Among 1,717,246 SNPs identified in our exome capture data, SIFT predicted 111,125 deleterious amino acid changes, and 25,596 that either introduced stop codons or caused stop codon loss. Hereafter, we referred to these two categories collectively as 'deleterious' and to nonsysnonymous tolerated SNPs as 'tolerated'. As a complementary approach, we also used PROVEAN software to predict deleterious alleles. More negative PROVEAN scores imply a greater probability that an allele is deleterious, and we found a distinct shift to more negative scores for SIFT-predicted deleterious compared with tolerated sites (Figure S5.1). The allele frequency distribution showed a striking and significant difference between deleterious, tolerated, and synonymous sites (Figure 5.2A). Among deleterious sites, 88.0% were present at allele frequency <0.05, while only 84.0% of tolerated sites had frequencies <0.05 (Chi-square X=1079.8; df=1; P <2.2e-16; Table 5.1). More generally, with increasing derived allele frequency, the deleterious to tolerated ratio decreased (Figure 5.2B). At the same time, while most deleterious alleles were present at low frequency, 0.42% had frequencies >0.5 across all samples. We compared the genotypic state between deleterious and tolerated SNPs of matched allele frequencies, and found that homozygous deleterious genotypes were less common than tolerated homozygous genotypes for derived allele frequencies less than 0.2 (P<0.001 based on Mann-Whitney U-test for each contrasts; Table S5.1). To assess whether allele frequency alone or a combination of allele frequency and deleterious/tolerated status explains these differences in

genotypic state, we regressed observed heterozygosity across all SNPs on the above factors and found that both allele frequency (P<2e-16 based on multiple linear regression) and deleterious/tolerated status (P<2e-16 based on multiple linear regression) contribute to variation in H_o.

Genomic distribution of deleterious variants: the role of recombination and selection

The genome-wide proportion of deleterious SNPs suggested peaks around putative centromere regions across the 19 poplar chromosomes (Figure 5.3A and Figure S5.2). We used a logistic mixed model to test whether recombination in general and centromeres in particular explained genomic variation in deleterious proportion, and found that deleterious variants were significantly enriched within centromeres compared with other regions (coefficient=1.02; P=0.0065 based on logistic model) (Figure 5.3B). While recombination rate more generally did not show a significant correlation with prevalence of deleterious sites, the negative coefficient suggests that areas of lower recombination harbor more deleterious variants (coefficient=-94.67; P>0.05 based on logistic model).

To test the hypothesis that deleterious SNPs have been driven to high frequency by hitchhiking with nearby adaptive alleles, we calculated integrated haplotype scores (Walter, et al. 2015), a measure of incomplete hitchhiking events, across the 19 poplar chromosomes (Voight, et al. 2006; Chun and Fay 2011). In general, high frequency deleterious SNPs (defined as those within the 99^{th} percentile of the empirical allele frequency distribution) were found in regions with larger iHS scores (mean=0.78; sd=0.26) than tolerated SNPs (mean=0.76; sd=0.24)(W=2994000, P=0.040). Similarly, the mean proportion of deleterious alleles within iHS hitchhiking windows (mean=0.15; sd=0.099) was significantly higher compared with that of non-hitchhiking regions (mean=0.14; sd=0.058) (W=3382300; P=0.028 based on Mann-Whitney U-test). Recombination rate within hitchhiking windows (mean=0.00093; sd=0.0096) was significantly smaller than that within non-hitchhiking windows (mean=0.0022; sd=0.0091) (W=2735400; P=0.0012 based on Mann-Whitney U-test). The 309 most deleterious enriched windows had a much higher proportion of deleterious alleles (mean=0.41; sd=0.14) compared with the genomewide average (mean=0.16; sd=0.050) (W=37356; P<2.2e-16 based on Mann-Whitney U-test), and these genomic intervals also showed significantly higher normalized iHS (0.83 for deleterious enriched

regions compared with genomewide mean of 0.77) (W=4608800; P=7.91e-5 based on Mann-Whitney U-test; Figure S5.3). These deleterious enriched genomic regions also had slightly lower recombination (mean=0.0061; sd=0.0038) compared with non-deleterious enriched windows (mean=0.0089; sd=0.0091) (W=4608800; P=2.39e-6 based on Mann-Whitney U-test). Several Gene Ontology (GO) terms were significantly overrepresented among deleterious enriched regions (Q <0.05), including "generation of precursor metabolites and energy" (GO:0006091), "light reaction" (GO:0019684), "defense response" (GO:0006952), "photosynthesis" (GO:0015979), "programmed cell death" (GO:0012501) and "apoptosis" (GO:0006915) (Table S5.4). Among regions enriched for high frequency deleterious alleles, interesting candidate genes included GIBBERELLIC-ACID INSENSITIVE protein and 1,3-BETA-GLUCOSIDASE, which may play roles in dormancy release (Bosch, et al. 2011; Rinne, et al. 2011), as well as the circadian clock-related gene TIC (TIME FOR COFFEE), and flowering integrator FRIGIDA (Xie, et al. 2014).

In general, deleterious SNPs had statistically smaller F_{ST} (mean F_{ST}=0.015; sd=0.068) than tolerated SNPs (mean F_{ST}=0.025; sd=0.086) (W=1.22e10; P<2.2e-16 based on Mann-Whitney U-test). Comparing F_{ST} of SNPs across DAF bins revealed that deleterious SNPs with DAF <0.1 were driving this relationship (Table S5.5). High frequency deleterious SNPs did not exhibit an increased mean level of population differentiation that would be expected if they were directly involved in local adaptation, with the exception of bin 0.4-0.5 (mean $F_{ST=}$0.052; sd=0.099 for high frequency deleterious compared with mean F_{ST}=0.051; sd=0.095 for the remainder; W=8820300; P>0.5 based on Mann-Whitney U-test; Table S5.5). A chi-square test of symmetry indicated there was an enrichment of tolerated alleles among loci in the top 1% of the genome-wide F_{ST} distribution (X=26.6, df=1, P=2.55e-7 based on chi-square test; Table S5.2). Genes containing deleterious SNPs that were among the top 1% of F_{ST} values were enriched for GO terms related to response to environmental and endogenous stimuli, including "reproductive process" (GO:0000003), "response to stress" (GO:0044699), "response to stimulus" (GO:0044763) and "anatomical structure development" (GO:0032502) (Table S5.3). Many sites of low allele frequency (<0.1) contained deleterious alleles of high F_{ST} (0.31-0.51), including the photoperiodic gene COP9 (CONSTITUTIVE PHOTOMORPHOGENIC 9), CRY2 (CRYPTOCHROME 2), PAP1 (PHYTHOCHROME-ASSOCIATED PROTEIN 1), and disease

resistance family proteins (LRR, NB-ARC). Thus, while high frequency deleterious alleles were not of higher F_{ST} on average, a number of candidates for divergent selection fall within this category.

Pattern of deleterious mutations across populations

To investigate the distribution of deleterious variants among populations, we compared the numbers of deleterious alleles and of homozygous deleterious genotypes for the 42 populations in our sample. We detected substantial differences in the distribution of deleterious alleles between peripheral and central populations. The mean derived homozygous deleterious genotypes and mean number of derived deleterious alleles per individual significantly increased with distance of a population from the range center (R^2=61.0%, P=6.32e-10 and R^2=65.8%, P=4.43e-11, respectively) (Figure 5.4A and Figure S5.4). This indicates that peripheral populations harbor more deleterious variants, which are more often present in the homozygous state, compared with central populations. Furthermore, we explored the effect of population size (N_e) on the distribution of deleterious homozygosity and number of deleterious alleles, and found that N_e explained 24% of variation in mean deleterious homozygosity (P=0.00047; Figure 5.4B). As the individual from which the reference genome is derived originates near the center of the species range, we sought to determine whether restricting our analyses to SNPs for which the reference was ancestral may explain the relationship between deleterious load and the position of populations in the species range (i.e., if reference derived alleles are more common in the center of the range, we may undercall deleterious alleles in these areas). If such an effect were driving this relationship, we would expect a higher level of segregating deleterious alleles in peripheral populations compared with the center (more deleterious alleles being called from SNPs that were polymorphic only in range peripheries). To test this, we regressed total number of deleterious SNP within each population against distance to center, and found no relationship (coefficient=-3.53, P=0.741).

Fitness effects of deleterious alleles

To investigate whether accumulation of predicted deleterious alleles has fitness consequences, we regressed plant height on the proportion of homozygous deleterious alleles. Because the proportion of deleterious alleles co-varied with distance from the range center, as does plant

height, we first adjusted the model for the effect of local adaptation on height, and further corrected the model for the effects of population structure using PCA (see Methods). The proportion of homozygous deleterious alleles explained 2.40% of residualized height variation (P=0.0016; Figure 5.5). It should be noted that SIFT fails to account for multiple mutations within the same codon, which may contribute some error to estimates of the relationship between deleterious load and phenotypes for individual trees where such adjacent non-synonymous polymorphisms are present. To further explore the possible fitness consequences of deleterious alleles, we also performed SNP-based and gene-based genome-wide association analyses on plant height. More than 81% of height associated genes contained predicted deleterious mutations, while 94.7% contained tolerated mutations. At the SNP level, P-values of tolerated sites were on average smaller than for deleterious sites (W=6410600000, P <2.2e-16 based on Mann-Whitney U-test) (Figure 5.6A). At the gene level, genes containing deleterious SNPs had significantly smaller P-values than genes containing only tolerated SNPs (W=29633000; P=0.00041 based on Mann-Whitney U-test) (Figure 5.6B). However, the chi-square test showed that height-associated genes had no preference for deleterious genes over tolerated genes (X=0.50, df=1, P=0.48; Table S5.6) and a comparison of the ratio of deleterious alleles within height-associated genes versus that within non-associated genes also showed no significant enrichment within associated genes (W=2127100, P=0.1751 based on Mann-Whitney U-test). Several interesting candidate genes stood out among the 178 height associated genes containing deleterious alleles, including PAP1 (PHYTOCHROME ASSOCIATE PROTEIN 1), which regulates IAA signaling pathway through interactions with phytochromes (Liscum and Reed 2002); PKS4 (PHYTOCHROME KINASE SUBSTRATE 4), which is a phytochrome signaling component that regulates phototropism and shade avoidance (Sellaro, et al. 2011; Demarsy, et al. 2012); and several genes related to lipid transfer and metabolism (e.g., LIPID TRANSFER PROTEIN 1, SYNAPTOTAGMIN 3) that may play a role in bud dormancy (Leida, et al. 2012; Pang, et al. 2007).

5.4 Discussion

Genomic factors contributing to deleterious load

The genomic distribution and effects of deleterious mutations have recently been a topic of intense interest across taxa (Lu, et al. 2006; Gunther and Schmid 2010; Chun and Fay 2011;

Lohmueller 2014; Mezmouk and Ross-Ibarra 2014; Henn, et al. 2015; Renaut and Rieseberg 2015). Our results showed that in wild populations of poplar trees, the vast majority of deleterious SNPs segregated at lower frequency than either tolerated or synonymous SNPs, which is consistent with reports in *Arabidopsis*, maize and rice (Gunther and Schmid 2010; Mezmouk and Ross-Ibarra 2014). This pattern is likely the result of purifying selection against deleterious SNPs (Fay, et al. 2001; Wong, et al. 2003), which should be relatively efficient in poplar and other tree species due to generally large effective population sizes (Petit and Hampe 2006). The fitness effects of these low frequency deleterious alleles may be masked by heterozygosity, an effect that was supported by our data – deleterious alleles in the left tail of the frequency distribution were more likely to present as heterozygotes compared with nonsynonymous tolerated alleles in the same frequency bin.

While most deleterious alleles in poplar segregated at low frequency, a subset of sites were present at relatively high frequency in our rangewide collection. A number of factors likely contribute to the presence of these high frequency deleterious alleles. It is possible that some are conditionally advantageous and have been directly targeted by positive selection. Alternatively, deleterious alleles may reach high frequency due to hitchhiking associated with positive selection on a linked adaptive allele. Evidence from other systems suggests such hitchhiking may be common. For example, artificial selection has increased the frequency of deleterious alleles during domestication (Chun and Fay 2011; Renaut and Rieseberg 2015; Marsden, et al. 2016), and hitchhiking also appears to be responsible for the presence of common deleterious SNPs in the human genome (Chun and Fay 2011). Our results are consistent with these studies: the proportion of deleterious alleles was higher in the top hitchhiking windows compared with the genome-wide mean, and the most deleterious enriched windows had higher average iHS scores. Several candidate genes that may be involved in seasonal dormancy – a process that is key to climatic adaptation in trees – were found among our hitchhiking windows, which suggest possible direct targets of selection that may in part be driving high levels of deleterious alleles in these regions.

Previous work has also shown that the accumulation of deleterious mutations occurs where recombination is suppressed (Charlesworth and Charlesworth 2000; Lu, et al. 2006; Campos, et

al. 2014; Hussin, et al. 2015; Rodgers-Melnick, et al. 2015). We found that recombination was lower in regions with high iHS, although it is difficult to disentangle the signature of recent selection from background recombination rate variation. As such, the distribution of deleterious alleles across the genome is likely a function of the combined effects of hitchhiking and weaker purifying selection efficiency in areas of the genome with reduced recombination (Betancourt, et al. 2009). This conclusion is supported by our finding that the proportion of deleterious SNPs was significantly correlated with proximity to putative centromeres. Taken together, these results suggest that while some 'deleterious' alleles may be advantageous and directly targeted by selection, linked selection and recombination rate variation likely play more important roles in governing the distribution of deleterious alleles across the poplar genome.

Demographic constraints on deleterious alleles

While selection and recombination clearly influence deleterious allele frequencies across the genome, demographic history has also been shown to affect the distribution of deleterious mutations among populations (Lohmueller, et al. 2008; Fu, et al. 2014; Lohmueller 2014; Henn, et al. 2016; Marsden, et al. 2016). We found a clear relationship between population position in the range and the frequency of deleterious alleles, with edge populations having a larger number of deleterious alleles and higher deleterious homozygosity compared with central populations. To test whether deleterious load in the edge population was a result of local adaptation, we calculated F_{ST} for these variants. Mean F_{ST} was lower for deleterious SNPs, suggesting they are on average less likely than tolerated sites to be targets of divergent selection. At the same time, high F_{ST} values for several SNPs within genes that may have roles in climatic adaptation hint at a possible role adaptive potential of a small proportion of deleterious sites, though we cannot rule out the possibility of hitchhiking in these cases. Given the overall trend toward lower F_{ST} for deleterious SNPs, and the genomic constraints discussed above, demographic forces are likely key to the overall relationship between deleterious homozygosity and central-peripheral status of our populations. To test this, we compared among-population deleterious allele proportions with N_e estimates, and found a significant effect, which suggests that drift in edge populations with smaller N_e may lead to greater accumulation of deleterious alleles compared with central population that have larger N_e, and where selection is therefore more efficient. Another explanation for the relationship between deleterious load central-peripheral structure is the

colonization history of *P. trichocarpa*. Particularly in the north, peripheral populations likely experienced recent bottlenecks and subsequent population expansion associated with postglacial re-colonization (Zhou, et al. 2014), and individuals from these areas also tend to carry more deleterious alleles. As such, the spectrum and distribution of deleterious variants across the poplar range are consistent with variable demographic history and N_e among populations.

Mutational load of deleterious alleles

Nonsynonymous mutations may affect protein structure and function and thus decrease the fitness of individuals carrying them. While severe deleterious mutations that alter the function of important genes are likely purged rapidly, mildly deleterious mutations may persist for many generations before being eradicated by selection or becoming fixed (Agrawal and Whitlock 2012; Henn, et al. 2015). For example, the average human carries over a thousand deleterious alleles, including ~300 loss-of-function mutations (Agrawal and Whitlock 2012). This mutation load reduces fitness compared with a hypothetical mutation-free genotype (Crow 1958; Keightley 2012), and our analysis shows that a significant reduction in plant height, a surrogate for fitness, can be explained by accumulation of deleterious alleles. At the same time, we did not find evidence for direct genotype-phenotype associations for deleterious SNPs, which may be due to a lack of power attributable to the low frequency of most deleterious alleles and their individually small effect sizes. However, gene-level association tests show that genes containing deleterious alleles were on average more significant (smaller P-values) than genes containing only tolerated loci. An analogous effect was reported in maize, wherein genes with significant associations were enriched for deleterious alleles (Mezmouk and Ross-Ibarra 2014). One possible explanation for this result is that linkage disequilibrium between common (tolerated) variants and deleterious alleles generates synthetic associations (Mezmouk and Ross-Ibarra 2014; Dickson et al. 2010), though deeper resequencing of larger numbers of samples in the relevant genomic intervals is needed to test this hypothesis.

5.5 Conclusions

Whereas most previous genome-wide studies have focused on the accumulation and effects of deleterious alleles as a result of the domestication process, our results reveal a significant role for these alleles in the evolutionary trajectory of natural populations. In poplar, deleterious alleles

accumulate as an indirect result of natural selection, due to recombination rate variation across the genome, and as a result of population-specific demographic factors. Deleterious alleles in poplar also show significant effect on fitness, which may partly explain the strong inbreeding depression observed in most temperate and boreal tree species. Importantly, deleterious load was most apparent in edge populations, and may reduce population fitness in the same areas that will be most challenged by anthropogenic climate change. Some caution, however, is warranted when interpreting the causes and consequences of deleterious alleles. Studies of adaptive variation typically use minor allele frequency cutoffs that mitigate the potential for false-positive SNPs as a result of sequencing errors. By contrast, the shift in the frequency distribution of deleterious alleles toward rare variants means that low frequency alleles must be included. We attempted to exclude such false positives through SNP quality and depth thresholds, and by only including SNPs called by two independent algorithms. Nevertheless, future studies for which deleterious alleles are of central interest would benefit from sequencing to higher average depth or by validating putative deleterious alleles with an independent SNP genotyping platform.

5.6 Reference Cited

Agrawal AF, Whitlock MC. 2012. Mutation Load: The Fitness of Individuals in Populations Where Deleterious Alleles Are Abundant. Annual Review of Ecology, Evolution, and Systematics, 43:115.

Betancourt AJ, Welch JJ, Charlesworth B. 2009. Reduced effectiveness of selection caused by a lack of recombination. Current Biology 19:655-660.

Benjamini Y, Hochberg Y. 2000. On the Adaptive Control of the False Discovery Rate in Multiple Testing With Independent Statistics. Journal of Educational and Behavioral Statistics 25:60-83.

Bosch M, Mayer CD, Cookson A, Donnison IS. 2011. Identification of genes involved in cell wall biogenesis in grasses by differential gene expression profiling of elongating and non-elongating maize internodes. Journal of Experimental Botany 62:3545-3561.

Boyko AR, Williamson SH, Indap AR, Degenhardt JD, Hernandez RD, Lohmueller KE, Adams MD, Schmidt S, Sninsky JJ, Sunyaev SR, et al. 2008. Assessing the evolutionary impact of amino acid mutations in the human genome. Plos Genetics 4:e1000083.

Campos JL, Halligan DL, Haddrill PR, Charlesworth B. 2014. The Relation between Recombination Rate and Patterns of Molecular Evolution and Variation in Drosophila melanogaster. Molecular Biology and Evolution 31:1010-1028.

Cao J, Schneeberger K, Ossowski S, Gunther T, Bender S, Fitz J, Koenig D, Lanz C, Stegle O, Lippert C, et al. 2011. Whole-genome sequencing of multiple Arabidopsis thaliana populations. Nature Genetics 43:956-963.

Charlesworth B. 2007. Mutation-selection balance and the evolutionary advantage of sex and recombination. Genet Res 89:451-473.

Charlesworth B, Charlesworth D. 2000. The degeneration of Y chromosomes. Philos Trans R Soc Lond B Biol Sci 355:1563-1572.

Charlesworth D, Willis JH. 2009. The genetics of inbreeding depression. Nature Reviews Genetics 10:783-796.

Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. 2012. Predicting the Functional Effect of Amino Acid Substitutions and Indels. Plos One 7.

Chun S, Fay JC. 2011. Evidence for hitchhiking of deleterious mutations within the human genome. Plos Genetics 7:e1002240.

Cirulli ET, Lasseigne BN, Petrovski S, Sapp PC, Dion PA, Leblond CS, Couthouis J, Lu Y-F, Wang Q, Krueger BJ, et al. 2015. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. Science 347:1436-1441.

Crow JF. 1958. Some possibilities for measuring selection intensities in man. Hum Biol 30:1-13.

Dabney A and Storey JD. 2015. qvalue: Q-value estimation for false discovery rate control. R package version 1.38.0.

Demarsy E, Schepens I, Okajima K, Hersch M, Bergmann S, Christie J, Shimazaki K-i, Tokutomi S, Fankhauser C. 2012. Phytochrome Kinase Substrate 4 is phosphorylated by the phototropin 1 photoreceptor. The EMBO Journal 31:3457-3467.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics 43:491-498.

Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. 2010. Rare variants create synthetic genome-wide associations. PLoS Biol 8:e1000294.

Do R, Stitziel NO, Won H-H, Jorgensen AB, Duga S, Merlini PA, Kiezun A, Farrall M, Goel A, Zuk O, et al. 2015. Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. Nature 518:102-+.
Evans LM, Slavov GT, Rodgers-Melnick E, Martin J, Ranjan P, Muchero W, Brunner AM, Schackwitz W, Gunter L, Chen JG, et al. 2014. Population genomics of Populus trichocarpa identifies signatures of selection and adaptive trait associations. Nature Genetics 46:1089-1096.
Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. Nature Reviews Genetics 8:610-618.
Felsenstein J. 1974. The evolutionary advantage of recombination. Genetics 78:737-756.
Fu W, Gittelman RM, Bamshad MJ, Akey JM. 2014. Characteristics of neutral and deleterious protein-coding variation among individuals and populations. Am J Hum Genet 95:421-436.
Gossmann TI, Song BH, Windsor AJ, Mitchell-Olds T, Dixon CJ, Kapralov MV, Filatov DA, Eyre-Walker A. 2010. Genome wide analyses reveal little evidence for adaptive evolution in many plant species. Molecular Biology and Evolution 27:1822-1832.
Goudet J. 2014. hierfstat: Estimation and tests of hierarchical F-statistics: R package version 0.04-14.
Gunther T, Schmid KJ. 2010. Deleterious amino acid polymorphisms in Arabidopsis thaliana and rice. Theoretical and Applied Genetics 121:157-168.
Henn BM, Botigue LR, Bustamante CD, Clark AG, Gravel S. 2015. Estimating the mutation load in human genomes. Nature Reviews Genetics 16:333-343.
Henn BM, Botigue LR, Peischl S, Dupanloup I, Lipatov M, Maples BK, Martin AR, Musharoff S, Cann H, Snyder MP, et al. 2016. Distance from sub-Saharan Africa predicts mutational load in diverse human genomes. Proceedings of the National Academy of Sciences of the United States of America 113:E440-E449.
Holliday JA, Zhou L, Bawa R, Zhang M, Oubida RW. 2016. Evidence for extensive parallelism but divergent genomic architecture of adaptation along altitudinal and latitudinal gradients in Populus trichocarpa. New Phytologist 209:1240-1251.
Hussin JG, Hodgkinson A, Idaghdour Y, Grenier J-C, Goulet J-P, Gbeha E, Hip-Ki E, Awadalla P. 2015. Recombination affects accumulation of damaging and disease-associated mutations in human populations. Nature Genetics 47:400-U162.
Jackson DA. 1993. Stopping Rules in Principal Components Analysis: A Comparison of Heuristical and Statistical Approaches. Ecology 74:2204-2214.
Jansson S, Douglas CJ. 2007. Populus: a model system for plant biology. Annual review of plant biology 58:435-458.
Keightley PD. 2012. Rates and fitness consequences of new mutations in humans. Genetics 190:295-304.
Kimura M, Maruyama T, Crow JF. 1963. The Mutation Load in Small Populations. Genetics 48:1303-1312.
Lande R. 1994. Risk of population extinction from fixation of new deleterious mutations. Evolution 48:1460-1469.
Lange LA, Hu Y, Zhang H, Xue C, Schmidt EM, Tang Z-Z, Bizon C, Lange EM, Smith JD, Turner EH, et al. 2014. Whole-Exome Sequencing Identifies Rare and Low-Frequency Coding Variants Associated with LDL Cholesterol. American Journal of Human Genetics 94:233-245.
Leida C, Conesa A, Llácer G, Badenes ML, Ríos G. 2012. Histone modifications and expression of DAM6 gene in peach are modulated during bud dormancy release in a cultivar-dependent manner. New Phytologist 193:67-80.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-1760.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-2079.
Liscum E, Reed JW. 2002. Genetics of Aux/IAA and ARF action in plant growth and development. Plant Molecular Biology 49:387-400.
Loewe L, Hill WG. 2010. The population genetics of mutations: good, bad and indifferent. Philos Trans R Soc Lond B Biol Sci 365:1153-1167.
Lohmueller KE. 2014. The distribution of deleterious genetic variation in human populations. Current Opinion in Genetics & Development 29:139-146.
Lohmueller KE, Indap AR, Schmidt S, Boyko AR, Hernandez RD, Hubisz MJ, Sninsky JJ, White TJ, Sunyaev SR, Nielsen R, et al. 2008. Proportionally more deleterious genetic variation in European than in African populations. Nature 451:994-997.
Lu J, Tang T, Tang H, Huang J, Shi S, Wu CI. 2006. The accumulation of deleterious mutations in rice genomes: a hypothesis on the cost of domestication. Trends Genet 22:126-131.
Lynch M, Conery J, Burger R. 1995. Mutation accumulation and the extinction of small populationS. American Naturalist 146:489-518.
Marsden CD, Ortega-Del Vecchyo D, O'Brien DP, Taylor JF, Ramirez O, Vila C, Marques-Bonet T, Schnabel RD, Wayne RK, Lohmueller KE. 2016. Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. Proceedings of the National Academy of Sciences of the United States of America 113:152-157.
McKown AD, Guy RD, Klapste J, Geraldes A, Friedmann M, Cronk QC, El-Kassaby YA, Mansfield SD, Douglas CJ. 2014. Geographical and environmental gradients shape phenotypic trait variation and genetic structure in Populus trichocarpa. New Phytologist 201:1263-1276.
McEvoy BP, Powell JE, Goddard ME, Visscher, PM. 2011. Human population dispersal "Out of Africa" estimated from linkage disequilibrium and allele frequencies of SNPs. Genome Research 21 (6): 821-829.
Mezmouk S, Ross-Ibarra J. 2014. The pattern and distribution of deleterious mutations in maize. G3 (Bethesda) 4:163-171.
Ng PC, Henikoff S. 2006. Predicting the effects of amino acid substitutions on protein function. Annu Rev Genomics Hum Genet 7:61-80.
Ng PC, Henikoff S. 2003. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 31:3812-3814.
Ng PC, Henikoff S. 2001. Predicting deleterious amino acid substitutions. Genome Research 11:863-874.
Oubida RW, Gantulga D, Zhang M, Zhou L, Bawa R, Holliday JA. 2015. Partitioning of multivariate phenotypes using regression trees reveals complex patterns of adaptation to climate across the range of black cottonwood (Populus trichocarpa). Front Plant Sci 6:181.
Paige KN. 2010. The Functional Genomics of Inbreeding Depression: A New Approach to an Old Problem. BioScience 60:267-277.
Pang X, Halaly T, Crane O, Keilin T, Keren-Keiserman A, Ogrodovitch A, Galbraith D, Or E. 2007. Involvement of calcium signalling in dormancy release of grape buds. Journal of Experimental Botany 58:3249-3262.
Peng G, Luo L, Siu H, Zhu Y, Hu P, Hong S, Zhao J, Zhou X, Reveille JD, Jin L, et al. 2009. Gene and pathway-based second-wave analysis of genome-wide association studies. European Journal of Human Genetics 18:111-117.

Petit RJ, Hampe A. 2006. Some evolutionary consequences of being a tree. In. Annual Review of Ecology Evolution and Systematics. p. 187-214.
Pounds S, Cheng C. 2006. Robust estimation of the false discovery rate. Bioinformatics 22:1979-1987.
R Core Team. 2015. R: A language and environment for statistical computing: R Foundation for Statistical Computing.
Renaut S, Rieseberg LH. 2015. The Accumulation of Deleterious Mutations as a Consequence of Domestication and Improvement in Sunflowers and Other Compositae Crops. Molecular Biology and Evolution 32:2273-2283.
Rinne PL, Welling A, Vahala J, Ripel L, Ruonala R, Kangasjarvi J, van der Schoot C. 2011. Chilling of dormant buds hyperinduces FLOWERING LOCUS T and recruits GA-inducible 1,3-beta-glucanases to reopen signal conduits and release dormancy in Populus. Plant Cell 23:130-146.
Rodgers-Melnick E, Bradbury PJ, Elshire RJ, Glaubitz JC, Acharya CB, Mitchell SE, Li CH, Li YX, Buckler ES. 2015. Recombination in diverse maize is stable, predictable, and associated with genetic load. Proceedings of the National Academy of Sciences of the United States of America 112:3823-3828.
Sellaro R, Yanovsky MJ, Casal JJ. 2011. Repression of shade-avoidance reactions by sunfleck induction of HY5 expression in Arabidopsis. The Plant Journal 68:919-928.
Simons YB, Turchin MC, Pritchard JK, Sella G. 2014. The deleterious mutation load is insensitive to recent population history. Nature Genetics 46:220
Slavov GT, DiFazio SP, Martin J, Schackwitz W, Muchero W, Rodgers-Melnick E, Lipphardt MF, Pennacchio CP, Hellsten U, Pennacchio LA, et al. 2012. Genome resequencing reveals multiscale geographic structure and extensive linkage disequilibrium in the forest treePopulus trichocarpa. New Phytologist 196:713-725.
Smithson M, Verkuilen J. 2006. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. Psychological Methods 11:54-71.
Stephens M, Donnelly P. 2003. A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet 73:1162-1169.
Stephens M, Scheet P. 2005. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. Am J Hum Genet 76:449-462.
Stephens M, Smith NJ, Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68:978-989.
Sunyaev S, Ramensky V, Koch I, Lathe W, 3rd, Kondrashov AS, Bork P. 2001. Prediction of deleterious human alleles. Hum Mol Genet 10:591-597.
Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. 2016. SIFT missense predictions for genomes. Nature Protocols 11:1-9.
Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. PLoS Biol 4:e72.
Walter K, Min JL, Huang J, Crooks L, Memari Y, McCarthy S, Perry JRB, Xu C, Futema M, Lawson D, et al. 2015. The UK10K project identifies rare variants in health and disease. Nature 526:82-+.
Whitlock MC. 2000. Fixation of new alleles and the extinction of small populations: drift load, beneficial alleles, and sexual selection. Evolution 54:1855-1861.

Whitlock MC, Bürger R. 2004. Fixation of New Mutations in Small Populations. In: Ferrière R, Dieckmann U, Couvet D, editors. Evolutionary Conservation Biology. Cambridge, UK: Cambridge University Press. p. 155-170.
Whitlock MC, Griswold CK, Peters AD. 2003. Compensating for the meltdown: The critical effective size of a population with deleterious and compensatory mutations. Annales Zoologici Fennici 40:169-183.
Wojcik GL, Kao WHL, Duggal P. 2015. Relative performance of gene- and pathway-level methods as secondary analyses for genome-wide association studies. BMC Genetics 16.
Wong GK, Yang Z, Passey DA, Kibukawa M, Paddock M, Liu CR, Bolund L, Yu J. 2003. A population threshold for functional polymorphisms. Genome Res 13:1873-1879.
Xie Q, Wang P, Liu X, Yuan L, Wang L, Zhang C, Li Y, Xing H, Zhi L, Yue Z, et al. 2014. LNK1 and LNK2 are transcriptional coactivators in the Arabidopsis circadian oscillator. Plant Cell 26:2843-2857.
Xue Y, Chen Y, Ayub Q, Huang N, Ball EV, Mort M, Phillips AD, Shaw K, Stenson PD, Cooper DN, et al. 2012. Deleterious- and disease-allele prevalence in healthy individuals: insights from current predictions, mutation databases, and population-scale resequencing. Am J Hum Genet 91:1022-1032.
Zhou L, Bawa R, Holliday JA. 2014. Exome resequencing reveals signatures of demographic and adaptive processes across the genome and range of black cottonwood (Populus trichocarpa). Mol Ecol 23:2486-2499.
Zhou L, Holliday JA. 2012. Targeted enrichment of the black cottonwood (Populus trichocarpa) gene space using sequence capture. Bmc Genomics 13:703.

Table 5.1. Counts of rare versus common deleterious, synonymous, and tolerated variants.

Class	Rare	Common	Ratio
Deleterious	97,750	13,375	7.31
Synonymous	267,427	108,734	2.46
Tolerated	244,310	47,150	5.18

Pearson Chi-square test of symmetry: deleterious SNPs vs. synonymous SNPs ($X=13000$; $df=1$; $P < 2.2e-16$); deleterious SNPs vs. tolerated SNPs ($X=1079.8$; $df=1$; $P < 2.2e-16$)

Table S5.1. Number of homozygous and heterozygous genotypes for deleterious and tolerated variants across derived allele frequency (DAF) bins.

DAF bin	Class	Mean number of homozygous genotypes	Std. Error	P	Mean number of heterozygous genotypes	Std. Error	P
0-0.025	Deleterious	0.20	0.62	<2.2e-16	3.63	4.10	<2.2e-16
	Tolerated	0.22	0.66		3.79	4.22	
0.025-0.05	Deleterious	2.11	2.18	<0.001	27.60	7.06	<0.001
	Tolerated	2.31	2.30		27.25	7.19	
0.05-0.1	Deleterious	4.59	3.87	<0.001	54.46	12.58	>0.05
	Tolerated	4.79	3.81		54.58	12.62	
0.1-0.2	Deleterious	12.16	8.29	<2.2e-16	102.07	23.50	>0.05
	Tolerated	12.79	8.09		101.56	22.99	
0.2-0.3	Deleterious	28.30	15.69	>0.05	160.71	32.54	>0.05
	Tolerated	28.99	15.42		161.08	32.38	
0.3-0.4	Deleterious	51.07	26.01	>0.05	206.42	51.76	>0.05
	Tolerated	51.28	25.17		206.63	50.15	
0.4-0.5	Deleterious	79.32	39.12	>0.05	238.64	75.63	>0.05
	Tolerated	79.08	39.58		241.66	79.50	
0.5-0.6	Deleterious	135.25	29.73	>0.05	214.93	50.35	>0.05
	Tolerated	135.68	32.67		216.07	53.48	
0.6-0.7	Deleterious	193.57	21.60	>0.05	190.79	30.74	>0.05
	Tolerated	193.83	21.01		188.93	26.50	
0.7-0.8	Deleterious	255.85	19.16	>0.05	154.03	18.61	>0.05
	Tolerated	257.42	21.26		152.00	21.57	
0.8-0.9	Deleterious	322.78	21.39	>0.05	104.84	16.76	>0.05
	Tolerated	325.60	22.04		104.41	20.51	
0.9-1.0	Deleterious	408.60	24.53	>0.05	33.80	21.22	>0.05
	Tolerated	393.52	21.04		48.08	18.44	

Table S5.2. Counts of deleterious and tolerated variants in the top 1% of the combined empirical F_{ST} distribution.

Class	Top 1% F_{ST}	Remainder	Ratio
Deleterious	852	98,063	0.009
Tolerated	2,712	254,607	0.011
Pearson Chi-square test of symmetry for deleterious vs tolerated (X=26.57, df=1, P=2.55e-7)			

Table S5.3. Selected gene ontology (GO) terms of genes containing deleterious SNPs that are overrepresented in top 1% F_{ST} and their Qvalue with FDR control in Gene Enrichment Analysis.

GO term	Ontology	Description	Q
GO:0044702	P	single organism reproductive process	0.018
GO:0022414	P	single-organism process	0.000
GO:0000003	P	reproductive process	0.009
GO:0048856	P	reproduction	0.009
GO:0032502	P	anatomical structure development	0.004
GO:0032501	P	developmental process	0.005
GO:0044767	P	multicellular organismal process	0.016
GO:0051716	P	single-organism developmental process	0.020
GO:0006950	P	cellular response to stimulus	0.031
GO:0044763	P	response to stimulus	0.001
GO:0050896	P	cellular process	0.000
GO:0044699	P	response to stress	0.035
GO:0065007	P	single-organism cellular process	0.000
GO:0009987	P	biological regulation	0.044
GO:0071704	P	organic substance metabolic process	0.007
GO:0008152	P	metabolic process	0.001
GO:0044237	P	cellular metabolic process	0.047

1 Ontology P refers to biological process

Table S5.4. Table of selected gene ontology (GO) terms of genes overrepresented within deleterious enriched regions and Q-value with FDR control in Gene Enrichment Analysis.

GO term	Ontology*	Description	Q
GO:0006091	P	generation of precursor metabolites and energy	0.0011
GO:0009987	P	cellular process	0.0049
GO:0019684	P	photosynthesis, light reaction	0.0089
GO:0015979	P	photosynthesis	0.016
GO:0016265	P	death	0.024
GO:0008219	P	cell death	0.024
GO:0012501	P	programmed cell death	0.029
GO:0006952	P	defense response	0.031
GO:0006915	P	apoptosis	0.038

1 Ontology P refers to biological process

Table S5.5. F_{ST} of deleterious versus tolerated variants.

DAF bins	Class	mean F_{ST}	Standard Error	P
0-0.025	Deleterious	0.016	0.072	<2.2e-16
	Tolerated	0.024	0.087	
0.025-0.05	Deleterious	0.018	0.079	<0.01
	Tolerated	0.020	0.087	
0.05-0.1	Deleterious	0.019	0.073	<0.01
	Tolerated	0.022	0.082	
0.1-0.2	Deleterious	0.021	0.078	>0.05
	Tolerated	0.024	0.082	
0.2-0.3	Deleterious	0.025	0.077	>0.05
	Tolerated	0.026	0.078	
0.3-0.4	Deleterious	0.035	0.077	>0.05
	Tolerated	0.034	0.074	
0.4-0.5	Deleterious	0.039	0.070	>0.05
	Tolerated	0.038	0.071	
0.5-0.6	Deleterious	0.051	0.083	>0.05
	Tolerated	0.054	0.084	
0.6-0.7	Deleterious	0.073	0.123	>0.05
	Tolerated	0.057	0.104	
0.7-0.8	Deleterious	0.063	0.125	>0.05
	Tolerated	0.061	0.120	
0.8-0.9	Deleterious	0.080	0.169	>0.05
	Tolerated	0.069	0.148	
0.9-1.0	Deleterious	0.060	0.102	>0.05
	Tolerated	0.067	0.123	

Mann Whitney U-test test compare the difference in mean F_{ST} of deleterious SNPs vs. tolerated SNPs

Table S5.6. Counts of deleterious and tolerated genes in associated genes versus non-associated genes from gene-based association study.

Class	Associated	Non-associated	Ratio
Deleterious	178	16,713	0.011
Tolerated	33	3,606	0.009

Pearson Chi-square test of symmetry for associated genes among deleterious vs. tolerated (X=0.50, df=1, P=0.48)

Figure 5.1. Sampling locations of the 449 poplar clones used in this study (red points) and species distribution for P. trichocarpa (green shaded area).

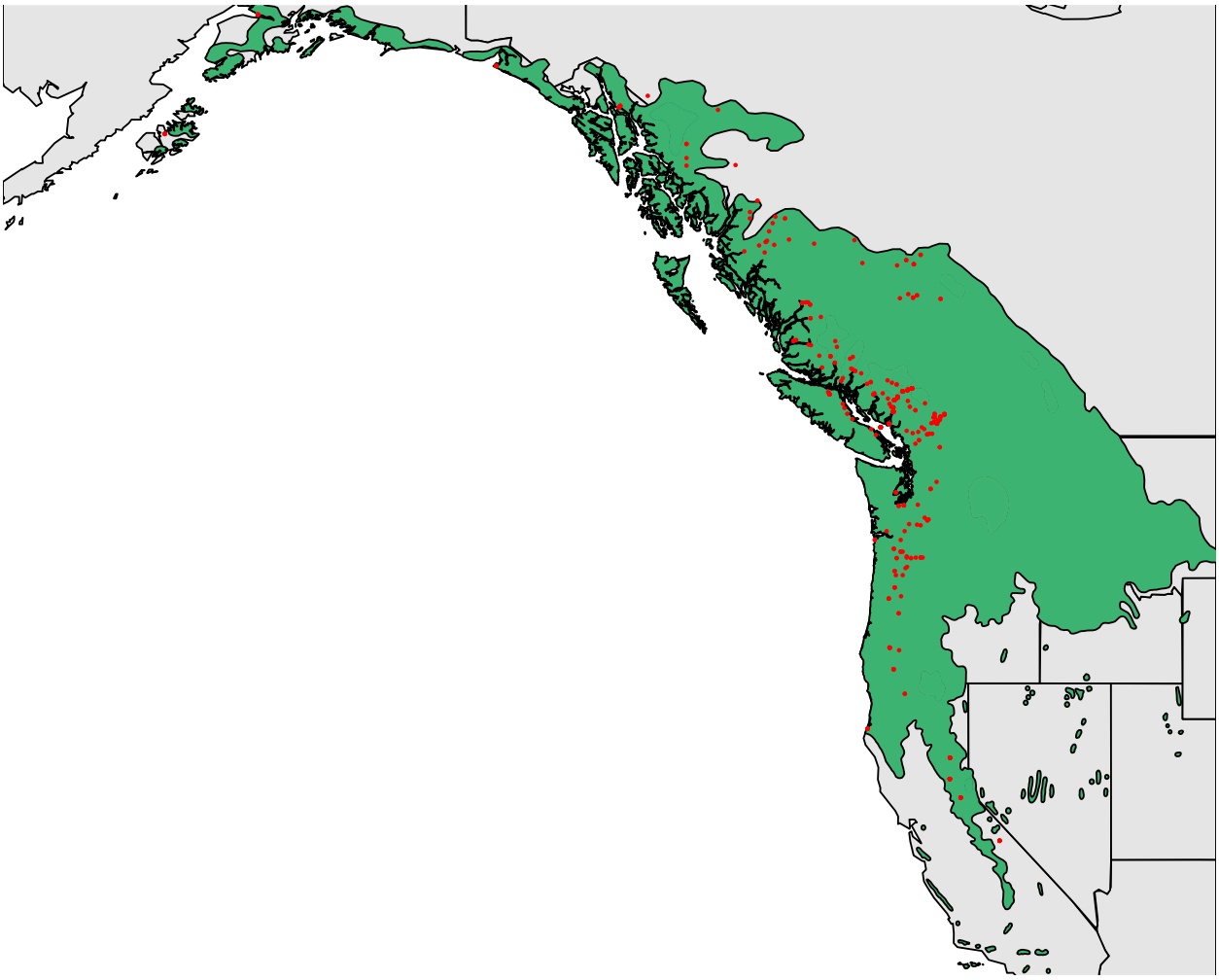


Figure 5.2. Derived allele frequency distribution of deleterious, tolerated, and synonymous SNPs (A) and barplot of proportions of deleterious, tolerated, and synonymous sites within DAF bins (B).

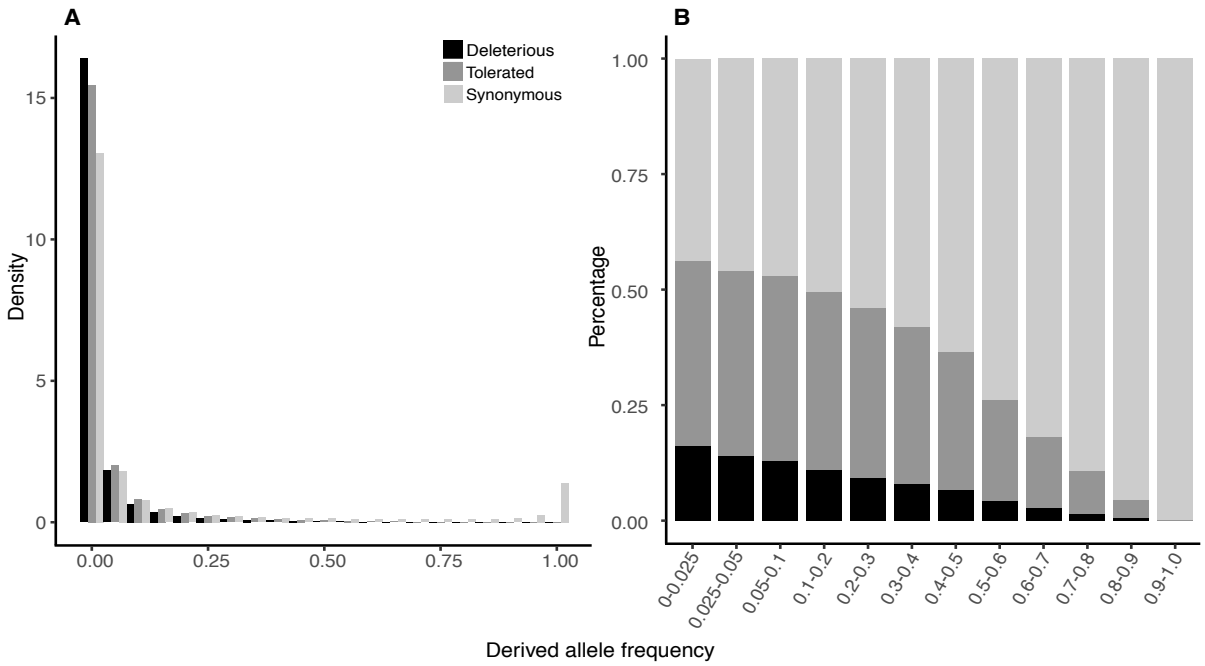


Figure 5.3. Hexbin plot showing an example of the correspondence between deleterious:tolerated ratio and centromere regions (shaded area), where the color of each hexagon indicates the density of points in that part of the graph (A). Bootstrapped mean proportion of deleterious SNPs within centromere regions and non-centromere regions, assessed using 100kb overlapping windows with a 10kb slide (B).

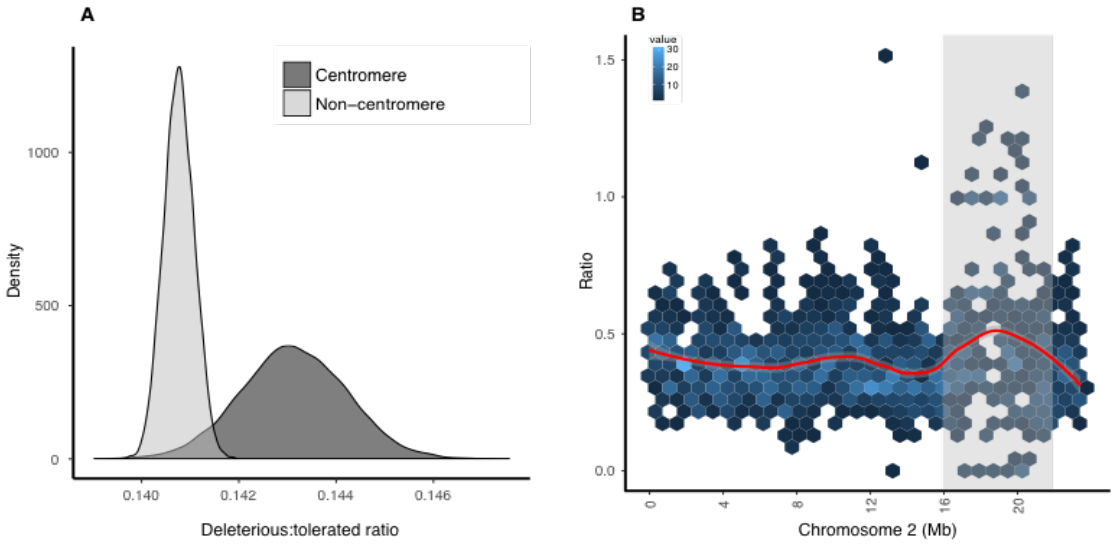


Figure 5.4. Geographical distribution of deleterious homozygosity across 42 populations in relation to distance from the center of the species range (centralized clone distance) (A) and relationship between effective population size and number of mean deleterious homozygotes per population (B). For (B), the explanatory variable (effective population size) is the residuals taken from a multiple linear regression model with covariates for local adaptation (principal components+centralized clone distance) as independent variables and dependent variable being effective population size. The response variable is residuals taken from model with the number of mean deleterious homozygotes per population as dependent variable and the same covariates as independent variables.

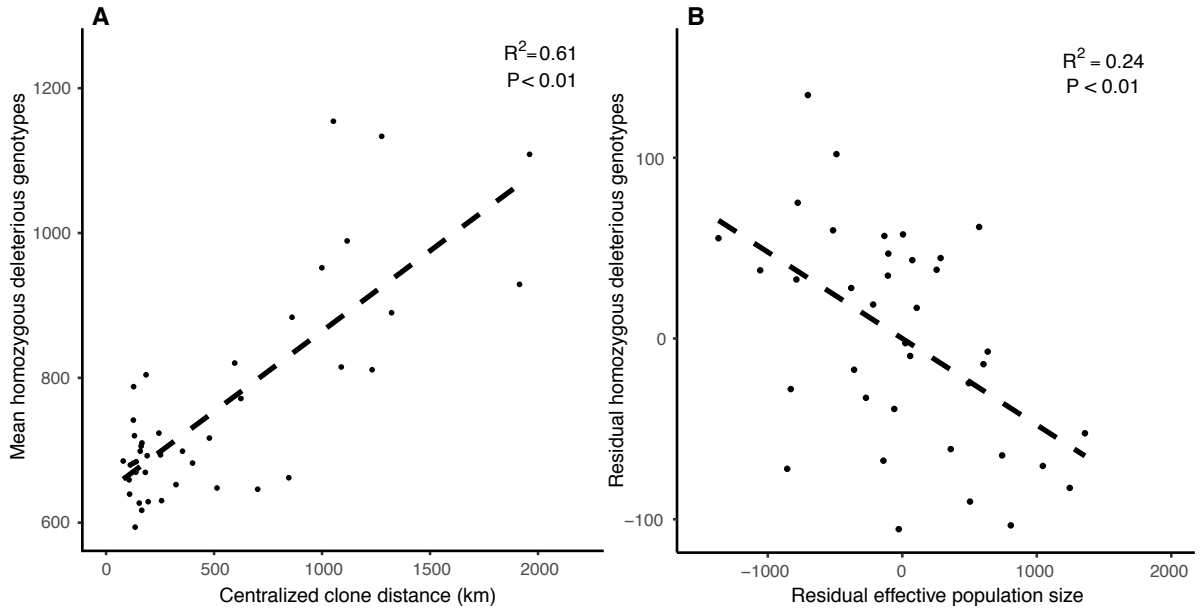


Figure 5.5. Relationship between plant height and number of homozygous deleterious genotypes in the common garden. In each panel, residuals from a regression of height BLUP on covariates (clone distance+principal components) are compared with residuals from a regression of deleterious homozygosity on the same covariates.

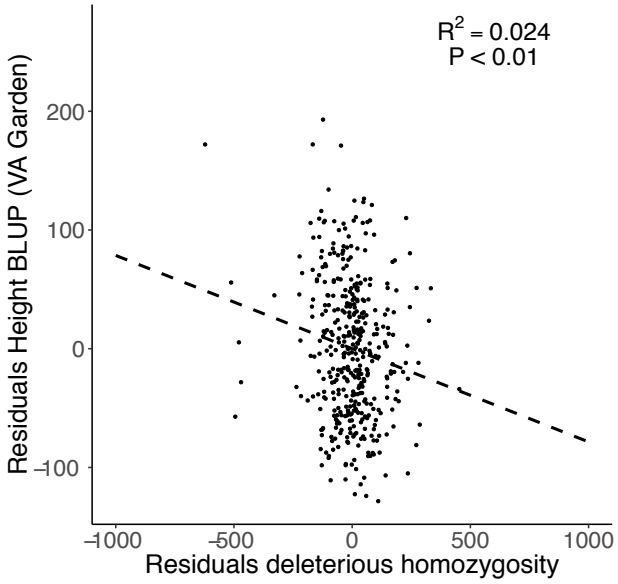


Figure 5.6. Distribution of P-values from SNP-level GWAS for deleterious and tolerated sites (A) and distribution of P-values for gene-level GWAS for genes containing deleterious sites compared with genes containing only tolerated sites (B).

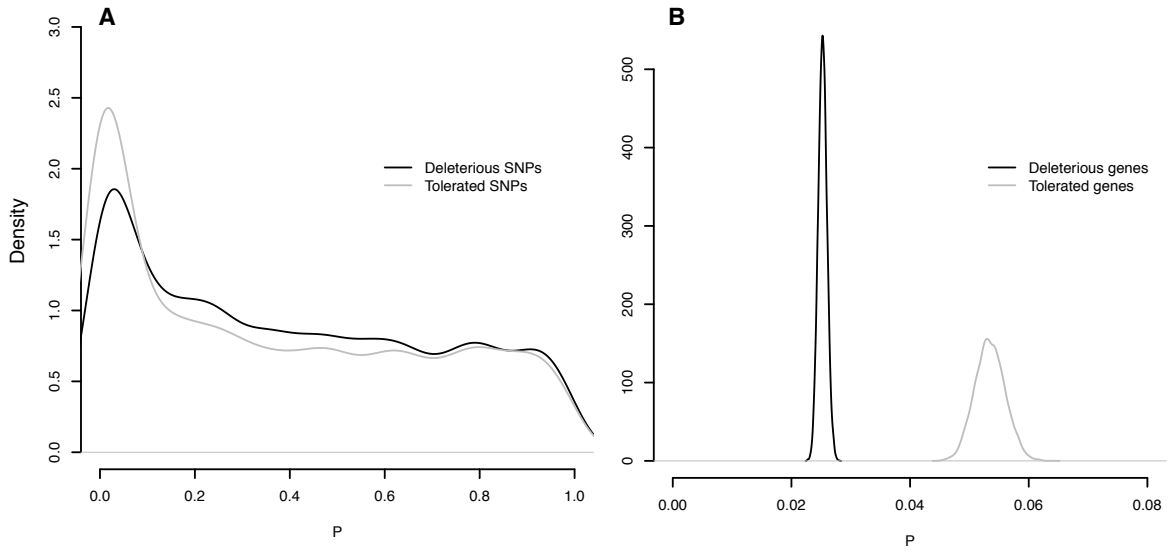


Figure S5.1. Comparison of PROVEAN scores for deleterious and tolerated sites called by SIFT.

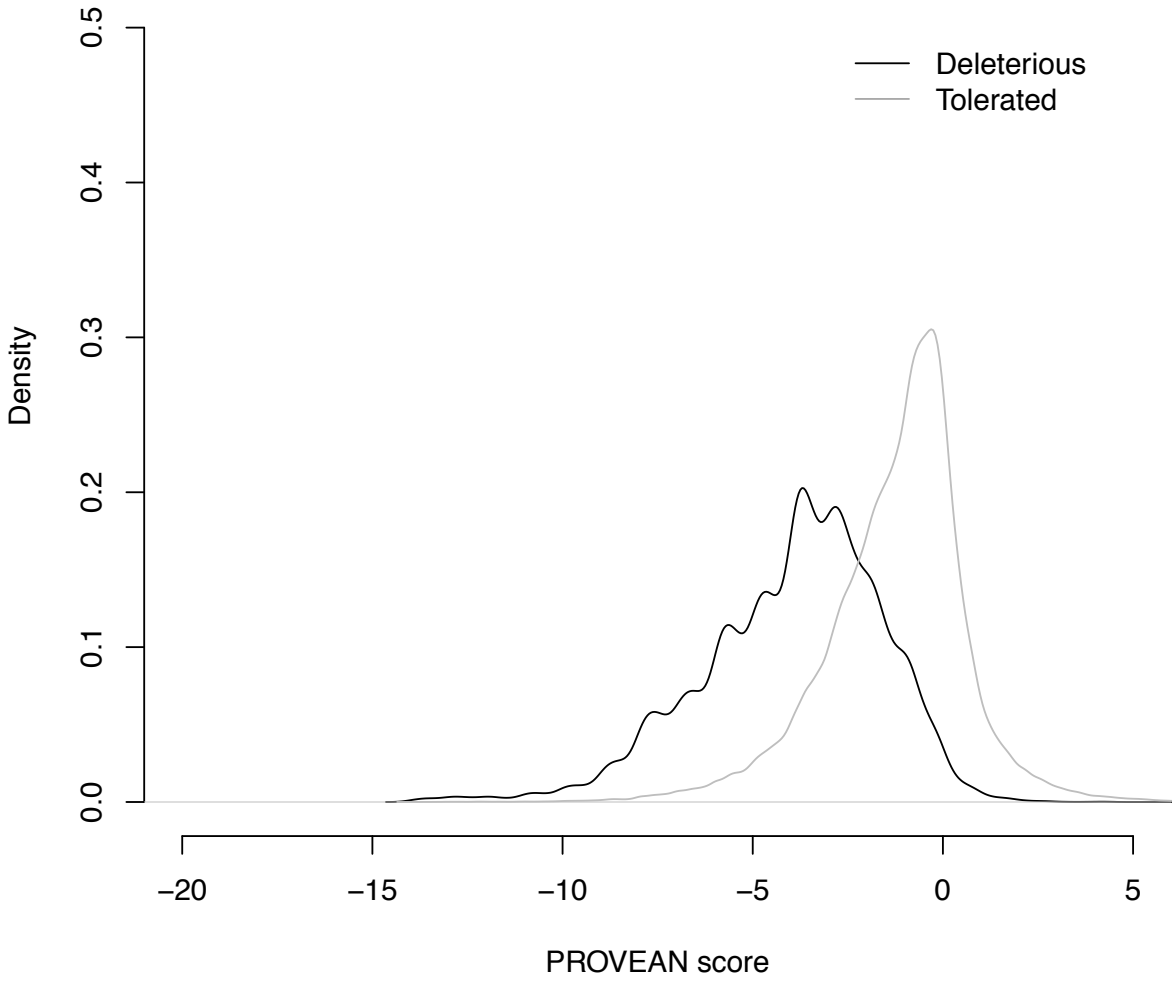


Figure S5.2. Hexbin plot for each poplar chromosome illustrating the correspondence between deleterious:tolerated ratio and centromere regions (shaded area).

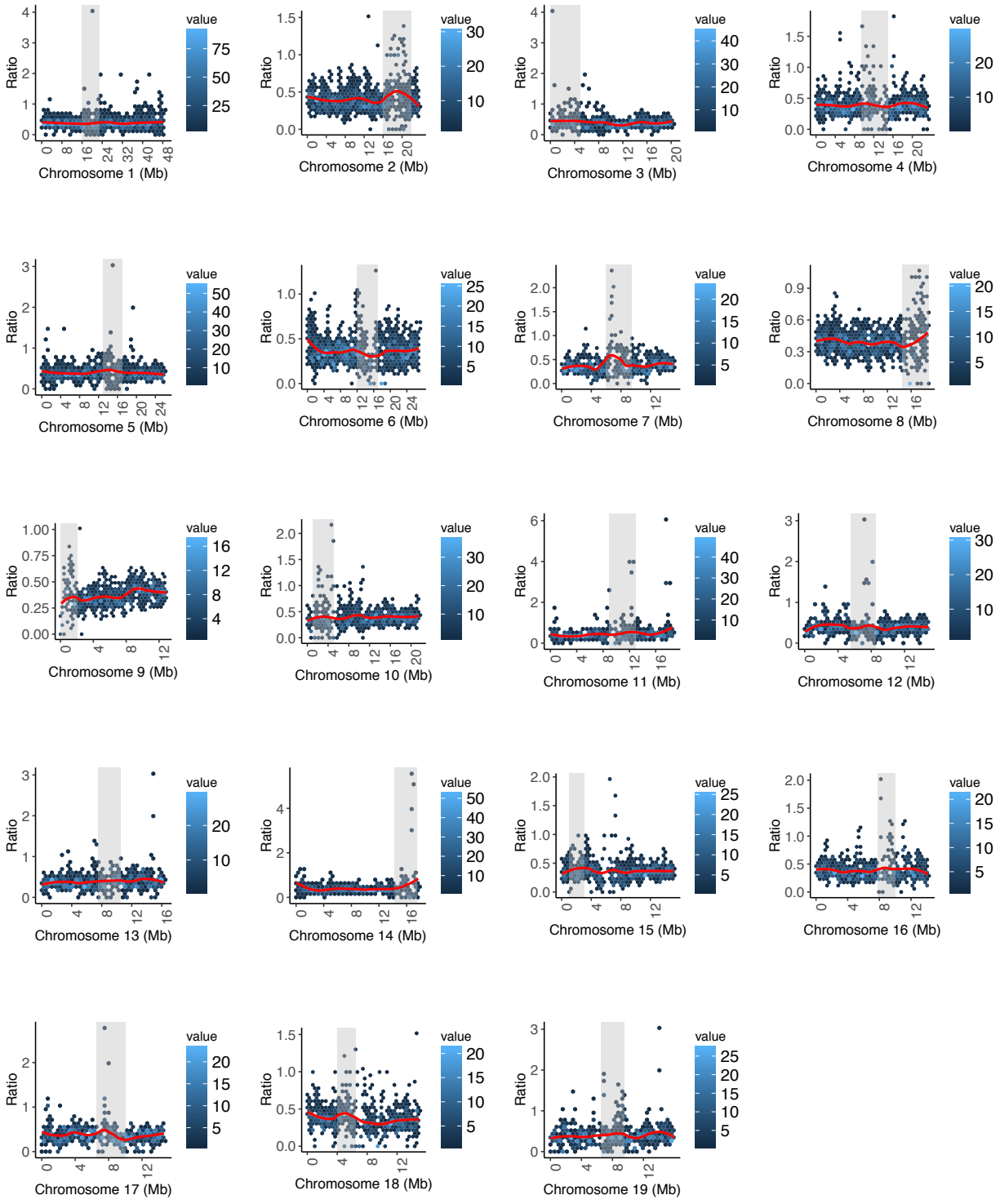


Figure S5.3. Distribution of iHS scores for genomic regions enriched for deleterious alleles compared with all other regions.

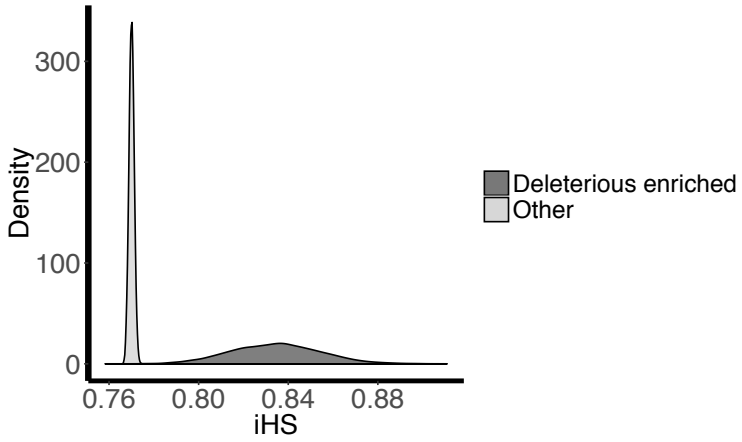
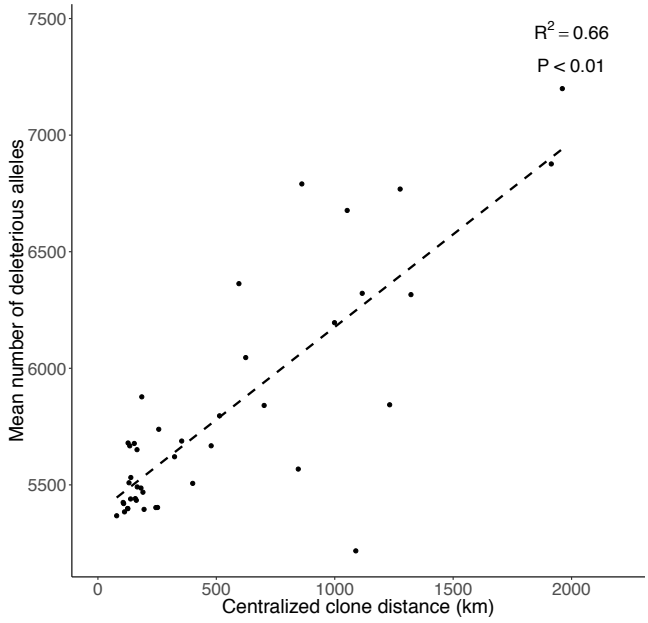


Figure S5.4. Geographical distribution of mean number of deleterious alleles across 42 populations in relation to distance from the center of the species range (centralized clone distance).



6. Conclusion

6. 1 Introduction

Forest tree species have tremendous economic and ecological value as foundation species in many ecosystems, as well as economic value, supplying biofuels and wood. However, rapidly changing global temperature and precipitation regimes are currently affecting the evolution, abundance, and spatial distribution of many species, including trees (Franks and Hoffmann 2012). The fate of local plant populations is largely determined on their capability to adapt and migrate (Aitken, et al. 2008). The maladaptation caused by uncoupling speed of adaptation or migration with climatic conditions threatens species' productivity and provisions to ecosystems and human community (Aitken and Whitlock 2013). A thorough understanding of local adaptation is needed to determine the potential for adaptation to climate change and facilitate breeding and conservation of tree species (Savolainen, et al. 2007). *P. trichocarpa* is the first woody perennial with a sequenced genome, and across its range this species exhibits substantial genetic and phenotypic variation, providing an ideal model for genomic studies of adaptive evolution and as a test case for translating this information to conservation and breeding practices (Ellis, et al. 2010). Several recent studies have combined samples across clines with genetic and genomic approaches, and revealed a number of candidate genes for local adaptation in tree species (Frewen, et al. 2000; Chen, et al. 2002; Evans, et al. 2014; McKown, Klápště, et al. 2014). However, several questions related to the genomics of adaptation remain.

The primary objective of this thesis is to identify the genetic variants associated with climatic adaptation in *P. trichocarpa*. We first demonstrated clinal variation in a set of adaptive traits associated with geography and climate, which is indicative of local adaptation. We then performed a genome-wide association study to identify the genomic determinants controlling adaptive traits. To further connect the genetic variants with environmental drivers of local adaptation, we scanned genome for selection targets with three selection scans. Candidate adaptive variants were compared between two parallel geographic clines to study the degree of parallel adaptation and to understand the evolutionary constraints on local adaptation. In addition to divergent selection driving local adaption, deleterious variation is one evolutionary process influencing ecologically important traits and population fitness (Mitchell-Olds, et al. 2007). We

therefore studied the how evolutionary forces shaped the distribution of deleterious alleles and the contribution of deleterious load to population fitness.

6.2 Clinal variation in adaptive traits indicate local adaptation

P. trichocarpa occupies a wide geographical range covering a variety of environmental conditions. We phenotyped seven adaptive traits in 451 *P. trichocarpa* genotypes that were collected from two natural transects spanning most species range. We observed strong clinal variation in bud phenology, height and diameter, cold hardiness, regeneration height, and branch number. The clinal pattern of adaptive traits is in agreement with observations from previous common garden studies (Savolainen, et al. 2007; McKown, Guy, et al. 2014). We demonstrated significant correlations between phenotypes, geography, and climate variables, suggesting heterogeneous environments drive phenotypic adaptation to local climates. These natural poplar accessions provide genetic materials for investigation of the evolutionary response to diverse climatic challenges and the genetic variants important for adaption processes. We also observed significant effects of genotype origin, growing environment, and their interaction on plant adaptive response, which is important for predicting response of natural populations to non-local climate.

6.3 Significant association between phenotype and genotype

We leveraged the substantial phenotypic variation noted above to link genetic variants to adaptive traits using a genome wide association approach. We combined sliding window analysis with traditional single marker association tests and identified a number of genomic regions with concentrated phenotypic association for timing of bud set and bud flush, height and diameter, cold hardiness, regeneration height and branch number. We found 65 genes associated with timing of bud set, 109 genes with timing of bud flush. 203 genes associate with height, and 14 genes with diameter, 28 genes associated with regeneration height, and 6 genes with regenerated branch number. These variants explained a moderate proportions of phenotypic variation individually and in combination. Many bud phenology associated genes were photoperiodic genes and hormone regulators, some of which were previously characterized in relevant dormancy cycling pathways. One particularly interesting gene, Potri.002G013100, was associated with both timing of bud flush and bud set. The Arabidopsis ortholog of this gene

regulates flowering autonomous pathways and promotes floral initiation through histone methylation of flowering repressors FLC (FLOWERING TIME LOCUS C) and FWA (FLOWERING WAGENINGEN) (Abou-Elwafa, et al. 2010). With flowering and dormancy sharing pathways, this gene is possibly a novel molecular mediator in dormancy cycling. Genes associated with height and diameter were involved in carbohydrate metabolic and secondary growth. Many regeneration-associated genes were annotated to function in embryo development. In general, these candidate genes provide insights into the molecular mechanisms of plant growth and adaptation processes and interesting mediators for reverse functional studies.

The GWAS scans separately performed on samples across latitude and altitude transects identified a number of SNP markers exhibiting parallel phenotypic associations in both transects, indicating conservation of genetic mechanisms controlling adaptive traits along parallel geographical gradients. There was no overlap among regions detected with the sliding windows approach, most likely due to limited power in small samples across the altitudinal transect. Sliding window approach performed well in identifying clusters of strong SNP associations from noisy background and control for false positives. However, this approach may be limited in distinguishing isolated true associations where LD is low. Numerous candidate genes were identified, and suggested the polygenic nature of adaptive traits. Nevertheless, it is likely that more important variants remain undetected, and future studies would benefit from larger sample sizes to increase power for variants or low minor allele frequency, as well as genotyping INDELs an structural variants that may contribute to adaptation.

6.4 Parallel adaptation under climate selection

As a complementary approach, we used three selection scans (iHS, F_{ST} and Bayenv) to further scan the genome for evidence of selection, with a specific focus on detecting evidence of parallelism between the altitudinal and latitudinal transects. For F_{ST}, more than 8% of genomic regions were shared between both transects, while 6% of Bayenv outliers overlapped between genomic regions associated with temperature related variables (MAT, MCMT), precipitation (MAP), and heat moisture (AHM), but little convergence less than 3% in regions associated with Hargreaves reference evaporation (Eref) and climate principal components. Combing findings from F_{ST} and Bayenv scans, we found more than 14% overlap in the candidate selection regions

between altitude and latitude, which is unlikely to have occurred by chance. The significant overlap between genomic regions associated with some climate variables (MAT, MCMT, MAP, and AHM) also supported that selection agents may target on similar genetic variants even across independent environmental gradients. However, other factors such as shared standing genetic variants and developmental constraints likely contribute to the genetic parallelism.

The discordant regions associated with same climate variables are likely the result of the different strengths and directions of selection agents that may favor different genetic variants across transects. The two transects also differ in some selection agents that may lead to different biological pathways being targeted. For example, photoperiodic pathways and UV-induced pigment accumulation were most overrepresented across latitude and altitudinal groups respectively. Finally, candidate selection regions captured with F_{ST} and Bayenv showed significantly stronger phenotypic association with adaptive traits compared with the genomic background, which suggested that regions under climatic selection contain genetic variants associated with adaptive traits. The parallel selection outliers shared across transects provide important candidate genes underlying climatic adaptation.

6.5 Evolutionary forces and demographic history shape deleterious load

As deleterious variation is predicted to affect fitness of individuals and populations, we studied the contribution of deleterious variation to population fitness, and factors shaping the distribution of deleterious alleles across genome and populations. We found that most deleterious alleles were kept at low frequency and in the heterozygous state, indicating purifying selection. However, a number of high frequency deleterious alleles were within regions of high iHS score, suggesting these deleterious alleles may hitchhike to high frequency with advantageous mutations. The distribution of deleterious alleles across the genome also demonstrated enrichment of deleterious alleles in regions of low recombination. Deleterious load reduced plant fitness, with individuals of high deleterious homozygosity having reduced height after accounting for local adaptation and population differentiation in deleterious homozygosity. Populations in the edge, especially at the northern range limit, held more deleterious alleles than central populations. As effective population size is smaller in edge populations, purifying selection is expected to be less efficient in purging deleterious mutations. Another likely

explanation is recent bottleneck at the populations that recolonized new habitats. During bottleneck, many rare alleles can be lost due to enhanced genetic drift, while some deleterious alleles can drift to higher frequencies, leading to increased deleterious load (Simons, et al. 2014). This is consistent with observed excess of rare deleterious alleles in European populations that arose during expansion out of Africa (Lohmueller, et al. 2008; Peischl, et al. 2013). Taken together, evolutionary forces and demographic history together shape the distribution of deleterious mutations across the genome and among populations.

6.6 Limitations and future research

The availability of high throughput sequencing and decreasing sequencing cost makes association mapping studies feasible on a genome-wide scale (Korte and Farlow 2013). Most ecologically important traits are of complex nature, with variation controlled by hundreds or thousands loci (Neale and Savolainen 2004). Our GWAS study was efficient in identifying a large number of candidate genes that explained a moderate proportion of phenotypic variation. However, there are many factors limiting the power and accuracy of GWAS, such as sampling, effect size of functional variants, sequencing error, and population structure. To realize the full potential of GWAS, sufficient sample size is suggested, especially when the effect size of most detected markers are small. False positives commonly arise in GWAS due to sequencing error and cofounding effect of population stratification (Korte and Farlow 2013). We controlled such false positives with SNP quality (quality score >30) and sequencing coverage (>10x), which reflects stringent cutoffs compared with most genomic studies (Tiffin and Ross-Ibarra 2014). In addition, our sliding window approach should be relatively robust to scattered false discoveries. Finding the proper adjustment for population structure is necessary to reduce false positives and avoid over-correction for true positives, and our principal components analysis was successful in controlling population structure (Novembre and Stephens 2008). However, with high gene flow, tree species form continuous spatial groups, which makes it challenging to determine the optimal method controlling population structure.

There is a growing interest in understanding the extent to which adaptive evolution is constrained, and one approach to this problem is to use closely related species or population to estimate the extent of parallelism in the genes contributing to adaptation (Conte, et al. 2012; Tiffin and Ross-

Ibarra 2014). Our study revealed parallel adaptation at the molecular level between two natural transects, though the degree of overlap is slightly less than previously reported (Holliday, et al. 2016). Repeated use of same genes during phenotypic evolution indicates their significance in local adaption and informs our understanding of the constraints on adaptive evolution (Conte, et al. 2012). We found a substantial degree of selected variants shared across latitude and altitude, which is unlikely to have occurred by chance. Such parallelism likely reflects both genetic constraint and the common ancestry of our transects, which leads to commonalities in shared standing variation at the outset of adaptation (Conte, et al. 2012). By associating environmental variables with genetic divergence, we were able to refine genes targeted by specific selective forces. To further consolidate the role of selection on driving parallel adaptation, studies can limit climate gradients in a finer scale and estimate the degree of gene overlap with more precise control of selective agents.

With a large number of target loci detected through genome selection scans, independent functional analyses are important in confirming whether candidate genes are causative. We demonstrated the adaptive significance of selection outliers by showing their overlap with strong phenotypic association signals. An alternative approach is to test if the candidate genes individually or as a group, can predict adaptive performance of independent samples. One example is that Fournier-Level et al. (2011), who first identified candidate loci associated with local adaptation using GWAS, and subsequently confirmed that these fitness alleles were more locally abundant than genomic controls using a set of accessions not used in initial analyses (Fournier-Level, et al. 2011). Future work can split mapping populations into samples used for initial genome-wide screen, and the other for targeted validation of candidate loci. The validation work can also be extended to related species. Knowledge of the extent to which adaptive variants are shared among related species can validate the detected candidate loci and also tell us if adaptation involves standing genetic variation or newly arisen mutations (Colosimo 2005). The identified shared adaptive variants would also be useful for marker-assisted breeding and conservation projects in related tree species that have not been studied.

Our studies demonstrated the fitness reduction induced by deleterious load and informed the evolutionary and demographic history leading to difference in deleterious load among natural

populations of *P. trichocarpa*. During range expansion of species, populations at the expanding wave front mimic populations adapting to new environments. As global warming trigger range shifts of many temperate tree species, the understanding of evolutionary dynamic and consequences of deleterious allele accumulation in marginal populations is an important topic for future evolutionary studies. One limitation of current studies of deleterious alleles is the difficulty in evaluating the adaptive importance of deleterious variation in natural populations. The allelic effect of individual deleterious polymorphisms often is too small to be captured with traditional experimental approaches. We attempted to verify the fitness effect of deleterious alleles with GWAS. However, single-variant association approaches performed poorly on rare variants due to limited power and synthetic association with non-causative rare variants (Korte and Farlow 2013). Recently, many region-based or gene-based multi-locus tests, adaptive burden test, and variance-component tests were proposed, which may provide more functional information on rare variants (Lee, et al. 2014).

6.7 Reference cited

Abou-Elwafa SF, Buttner B, Chia T, Schulze-Buxloh G, Hohmann U, Mutasa-Gottgens E, Jung C, Muller AE. 2010. Conservation and divergence of autonomous pathway genes in the flowering regulatory network of Beta vulgaris. Journal of Experimental Botany 62:3359-3374.
Aitken SN, Whitlock MC. 2013. Assisted Gene Flow to Facilitate Local Adaptation to Climate Change. Annual Review of Ecology, Evolution, and Systematics 44:367-388.
Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S. 2008. Adaptation, migration or extirpation: climate change outcomes for tree populations. Evolutionary Applications 1:95-111.
Chen THH, Howe GT, Bradshaw HD. 2002. Molecular genetic analysis of dormancy-related traits in poplars. Weed Science 50:232-240.
Colosimo PF. 2005. Widespread Parallel Evolution in Sticklebacks by Repeated Fixation of Ectodysplasin Alleles. Science 307:1928-1933.
Conte GL, Arnegard ME, Peichel CL, Schluter D. 2012. The probability of genetic parallelism and convergence in natural populations. Proceedings of the Royal Society B: Biological Sciences 279:5039-5047.
Ellis B, Jansson S, Strauss SH, Tuskan GA. 2010. Why and How *Populus* Became a "Model Tree". In: Jansson Sea, editor. Genetics and Genomics of *Populus*, Plant Genetics and Genomics: Crops and Models 8: Springer Science+Business Media, LLC.
Evans LM, Slavov GT, Rodgers-Melnick E, Martin J, Ranjan P, Muchero W, Brunner AM, Schackwitz W, Gunter L, Chen JG, et al. 2014. Population genomics of *Populus trichocarpa* identifies signatures of selection and adaptive trait associations. Nature Genetics 46:1089-1096.
Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM. 2011. A Map of Local Adaptation in Arabidopsis thaliana. Science 334:86-89.
Franks SJ, Hoffmann AA. 2012. Genetics of Climate Change Adaptation. Annual review of genetics 46:185-208.
Frewen BE, Chen THH, Howe GT, Davis J, Rohde A, Boerjan W, Bradshaw HD. 2000. Quantitative trait loci and candidate gene mapping of bud set and bud flush in Populus. Genetics 154:837-845.
Holliday JA, Zhou L, Bawa R, Zhang M, Oubida RW. 2016. Evidence for extensive parallelism but divergent genomic architecture of adaptation along altitudinal and latitudinal gradients in *Populus trichocarpa*. New Phytologist 209:1240-1251.
Korte A, Farlow A. 2013. The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9:29.
Lee S, Abecasis Gonçalo R, Boehnke M, Lin X. 2014. Rare-Variant Association Analysis: Study Designs and Statistical Tests. The American Journal of Human Genetics 95:5-23.
Lohmueller KE, Indap AR, Schmidt S, Boyko AR, Hernandez RD, Hubisz MJ, Sninsky JJ, White TJ, Sunyaev SR, Nielsen R, et al. 2008. Proportionally more deleterious genetic variation in European than in African populations. Nature 451:994-997.
McKown AD, Guy RD, Klapste J, Geraldes A, Friedmann M, Cronk QC, El-Kassaby YA, Mansfield SD, Douglas CJ. 2014. Geographical and environmental gradients shape phenotypic trait variation and genetic structure in *Populus trichocarpa*. New Phytologist 201:1263-1276.
McKown AD, Klápště J, Guy RD, Geraldes A, Porth I, Hannemann J, Friedmann M, Muchero W, Tuskan GA, Ehlting J, et al. 2014. Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of *Populus trichocarpa*. New Phytologist 203:535-553.

Mitchell-Olds T, Willis JH, Goldstein DB. 2007. Which evolutionary processes influence natural genetic variation for phenotypic traits? Nature Reviews Genetics 8:845-856.
Neale DB, Savolainen O. 2004. Association genetics of complex traits in conifers. Trends in Plant Science 9:325-330.
Novembre J, Stephens M. 2008. Interpreting principal component analyses of spatial population genetic variation. Nature Genetics 40:646-649.
Peischl S, Dupanloup I, Kirkpatrick M, Excoffier L. 2013. On the accumulation of deleterious mutations during range expansions. Molecular Ecology 22:5972-5982.
Savolainen O, Pyhäjärvi T, Knürr T. 2007. Gene Flow and Local Adaptation in Trees. Annual Review of Ecology, Evolution, and Systematics 38:595-619.
Simons YB, Turchin MC, Pritchard JK, Sella G. 2014. The deleterious mutation load is insensitive to recent population history. Nature Genetics 46:220-224.
Tiffin P, Ross-Ibarra J. 2014. Advances and limits of using population genetics to understand local adaptation. Trends in Ecology & Evolution 29:673-680.